Characterisation of Glutathione Transferase from *Arabidopsis thaliana*

Laziana Ahmad

PhD

University of York

Biology

January 2017

For my Dad,
Utterly beloved,
Deeply missed,

Until we be together again.

Abstract

Glutathione transferases (GSTs) are soluble enzymes with activity towards a wide range of xenobiotic and endogenous compounds. The Arabidopsis genome encoded 54 GSTs, which have been classified into eight classes including Tau and Phi. Members of these Tau and Phi clades are strongly upregulated in response to abiotic stresses such as xenobiotics and biotic stresses including pathogen attack. The most characterised activity of GSTs is the transfer of glutathione to an electrophilic centre to form a polar glutathionylated conjugate. However, increasing number of research demonstrated a non-catalytic activity plants GSTs, especially in the transportation of flavonoids from the cytosol to the vacuole. Despite the wealth of investigations into GSTs, and probably as a result of overlapping substrate specificities, the endogenous roles for the vast majority of these plant GSTs remains unknown.

Here, the binding interaction of camalexin, indole-3-aldehyde, quercetrin and quercetin to GSTF2 in a non-catalytic fashion was observed in three different sites; two identical sites of **L1** and one **L2** from the X-ray crystallography data. Mutagenesis of the active residues, Q73L, H77A, Y97A and R154A were performed and using isothermal calorimetry (ITC) techniques, lower binding affinities were observed for all mutants towards all ligands except for Y97A and Q73L which showed higher binding affinities with indole-3-aldehyde. This unexpected finding was likely due to the conformational change of the mutant compared to the wild type, as observed in the structure of mutant Y97A.

On elucidating the catalytic activity of GSTs, the structure of GSTU25 in complex with disulphide glutathione was obtained. The GSTU25 has been recently identified to catalyse the denitration of TNT to form 2-glutathionyl- 4,6-dinitrotoluene, a potentially more amenable product for subsequent degradation. This structure complex provides insights of GSTU25 folding upon substrate binding.

The involvement of GSTU25 in the detoxification of TNT was further analysed using CRISPR/Cas9 technology. Subclades of *GSTU25*, including *GSTU24*, *GSTU21*, *GSTU19* were knockout with the aim to remove overlapping substrate specificities and to finally reveal TNT-specific phenotypes. Only *gstU25* Cas9 segregated Arabidopsis were obtained from the experiment that could potentially be optimised in the future study.

List of contents

List of tables

List of figures

Acknowledgements

The completion of this thesis could not have been possible without the assistance of so many people. I would like to start by thanking Majlis Amanah Rakyat (MARA), Malaysia for generously providing three years of support and funding which enabled me to pursue my PhD.

Special thanks to both my supervisors: Prof. Neil Bruce and Prof. Gideon Grogan, whose guidance, care and trust were sometimes all that kept me going. I have enjoyed learning in their lab and will always be grateful for this opportunity.

Many thanks to Dr. Liz Rylott for her enthusiastic encouragement and useful critiques of this research work.

To my Thesis Advisory Panel members: Prof. James Moir and Dr. Louise Jones for their supports and critical assessment of my work.

Thanks also to all of the members of GG group, NCB group (former and current) and to everyone on M2. Big thanks to Dave for his excellent help in the lab whenever needed and to Margaret who helped with my travel itineraries.

Thanks to Prof. Robert Edwards for making this journey real for me, and to his group members: Dr. Melissa Brazier Hicks, Bekki and Keir for their wonderful help in the first six months when I arrived.

Thanks also to the Department of Biology, University of York, for the PhD Studentship award.

To my big family in York Mosque and good friends; Nahed, Shikin and Maria, thank you for being there all the time, through happy and not so happy times.

And finally, thanks to my family, especially my mum for her almost unbelievable support. She is the most important person in my life, my main motivation and I dedicate this thesis for her.

Author's declaration

I declare that I am the sole author of the work in this dissertation and that it is original except where indicated by special reference in the text. No part of the dissertation has been submitted for any other degree to any other institution. All sources are acknowledged as References.

Publication arising from this research:

1. **Laziana Ahmad**, Elizabeth Rylott, Neil Bruce, Robert Edwards, Gideon J. Grogan. 2016. Structural evidence for Arabidopsis glutathione transferase *At*GSTF2 functioning as a transporter of small organic ligands. *FEBS Open Bio*. DOI: 10.1002/2211-5463.12168.
2. Kyriakos Tzafestas, **Laziana Ahmad**, Henry Man, Gideon Grogan, Elizabeth Rylott, Neil Bruce. Crystal structure of Arabidopsis glutathione transferase U25 and the catalytic features involved in its detoxification of the environmental pollutant and explosive, 2,4,6-trinitrotoluene. *Manuscript under preparation.*

1 Introduction

1.1 GLUTATHIONE (GSH)

Glutathione (GSH) is a tripeptide of *γ*-L-glutamyl-L-cysteinyl-glycine with molecular weight of 307 g/mol is considered as a low molecular mass thiol molecule (**Figure 1.1**). GSH carries an active thiol, a sulfhydryl group (SH) in the cysteine residue and acts as antioxidant directly by interacting with reactive oxygen or nitrogen species (ROS/RNS) (Sies 1999). Interaction with electrophiles as a cofactor for various enzyme activity has also been studied (Cooper *et al.*, 2011).

Several homologs of GSH, in which the C-terminal residue glycine been substituted by other amino acids can be found in some plant taxa. In legumes, the glycine is substituted by an alanine, forming a *γ*-L-glutamyl-L-cysteinyl-*β*-alanine, also known as homoglutathione (Rennenberg 1982; Klapheck 1988). Both homoglutathione and GSH are believed to have resulted from the evolution of different GSH synthetases or homoglutathione synthetases (Frendo *et al.*, 2001).

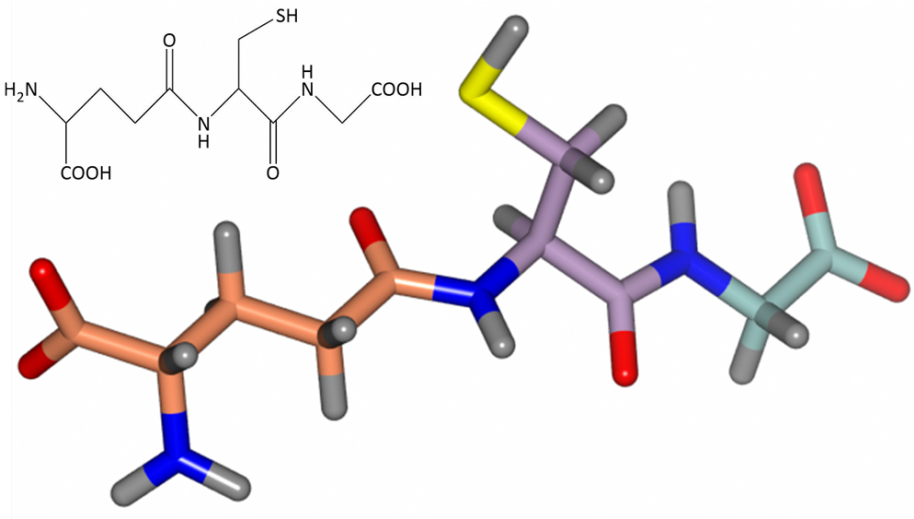


Figure 1.1. Structure of GSH in cylinder representation. The linear tripeptide comprises three amino acids, from left to right: *γ*-L-glutamine, L-cysteine and glycine and each are coloured in orange, lilac and green respectively. The nucleophile sulfhydryl (SH) with sulphur atom is coloured in yellow. Nitrogen, oxygen and hydrogen atom are coloured in blue, red and grey respectively. The insert is the chemical structure of GSH.

1.1.1 Biosynthesis of GSH

In rodent and human tissue, the GSH is synthesised *de novo* in the cytoplasm, in a sequential enzymatic reaction powered by adenosine triphosphate (ATP). The process involves the formation of ϒ-glutamylcysteine from cysteine and glutamate followed by the generation of GSH from ϒ-glutamylcysteine and glycine (**Figure 1.2**). The first step in the reaction is catalysed by glutamate cysteine ligase (GCL) (EC 6.3.2.2; formerly known ϒ-glutamylcysteine synthetase (ϒ-ECS)) (Gipp *et al.*, 1995; Huang *et al.*, 1995) while the second step is catalysed by GSH synthase (GS) (EC 6.3.2.3) (Griffith 1979).

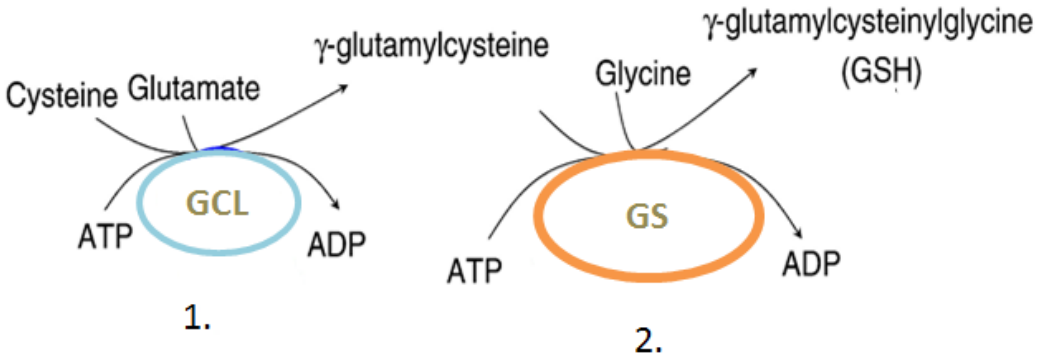


Figure 1.2. GSH biosynthesis in the cytoplasm of human and rodent tissue. The process involves two-step ATP-dependent reactions: 1: catalysed by GCL and 2: catalysed by GSH synthase. GCL: glutamate cysteine ligase, GS: glutathione synthase, ADP: adenosine diphosphate. Figure reproduced from Lu (2014).

Similarly in plants, the GSH is synthesised by two ATP-dependent steps (Noctor *et al.*, 2012). The first step of GSH synthesis in Arabidopsis is encoded by *GSH1* while the GSH synthase catalysed the formation of GSH in the second step is encoded by *GSH2* (May & Leaver 1994; Ullmann *et al.*, 1996). Using transcript analysis, *in vivo* targeting study with *GSH1* as reporter gene fusion proteins, immuno-cytochemical localisation and analysis of expressed His-tagged GSH1 protein, the first step of GSH synthesis in Arabidopsis was found to be located in plastids (mainly chloroplast), while the second step occurs in the plastid and the cytosol (**Figure 1.3**) (Wachter *et al.*, 2004; Rausch *et al.*, 2007). Several factors regulate the synthesis of GSH, but the components involved in the first step of GSH synthesis reaction such as cysteine availability and the activity of GCL were considered to be the most important (Noctor *et al.*, 2012).

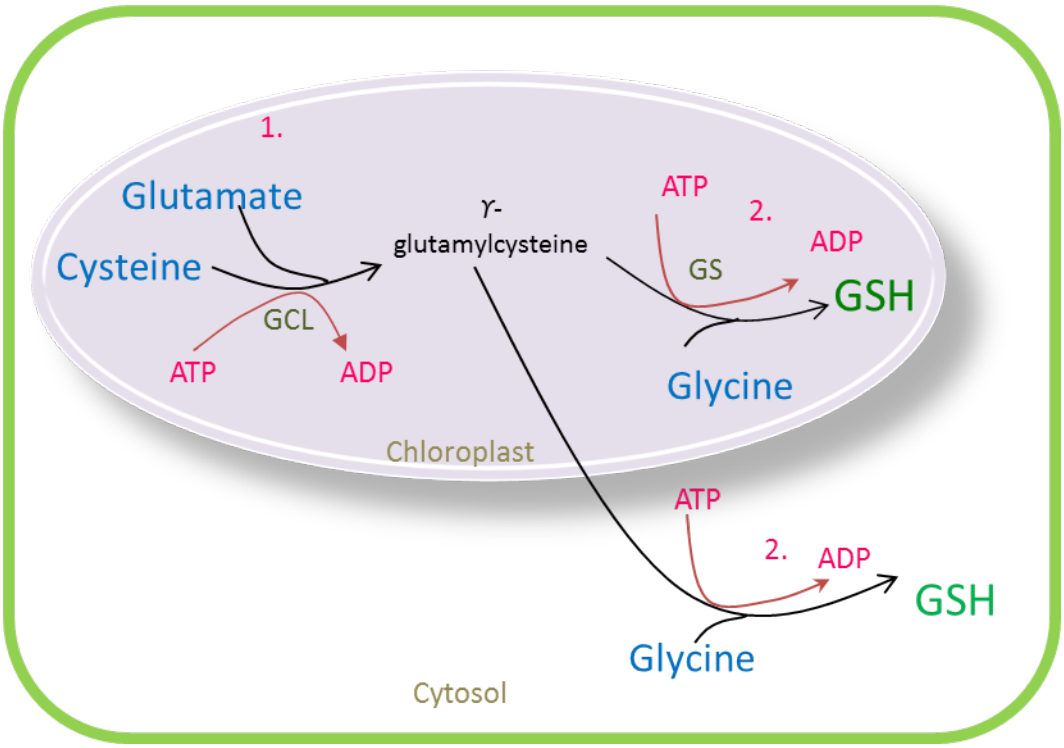


Figure 1.3. GSH biosynthesis in Arabidopsis. 1. and 2. represent the first and second step in GSH biosynthesis. Figure adapted from Rausch *et al.*, (2007).

1.1.2 The biological role of GSH in plants

The activity of GSH lies with the sulfhydryl group on the cysteinyl residue of GSH which provides the nucleophilic thiol for electron transfer (Ketterer *et al.*, 1983). As electrons are lost, the GSH is oxidised and two molecules of GSH are dimerised by a disulphide bridge to form a disulphide GSH (GSSG) (**Figure 1.4**). To maintain the homeostatic balance of the reduced (GSH) and oxidised form (GSSG), the GSSG is then rapidly recycled to GSH by NADPH-dependent glutathione reductase (GR), in a fluctuating ratio across different tissues and subcellular compartments (Noctor *et al.*, 2012; Meyer *et al.*, 2007; Queval *et al.*, 2011). In unstressed conditions, GSH is present predominantly in its reduced form with a relatively small fractions present in the oxidised, GSSG. During oxidative stress, higher level of the oxidised form accumulates. The ratio of GSH to GSSG is used as a critical redox regulator by the cell and the decline in this ratio is closely related to oxidative stress (Zhou *et al.*, 2014).

Under physiological conditions, experiments indicate that GSH accumulates to millimolar concentrations with the ratio of GSH:GSSG in leaf tissue of Arabidopsis maintained at least 20:1 (Mhamdi *et al.*, 2010). While it is difficult to accurately measure *in planta* GSH levels, immunogold cytochemistry and computer-supported transmission electron microscopy

have been used with varying degrees of success, and have shown that mitochondria contain the highest level of GSH followed by the nucleus, cytosol and peroxisomes in the leaves and roots of Arabidopsis (Zechmann *et al.*, 2008; Zechmann & Müller 2010). Genetically encoded biosensors redox-sensitive green fluorescent protein (roGFP) has enabled non-invasive and real-time dynamic redox measurement *in vivo* to measure the concentration of GSH in the cell. Recently, roGFP targeted to chloroplastic, mitochondriol, peroxisomal and cytosolic compartments has been successfully used to measure the GSH redox potential that is associated with cell death pathways in the different organelles (Bratt *et al.* 2016).

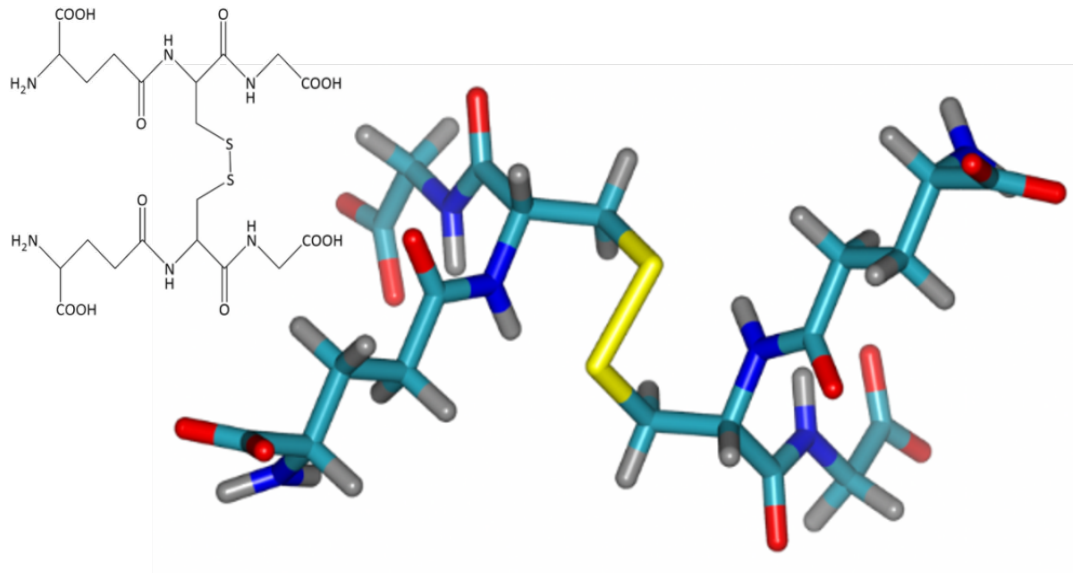


Figure 1.4. The structure of the oxidised form of glutathione, disulphide GSH (GSSG). The disulphide bond at the –SH group coloured in yellow. The insert is the chemical structure of GSSG.

In addition to its central role in antioxidant and detoxification metabolism, GSH is implicated in a number of other important functions during plant development and metabolism (Noctor *et al.*, 2012). Some of the most important functions of GSH include redox regulation, xenobiotic and heavy metal detoxification as well as in sulphur, carbon and nitrogen metabolism, as overviewed in **Figure 1.5**. That GSH is essential for plant survival has been demonstrated using T-DNA insertion mutants: the *gsh1* mutant displayed a recessive embryo-lethal phenotype and mutant of *gsh2* showed a seedling-lethal phenotype (Pasternak *et al.*, 2007; Cairns *et al.*, 2006).

In a yet to be fully-characterised mechanism, GSH is also believed to be essential for cell proliferation and root growth. GSH synthesis Arabidopsis mutants, lacking in *GSH1* activity,

showed 79% reduced primary root growth, 50% reduced of growth in the lateral root compared to wild-type (Bashandy *et al.*, 2010). Experiments by Koprivova *et al.* (2010) indicated that the inhibition of root growth by the depletion of GSH is linearly correlated to alterations in auxin homeostasis. The depletion of GSH content is also linked to the modification of the gene expression related to plant defence, cell signalling and stress tolerance (Ball *et al.* 2004; Fratelli *et al.* 2005).

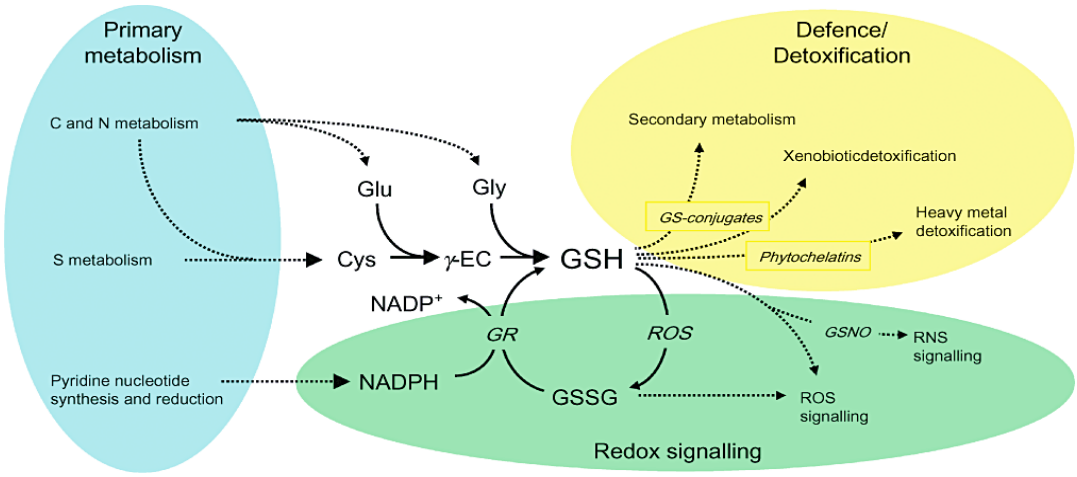


Figure 1.5. Overview of the GSH functions during plant development (synthesis, redox turnover, metabolism and signalling). *γ*-EC: *γ*-glutamylcysteine; GS-conjugates: glutathione S-conjugates; RNS: reactive nitrogen species; ROS: reactive oxygen species; GSNO: S-nitrosoglutathione. Figure from Noctor *et al.*, 2012.

In plants, the GSH-related detoxification system for xenobiotics occurrs in three phases (**Figure 1.6**). In phase I, the xenobiotics are activated by P450 monooxygenases, peroxidases or other enzymes that catalyse oxidation, reduction or hydrolysis of the compound. Conjugation to GSH through the activity of glutathione transferase (GST) proceeds in phase II. In phase III, the xenobiotic conjugates are sequestered from the cytosol by the activity of tonoplast Multidrug Resistance-associated Protein (MRP) transporters, a family of the ABC–transporter (Coleman *et al.*, 1997). This reaction is ATP-dependent and results in the effective removal of the conjugates from the cytosol into the vacuoles (Schröder *et al.*, 2007). Recently, cryo-electron microscopy data of MRP1 indicated a bipartite nature of the molecules enabling two substrate recognition sites. One site is positively charged and coordinates GSH moiety (P-pocket) and the second encompasses by hydrophobic lipid tail (H-pocket) providing site for hydrophobic substrates. The structure of MRP1 suggested that MRP1 recruits its substrate exclusively from

cytoplasm *via* its well-ordered transmembrane helices into the cytosol of organelle (Johnson & Chen 2017).

A study using the herbicide alachlor in the cereal crop barley (*Hordeum vulgare*) indicated that the breakdown of GSH conjugates in leaves was catalysed by vacuolar carboxypeptidase. This enzyme cleaves glycine from the GSH-conjugate to form *γ*-glutamylcysteine derivatives, which in turn were processed to cysteine conjugates by *γ*-glutamyl transpeptidases (Wolf *et al.*, 1996). However, a study in Arabidopsis (*Arabidopsis thaliana*), using monochlorobimane has identified two parallel processing pathways. In pathway 1, that was found to predominate in the roots, the GSH conjugates were sequentially processed in the vacuole to cysteinylglycine derivatives and other cysteine conjugates (Grzam *et al.*, 2007; Grzam *et al.*, 2006; Ohkama-Ohtsu *et al.*, 2007). In the second pathway, that was mainly active in the leaves, the GSH conjugates were processed to *γ*-glutamylcysteine derivatives in the cytosol, through the activity of phytochelatin synthase (Blum *et al.*, 2007). Subsequently, the *γ*-glutamylcysteine derivatives were converted to cysteine conjugates by a *γ*-glutamyltranspeptidase isoenzyme localised in the plasma membrane (Ohkama-Ohtsu *et al.*, 2007). In studies on onion (*Allium cepa*) and spruce tree (*Picea* spp.), the GSH conjugates have also shown to be metabolised to cysteinyl- and malonylcysteinyl-metabolites and rarely to sulfonic acids (Lamoureux & Rusness 1980; Lamoureux *et al.*, 1993; Aga *et al.*, 1996).

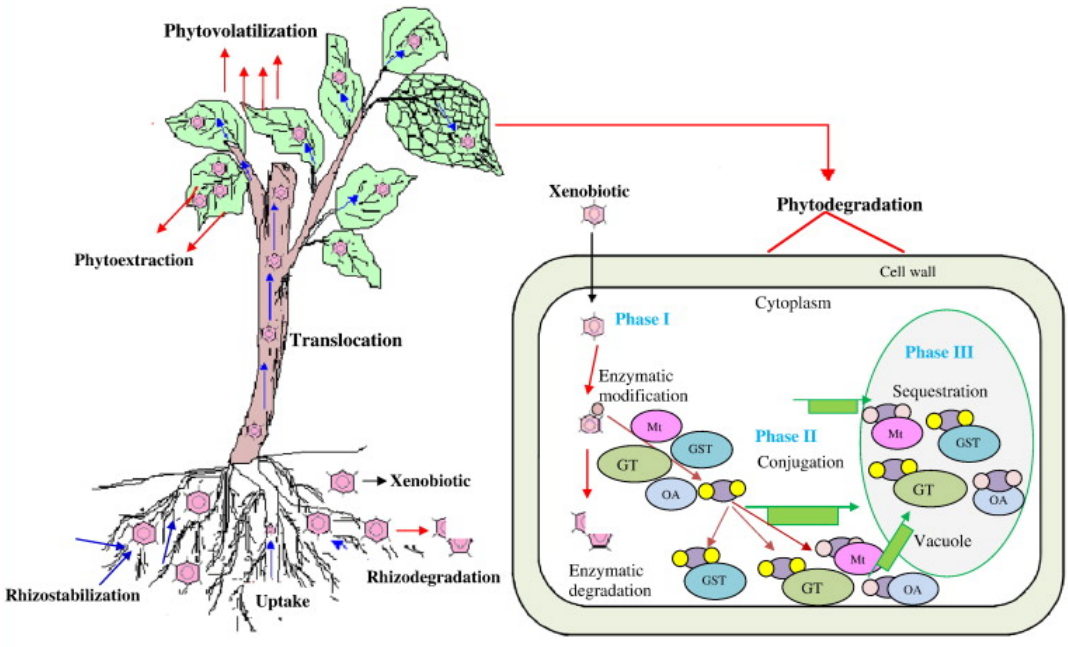


Figure 1.6. Summary of plant xenobiotic detoxification mechanism. In the rhizosphere, the xenobiotics can be stabilised or degraded and adsorbed or accumulated in the roots, transported to the aerial parts and degraded or volatilised in the plant tissue. The detoxification mechanism involved three phases: I: enzymatic modification, II: conjugation and III: active sequestration. GST: Glutathione transferases, GT: glucosyltransferase, Mt: malonyltransferase; OA: organic acids and active transporters in green boxes. Figure from Abhilash *et al.* (2009).

1.2 GLUTATHIONE TRANSFERASES (GSTS)

Glutathione transferases (GSTs) are an ancient superfamily of catalytic and ligand binding enzymes distributed in a diverse range of aerobic organisms, ranging from bacteria to humans (Frova 2003). GSTs are named for their GSH-dependent transferase activity towards toxic compounds. Synonyms include GSH *S*-aryl transferases, GSH *S*-epoxide transferases, GSH *S*-alkene transferases, GSH *S*-alkyl transferases and GSH *S*-alkaryl transferases, a reflection of their remarkable spectrum of substrate recognition (Boyland & Williams 1965; Boyland & Chasseaud 1969).

A wide range of abiotic and biotic factors induces the expression of GSTs. In mammals, the GSTs provide protection against the toxic effect of various exogenous chemicals (Hayes & Pulford 1995). For example, the *in vivo* study of human GST (hGSTP1) in rat demonstrated that the expression of hGSTP1 was easily regulated by 2-*tert*-butyl-4-hydroxyanisole (BHA), a compound that is known as a cancer chemo-preventive agent (Henderson *et al.*, 2014). In plants, high-level expression of GSTs can be observed after exposure to multiple inducers such as herbicides, heavy metals, wounding, pathogen attack, salicylic acids, growth hormones and environmental stresses like cold, dehydration and increased salinity (Sappl *et al.*, 2009; Kiyosue *et al.*, 1993).

Members of the GSTs superfamily are extremely diverse and although GSTs from a variety of species have been characterised, a large fraction of GST sequences deposited in the public databases remain of unknown function (Atkinson *et al.*, 2009). GSTs often functions as dimers either homodimers or heterodimers between different and similar isoenzyme classes (Pettigrew & Colman 2001; Dixon & Edwards 2010). GSTs that are grouped into same class usually have similar substrate recognition, while members of different classes have significantly different general substrate profiles (Hayes *et al.*, 2005; Honaker *et al.*, 2011). In plants, interactions with chloroacetanilide herbicides such as metalachor and herbicide safeners including fenclorim, fluxofenim and dichlormid have been well studied (Wu *et al.* 1999; Gronwald & Plaisance 1998; Dixon *et al.* 1997). More recently, knowledge of the activity of GSTs has increased enormously with catalytic and non-catalytic activities, also known as ligandin activity with diverse compounds in response to both biotic and abiotic stress.

1.3 CLASSIFICATION OF GSTS

The GSTs belong to a superfamily of multifunctional enzymes that evolved from a thioredoxin-like ancestor by multiple gene duplication and subsequent diversification (Hayes & Mclellan 1999).

Widely distributed in all eukaryotic and prokaryotic species that have been analysed, GSTs are categorised into at least four major protein families according to their cellular location; 1) cytosolic, 2) mitochondrial, 3) microsomal and 4) bacterial fosfomycin resistance proteins (Hayes *et al.*, 2005; Armstrong 2000). While representatives from the first three of the categories have been detected in eukaryotes, prokaryotes only contain the cytosolic and bacterial fosfomycin resistance proteins GSTs (Allocati *et al.*, 2009).

1.3.1 Prokaryotic GSTs

Many GSTs from bacteria such as *Proteous mirabilis* (Perito *et al.*, 1996), *Escherichia coli* (Conserved *et al.*, 1994)*, Pseudomonas* (Santos *et al.*, 2002)*, Ochrobactrum anthropi* (Allocati *et al.*, 2008) and *Xylella fastidiosa* (Travensolo *et al.*, 2008) have been studied in detail. Bacterial GSTs in the cytosolic family are classified into five main classes: Beta, Chi, Theta, Zeta and Eta according to sequence identity. The Beta class is the most abundant in bacteria and several representatives from this class have been characterised in *Proteus mirabilis* (Rossjohn *et al.*, 1998) and *O. anthropi* (Allocati *et al.*, 2008).

The Chi class, formerly identified in cyanobacteria, has been found in *Agrobacterium tumefaciens* (Wiktelius & Stenberg 2007; Skopelitou *et al.*, 2012). A distinct GST class, based on its sequence and structural characteristic, named as Eta was also recently found in *A. tumefaciens* (Skopelitou *et al.*, 2012).

Fosfomycin-specific glutathione transferases have been described in *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Staphylococcus aureus*, *Mesorhizobium loti* and *Listeria monocytogenes* (Rigsby *et al.*, 2005). This GST family was named based on the ability to catalyse the conjugation of GSH to the antibiotic fosfomycin, which then causes the inactivation of the antibiotic (Arca *et al.*, 1988).

1.3.2 Eukaryotic GSTs

1.3.2.1 Mammalian GSTs

In mammalian species, at least 16 cytosolic GSTs have been categorised into eight distinct isoenzyme classes: Alpha (A), Mu (M), Pi (P), Omega (O), Sigma (S), Theta (T), Zeta (Z) and

the mitochondrial GST, Kappa (K) (Jakobsson *et al.*, 1999). The Membrane-Associated Protein in Eicosanoids and Glutathione metabolism (MAPEG) class GSTs consist of six proteins, categorised according to their enzymatic activities, sequence motifs and structural properties represented by six proteins: 5-lipoxygenase-activating protein (FLAP), leukotriene C_4 (LTC_4) synthase, microsomal glutathione transferase 1 (MGST1), MGST2, MGST3 and MGST1-like 1 (MGST1-L1) (Jakobsson *et al.*, 2000; Jakobsson *et al.*, 1999).

1.3.2.2 Plant GSTs

Plant GSTs were first recognised in maize (*Zea mays*) where they were found to conjugate the herbicide chloro-*S*-triazine with GSH to protect the plant from the toxicity of the herbicide (Frear & Swanson 1970). In the species studies to date, plant GSTs comprise larger and more complex gene families than the mammalian GSTs (Edwards & Dixon 2005). For instance, there are 25 GST genes in soybean (*Glycine max*), 42 in maize, 54 in Arabidopsis, and 62 in sweet orange (*Citrus sinensis* (L) c.v. Osbeck), 81 in black cottonwood (*Populus trichocarpa*), 99 in Sorghum (*Sorghum bicolor)* and 40 in rice *(Oryza sativa)* (McGonigle *et al.*, 2000; Wagner *et al.*, 2002; Lan *et al.*, 2009; Licciardello *et al.*, 2014; Zhang *et al.*, 2013).

Initially, plant GSTs were divided into four categories (I, II, III and IV) according to sequence identity and conservation of gene structure such as number of introns and exons in the coding region (Droog *et al.*, 1995; Droog 1997). Type I GSTs have three exons and two introns; identified in numerous plants species with herbicide-detoxifying ability (Edwards & Dixon 2005; Conn *et al.*, 2008). Type II GSTs have ten exons and have been identified in wheat (*Triticum aestivum*) and carnation (*Dianthus caryophyllus*) (McGonigle *et al.*, 2000). Type III GSTs contain two exons and a single intron and form the largest family of plant GSTs including many that were originally identified as auxin-regulated proteins (Takahashi *et al.*, 1995; Edwards & Dixon 2005). Type IV was proposed for Arabidopsis genes that have similarity with the mammalian Theta class (Dixon *et al.*, 1998).

This classification method has since been refined and the most recent genomic and phylogenetic analyses have categorised plant GSTs into 14 classes: Phi (F),Tau (U), Zeta (Z), Theta (T), Lambda (L), dehydroascorbate reductase (DHAR), tetrachlorohydroquinone dehalogenase (TCHQD), Metaxin, Hemerythrin, Iota, Ure2p, elongation factor 1B gamma (EF1BG), glutathionyl-hydroquinone reductase (Xi) and microsomal prostaglandin E synthase type 2 (mPGES2). The Hemerythrin, Iota and Ure2p class GSTs are limited to nonvascular plants (Lallement *et al.*, 2014). The Phi class GST was previously known as Type

I GSTs, Tau was the Type III GSTs, Zeta was the Type II GSTs and Theta was the Type IV GSTs (Dixon & Edwards 2010).

1.3.2.3 GSTs in other organism

Following the increasing number of available sequence data, GSTs have been identified, and classified in other phyla. For instance in arthropods, six classes of GSTs have been identified: Zeta (Z), Theta (T), Beta (B), Lambda (L), Delta (D) and Epsilon (E) (Ketterman *et al.*, 2011).

1.4 NOMENCLATURE

The nomenclature for all GSTs is based on the naming system used for mammalian GSTs, *Xx*GSTYn. The *Xx* represents the species Latin name for instance *At* for *Arabidopsis thaliana*, Y is for the single letter code class (for example, L for Lambda and T for Theta) and n is the number of the enzyme within the class. For the two GSTs characterised in this thesis, both found as homodimers, *At*GSTF2 is the second to be named in *Arabidopsis thaliana* Phi class and A*t*GSTU25 is the number 25^{th} to be named in the Tau class *Arabidopsis thaliana* GST.

Since GST isoenzymes of the same classes can form heterodimers, the nomenclature suggested is *XX*GSTYn1-n2 where the n1 and the n2 are the numbers of the two subunits. For example, *Zm*GSTF1-2, denotes a *Zea mays* (maize) Phi class of two subunits: GSTF1 and GSTF2 (Dixon & Edwards 2010).

1.5 THE GST-GSH INTERACTION

The GSTs catalyse a broad range of reactions that involve the addition of GSH to substrate compounds, but their primary role appears to be in the enzymatic detoxification of xenobiotics. Conjugation to the compounds provides cellular protection against free radicals, reactive oxygen species (ROS), endogenous and exogenous electrophiles. Additional GST activities have also been determined (see **Section 1.9**).

The common structural feature of nearly all cytosolic GSTs is the conserved hydrogen bond between hydroxyl group of the catalytic residue of either serine or tyrosine with the thiol group of GSH (Ibarra *et al.* 2003). The basis of this catalytic activity in GSH conjugation lies in its ability to lower the pK_a of the sulfhydryl group (SH) of GSH from 9.0 in aqueous solution to 6.5 at the active site of GSTs. For example in hGST A4-4, spectroscopic pK_a values of tyrosine is 8.1 compared to 10.3 for tyrosine in solution. General base catalysis can be considered wherein the decrease in the pK_a contributes to catalysis by altering the equilibrium of the proton shared between the active site tyrosine and GSH. The tyrosine may extract the GSH proton, because of its higher pK_a, to promote deprotonation of GSH prior attack of the resulting GS^- at the electrophiles (Atkins *et al.* 1993).

This reaction is accomplished by two reactions, either through a substitution reaction (A) or an addition reaction (B) where R is the xenobiotic and X is the leaving group:

$$\text{(A)}\ \text{R-X} + \text{GSH} \xrightarrow{\text{GST}} \text{R-GS} + \text{XH}$$

$$\text{(B)}\ \text{R= R'} + \text{GSH} \xrightarrow{\text{GST}} \text{HR'-R-GS}$$

A well described conjugation activity catalysed by GSTs is the substitution of GSH for the chloro group of 1-chloro-2,4-dinitrobenzene (CDNB), a xenobiotic that is widely used in the dyeing industry (Habig *et al.*, 1974; Booth 2000).

To date, this activity is commonly used as a generic assay to analyse GST conjugation activity in the laboratory (**Figure 1.7**). The GST catalyses the conjugation of GSH to CDNB to produce dinitrophenyl thioether that can be detected by a spectrophotometer at wavelength 340 nm.

However, not all GSTs show catalytic activity with CDNB, and the xenobiotic conjugation activity may not represent the endogenous roles of GSTs, which are not exposed to the

synthetic chemicals *in vivo* (Dixon & Edwards 2010). Other substrates that are commonly used for GST characterisation include cumene hydroperoxide (CHP), 1,2-epoxy-3-nitrophenoxypropane (EPNP), sulfobromophthalein (BSP), ethacrynic acid (EA), *p*-nitrobenzyl chloride (NBC), 1,2-dichloro-4-nitrobenzene (DCNB), and *trans*-4-phenyl-3-butene-2-one (PBO) (**Figure 1.8).**

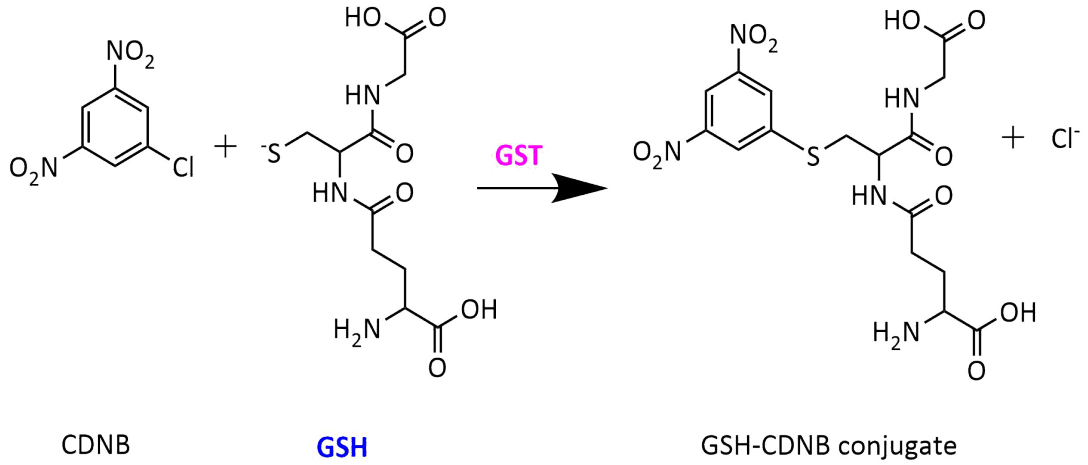


Figure 1.7. Conjugation of GSH to CDNB catalysed by GST.

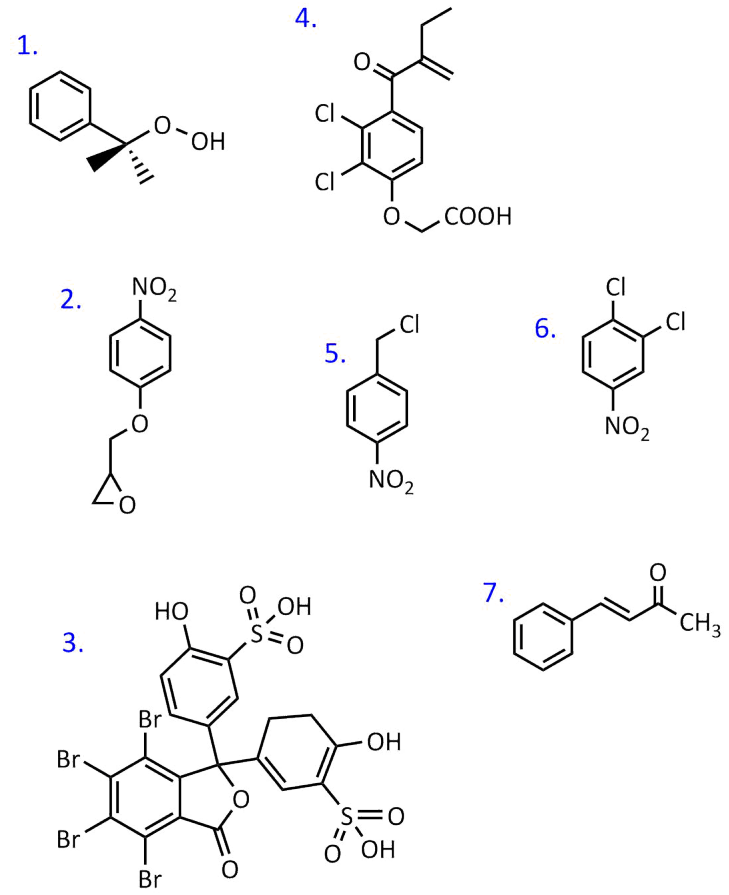


Figure 1.8. Standard substrates used to measure the activity of GSTs. 1: cumene hydroperoxide (CHP); 2: 1,2-epoxy-3-nitrophenoxypropane (EPNP); 3: sulfobromophthalein (BSP); 4: ethacrynic acid (EA), 5: *p*-nitrobenzyl chloride (NBC); 6: 1,2-dichloro-4-nitrobenzene (DCNB); and 7: *trans*-4-phenyl-3-butene-2-one (PBO).

1.6 3D STRUCTURE OF GSTs

Representative structures of cytosolic, mitochondrial and MAPEG GSTs from every class are deposited in the protein structure database Protein Data Bank (www.rcsb.org). The GSTs are often crystallised as a dimer with two subunits of 25-30 kDa consisting of about 200-250 amino acid residues. The N-terminus of each subunit, approximately from residue 1 to 80, provides a binding site for GSH, known as G-site. The C-terminal domain which is relatively more hydrophobic *α*-helical domain is joined to the N-terminus by a short linker sequence. The N-terminal domain adopts a similar topology of thioredoxin also found in disulphide isomerase, glutaredoxin and glutathione peroxidase (Martin *et al.*, 1993; Bushweller *et al.*, 1994; Epp *et al.*, 1983). The thioredoxin domain contains four *β*-sheets strands with three flanking *α*-helices in a *β1-α1-β2-α2-β3-β4-α3* structural motif (Sheehan *et al.*, 2001). The C-terminal domain is located approximately from residues 87-210, usually consists of five to six *α*-helices as the numbers varies for each individual GSTs is more flexible and important in defining the GST isoenzymes specificity by the ability to recognise broad range of hydrophobic substrates (Ji *et al.*, 1992; Reinemer *et al.*, 1991).

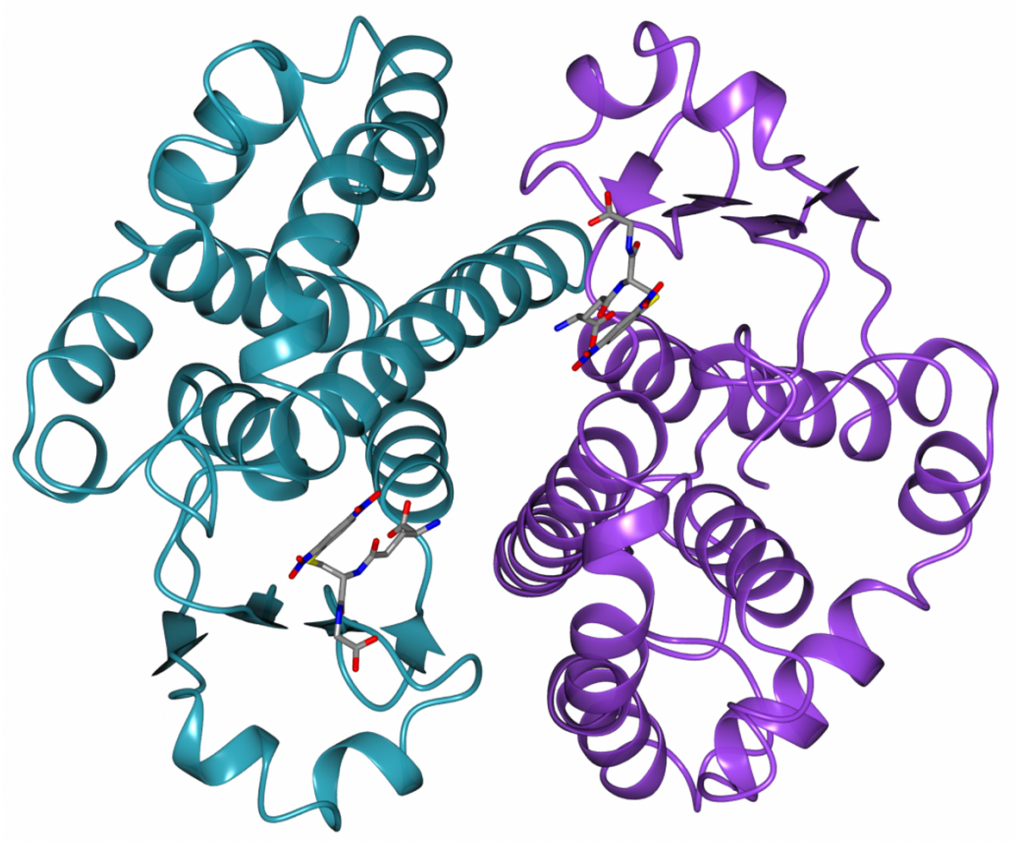


Figure 1.9. Overview of overall tertiary architecture of GST using hGSTP1-1 dimer in complex with GSH-conjugate (1-*S*-glutathionyl-2-4-dinitrobenzene) (PDB ID: 18GS) as model. Model is represented in ribbon configuration with cyan representing subunit A and purple representing subunit B.

1.6.1 Glutathione binding site (G-site)

The GSH binding site (G-site) is conserved in all GSTs throughout all kingdoms, providing the essential catalytic activity of GSTs by the interaction with GSH. The G-site is located at the thioredoxin domain of GST (**Figure 1.10**). The interaction of the G-site with GSH mainly occurs by several hydrogen bonds (H-bond) between the residues on the main chain of GSTs to the peptides of GSH. The H-bond mainly forms through: i) interaction between the active residues of GST, usually either serine to the *γ*-glutamyl residues of GSH and ii) interaction between the conserved active site residues of GST and the sulfur atom of the GSH cysteinyl moiety (Sheehan *et al.*, 2001; Oakley 2011).

The cytosolic GSTs are classified into two major groups by the type of active residue in the G-site that is interacting with the sulfur atom of the GSH, as evidenced from their protein sequence, and shown by a sequence similarity network formulated by Atkinson and Babbitt (2009). The network classified the cytosolic GST based on whether the interaction is *via* a tyrosine, cysteine or serine to the sulfur atom of GSH (Atkinson & Babbitt 2009). The tyrosine group of cytosolic GSTs, although rare, are found mostly in mammalian GSTs including Alpha, Mu, Pi and Sigma class while the serine/ cysteine GSTs are far more populated and present in a broader taxonomic diversity (Atkinson & Babbitt 2009).

1.6.2 Hydrophobic substrate binding site (H-site)

In contrast to the G-site, the H-site is more flexible and provides broad recognition for a range of substrate of different hydrophobicity, shape and size (Prade *et al.* 1998). The H-site is located adjacent to the G-site where the size and shape is determine by hydrophobic residues methionine, tryptophan, phenylalanine and isoleucine. In maize GST, the flexibility lies in the upper part of the H-site, regulated by the two glycine residues that are located next to each other (G123; G124) (Labrou *et al.* 2001). For the GST model used in **Figure 1.10**, the H-site is located in the loop between strand *β*1 and the helix *α*1, helix *α*4 of the thioredoxin domain and the *α*-helical domain at the C-terminal tail (Oakley *et al.* 1999).

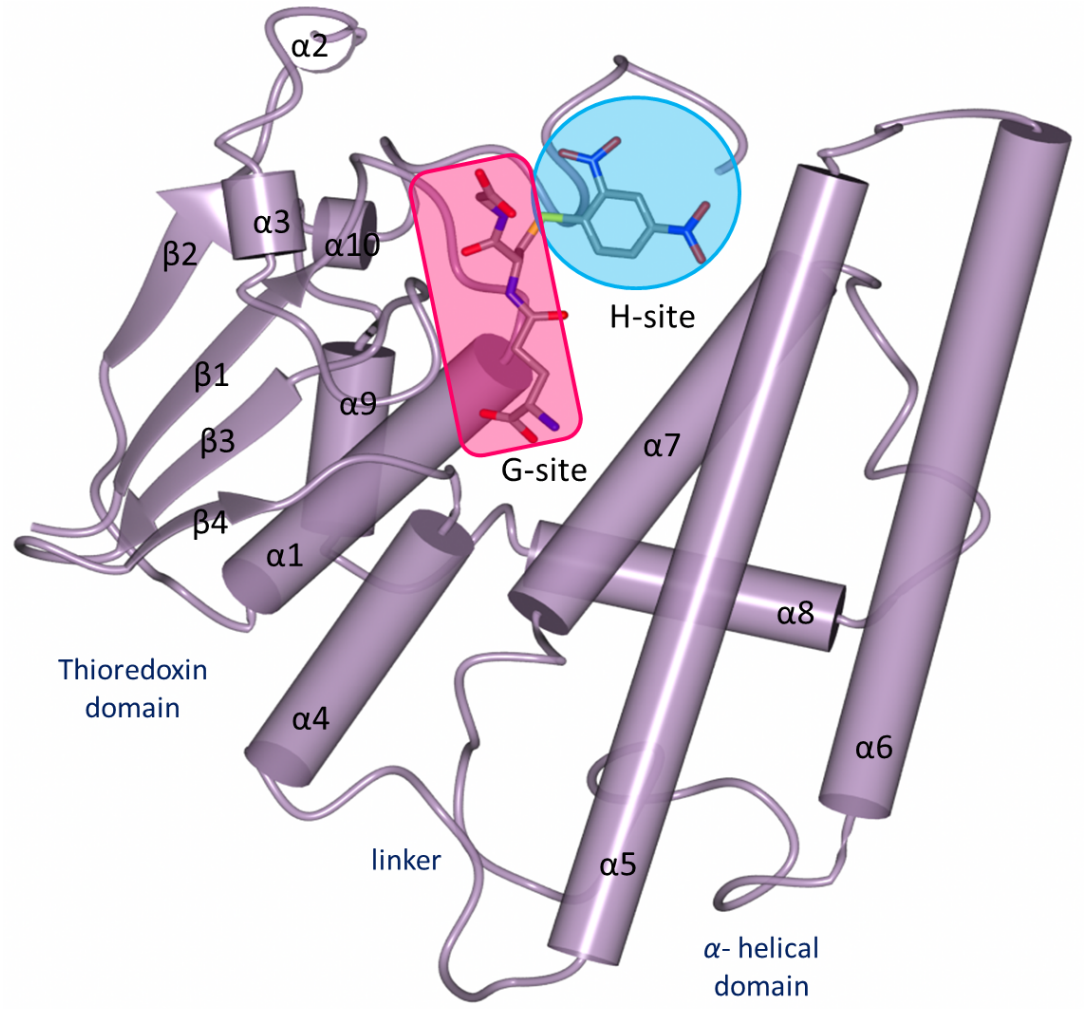


Figure 1.10. The location of G-site and H-site on a GST subunit (PDB ID: 18GS).

1.6.3 Ligandin binding site (L-site)

Besides their widely known catalytic activity in GSH-conjugation, one emerging function of GSTs, a non-catalytic binding property towards a wide range of endogenous and exogenous ligands, has gained considerable interest. This activity is thought to contribute to intracellular sequestration and transport of xenobiotics and hormones. The exact location of the L-site varies, depending on the GST class and the interacting ligands. For example, the L-site identified in Arabidopsis GSTF2 is located next to the G-site , whereas the L-site of human Pi class, hGSTP1-1, with various ligands: sulfasalazine, cibacron blue and bromosulfophthalein are in the H-site (Oakley *et al.* 1999). Binding site at the dimer interface has also been determined for Mu class GST from *Schistosoma japonica Sj*GST in complex with praziquantel (Mctigue *et al.*, 1995). These data clearly shows that

observation of ligandin-binding interactions of one GST cannot be easily extrapolated to other individual GSTs (Oakley 2011).

1.6.4 Structure representation of the GSTs subunits

Cytosolic and mitochondrial GSTs share similarities in their structural fold of the thioredoxin N-terminal domain and α-helical C-terminal domain. One finding indicated that the mitochondrial GSTs contains an additional α-helix in between helix-α2 and strand-β3 in comparison to the cytosolic GSTs (Reinemer *et al.*, 1991). The structure of the Membrane-Associated Protein in Eicosanoids and Glutathione metabolism (MAPEG) class GSTs represents the membrane related protein structure of different topological features compared to the cytosolic GSTs protein structure. The MAPEG GSTs consist of four transmembrane α-helix bundles assembled into trimers (Takusagawa 2013). The overall topological similarities of cytosolic and mitochondrial GST in all organisms and the structure of MAPEG GSTs are shown in **Figure 1.11**.

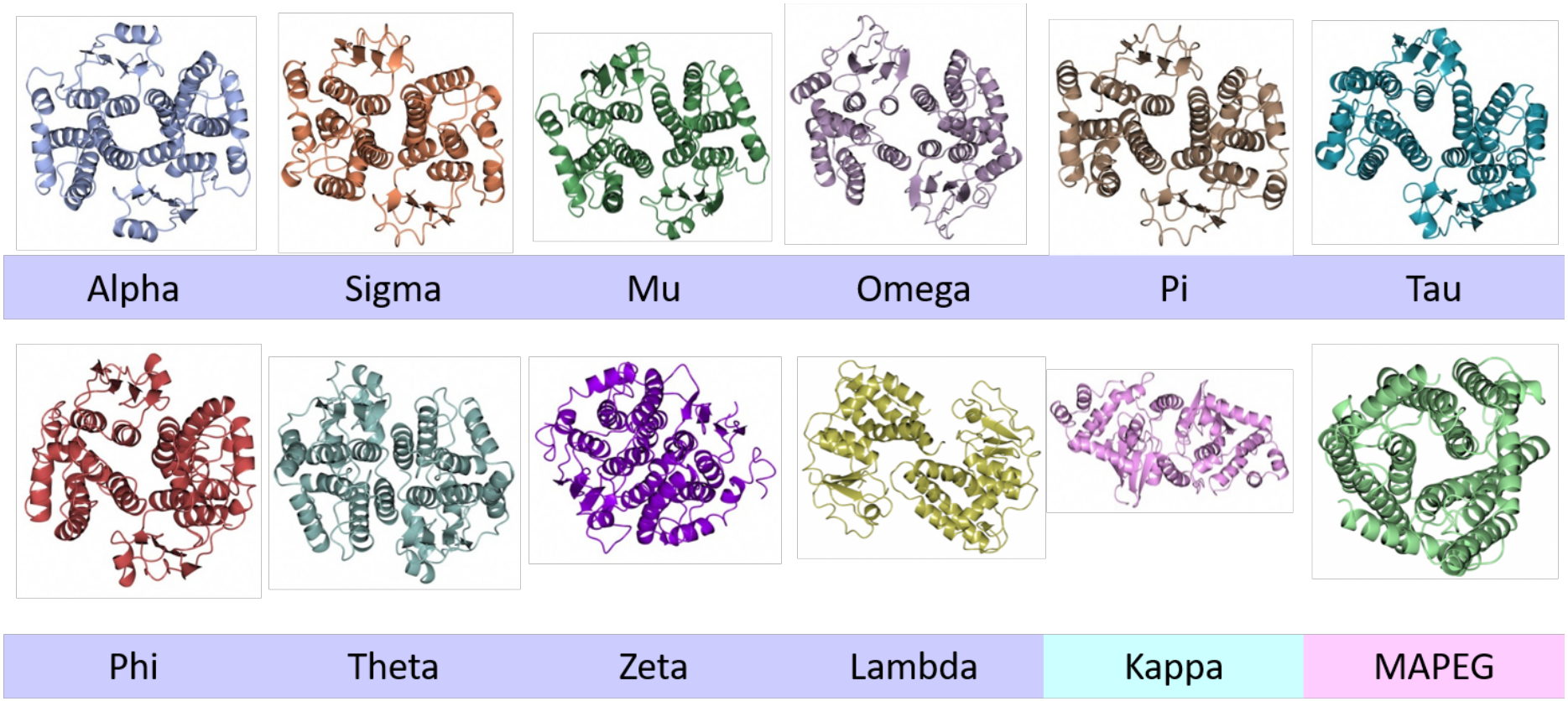


Figure 1.11. Ribbon representation of GST structures obtained from the PDB website. Significant conservation can be observed across all classes. The structures compiled in this figure include PDB ID: 3ZFB (Alpha class GSTs), 4PQH (Lambda class GSTs), 3GUR (Mu class GSTs), 1EEM (Omega class GSTs), 5D73 (Pi class GSTs), 5AGY (Tau class GSTs), 1GNW (Phi class GSTs), 4MPF (Theta class GSTs), 3N5O (Zeta class GSTs), 3VUR (Sigma class GSTs), 3RPN (Kappa class GSTs) and 4AL0 (MAPEG class GSTs). The purple label indicates cytosolic GSTs, blue for mitochondrial GSTs and pink for membrane-related GSTs.

1.7 ARABIDOPSIS GSTs

To date, 54 GSTs have been reported as present in the Arabidopsis genome and these have been categorised into eight classes. The eight classes of Arabidopsis GSTs are: Phi (F), Tau (U), Lambda (L), Theta (T), Zeta (Z), DHAR and TCHQD. A single membrane-related MAPEG GST was also found in Arabidopsis (Dixon & Edwards 2010). Most classes are represented by one to four GSTs, whereas the Tau and Phi classes are represented by a higher number of GSTs: 28 and 13 members of GSTs respectively, most likely resulted from gene duplication events (Frova 2003).

The GST genes are scattered throughout all five chromosomes of Arabidopsis but the location of each genes is often gathered as part of a cluster (**Figure 1.12**). A phylogenetic tree of the GSTs, based on their polypeptide sequences, shows the clustering of the GSTs (**Figure 1.13**) (Dixon & Edwards 2010).

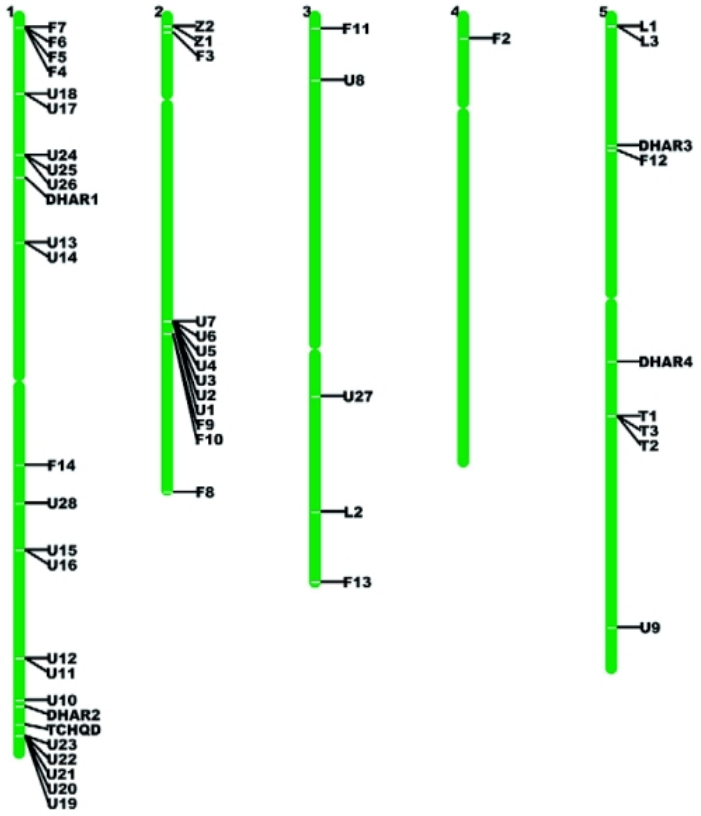


Figure 1.12. Distribution of GST genes in the Arabidopsis genome. The location of each gene is plotted using TAIR chromosome map utility. F: Phi, U: Tau, Z: Zeta, T: Theta, L: Lambda, DHAR: Dehydroascorbate reductase, TCHQD: Tetrachlorohydroquinone dehalogenase. Figure from Dixon & Edwards (2010).

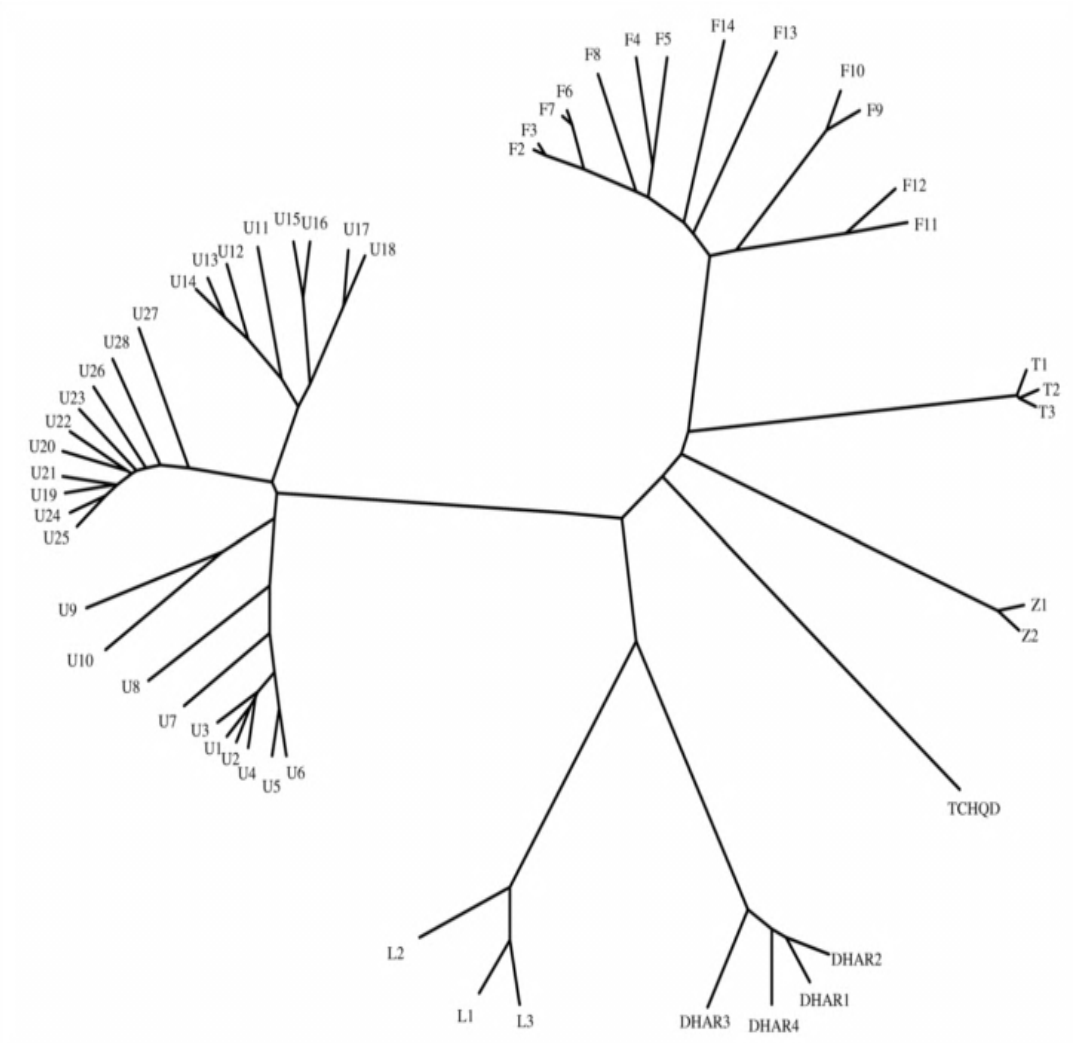


Figure 1.13. Phylogenetic tree of cytosolic GSTs in Arabidopsis. F: Phi, U: Tau, Z: Zeta, T: Theta, L: Lambda, DHAR: Dehydroascorbate reductase. Figure from Dixon & Edwards (2010).

1.7.1 The Phi Class (GSTF)

The Phi (F) GSTs are the most studied among all plant GSTs in terms of their protein structure and biochemical activities. They are the second largest class of Arabidopsis GSTs with members identified as GSTF2 to GSTF14.

Along with the other plant GSTs such as DHAR, Lambda and Tau, the Phi class had been regarded as plant specific until recently, when amino acid sequences from the recently released basidiomycete fungus *Phanerochaete chrysosporium* genome showed 75% similarity with Poplar and Arabidopsis Phi class GST (Morel *et al.*, 2013; Munyampundu *et al.*, 2016).

The GSTFs possess serine as the active residue in the G-site to stabilise the thiolate ion of GSH. Apart from typically assisting GSH conjugation to electrophiles, some members of this class have functions in ethylene and auxin signalling response, along with the ability to transport flavonoid groups (Wangwattana *et al.*, 2008; Smith *et al.*, 2003). Interestingly, Dixon and co-workers also recently found that different small heterocyclic ligands could selectively bind to the Arabidopsis Phi class GSTs, as detected by mass-spectrometry (MS) analysis (Dixon *et al.*, 2011).

1.7.2 The Tau Class (GSTU)

The Tau GSTs is the largest class of Arabidopsis GSTs, comprising 28 members that are sub-divided into three distinct clades on the phylogenetic tree according to their sequence identity (Dixon & Edwards 2010).

The Arabidopsis Tau class GSTs showed clear preference towards various acyl moieties of different chain length and hydroxyl groups. For example, GSTU9, GSTU10 and GSTU13 showed preference for long chain (C16, C18) acyl derivatives. Conversely, GSTU6 and GSTU16 showed preference towards shorter chain length acyl moieties (C8 to C14) (Dixon & Edwards 2009).

Several Tau class GSTs genes were also upregulated in response to the toxic explosive 2,4,6-trinitrotoluene (TNT) (Ekman *et al.*, 2003; Mezzari *et al.*, 2005; Gandia-Herrero *et al.*, 2008). More recently, their involvement in the detoxification of TNT via the formation of GST-TNT conjugates has been demonstrated in Arabidopsis (Gunning *et al.*, 2014).

1.7.3 The Zeta Class (GSTZ)

The GSTZ class of Arabidopsis is unique in that members exhibit dechlorination activity of the dichloroacetic acid (DCA) to glyoxylic acid. The DCA is a potential carcinogenic contaminant of chlorinated drinking water (Tao *et al.*, 2008). The Arabidopsis GSTZ crystal structure has been published in the PDB, with a PDB ID: 1E6B, exhibiting the typical dimeric topology of GST structure (Thom *et al.*, 2001).

Although the Zeta class showed low protein similarity, at about 15 %, with other classes, it carries a typical serine residue at the G-site for interaction with GSH. The mutation of the serine resulted in a lower binding affinity not only to GSH, but also to DCA (Tao *et al.*, 2008).

1.7.4 The Theta Class (GSTT)

The Theta GSTs in Arabidopsis exhibit particularly high GPOX activity towards cumene hydroperoxide and endogenous fatty acid oxidation products, such as linoleic acid hydroperoxide and linolenic acid hydroperoxide (Dixon *et al.*, 2009). The Theta class GSTs have serine as the active residue for the interaction with GSH.

1.7.5 The Lambda Class (GSTL)

The Lambda GSTs are active as monomers and are found in the cytosol and the chloroplast (Dixon *et al.*, 2002). The Lambda class is represented by three members, all have cysteine as the active residue at the G-site and have the ability to function as thiol transferases (Dixon *et al.*, 2002). The GSTs of this class showed high affinity towards flavonoids in a GSH-dependent manner which is crucial for flavonoid recycling in the cells (Dixon & Edwards 2010).

1.7.6 The Dehydroascorbate Reductase Class (DHAR)

The DHAR class GSTs is known to catalyse the reduction of dehydroascorbate to ascorbate by the conjugation activity of GSH (Urano *et al.*, 2000). Like the Lambda GSTs, the G-site of the DHAR class GSTs contains cysteine as the active residue. While the DHAR GSTs lack detectable GSH conjugating activity towards CDNB, they exhibit significant thiol transferase activity, similar to the Lambda GSTs (Dixon *et al.*, 2002).

In Arabidopsis, four transcribed genes that encode functional DHAR GSTs and one pseudogene have been found. They are expressed as monomers and often found in the chloroplast proteome (Zybailov *et al.*, 2008; Dixon *et al.*, 2002).

1.7.7 The Tetrachlorohydroquinone Dehalogenase Class (TCHQD)

There is only one TCHQD GST in Arabidopsis and not much is yet known about it apart from its location in the plasma membrane (Dixon *et al.*, 2009). The TCHQD GST can catalyse GSH conjugation, and contains serine in the G-site (Dixon & Edwards 2010).

1.7.8 The Membrane-Associated Protein in Eicosanoids and Glutathione metabolism Class (MAPEG)

So far only one transmembrane microsomal MAPEG GST (*At*1g65820) has been identified in Arabidopsis. Preferred substrates of this enzyme are hydrophobic such as halogenated hydrocarbons or phospholipid hydroperoxides which are in accordance with the membrane location of this enzyme (Dekant *et al.*, 1990).

1.8 ROLES OF PLANT GSTS

In plants, much of the early work on GSTs focused on their important role in herbicide detoxification, by the conjugation activity of GSH (Edwards and Dixon 2002). Proven experimentally, the GSTs showed additional abilities including GSH-dependent peroxidase activity (GPOX), GSH-dependent thiol transferase activity, GSH-dependent *cis-trans* isomerisation activity and GSH-dependent hydrolytic dehalogenation. The variations in catalytic efficiency of the isoenzymes towards different substrates or types of reactions are perhaps the most crucial reason to understand the biological importance of GSTs.

1.8.1 GSH-dependent peroxidase (GPOX) activity

The GSH-dependent GPOX activity reduces organic hydroperoxides to their less toxic alcohol forms, leading to the reduction of cytotoxicity level in the cell. The Theta class of Arabidopsis GSTs was one of the earliest classes found to have GPOX activity. The Theta class can catalyse the conversion of 13-hydroperoxy-9,11,15-octadecatrienoic acid and 13-hydroperoxy-9,11-octadecadienoic acid to the corresponding hydroxyl derivatives, together with the formation of disulphide glutathione (GSSG) (Bartling *et al.*, 1993). The GPOX activity of GSTs on cumene hydroperoxide is illustrated in **Figure 1.14**.

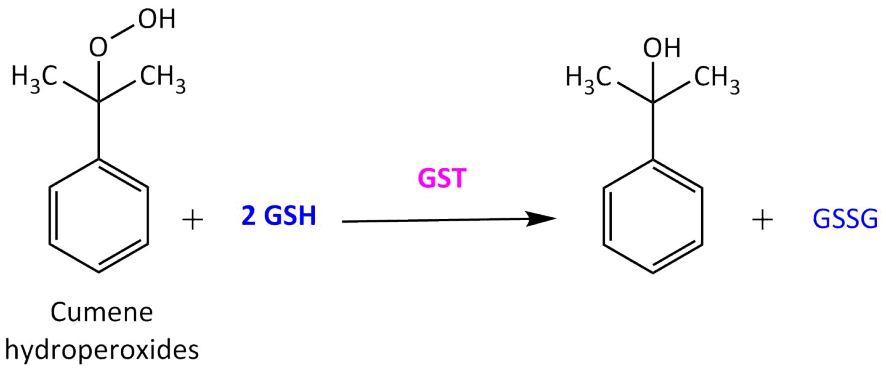


Figure 1.14. The GPOX activity of GST towards cumene hydroperoxide in the presence of GSH. Figure from Dixon & Edwards (2010).

1.8.2 GSH-dependent thiol transferase activity

The GSTs with cysteine as the active residue can activate thiol transferase activity in a similar fashion to glutaredoxins when assayed with substrates such as β-hydroxyethyl disulphide (HED) (**Figure 1.15**). In Arabidopsis, DHAR and Lambda class GSTs exhibit thiol transferase activity, in addition to their ability to assist GSH-conjugating activity. The thiol transferase activity however, does not involve GSH conjugating to the substrate but rather two molecules of GSH are oxidised to form disulphide GSH (GSSG).

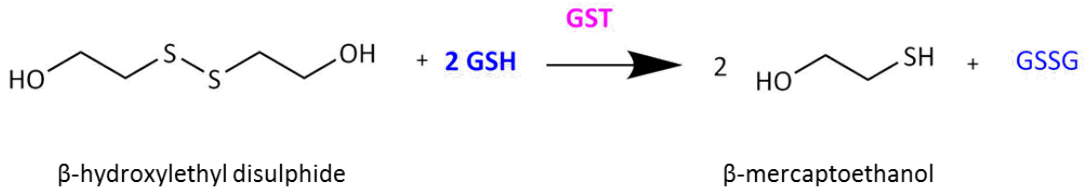


Figure 1.15. Thiol transferase assay using HED catalysed by GST in presence of GSH. Figure from Dixon & Edwards (2010).

1.8.3 GSH-dependent isomerisation activity

The isomerisation activity of GST has been demonstrated in the enzymatic conversion of the herbicide thiadiazolidines I to triazolidines II. The reaction product, triazolidine II is a potent inhibitor of protoporphyrinogen oxidase, an enzyme that produces protoporphyrin IX that is an important precursor for chlorophyll (Jablonkai & Ko 1999; Hao *et al.*, 2011; Edwards *et al.*, 2000).

The isomerisation reaction involves nucleophilic attack at the carbonyl group of the thiadiazolidines I by GSH, catalysed by GST (**Figure 1.16**). The attack causes ring opening which allows rotation around the N-bond and the C=N double bond transferred to the C-S group. The elimination of GSH occurred when the nitrogen atom attacks the carbonyl group to reform the ring (Edwards *et al.*, 2000; Jablonkai & Ko 1999). This activity is predominantly found in members of the Zeta class of GSTs.

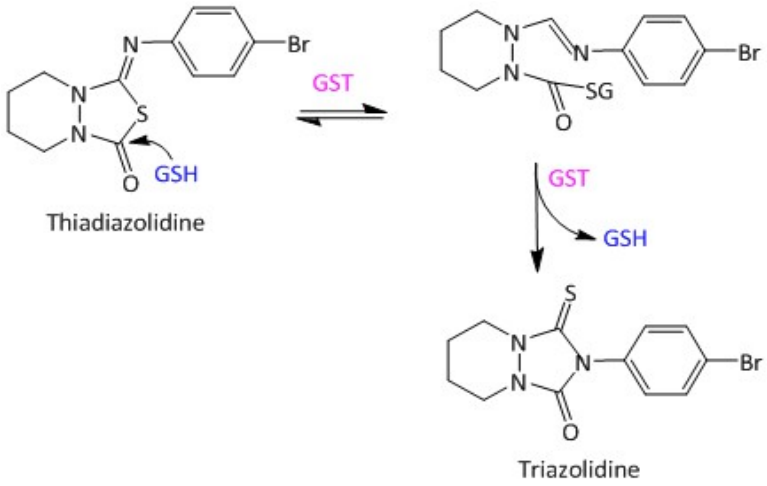


Figure 1.16. The proposed reaction mechanism of isomerisation of thiadiazolidine proherbicide to triazolidine catalysed by Zeta class GST. Figure from Edwards *et al.* (2000).

1.8.4 GSH-dependent hydrolytic dehalogenation

Arabidopsis Zeta class GSTs, particularly GSTZ1, have been shown to catalyse GSH dependent dechlorination of dichloroacetic acid (DCA) to glycoxylic acid (Dixon *et al.*, 2000). In the DCA dehalogenation reaction, GSTZ1 assists the GSH to replace two chlorine substituents with hydroxyl groups (**Figure 1.17**). This mechanism does not consume GSH, which is eliminated at the end of the reaction.

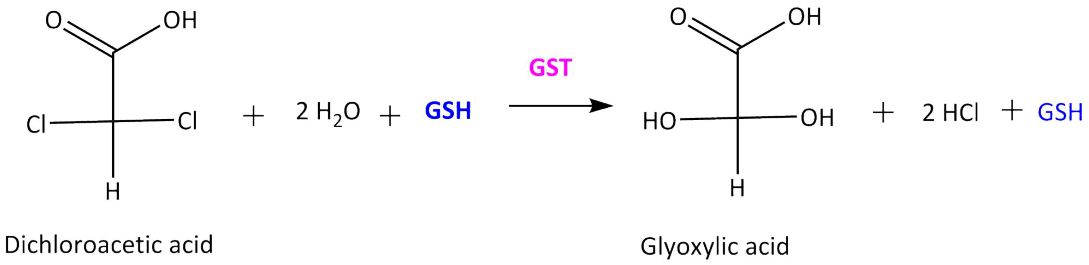


Figure 1.17. Dehalogenation of dichloroacetic acid to glyoxylic acid catalysed by GSTZ1 with activation of GSH. Figure from Dixon *et al.* (2000).

1.9 NON-CATALYTIC ACTIVITY OF GSTS

In addition to their catalytic roles, GSTs act as ligand-binding proteins of various non-substrate hydrophobic molecules including azo-dyes, bilirubin, hemin, fatty acids, bile salts, steroids and thyroid (Litwack *et al.*, 1971; Habig *et al.*, 1974). This property is known as ligandin activity, and may be important for transport of the ligand within different compartments in the cell and for storage in the vacuole (Oakley 1999).

In plants, the maize recombinant proteins, *Zm*GSTF1, *Zm*GSTF2 and *Zm*GSTF3, bind to porphyrins including mesoporphyrin, coproporphyrin, uroporphyrin and magnesium (Mg)-protoporphyrin. The binding of these GSTs to the porphyrin derivatives did not involve the formation of GSH-conjugates but rather a non-covalent binding that inhibits GSTs activity in GSH-CDNB conjugating assays (Lederer & Böger 2003). Purified *Zm*GSTU1 and *Zm*GSTU2 were detected with bound porphyrin precursors identified as uroporphyrin and coproporphyrin, confirmed by HPLC analysis (Dixon *et al.*, 2008).

The GST ligandin activity in Arabidopsis has been recorded in Phi and Tau class GSTs with ligands such as labelling dyes, flavonoids, auxin, ethylene, phytoreceptors, phytohormones and fatty acid derivatives as listed in **Table 1.1**. From a structural perspective, the ligandin site (L-site) was reported in GSTF2 structure in complex with two S-hexylglutathione molecules with little characterisation on the binding activity towards the ligand. The L-site was shown to be closed to G-site and occupied by the *γ*-glutamate and glycine moiety of one of the S-hexylglutathione molecules (see **Section 3.1.3**) (Reinemer *et al.*, 1996).

Table 1.1. The list of identified Arabidopsis GSTs with ligandin activity in plants.

GST	Ligand	Binding activity	Source
GSTF2	Ethylene and auxin	GSTF2 directly binds to molecules in ethylene and auxin pathway and flavonoids.	Smith *et al.*, 2003
	Analog of IAA	GSTF2 binds directly to the labelling molecule.	Zettl *et al.*, 1994
	Phytoalexins and flavonoids	Binding of small heterocyclic compounds from Arabidopsis extracts and Luria – Bertani (LB) rich media.	Dixon *et al.*, 2011
GSTF12 (TT19)	Anthocyanin	Transport anthocyanins from cytosol to tonoplast.	Sun *et al.*, 2012; Kitamura *et al.*, 2004
GSTU1	Indole-3-acetic acid (IAA)	GSTU1 bind directly to IAA and eluted from GSH-immobilised agarose column by addition of IAA	Watahiki *et al.*, 1995
GSTU17	Phytoreceptors phytochrome (*phyA)* and multiple phytohormones	GSTU17 is a responsive gene after induction by far-red (FR) light irradiation in gene expression study. The gene expression was inhibited by mutation of *phyA* gene.	(Chen *et al.*, 2012; Tepperman *et al.*, 2001; Jiang *et al.*, 2010)
GSTU19	Fatty acid derivatives	Binding of fatty acid, which are varied in chain length (C_6 to C_{18}) and later identified as oxophytodienoic acid and oxygenated fatty acid.	(Dixon & Edwards 2009)
GSTU20	Far-Red Insensitive 219 (FIN219)	Regulate FIN219 and phyA light signalling pathway, which are crucial in cell elongation and plant development. GSTU20 was found to bind at C-terminal of FIN219.	(Chen *et al.*, 2007)

1.10 GST INVOLVEMENT IN THE DETOXIFICATION OF 2,4,6-TRINITROTOLUENE (TNT)

The nitroaromatic explosive 2,4,6-trinitrotoluene (TNT) is an organic pollutant and highly resistant to biodegradation in the environment. High levels of contamination occur at military sites and manufacturing facilities (Rylott *et al.*, 2015; Rylott *et al.*, 2011). Remediation of the contaminated sites using conventional methods such as excavation and burial in landfill or incineration is costly, only applicable to relatively small areas of land and, in the case of incineration, it destroys the microenvironment of the soil (Peuke & Rennenberg 2005). Hence, an alternative method, known as phytoremediation, which utilises plants for removal, degradation or containment of contaminants in soils, is generating significant interest.

Plant species including tobacco (*Nicotiana tabacum)*, bean (*Phaseolus vulgaris)*, wheat (*Triticum aestivum)*, poplar (*Populus* spp.), and switchgrass (*Panicum virgatum*) have been tested to analyse the uptake of TNT, which has been shown to accumulate in the roots (Sens *et al.*, 1998; Hannink *et al.*, 2007; Dillewijn *et al.*, 2008; Brentner *et al.*, 2010). Gene expression studies including the serial analysis of gene expression (SAGE) (Ekman *et al.*, 2003), transcriptomic profiling using reverse transcription (RT)-PCR (Mezzari *et al.*, 2005) and microarray analysis (Gandia-Herrero *et al.*, 2008) all using Arabidopsis as the model plant treated with TNT revealed that GSTs are among the genes responsive towards the TNT treatment. The GSTs identified include GSTU1, GSTU22 and GSTU24 (Ekman *et al.*, 2003), GSTU3, GSTU4, GSTU24, GSTU25 (Gandia-Herrero *et al.*, 2008), GSTU24, GSTF2 and GSTU1 (Mezzari *et al.*, 2005).

Following the microarray study by Gandia-Herrero *et al.* (2008), quantitative PCR analysis further verified the TNT-responsive up-regulation of GSTU24 and GSTU25 by 252-fold and 177-fold respectively with activity in GSH-TNT conjugating assays recorded at K_m: 1.6 mM and 1.2 mM respectively. Together, GSTU24 and GSTU25 produced three TNT-GSH conjugates, known as C-glutathionylated 4-HADNT (conjugate **1**), C-glutathionylated 2-HADNT (conjugate **2**) and 2-glutathionyl-4,6-dinitrotoluene (conjugate **3**).

Conjugate **3** is particularly of interest due to the denitration of one of the nitro groups. It is the electron-withdrawing properties of the three nitro groups of TNT that make the aromatic ring particularly resistant to oxidative attack and ring cleavage. Loss of a nitro group would reduce the stability of the TNT ring and thus conjugate **3** may be potentially more amenable to subsequent biodegradation and mineralisation in the environment.

Using Arabidopsis lines overexpressing *GSTU24* and *GSTU25* in hydroponic and soil based experiments, significant uptake of TNT from the TNT-treated liquid medium and soil in comparison to the unmodified plants was observed. In addition, the overexpressing lines also produced significantly larger root-surface areas and increased biomass in TNT-treated media and soil respectively suggesting a significant role of plant GSTs in TNT detoxification.

1.11 THESIS OBJECTIVES

It is clear that GSTs have multiple roles in the cell involving both catalytic and non-catalytic mechanisms. The aim of this thesis is to use X-ray crystallography, thermodynamic analysis and molecular biological methods such as site-directed mutagenesis and CRISPR/Cas9 technology to further elucidate the roles GSTs play in Arabidopsis metabolism, as a model plant. Collectively, the objectives of this study are to:

1. investigate the ligandin binding site of GSTF2,
2. elucidate the protein structure of GSTU25, and
3. uncover potential contributions of *GSTU25* and related genes within the same clade on the phylogenetic tree including *GSTU24*, *GSTU21*, and *GSTU19* in TNT detoxification.

2 General material and methods

This chapter describes the general material and methods used throughout this project. Specific method for each experiments is in the relevant results chapters; Chapter 3, 4 and 5.

2.1 MATERIALS

Unless stated otherwise, all chemicals used in this work were of analytical grade from Sigma Aldrich Company Ltd. (Poole, UK).

Oligonucleotide primers were synthesized and purchased from Integrated DNA Technology (IDT) (Interleuvenlaan, Belgium). Protein gel markers were obtained from New England BioLabs Ltd. (NEB) (Herts, UK) and Promega (Southampton, UK). DNA polymerases and restriction enzymes were purchased from NEB (Herts, UK), Promega (Southampton, UK) and Invitrogen (Paisley, UK).

All buffers were formulated in ultrapure water (18.2 MΩ cm^{-1}) followed by filtration and degassing. All growth media was autoclaved prior to use.

The explosive TNT (>95% purity) was kindly provided by the Defence Science and Technology Laboratory (DSTL), Fort Halstead, UK.

2.2 INSTRUMENTS

Polymerase chain reactions (PCR) were performed using Primer Thermal Cycler (Techne, UK).

Protein purification was performed using ÄKTA-FPLC system and ÄKTA start system, GE Healthcare (Little Chalfont, UK).

Spectrophotometric assays were measured using a Varian Cary® 50 UV-Vis, Agilent Technologies (Cheshire, UK).

Isothermal titration calorimetry was performed using a MicroCal VP-ITC system (Malvern Instrument Limited, UK).

2.3 PLASMIDS, BACTERIA AND GROWTH CONDITION

2.3.1 Plasmid

The plasmids used for gene cloning and enzyme expression are listed in **Table 2.1.**

Table 2.1. Plasmids used for protein expression and knockout study. Asterisks (*) represent plasmid produced by Xing *et al.* (2014) and the plasmid map are provided in the supplementary section.

Plasmid	Antibiotic resistance	Antibiotic concentration (μg mL^{-1})	Source
pET-24a	Kanamycin	50	Novagen, UK
pET-22b	Kanamycin	50	Novagen, UK
pDT1DT2*	Chloramphenicol	25	Addgene (MA,USA)
pDT2DT3*	Chloramphenicol	25	Addgene (MA,USA)
pDT3DT4*	Chloramphenicol	25	Addgene (MA,USA)
pHSE401*	Kanamycin	50	Addgene (MA,USA)

pET vector system

The pET vectors are short, circular plasmid containing: 1) a gene coding for antibiotic resistance, 2) a *LacI* gene that code for the lac repressor, and 3) an insertion site for the gene of interest downstream of T7 promoter DNA sequence, lac operator DNA sequence and the ribosome binding site. The system uses T7 promoter that is highly specific to T7 RNA polymerase that has a high translation efficiency. The T7 RNA polymerase is engineered into many commercially available *E. coli* strains by the modification of the lac operon system. Compare to the native *E. coli*, the RNA polymerase bind to a lac promoter sequence instead of a T7 promoter sequence in front of the lac operator sequence in the commercially available *E. coli* strains.

Protein expression is induced by isopropyl-*β*-D-1-thiogalactopyranoside (IPTG), a molecule that is structurally identical to lactose, that can bind to the lac repressor (*LacI*) and induces the expression of protein. Once IPTG binds to *LacI*, it induces a conformational change in the protein structure that inhibit the binding to the operator DNA sequence. Once the binding is inhibited, native *E. coli* RNA polymerase begins transcribing the T7 RNA polymerase in high number. The expressed T7 RNA polymerase protein will bind to the T7 promoter sequence upstream of the gene of interest on the transformed pET plasmid and transcribe the target gene.

The schematic form of host and vector elements for control of T7 RNA polymerase levels and the subsequent transcription of a target gene in a pET vector is illustrated in **Figure 2.1**.

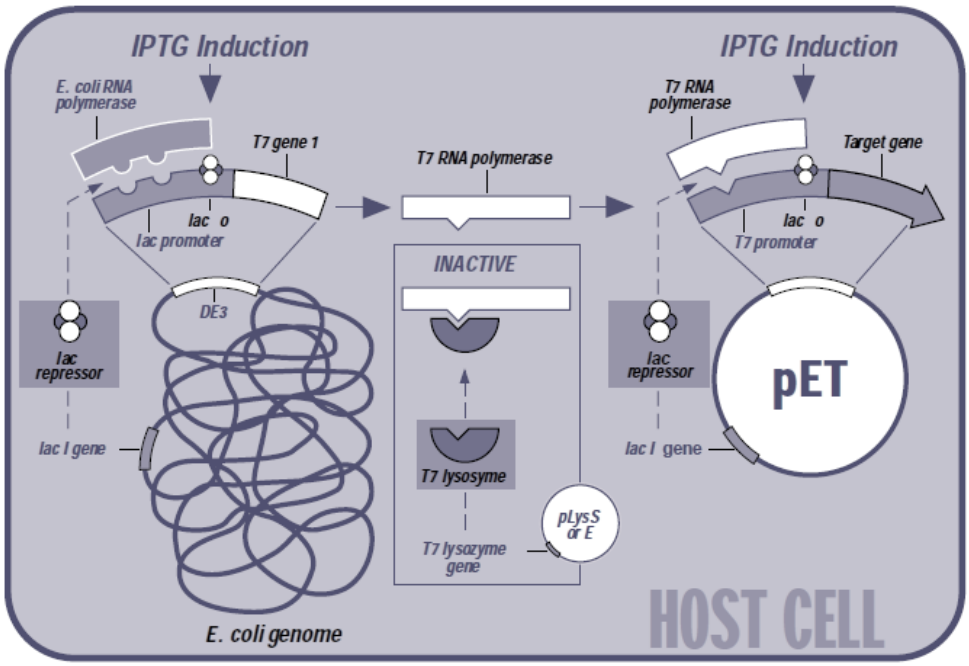


Figure 2.1. Important elements of the pET system during IPTG induction in the host cell. Figure from Novagen.com.

2.3.2 Bacterial strain

The bacterial strains used in this study are listed in **Table 2.2.**

Table 2.2. Bacterial strains used for protein expression and Arabidopsis transformation.

Bacteria	Strain	Resistance	Purpose	Source
Escherichia coli	Tunetta	Kanamycin Chloramphenicol	Expression host	Edward group stock
Agrobacterium tumefaciens	GV3101	Gentamycin (50 µL mL^{-1}) Spectinomycin (25 µL mL^{-1})	Transformation of Arabidopsis	Bruce group stock

2.4 MOLECULAR BIOLOGY TECHNIQUES

2.4.1 Polymerase chain reaction (PCR)

A reaction mix containing 1x Phusion DNA polymerase buffer (supplied by the manufacturer), 1 µL template DNA, 500 nM forward primer, 500 nM reverse primer, 0.004 U Phusion DNA polymerase and 200 µM dNTPs were prepared in a total volume of 20 µL. Typically, samples were denatured at 95°C for 30 s and then exposed to 33 heating cycles of: 95°C for 10 s, 55°C for 30 s, and 72°C for 30 s kb^{-1}. This was followed by a final extension step at 72°C for 7 min.

2.4.2 Separation of DNA using agarose gel electrophoresis

The DNA fragments were separated according to their molecular weight on an agarose gel. The 0.8% (w/v) agarose gels were prepared by dissolving 0.8% (w/v) agarose in TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8.0). The solution was cooled to room temperature prior to the addition of 0.6 µM of ethidium bromide (Sigma Aldrich, Poole, UK). The solution was poured into a preassembled cast and allowed to set. The comb was carefully removed to reveal the wells once the gel was set and the gel is placed in a tank filled with TAE buffer. The DNA sample was mixed with an appropriate volume of 6x loading buffer (10 mM Tris-HCl pH 7.6, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, 60 mM EDTA). Samples were run alongside 5 µL DNA 1 kb Hyperladder (Bioline, UK) to allow the estimation of the nucleotide size. Samples were run at 110 V for 40 min. Visualisation of ethidium bromide stained DNA was achieved through exposure to UV light.

2.4.3 Purification of DNA from agarose gel

The DNA bands of interest were excised from 0.8% agarose gel and were purified using Wizard® SV Gel and PCR Clean-Up System (Promega, Southampton, UK) or the QIAquick (Qiagen West Sussex, UK) according to the manufacturer's instruction. The elution was done with 50 µL sterile nuclease free H_2O which was preheated at 60°C.

2.4.4 Purification of plasmid DNA

Purification of plasmid DNA for sequencing and cloning was carried out using QIAprep Spin Miniprep Kit (Qiagen West Sussex, UK) according to the manufacturer's protocol. The concentration of the DNA was measured at absorbance at 260 nm and purity level of the DNA was determined at value $A_{260}/_{280}$ using a NanoDrop spectrophotometer (ThermoFisher Scientific, UK).

2.4.5 DNA restriction digests

A reaction mix prepared for DNA digestion contained 7 µL of DNA sample (100 ng), 5 U of restriction enzyme, 1 x restriction buffer (recommended by manufacturer) in a total volume of 10 µL. Samples were incubated for 60 min at room temperature as recommended by the manufacturer for the specific restriction enzymes being used.

2.4.6 Dephosphorylation of the linearized plasmid

In case when the plasmid is single-digested, the purified plasmid was dephosphorylated in a reaction volume of 50 µL. The reaction was carried out in SimpliAmp Thermal Cycler

(Applied Biosystem, UK) at 37°C for 60 min, followed by enzyme deactivation at 65°C for 5 min.

2.4.7 DNA ligation reactions

A reaction mix containing 40 U T4 DNA ligase, 1x T4 DNA ligase buffer (Promega, Southampton, UK), 1 µL linearized vector (50 ng) and 7 µL insert DNA (30 ng) was prepared and incubated at 22°C for 60 min.

2.4.8 Bacterial media

Luria Bertani (LB) broth was prepared as the following formulation: 1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast granulated extract. LB-agar was prepared as above with the addition of 1.5% (w/v) agar.

2.4.9 *E. coli* host transformation

Unless stated otherwise, a mix of 2 µL ligated plasmid DNA (80 ng) and 50 µL of the desired bacterial host strain cells were incubated on ice for 10 min. Cells were heated at 42°C for 30 s followed by rapid cooling on ice for 2 min. Following heat treatment, 100 µL of warmed LB was added and the sample was shaken at 200 rpm for 60 min at 37°C. After shaking, 70 µL of transformed cells were inoculated on LB agar containing relevant antibiotic and incubated overnight at 37°C. For cloning application, the plasmid DNA was transformed into XL-Gold Ultracompetent cells (Agilent Technologies, Cheshire, UK). For recombinant protein expression, plasmid DNA was transformed into *E. coli* Tuner (DE3) cells that also contained the pRARE plasmid derived from *E. coli* Rosetta, subsequently termed *E. coli* Tunetta cells (Taylor 2012; Zhao & Dixon 2010).

2.4.10 DNA Sanger sequencing

DNA sequencing was performed by GATC Biotech (Germany). The sequence profile was analysed using BLASTn and DNA alignments were generated using ClustalX Omega.

2.5 PROTEIN EXPRESSION AND PURIFICATION

2.5.1 Cell lysis by sonication

Cell pellets were resuspended to 1 g mL^{-1} with 20 mM Tris HCl pH 7.4 and 0.1 mM DTT. Sonication was carried out by ultrasonication for 3 x 30 s bursts at 4°C with 1 min intervals using Vibra Cell VC375 (Cell™ Sonic Materials). The soluble and insoluble material fractions

separated by centrifugation at 15,000 rpm for 30 min using Jouan CR312 centrifuge. Supernatants were filtered through 0.45 µm syringe filters prior protein purification.

2.5.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE experiments were performed using polyacrylamide gel system and sodium dodecyl sulfate (SDS) to denature the proteins, with Mini-Protean Tetra cell apparatus (Bio-Rad, USA). The resolving gel was a composition of 10 mL of 1.5 M, Tris-HCl pH 9, 0.4% (v/v) tetramethylethylenediamine (TEMED), 0.4% (w/v) SDS, 4.2 mL H_2O and 3.2 mL 40% acrylamide/bis acrylamide. The gel solution was de-gassed and acrylamide polymerisation was induced by the addition of 0.1 mL 10% (w/v) ammonium persulfate. The solution was quickly transferred into a pre-assembled gel apparatus and allowed to solidify. The stacking gel contained 5 mL stacking buffer (0.14 M Tris-HCl pH 6.8, 0.11% (v/v) TEMED, 0.11% (w/v) SDS) and 0.5 mL 40% acrylamide/bis acrylamide. The gel solution was de-gassed and 0.05 mL 10% (w/v) ammonium persulfate was added to aid gel polymerization. The solution was quickly transferred on top of the resolving solution in the pre-assembled gel apparatus, the well comb added and the gel was allowed to solidify. Protein samples were mixed with appropriate volume of 2X loading buffer composed of 100 mM Tris-HCl pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 0.2 M dithiothreitol (DTT), 0.2% (w/v) bromophenol blue, boiled at 95°C for 5 min and loaded into assembled tank filled with running buffer (25 mM Tris pH 8.3, 192 mM glycine, 0.1% (w/v) SDS). Protein samples were run alongside 10 µL pre-stained broad range protein marker to allow estimation of protein weights. Once the samples were loaded, gels were run at 100 V as samples moved through stacking gel and 200 V hereafter, until the marker eluted from the gel. To visualize the protein bands, gel was washed twice with H_2O and then stained with Instant-Blue dye reagent (Expedeon Inc., USA) according to the manufacturer's recommendation.

2.5.3 Recombinant protein quantification

The concentration of purified recombinant GSTs was determined by UV-vis spectrophotometry in 0.1 mL cuvette with a Varian Carry® 50 Bio UV-Vis Spectrophotometer. The absorbance at 280 nm of 10 µL protein sample was determined using the estimated extinction coefficient of the recombinant protein (ProtParam, ExPaSy web program, Swiss Institute of Bioinformatics).

2.6 PLANT WORK

2.6.1 Seed sterilisation

Seeds were sterilised by chlorine gas generated by the addition of 3 mL concentrated hydrochloric acid in 100 mL bleach in an airtight container. The seeds were incubated with the chlorine gas for 4 h. After sterilisation, the lid of the container was opened in a flow hood for 10 min to release residual chlorine gas.

2.6.2 Seed stratification

Seeds were applied to ½ MS agar or damp soil and were imbibed in the dark (covered with foil) at 4°C for a minimum 72 h.

2.6.3 Growth room conditions

The growth room for seeds grown on ½ MS agar plates had low lighting (20 $\mu mol.m^{-2}.s^{-1}$), with a 16 h light, 8 h dark cycle at 22°C.

2.6.4 Growth conditions for soil

Soil used in this experiment was Levington's F2 compost. Non-sterile seeds were evenly spread on top of pots filled with F2 compost treated with active substance: imidacloprid and stratified (**Section 2.6.2**). Plants were allowed to propagate in the greenhouse. Plants for the purpose of floral dipping were grown in 3 inch pots and were weeded down to 10-15 plants per pot.

2.6.5 Genomic DNA isolation from plants

Plant tissue usually large leaf (3 to 4 cm) was ground with pestle within a 1.5 mL Eppendorf tube with 500 µL 2x CTAB buffer (2% cetyl trimethylamin bromide, 1.4 M NaCl, 100 mM Tris-HCl pH 8, 20 mM Na_2EDTA) and incubated at 65°C for 1 h. The sample was vortexed with 300 µL of chloroform: iso-amyl-alcohol and centrifuged at 13,000 rpm at 10 min. About 300 µL of the aqueous layer was transferred to a new 1.5 mL Eppendorf tube containing 960 µL ethanol and 40 µL of 3 M sodium acetate. The sample was mixed and were precipitated at 23°C for 40 min and pelleted by centrifugation at 13,000 rpm for 5 min at 4°C. The supernatant was discarded and the pellet was rinsed in 70% ethanol prior another centrifugation for 5 min. The ethanol was removed with pipette and the pellet was dried in a Savant DNA Speed-Vac, high temperature setting for 10 min to remove any residual ethanol. The pellet was resuspended in 50-100 µL of H_2O.

3 Evidence of non-catalytic binding of GSTF2

The aim of the work in this chapter was to characterize the non-catalytic binding of small ligands including indole-3-aldehyde, camalexin, quercetrin and quercetin which were selected from the study carried out by Dixon *et al.* (2011), where they were shown to interact with GSTF2. Data from X-ray crystallography indicated that there are three main binding sites, two symmetrically identical of **L1** and one **L2.** Indole-3-aldehyde and quercetrin bound at all three sites whereas camalexin and quercetin were only found at **L1** and **L2** respectively. Validation of the active sites using calorimetry studies revealed that all mutants showed a lower binding affinity compared to the wild type GSTF2 except for mutant Q73L and Y97A interacting with indole-3-aldehyde. This was further investigated by structural data of the mutant Y97A in complex with indole-3-aldehyde which indicate a conformational change when superimpose with the wild type complex of the same ligand. The data in this chapter represented the first structural evidence of non-catalytic activity of Arabidopsis GSTs with non-substrate molecule.

3.1 INTRODUCTION

3.1.1 The Arabidopsis GSTF2

Arabidopsis *GSTF2* (*At*4g02520), is a member of the Arabidopsis Phi (F) class glutathione transferase (GSTs) and the gene is located between base pair 1110452 and 1111660 on chromosome 4 of Arabidopsis Columbia ecotype 0 (Col-0) (The Arabidopsis Information Resources database (TAIR)). The *GSTF2* is flanked by both a chloroplast protein import receptor named Translocase Of Chloroplast 159 (*TOC159 (At*4g02510), located 0.9 kb upstream) and a chloroplast thylakoid lumen protein (*At*4g02530, 0.5 kb downstream). The genomic sequence of *GSTF2* is 1.2 kb and includes two ~0.1 kb intron regions. At the amino acid level, GSTF2 comprises 212 amino acids, with a predicted molecular weight of 26 kDa.

Microarray data available through Genevestigator (Hruz *et al.*, 2008) indicate that *GSTF2* is highly induced by biotic attack, treatment with cold stress, nutrient deficiency, chemicals and hormones such as salicylic acid, abscisic acid (ABA) and naphthaleneacetic acid. *GSTF2* is also highly expressed in leaves followed by seedlings and roots particularly during senescence and matured rosette developmental stage (**Figure 3.1**).

3.1.2 Small molecule interaction with GSTF2

The recombinant GSTF2 purified from *E. coli* was identified by tandem mass spectrometry (MS/MS) bound with N-heterocyclic ligands: harmane, norharmane, lumichrome, indole-3-aldehyde and 1-acetyl-*β*-carboline (Dixon *et al.*, 2011). When the purified GSTF2 was incubated with Arabidopsis extracts, the tandem mass spectrometry analysis revealed that plant hormones quercetrin and camalexin were also found to attach to GSTF2 (Dixon *et al.*, 2011). Isothermal titration calorimetry (ITC) analysis of the binding indicated that indole-3-aldehyde had the lowest affinity, with a K_D value of 11.10 µM towards GSTF2, followed by quercetrin and camalexin (6.25 µM and 1.20 µM, respectively) (Dixon *et al.*, 2011). GSTF2 has also been found to bind to 5-azido-(7-^{3}H)indole-3-acetic acid, aminopeptidase *N*-1-naphthylthalamic (NPA) (Zettl *et al.*, 1994; Murphy *et al.*, 2002; Smith *et al.*, 2003).

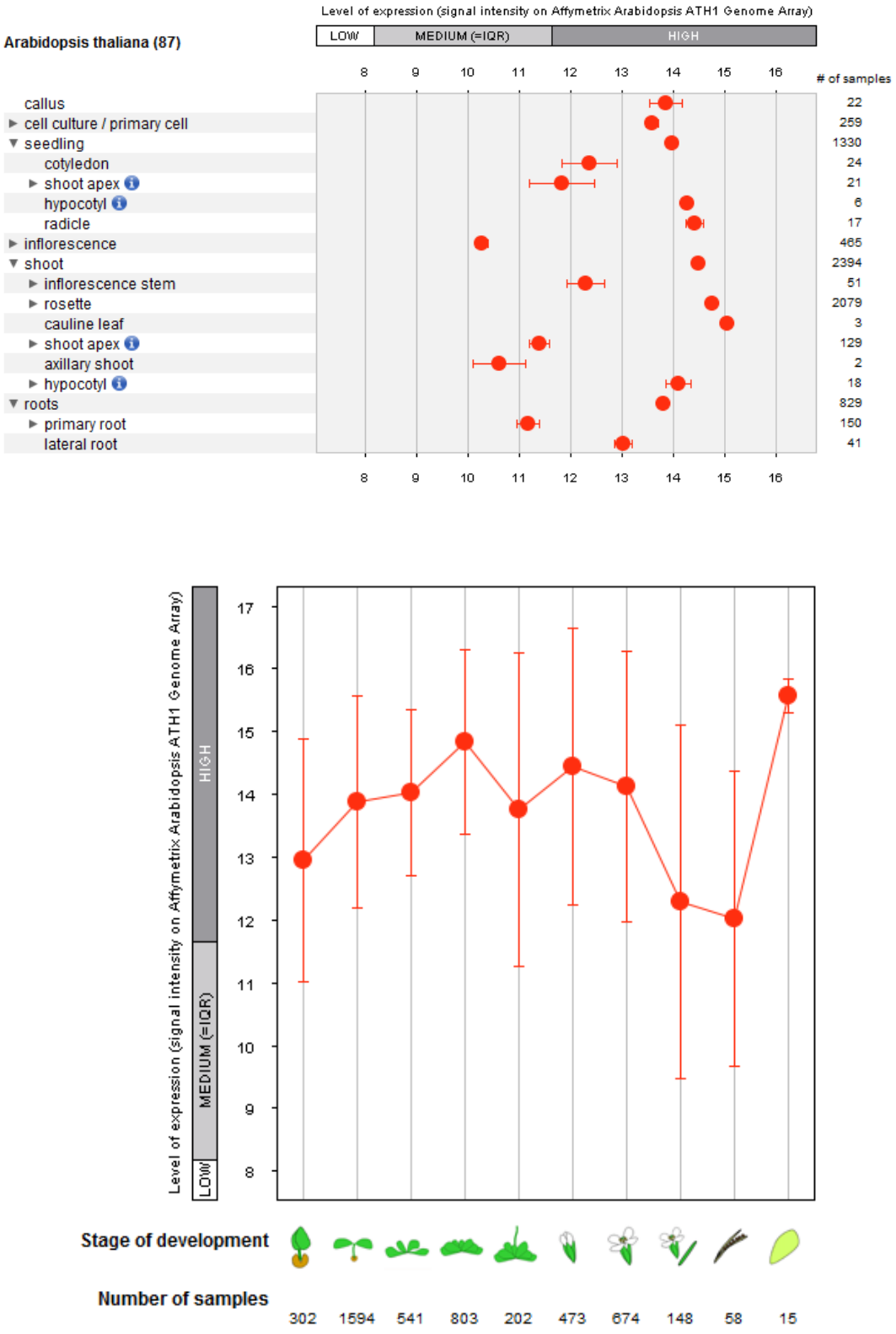


Figure 3.1. Expression of *GSTF2* gene in Arabidopsis. **Top panel**: *GSTF2* expression in different Arabidopsis compartment and; **Bottom panel**: *GSTF2* expression throughout different stage of Arabidopsis development. Level of expression in log2 scale. The stage of development is defined as; (from left) germinated seed, seedling, young rosetta, developed rosetta, bolting, young flower, developed flower, flowers and siliques and senescence. Source: Genevestigator (Hruz *et al.*, 2008), accessed 9th January 2017.

3.1.3 Structure of GSTF2

Structures of GSTF2 were solved in two forms; first in complex with a glutathione derivative, *S*-hexylglutathione at 2.2 Å (PDB ID: 1GNW) (Reinemer *et al.*, 1996), and second, in the presence of the acetamide group-herbicide conjugate, FOE-4053-GSH, at 2.6 Å (PDB ID: 1BX9) (Prade *et al.*, 1998). The thioredoxin domain at the N-terminal of GSTF2 was connected to the *α*-helical network at the C-terminal by a short linker of 15 residues. The GSTF2 dimer for both structures contained two G-sites and two H-sites at an identical location on each monomer (**Figure 3.2**).

In the first structure, two *S*-hexylglutathione molecules bound to the active site at each subunit of the GSTF2 structure (Reinemer *et al.*, 1996). The GSH backbone of one of the *S*-hexylglutathione molecules formed multiple H-bond interactions in the G-site. Residues K41, E53, V54, E66, S6 and R68 formed the G-site in this structure. The GSH backbone of the second *S*-hexylglutathione molecule was located next to the G-site and bound weakly due to forming few interactions with polar residues. The *S*-hexyl moiety of both *S*-hexylglutathione molecules bound in a region termed as the H-site site, which is formed by the segment of strand *β*1 connecting to helix *α*8, *α*10 and *α*11 at the C-terminal. Residues involved in the H-site interaction included H8, A10, S11, L35, F119, F123, Y127 and Y178. Located next to the G-site, the binding site occupied by the *Y*-glutamate and glycine of the second *S*-hexylglutathione molecule was defined as the L-site (**Figure 3.3**).

The second structure of GSTF2 (PDB ID: 1BX9) showed that the aromatic ring of FOE-4053 was located in the hydrophobic pocket of the H-site and the GSH moiety at the G-site. The FOE-4053 moiety resides in the pocket built up by residue I12, A13, S115 and F123. In the G-site, the carboxylate of the glycine group directly bound to K41 and H40, in a similar G-site region to the first structure. The L-site was not identified in this structure (Prade *et al.*, 1998).

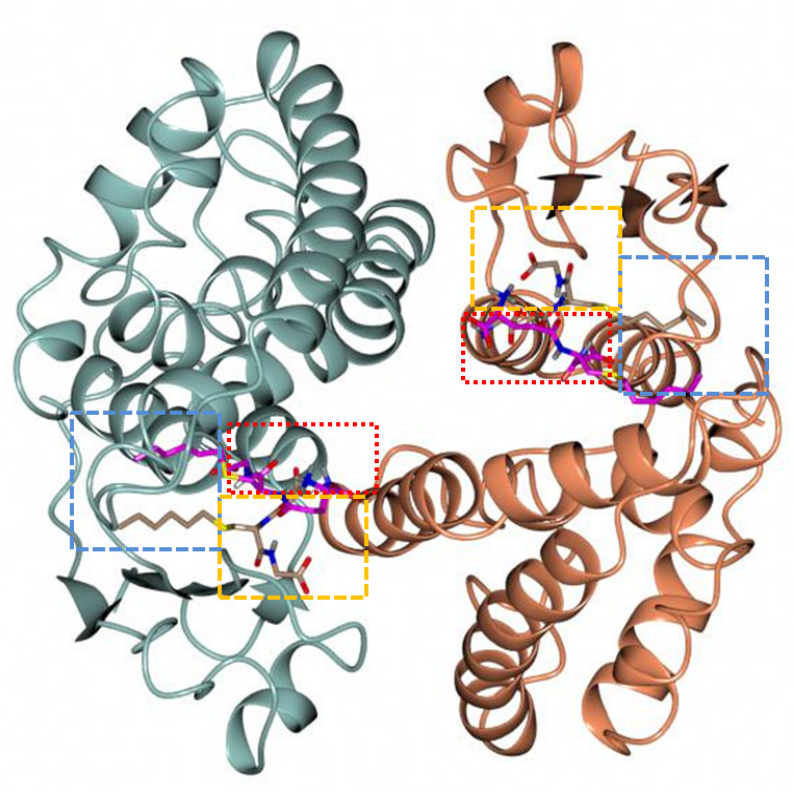


Figure 3.2. GSTF2 crystal structure in complex with *S*-hexylglutathione (PDB ID: 1GNW) as determined by Reinemer *et al.* (1996). The protein dimer was bound with two ligand molecules, one per monomer. The region surrounding the hexyl moiety is known as the H-site (blue box), the region bound by GSH moiety of the first *S*-hexylglutathione molecules is the G-site (orange box) and the observed L-site (red box).

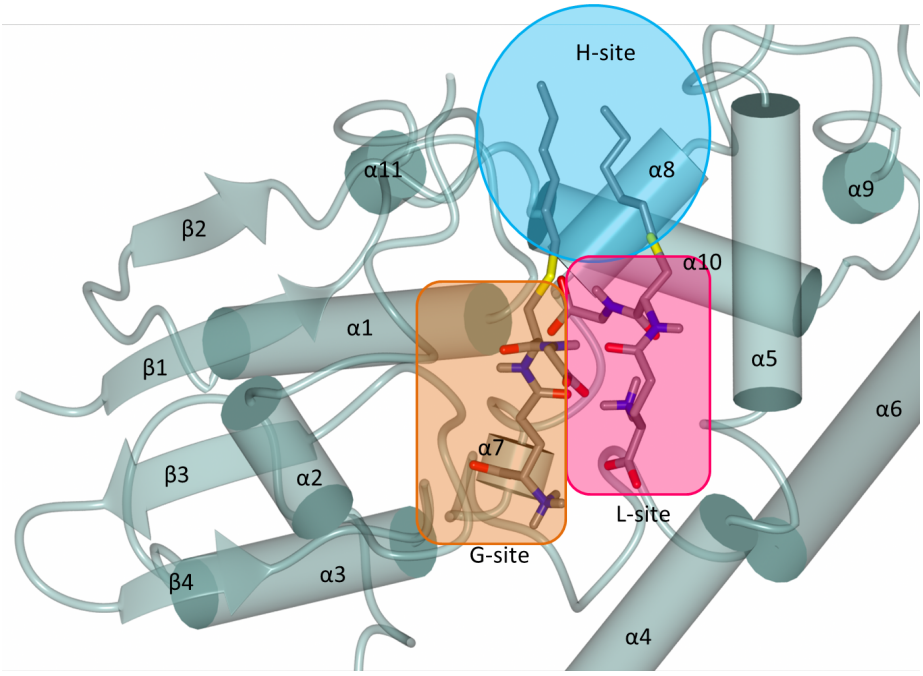


Figure 3.3. Close up view of the active sites identified in the structure of GSTF2 in complex with *S*-hexylglutathine, showing the location of G-site, H-site and L-site (PDB ID: 1GNW).

3.1.4 Ligands used in this study

The ligands indole-3-aldehyde, camalexin, quercetrin and quercetin, all identified previously by Dixon *et al.* (2011), were selected to further the understanding of the non-catalytic activity of GSTF2 (**Figure 3.4**). The ligands were chosen based on their distinctive chemical structure for easier subsequent identification by X-ray diffraction. It must be noted that these ligands are not reported to be GST substrates, but are important in plants as the plant hormones, phytoalexins.

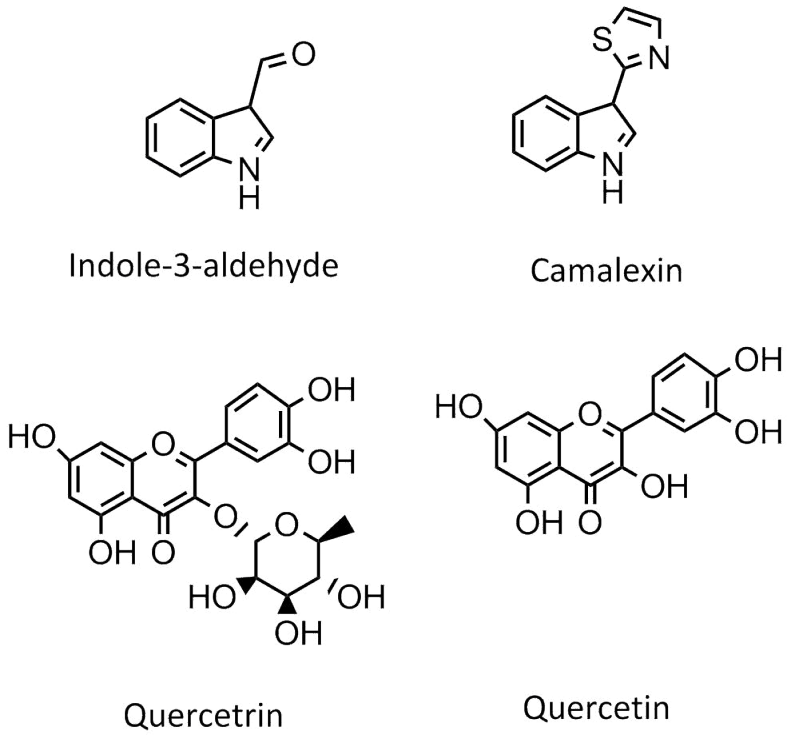


Figure 3.4. Chemical structures of the ligands used in this study.

3.1.4.1 Phytoalexins

Phytoalexins are related to defence mechanisms in plants and are only elicited in response to stresses, such as pathogen attack, fungal infection in *Botrytis cinerea* and *Alteria brassicola*, or in response to high levels of endogenous molecules such as auxins and ethylene (Jeandet *et al.*, 2014; Thomma *et al.*, 1999; Rowe *et al.*, 2010). The phytoalexins found in Arabidopsis are mostly indolic phytoalexins and flavonoids (Ayert & Tewariy 1991; Pedras *et al.*, 2000).

3.1.4.1.1 Indolic phytoalexins

Camalexin and indole-3-aldehyde are indolic phytoalexins that are synthesized from tryptophan *via* indole-3-acetaldoxime (IAOx) and indole-3-acetonitrile (IAN). The first step in the camalexin biosynthesis pathway includes the formation of IAOx from tryptophan by two Cytochromes P450 (CYPs); CYP79B2 and CYP79B3, to form indole -3-acetonitrile (IAN). After this step, a GSH conjugate is formed via catalysis by GSTF6. The conjugate is subsequently degraded to a cysteine-indole-3-acetonitrile conjugate (Cys(IAN)) (Geu-Flores *et al.*, 2011; Su *et al.*, 2011; Møldrup *et al.*, 2013). The Cys(IAN) conjugate is the substrate for a unique bifunctional P450 monooxygenase, CYP71B15, also known as PHYTOALEXIN-DEFICIENT3 (PAD3). PAD3 is responsible for converting Cys(IAN) to dihydrocamalexic acid and camalexin (**Figure 3.5**) (Zhou *et al.*, 1999; Schuhegger *et al.*, 2006; Bottcher *et al.*, 2009).

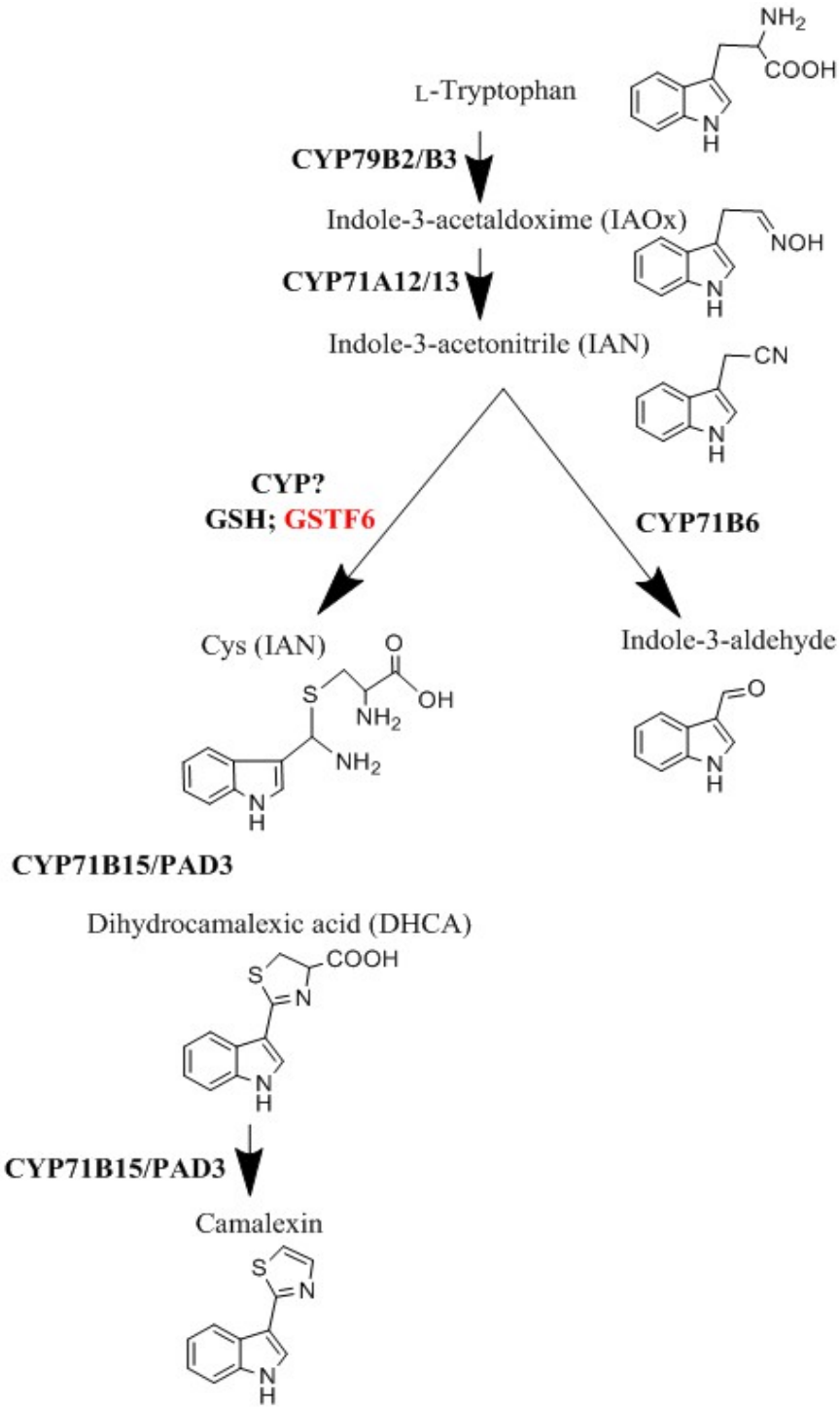


Figure 3.5. Putative indolic phytoalexins biosynthesis pathways, as adapted from Lemarié *et al.* (2015), Böttcher *et al.* (2014) and Su *et al.* (2011).

3.1.4.1.2 Flavonoids

Flavonoids are a large class of secondary metabolites involved in diverse physiological functions including UV protection, insect attraction, defence against pathogens and colour pigments (Winkel-Shirley 2001). Flavonoids also have protective functions against fungal infections. In sorghum, flavonoids act against the fungal plant pathogen *Colletotrichum* spp. by accumulating in the epidermal cells at the site of attack to restrict fungal proliferation (Ibraheem *et al.*, 2010; Snyder & Nicholson 1990).

Flavonoids have a benzo-*Ɣ*-pyrone structure and are synthesized *via* the phenyl-propanoid pathway (Kumar & Pandey 2013). Chemically, a benzo-*Ɣ*-pyrone consists of two benzene rings (A and B) linked through a heterocyclic pyrene ring (C), as shown in **Figure 3.6**. Flavonoids can be divided into a variety of classes according to the degree of oxidation and substitution at the pyrene ring (C), in which the individual compounds within the same class differ on the basis of the substitution pattern at the benzene rings (Middleton Jr. 1998; Kumar & Pandey 2013).

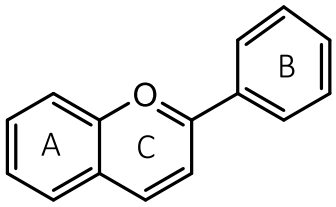


Figure 3.6. Basic chemical structure of flavonoids.

Interactions between GSTs AN9 from Petunia, GST1 and GSTIII from *Zea mays* with labelled flavonoids [^{3}H]isoquercetrin and [^{3}H]luteolin was analysed using equilibrium dialysis. The analysis showed that both luteolin and isoquercetrin bound to all tested GSTs. Eight molecules of isoquercertin bound to each GST molecules in the absence of GSH, with apparent K_D of 90 µM and four isoquercetrin molecules were bound to GST dimer, with K_D of 66 µM in the assays that contained GSH (Mueller *et al.*, 2000). This was the first study demonstrating that GSTs can possess multiple-binding sites of GSTs for flavonoids.

3.2 MATERIAL AND METHODS

3.2.1 Generation of pET24a-*GSTF2*

To produce GSTF2 without a tag and suitable for crystallisation, the construct of pETStrep3-GSTF2 with a strep tag, available in Robert Edwards lab, and vector pET24a without a tag, were digested with *Nde*I and *Xho*I. The treated products were analysed on agarose gels and the DNA band with the correct estimated size for pET24a and *GSTF2* were excised and purified. Following the purification, the products: 50 ng pET24a and 30 ng *GSTF2* were mixed with T4 DNA Ligase for ligation and incubated for 1 h at 22°C. The ligation product was transformed into XL-10 Gold Ultra-competent *E. coli* cells for propagation, followed by plasmid purification and DNA sequencing. After sequence confirmation, the constructs were transformed into *E.coli* Tunetta cells, as mentioned in **Section 2.4.9** (Taylor 2012; Zhao & Dixon 2010).

3.2.2 Recombinant protein expression

For gene expression, one colony from the antibiotic-containing plate was picked, inoculated and grown overnight in 5 mL LB media containing 100 μg mL^{-1} kanamycin and 35 μg mL^{-1} chloramphenicol (prepared as a 35 mg mL^{-1} stock in 100% EtOH) with shaking at 200 rpm at 37°C. The overnight culture was then transferred to 0.5 L LB media containing the same amount of antibiotic and left shaking at 200 rpm at 37°C. The culture was maintained under this condition until it reached an optical density of 0.6 at absorbance wavelength of 600 nm (OD_{600}, as determined against an LB blank). The culture was cooled to room temperature and 0.1 mM of isopropyl-*D*-thiogalactopyranoside (IPTG) was added to the growing culture to induce protein expression. The culture was left for a further 18 h with shaking at 200 rpm at 20°C. Cells were then harvested by centrifugation (at 4,000 x g, 15 min, 4°C) and the bacterial cell pellet was stored at -20°C until required.

3.2.3 Protein purification using GSH-affinity chromatography and size-exclusion chromatography

Frozen cells were resuspended collectively in 20 mM Tris-HCl buffer pH 7.5 with the addition of 2 μM dithiothreitol (DTT) to reduce the formation of disulphide bonds. Cells were disrupted by ultrasonication (70% amplitude, 5 min total sonication time, 10 s sonication and 5 s cooling). The soluble and insoluble material fractions were separated by centrifugation at 2,000 x g for 15 min. The supernatant, containing the soluble GSTF2, was further clarified using a 0.45 μm Millex-HA syringe filter unit (Merck Millipore) prior to loading onto a 10 mL GSH sepharose 4B prepacked column (GE Healthcare). The column

was first pre-equilibrated with 20 mM Tris-HCl buffer pH 7.5 and the cell lysate was loaded at flow rate of 0.02 mL min^{-1}. Once the unbound protein has been removed from the column, recombinant protein was eluted with buffer containing 10 mM GSH. Eluted protein fractions were analysed by SDS-PAGE and the fractions containing purified protein were pooled and concentrated using a Centricon® filter membrane (10 kDa cut-off). Concentrated protein was loaded onto a S75 Superdex™ gel filtration column that had been equilibrated with 20 mM Tris-HCl buffer pH 7.5 with the addition of 150 mM NaCl. Fractions containing pure protein, as judged by SDS-PAGE, were pooled, flash frozen and stored at -80°C until required. Both the GSH sepharose 4B premade column and S75 Superdex™ gel filtration column were regenerated by washing with filtered water and 20% (v/v) ethanol prior to storage at 4°C. Final protein concentration used in this study was recorded at 18 mg mL^{-1} from 1 L of starting culture.

3.2.4 Site-directed mutagenesis

The recombinant plasmid pET24a-GSTF2 was used as the DNA template for the preparation of mutants. Mutagenesis rounds were carried out following the QuickChange (Stratagene) protocol. The primers employed for each mutation are listed in **Table 3.1**. The thermal cycle for PCR amplification included initial denaturation of double stranded DNA at 98°C for 1 min, 14 cycles of denaturation, annealing and elongation at 30 s 98°C, 1 min 60°C and 4 min 72°C respectively. Final extension was carried out for 4 min at 72°C. The PCR product was treated with *Dpn*I endonuclease (target sequence 5'-Gm^6ATC-3') to digest the methylated and hemimethylated parental DNA template and to select for newly-synthesised DNA, containing the desired mutation. The mutated construct was then transformed into XL1-Blue supercompetent cells, using the same protocol as described in section 2.4.9. The plasmid was purified using a QIAprep Spin Miniprep Kit (Qiagen) and the presence of mutation was validated by DNA sequencing performed by GATC Biotech. The mutant genes were expressed and purified using the same protocols as described for the wild-type enzyme.

Table 3.1. Primers used for PCR amplification of wild-type *GTSF2* and the generation of mutant in site-directed mutagenesis analysis.

Primer	Nucleotide	TM (°C)	% GC
WT_F	ATGGCGAAATCACTCCTTTG	56.4	45.0
WT_R	TCACAGTCCCTTAAGCTGTTC	59.5	48.0
Q73L_F	CAAGCTCTTCGAATCAAGAGCGATTACTCTGTACATAG	61.6	42.1
Q73L_R	GGTTTTCATATCGGTGAGCTATGTACAGAGTAATCGC	61.9	43.2
Y97A_F	CAAACCGACTCCAAGAACATATCTCAGGCCGCAATC	65.2	50.0
Y97A_R	CTTCTACTTGCATTCCAATGGCCATGATTGCGGCCTGAG	67.4	51.3
H77A_F	GAATCAAGAGCGATTACTCAGTACATAGCTGCCCGATATG	63.6	45.0
H77A_R	GAAGGTTGGTTCCTTGGTTTTCATATCGGGCAGCTATG	64.9	47.4
R154A_F	CTTGATGTCTACGAGGCTGCGCTCAAGGAGTTCAAG	66.0	52.0
R154A_R	GAAAGTTTCACCAGCCAAATACTTGAACTCCTTGAGCGCAGCCTC	67.8	48.9

3.2.5 Protein crystallisation

For protein crystallisation, a co-crystallisation method was employed. The pure GSTF2 was incubated with the selected ligands (5 mM to 10 mM) for 1 h on ice followed by microcentrifugation at 4,000 x g to remove any insoluble precipitates resulting from the complexation. The incubation mixture was then subjected to crystallisation trials using a Mosquito® ROBOT (TTP LabTech) and a range of commercially available crystallisation screens, including PACT premier™ HT-96 (Molecular Dimension), Index HT (Hampton Research) and Crystal Screen 1 and 2 (Hampton Research). The screening was carried out in 96-well plate sitting drop format with which each drop consisting of 150 nL protein and 150 nL precipitant solution of the screening buffer. Larger crystals for diffraction analysis were obtained from the hanging drop vapour diffusion method in 24-well plate Linbro dishes. The 2 µL drop also contained precipitant reservoir solution and the protein solution (in 1:1 ratio), pipetted onto a premade siliconized cover slip and sealed onto the well using high-vacuum grease. The droplet was allowed to equilibrate with the larger reservoir solution containing similar buffers and precipitants. As water vaporised from the drop to the reservoir, the precipitant concentration increased to a level optimal for crystallisation, allowing the crystal to form and grow. Prior to the analysis on the in-house X-ray equipment, the crystals were washed with the same precipitant reservoir solution containing 20% (v/v) ethylene glycol as the cryoprotectant, followed by flash-cooling in liquid nitrogen. Crystals were tested for diffraction using a Rigaku Micromax-007HF X-ray generator fitted with Osmic multilayer optics and a MARRESEARCH MAR345 imaging plate detector. The crystals that diffracted to a resolution equal to, or better than, 3 Å were retained for dataset collection at the Diamond Light Source synchrotron, Didcot, Oxfordshire, UK synchrotron.

3.2.6 Data collection, structure solution, model building and refinement

The dataset for the crystal structure reported in this thesis was collected at the Diamond Light Source. Data were processed and integrated using the X-ray Detector Software (XDS) program package (Kabsch 2010) and scaled using SCALA (Evans 2006) included in the Xia2 processing system (Winter 2010). Data collection and refinement statistics are summarised in **Table 3.2**. The structures of GSTF2 in complex with ligands were solved by Prof. Gideon Grogan, using a monomer model of GTSF2 in complex with *S*-hexylglutathione (PDB ID: 1GNW) as the model with Molecular Replacement program (MOLREP) (Vagin & Teplyakov 1997). The structures were built and refined through iterative cycles of Crystallographic Object-Oriented Toolkit (Coot) (Emsley & Cowtan 2004) and REFMAC 5 (Murshudov *et al.*, 1997), employing local non-crystallographic symmetry (NCS) restraints in the refinement cycles. Ligands and associated refinement libraries were prepared using the PRODRG program (Schüttelkopf & Van Aalten 2004). All structures were finally validated using the PROCHECK program (Laskowski *et al.*, 1993). The Ramachandran plot was employed to confirm the most energetically-favourable locations of the residues. The coordinates of GSTF2 with indole-3-aldehyde, camalexin, quercetrin and quercetin have been deposited to the PDB database under the accession numbers 5A4U, 5A5K, 5A4W and 5A4V, respectively.

Table 3.2. Data collection and refinement statistics for GSTF2 in complex with indole-3-aldehyde, camalexin, quercetrin and quercetin. Data for the highest resolution shells are given in parentheses.

	GSTF2 with indole-3-aldehyde	**GSTF2 with camalexin**	**GSTF2 with quercetrin**	**GSTF2 with quercetin**
Beamline	Diamond I03	Diamond I03	Diamond I04-1	Diamond I04-1
Wavelength (Å)	0.97625	0.97625	0.92000	0.92000
Resolution (Å)	94.41-2.00 (2.05-2.00)	87.58-2.77 (2.84-2.77)	59.09-2.25 (2.31-2.25)	59.59-2.38 (2.44-2.38)
Space Group	$P2_12_12_1$	P1	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	a = 87.86; b = 94.41; c = 152.38 a = b = g = 90°	a = 97.10; b = 113.72; c = 132.02 a = 83.7 b = 79.5 g = 65.9°	a = 87.35; b = 93.57; c = 152.42 a = b = g = 90°	a = 88.03; b = 94.83; c = 153.20 a = b = g = 90°
No. of molecules in the asymmetric unit	6	24	6	6
Unique reflections	86285 (6308)	126932 (9333)	60022 (4365)	52158 (3838)
Completeness (%)	100.0 (100.0)	98.6 (98.1)	100.0 (100.0)	99.9 (100)
Rmerge (%)	0.08 (0.63)	0.10 (0.72)	0.09 (0.72)	0.11 (0.68)
Rp.i.m.	0.04 (0.34)	0.10 (0.72)	0.04 (0.32)	0.07 (0.41)
Multiplicity	8.1 (8.4)	2.2 (2.2)	6.8 (7.1)	6.7 (7.0)
$<I/\sigma(I)>$	17.2 (3.3)	6.8 (1.8)	16.5 (3.2)	14.9 (2.8)
CC1/2	1.00 (0.89)	0.99 (0.74)	1.00 (0.88)	1.00 (0.84)
Overall B factor from Wilson plot (Å2)	28	35	30	21
Rcryst/ Rfree (%)	19.9/23.4	25.0/28.4	21.4/25.2	20.4/24.6
r.m.s.d 1-2 bonds (Å)	0.017	0.014	0.012	0.012
r.m.s.d 1-3 angles (o)	1.85	1.98	1.69	1.46
Avge main chain B (Å2)	32	49	37	34
Avge side chain B (Å2)	35	51	40	37
Avge water B (Å2)	33	29	37	33
Avge ligand B (Å2)	26	53	51	44

3.2.7 Ligand binding analysis by isothermal titration calorimetry (ITC) of GSTF2 wild-type and mutants

The binding of the wild-type GSTF2 and the mutants to each ligand was measured in phosphate buffered saline (PBS; 36 mM Na_2HPO_4, 64 mM Na_2HPO_4, 50 mM NaCl, 1% (v/v) DMSO, pH 7.0) at 25°C using a VP-ITC microcalorimeter (MicroCal). In the standard experiment, the cell contained 1.4 mL of a solution of protein and the syringe contained 280 µL of ligands at a concentration that was 15-20 times higher than the protein concentration in the cell. Both the cell and syringe solutions were degassed at 23°C for 20 min. The titrations were performed as follows: one preliminary injection of 2 µL of ligand, and subsequently, 29 injections of 10 µL ligand at an injection speed of 0.5 µL s^{-1}. The stirring speed was 307 rpm, with a delay time between the injections of 4 min. The blank titration was performed by injecting the ligand solution into buffer without the protein, and the average heat of dilution of the control was subtracted from the experiment with protein. Data was analysed using the MicroCal Origin software and fitted to a one-site set of binding model or two-sites set of binding model using non-linear regression analysis. Ligand concentrations were determined by weighing method. Protein concentrations were determined by UV-VIS spectrophotometry, with the value of absorbance at 280 nm and the estimated extinction coefficient (ε_{280}) of 15.93 mM^{-1} (calculated with the ProtParam tool, ExPaSy website, Swiss Institute of Bioinformatics).

3.3 RESULTS

3.3.1 Cloning, expression and purification of GSTF2

The GSTF2 was previously cloned into pET-Strep3 (Novagen, UK) expression vector by Dixon *et al.* (2011) which enabled the expression of an N-terminal strep-tagged protein. To express a protein suitable for X-ray crystallography studies, the tag was removed by sub-cloning GSTF2 into the pET24a vector (Novagen, UK), between the *Nde*I and *Xho*I sites (**Figure 3.7**). After GSTF2 sequence validation, the protein was expressed in *E. coli* Tunetta cells (Novagen, UK), following the protocol recommended by Taylor (2012). The expressed protein was purified with GSH sepharose affinity chromatography followed by size-exclusion chromatography (SEC) using an S200 column (GE Healthcare, UK) (**Figure 3.8; Figure 3.9**). The purified GSTF2 was concentrated to 10 mg mL^{-1} in 20 mM Tris-HCl pH 7.4 and stored at -80°C prior to the crystallography screening.

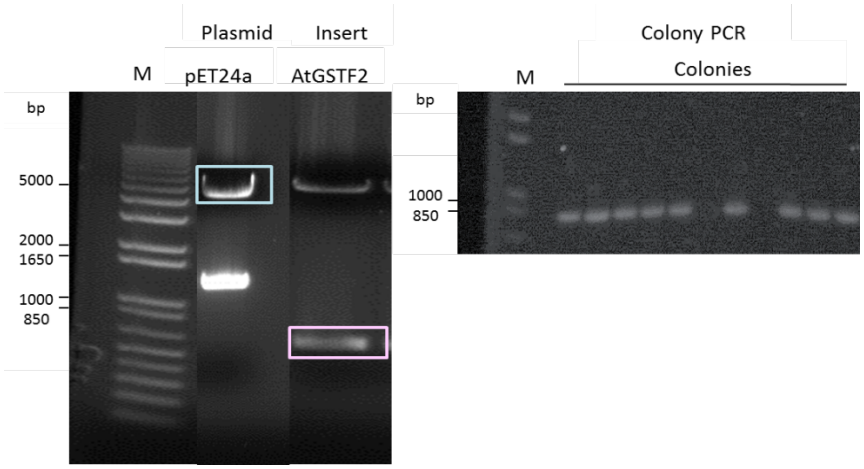


Figure 3.7. Ligation of GSTF2 into pET24a. From left: linearization of pET24 plasmid carrying other gene insert and GSTF2 initially ligated with pET-Strep3. Colony PCR showing positive clones with the estimated gene band at 640 bp.

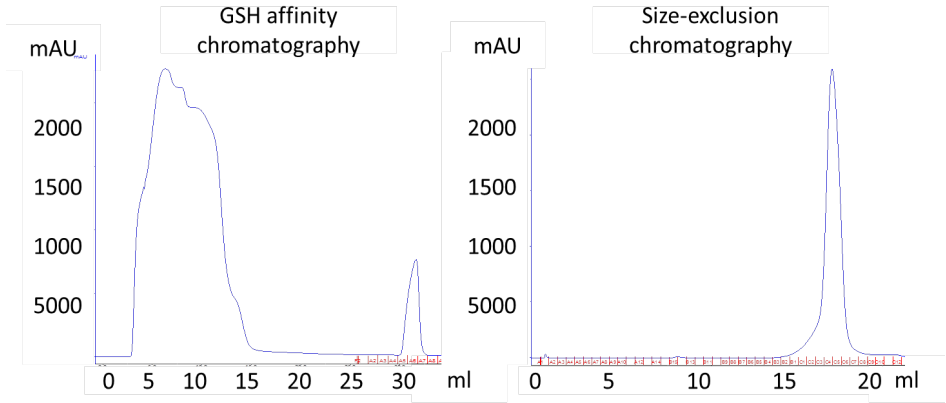


Figure 3.8. Protein purification profile of GSTF2. Left: GSH affinity chromatography showing a single peak of protein eluted; Right: Size exclusion chromatography profile, showing a single peak for the eluted protein.

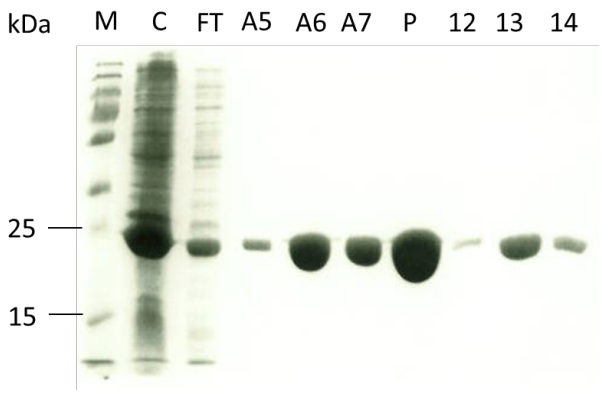


Figure 3.9. Protein purification profile of GSTF2. SDS-PAGE gel showing: M: Broadrange protein marker; C: Crude lysate of total protein fraction; FT: Flow-through of the unbound protein fractions; Fractions A5-A7 are the elution fractions after the addition of 10 mM GSH in GSH column; P: pooled fractions from the gel filtration column; Fractions 12-14 are the elution fractions from the size-exclusion chromatography.

3.3.2 Co-crystallisation of GSTF2 with target ligands.

The co-crystallisation screens yielded rod-like crystals within a week of incubation at 18°C (**Figure 3.10**). The buffer condition was optimised on a larger scale (1 mL) with the addition of several precipitants including glycerol, ethylene glycol, DMSO, propanol, polyethylene glycol (PEG 3350) and 2-methyl-2,4-pentanediol (MPD) by hanging drop vapour diffusion method on a 24-well Linbro plate (Hampton Research, UK). The addition of precipitants aimed to induce supersaturation by reducing protein solvation and electrostatic shielding, forcing the protein molecules out of the solution and increasing the electrostatic effect to initiate crystallisation.

The largest single crystals of GSTF2 with indole-3-aldehyde and camalexin were collected from drops containing 0.2 M sodium acetate, 20% (w/v) PEG 3350 and 10% (v/v) propanol. The crystals of GSTF2 with quercetrin and quercetin were collected from drops containing PCTP buffer system (sodium propionate, sodium cacodylate trihydrate and bis-tris propane pH 7.0, in the molar ratios of 2:1:2), 25% (w/v) PEG 1500 and PEG 3350. The diffraction image for GSTF2 with indole-3-aldehyde was collected at 2 Å, while images for GSTF2 with camalexin, quercetrin and quercetin were collected at 2.77 Å, 2.25 Å and 2.38 Å respectively.

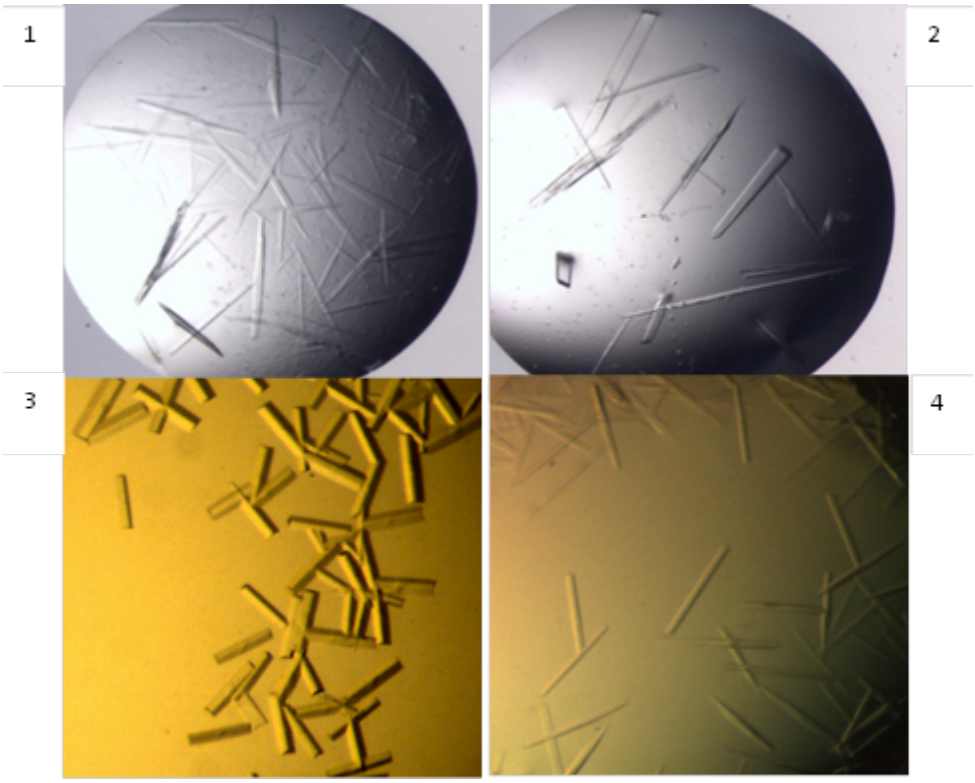


Figure 3.10. Crystals of GSTF2 after incubation with 1: indole-3-aldehyde, 2: camalexin, 3: quercetrin and 4: quercetin.

3.3.3 Ligands bound within the interface of the GSTF2 dimer

In the asymmetric unit GSTF2, in complex with indole-3-aldehyde, quercetrin and quercetin, was observed as a hexamer of three dimers in the $P2_12_12_1$ space group while GSTF2 with camalexin was observed in a $P1$ space group of three dimers stacking together in four hexamers (**Figure 3.11; Figure 3.12**). All ligands were unambiguously present in the crystal structure, as observed by clear peaks in omit maps (F_o-F_c map) contoured at 3σ after protein and solvent refinement.

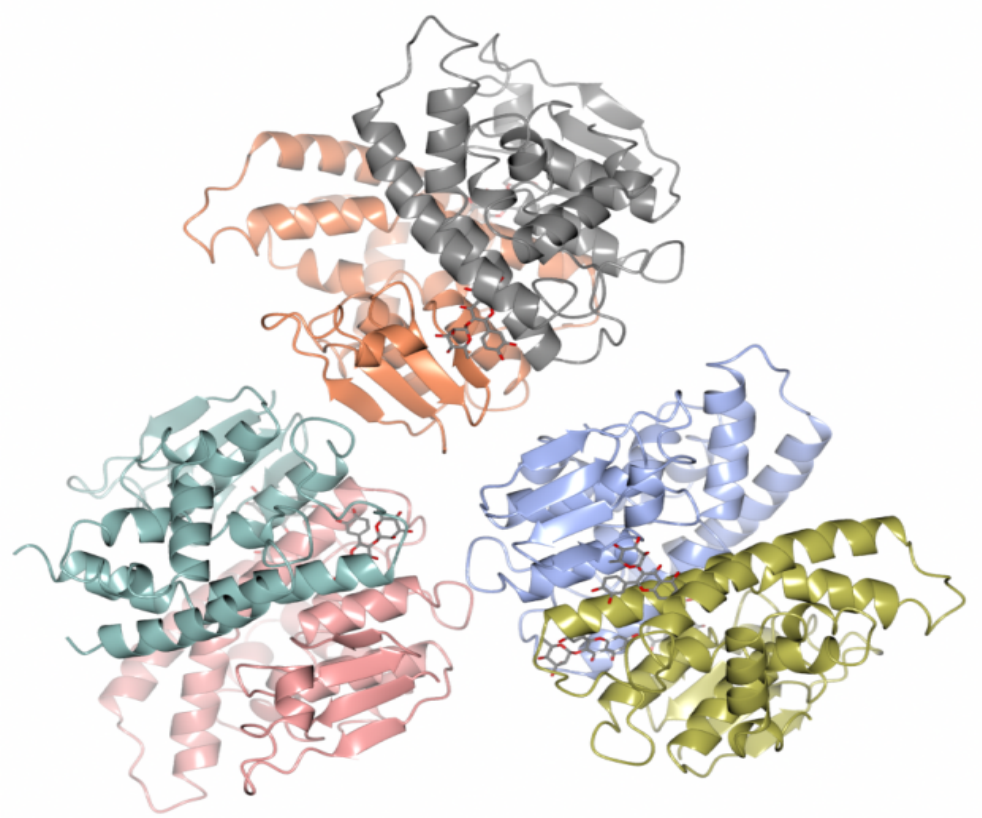


Figure 3.11. Asymmetric unit of GSTF2 as trimer of dimers when in-complex with ligand indole-3-aldehyde, quercetrin (in the figure) and quercetin.

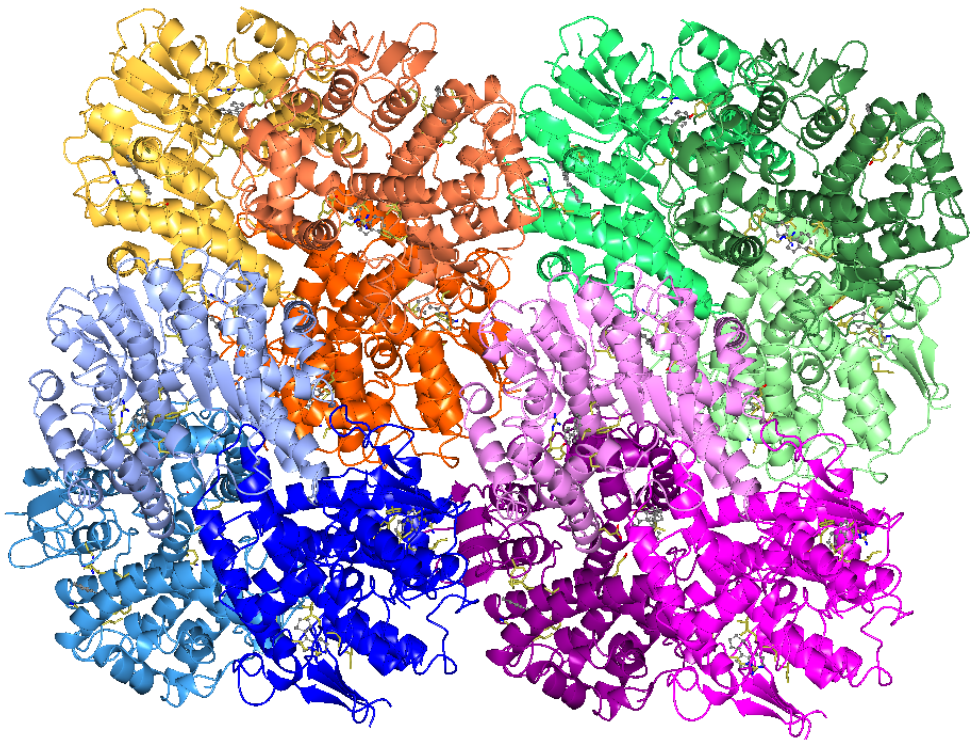


Figure 3.12. Asymmetric unit of GSTF2 as four hexamers of three dimers, shown as subunits of similar colour, in-complex with camalexin.

In the GSTF2 protein dimer, three main binding sites were identified: two **L1** and an **L2**. The **L1** sites were symmetrically equivalent and located in between helices α-3, α-6 and the loop region between K159 and E164 at periphery of the dimer interface. The other site, **L2** was located at the centre of the **L1** sites, in between helix α-2 of one monomer and α-3 of its neighbour. For each ligand, indole-3-aldehyde and quercetrin were observed at all three locations of the dimer interface while camalexin was only observed at **L1** and quercetin only at **L2** (**Figure 3.13**).

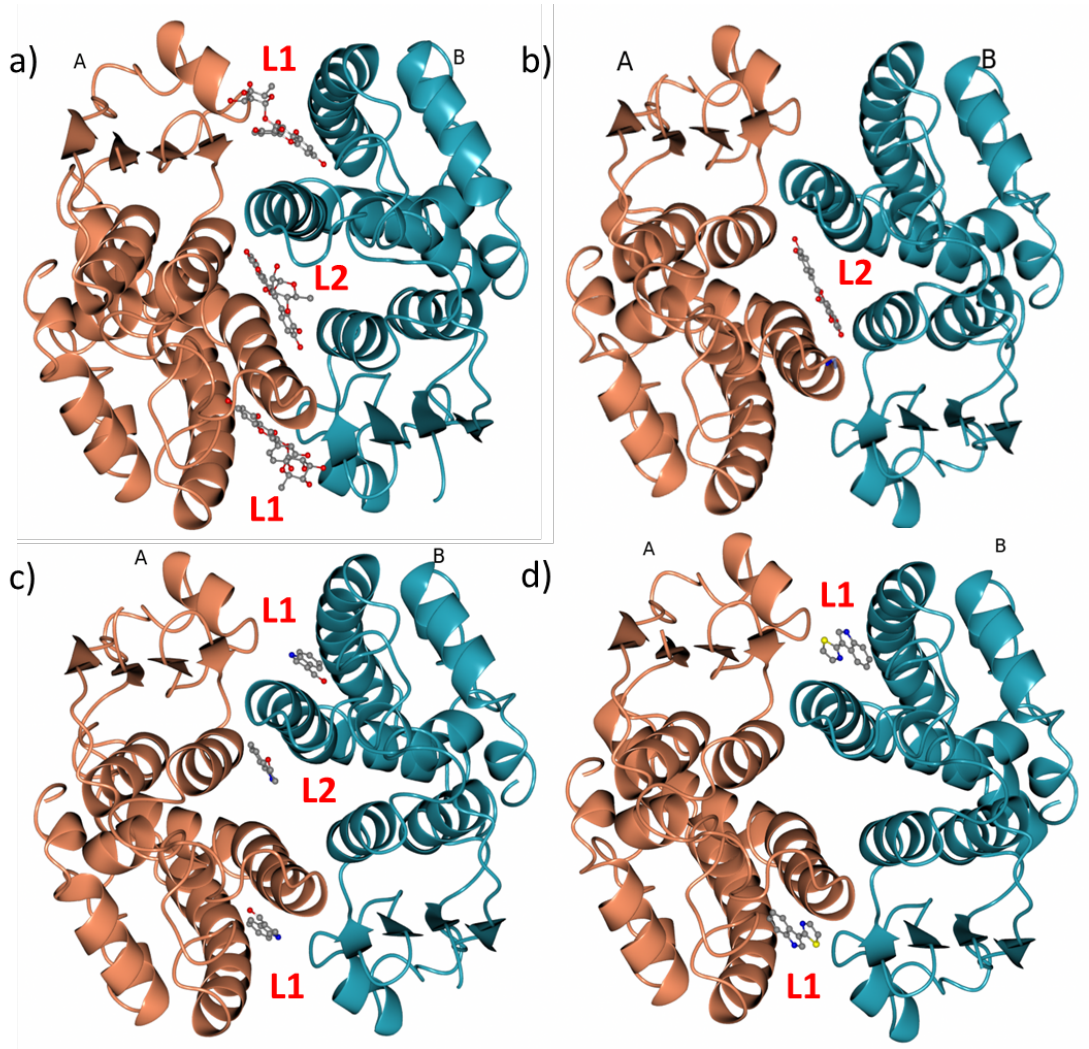


Figure 3.13. The dimers labelled as 'A and B' from ligand complex structures and location of ligands in binding sites **L1** and **L2**. a) GSTF2 with quercetrin; b) GSTF2 with quercetin; c) GSTF2 with indole-3-aldehyde; d) GSTF2 with camalexin.

3.3.4 Hydrophobic pocket and π-stacking interactions in the L1 site

Three ligands, indole-3-aldehyde, quercetrin and camalexin were observed at the **L1** site. The interactions at the **L1** site are mainly contributed by the hydrophobic pocket formed by V150, Y151, R154, V106 and I102 and π-staking interactions with the guanidinium group of R154.

With quercetrin, the resorcinol ring of bicyclic chromanone occupies the equivalent site to the benzene ring of indole-3-aldehyde. The planar bicyclic chromanone arranged between R154 and the hydrophobic pocket of V106, I102, F66 and F52. The R154 is slightly shifted, relative to the binding with indole-3-aldehyde, owing to the presence of catechol ring that is rotated approximately 45° relative to the chromanone system. Particularly, the OAE catechol hydroxyl is distanced 4 Å from the guanidinium group of the R154 side chain. In the chromanone system, the OAC hydroxyl forms a hydrogen bond with the water molecule, which in turn formed a hydrogen bond with the phenolic hydroxyl of Y151. The rhamnose sugar is parallel to that of the catechol ring, forming a hydrogen bond from O2 and O3 of the hydroxyl groups with the backbone carbonyl group of S48.

In the complex with indole-3-aldehyde, the ligand aromatic ring is sandwiched between R154 and hydrophobic residues in the pocket, including V106, I102, F52 and F66. The cation guanidinium side chain of R154 forms π-stacking interaction with the aromatic ring of the ligand, a favourable non-covalent bond that contributes to protein stability. The aldehyde moiety of indole-3-aldehyde forms a hydrogen bond at a 3.3 Å from the peptidic carbonyl of L161, 3.4 Å from I99 and 4 Å from T169. In addition, the indole nitrogen was not observed to make hydrogen bond with any side-chains in the site.

Similarly to indole-3-aldehyde, the bicyclic and thiazole rings present in camalexin made π-stacking interactions with the guanidinium ion of R154 in the hydrophobic pocket, as they located beneath the side chain of R154. The indole ring rotated approximately 60° relative to the orientation observed with indole-3-aldehyde, and pushed the indole nitrogen within a distance of 4.7 Å of the backbone carbonyl of V150.

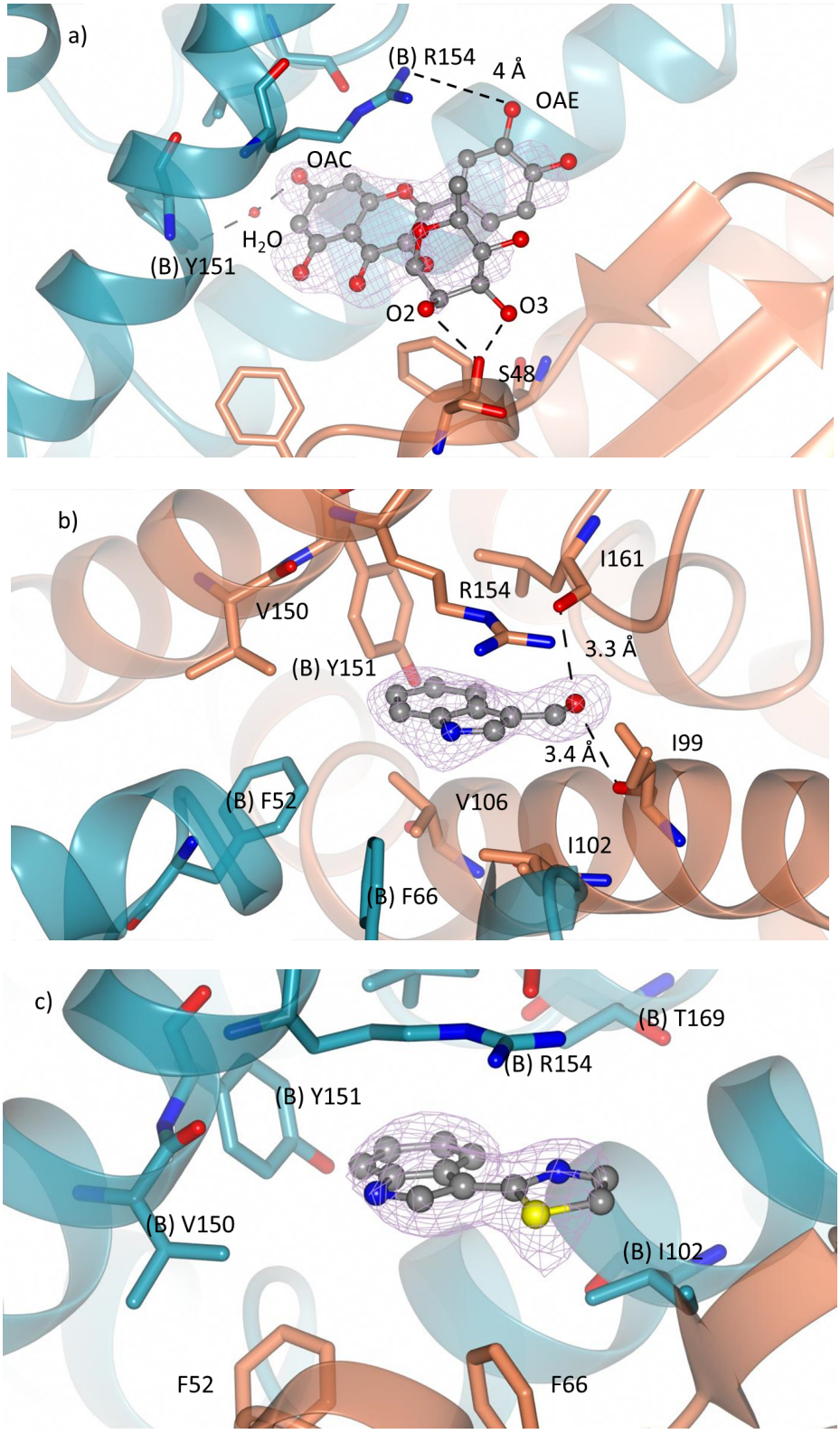


Figure 3.14. The binding interaction at the **L1** site for a) quercetrin, b) indole-3-aldehyde, and c) camalexin. Backbone and side-chains of monomers A and B of a dimer of GSTF2 are shown in ribbon and cylinder format, in blue and orange respectively. All ligands are shown in ball-and-stick format with the carbon atoms in grey. The dashed lines indicated hydrogen bond interactions between protein and ligand with a distance <3.5 Å. Electron density maps are shown in lilac and correspond to the F_o-F_c omit contoured at levels of 3 σ, which was obtained prior to the refinement of the ligand atoms. Ligand atoms from the ligand complex structure have been added afterwards for clarity.

3.3.5 Hydrophobic stacking interactions and H-bonds at L2 site

The **L2** site is a symmetrical site owing to the location at the two-fold axis of the monomer interface. In this site, the bicyclic ring of indole-3-aldehyde is located in the middle of H77 and Y97 side chains. The nitrogen in the aromatic side chain of the ligand forms a hydrogen bond with Q73 within a distance of 3.9 Å.

For quercetrin, the bicyclic chromanone is stacked between H77 and Y97. In the chromanone ring, the resorcinol hydroxyl is distanced 2 Å from the Q73 side chain and the OAF resorcinol hydroxyl is 2.9 Å distanced away from the backbone carbonyl of I94. The three rings of the flavone system are co-planar in **L2** site when compared to **L1** site. The rhamnose ring in this site occupies at the periphery of the dimer, forming H-bonds with endocyclic oxygen and the backbone carbonyl of S91. The O2 hydroxyl of the rhamnose is also H-bonded to the backbone carbonyl of K92.

Quercetin was also found within the **L2** site, governed by similar hydrophobic stacking interactions, in which the tricyclic flavone superimposes exactly with that of the rhamnosylated quercetrin.

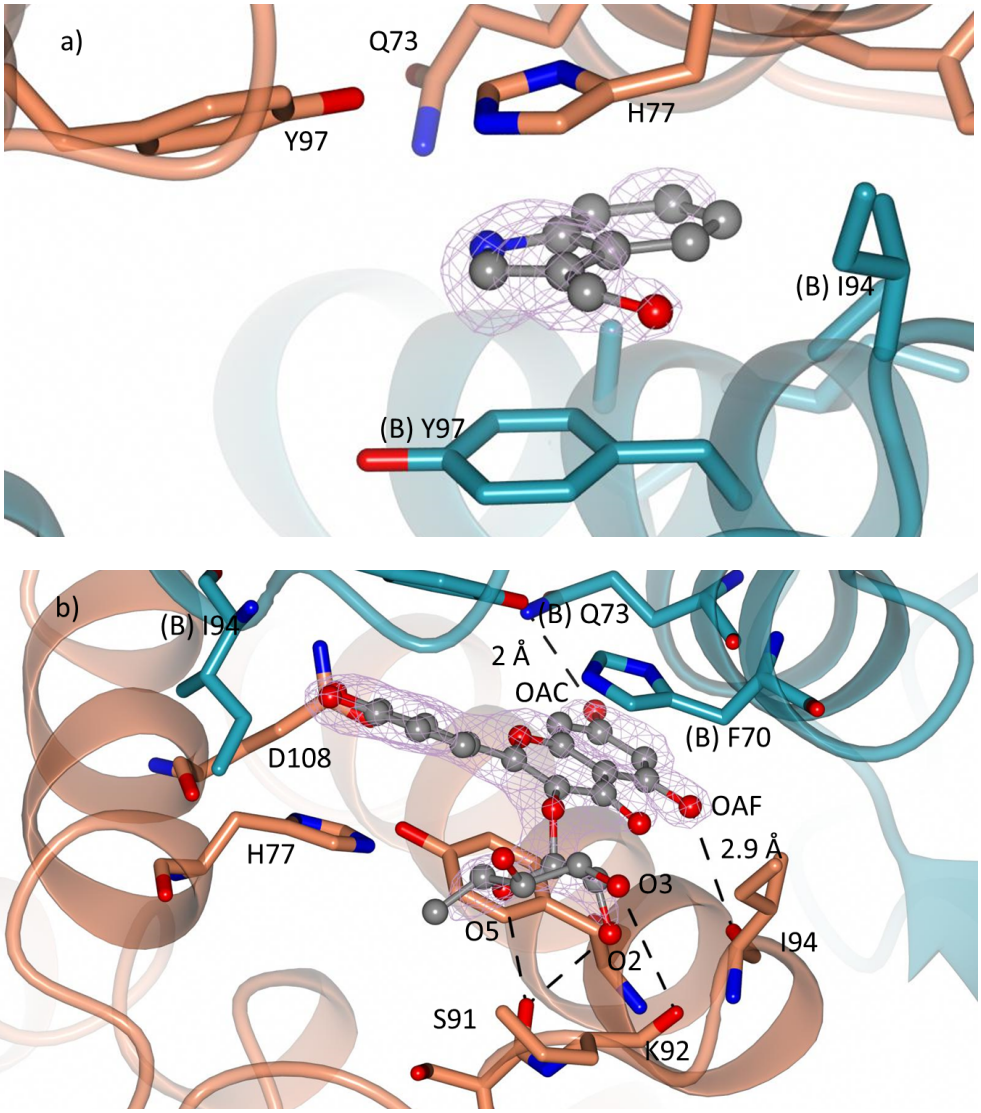


Figure 3.15. The binding interaction at the **L2** site for a) indole-3-aldehyde and b) quercetrin. Backbone and side-chains of monomers A and B of a dimer of GSTF2 are shown in ribbon and cylinder format in blue and orange respectively. All ligands are shown in ball-and-stick format with the carbon atom in grey. The dashed lines indicated hydrogen bonding interactions between protein and ligand with a distance of <3.5 Å. Electron density maps are shown in lilac and corresponded to the F_o-F_c omit contoured at levels of 3 σ, which was obtained prior to the refinement of the ligand atoms. Ligand atoms from the ligand complex structure have been added afterwards for clarity.

3.3.6 Site-directed mutagenesis of GSTF2 active residues

Key amino acids in GSTF2 that interacted with quercetrin, the largest and more functionalised ligand, were selected for mutagenesis analysis. For the **L1** site, the role of R154 was investigated by mutating this residue to alanine, while for the **L2** site, Q73, H77 and Y97 were selected and mutated to leucine, alanine and alanine respectively. Q73 was changed to leucine to investigate the significance of hydrogen bond interaction over the steric factors including orientation of the side chain.

Figure 3.16. Multiple sequence alignment between the sequence of wild-type and the mutants. Residues mutated are labelled as triangle (▼). Alignment was generated using ClustalOmega where red residue indicates hydrophobic residues, blue for acidic, magenta for basic and green for residues containing hydroxyl, sulfhydryl and amine (Sievers *et al.*, 2011).

3.3.7 Isothermal titration calorimetry (ITC) analysis

To analyse the binding activity of quercetrin, indole-3-aldehyde, camalexin and quercetin to wild-type GSTF2 and to the mutants GSTF2_Q73L, GSTF2_H77A, GSTF2_Y97A and GSTF2_R154A, ITC was used. Data for quercetin was not obtained due to the low solubility of the ligand. On the basis of crystallography data, at least two binding sites were identified within the dimer molecules. Only the ITC data from GSTF2_Q73L titrated with quercetrin could be analysed using a two-site model (two independent binding sites). The other experimental data, including those from wild-type protein and the mutants, were analysed using a one-site binding model. To ensure the integrity of the comparison between wild-type and the mutants, the same concentration of all proteins was used to study each ligand.

The binding stoichiometry (N) of quercetrin, indole-3-aldehyde and camalexin to the wild-type were recorded at 1.00, 0.69 and 3.75 respectively (**Table 3.3**). The binding activity (K_D) for each ligand was recorded at 5.9 µM for quercetrin, 24.3 µM for indole-3-aldehyde and 16.4 µM for camalexin. All three binding reactions occurred spontaneously according to Gibbs energy of binding (ΔG) where $\Delta G<0$ when $\Delta G = \Delta H - T(\Delta S)$ at the analysis temperature (298.15 K).

The binding activity of ligands to the mutants was investigated by fixing the stoichiometries to match the wild-type binding. The results showed that the K_D for the binding of quercetrin to GSTF2_H77A, GSTF2_Y97A and GSTF2_R154A were four times higher than the wild-type. There was no significant difference in terms of the binding enthalpy but more-negative values of entropy were observed for the mutants in comparison to the wild-type. The reactions for each mutant, as according to Gibbs binding energy, were favourable and occurred spontaneously with $\Delta G<0$.

The analysis for GSTF2_Q73L was carried out using the two-sites binding model because the data could not be fitted into the one-set model analysis. The results showed that the first site has lower stoichiometry than the second site. The quercetrin molecule bound more strongly to the first site, compared to the second site, by 20-fold, based on the K_D value: 1.45 µM for the first site compared to 30.21 µM for the second site. The enthalpy for the binding sites corresponded with the binding affinity for each site, where a more exothermic reaction occurred at the first site. The binding entropy however, suggested that the second site had an increase in binding disorder ($\Delta S>0$). Nevertheless, the Gibbs binding energy suggested that, at the analysis temperature, the reactions at the first and second site were both favourable ($\Delta G<0$).

The binding of indole-3-aldehyde to GSTF2_H77A showed a reduced binding activity, with K_D: 28.90 µM; less exothermic than the wild-type. With the entropy value less than 0 ($\Delta S<0$), the Gibbs energy law indicated that the reaction was favourable ($\Delta G<0$). The GSTF2_Q73L and GSTF2_Y97A exhibited a stronger binding activity, with K_D: 1.96 µM and 5.29 µM respectively. Compared to the wild-type, binding enthalpies for both mutants were recorded at less exothermic values but varied in entropy values. The binding reaction of GSTF2_Q73L was favourable with entropy value of $\Delta S>0$ in an exothermic condition of $\Delta H<0$. The entropy of indole-3-aldehyde binding to GSTF2_Y97A was less than 0 ($\Delta S<0$). However, the binding reaction was a favourable reaction, with $\Delta G<0$. Binding activity was not detected for GSTF2_R154A in assays with indole-3-aldehyde, as no detectable heat changes were observed from each titration (**Figure 3.26**).

Regarding the binding of camalexin, the only mutant that interacted with the ligand was GSTF2_R154A. The titration resulted in no detectable heat change during the binding of camalexin to the mutant. The representative figures for ITC titration for each experiment are shown in **Figure 3.17** to **Figure 3.28**.

Table 3.3. ITC analysis of GSTF2 wild-type and mutants with three ligands: quercetrin, indole-3-aldehyde and camalexin. Assays were carried out in phosphate buffered saline (PBS; 36 mM Na_2HPO_4, 64 mM Na_2HPO_4, 50 mM NaCl, 1% (v/v) DMSO, pH 7.0) at 25°C. The parameters were derived from one-site binding model except for GSTF2_Q73L titration with quercetrin, for which two-sites binding analysis was used instead. The error values represent standard deviation from two or three replicates. ND means non-detectable binding activity.

Interaction		Parameter				
		Stoichiometry (N)	Binding constant (K_D) (μM)	Enthalphy (-ΔH) kcal mol^{-1}	Entropy (ΔS) kcal mol^{-1}	Gibbs free energy (ΔG) kcal mol^{-1}
Quercetrin						
GSTF2		1.0	5.99	9.71E3 ± 2.5E3	-8.655 ± 8.0	-7125
GSTF2_Q73L	1	0.62	1.45	5.77E3 ±2.6E2	7.37 ± 0.71	-7967
	2	4.65	30.21	8.45E2 ± 1.0E2	17.8 ± 0.65	-6152
GSTF2_H77A		1.0	26.46	9.46E3 ± 4.69E2	-17.9 ± 1.7	-4123
GSTF2_Y97A		1.0	26.59	9.92E3 ± 5.7E2	-12.3 ± 2.2	-6253
GSTF2_R154A		1.0	27.32	9.60E3 ± 8.4E2	-11.3 ± 3.3	-6230
Indole-3-aldehyde						
GSTF2		0.692	24.27	1.18E4 ± 1.4E1	-18.55 ± 0.2	-6269
GSTF2_Q73L		0.692	1.97	7.07E3 ± 1.1E1	2.38 ±0.3	-7779
GSTF2_H77A		0.692	28.90	9.85E3 ± 1.3E2	-12.2 ± 0.6	-6212
GSTF2_Y97A		0.692	5.29	9.11E3 ± 3.1E2	-6.42 ± 1.3	-7196
GSTF2_R154A				**ND**		
Camalexin						
GSTF2		3.70	16.42	2.46E3 ± 4.9E2	13.6 ± 0.99	-6515
GSTF2_R154A				**ND**		

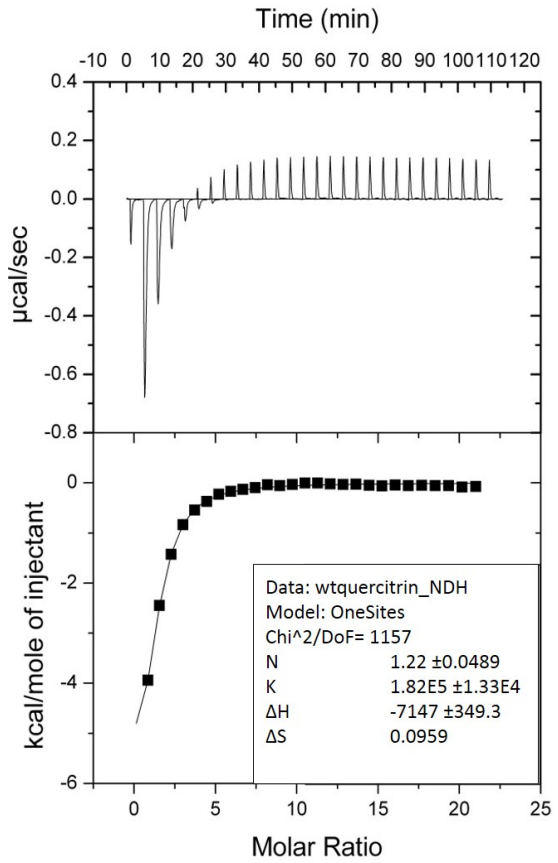


Figure 3.17. ITC of GSTF2 wild-type with quercetrin. Quercetrin (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2 in buffer. The binding affinity recorded was K_D: 5.41 µM.

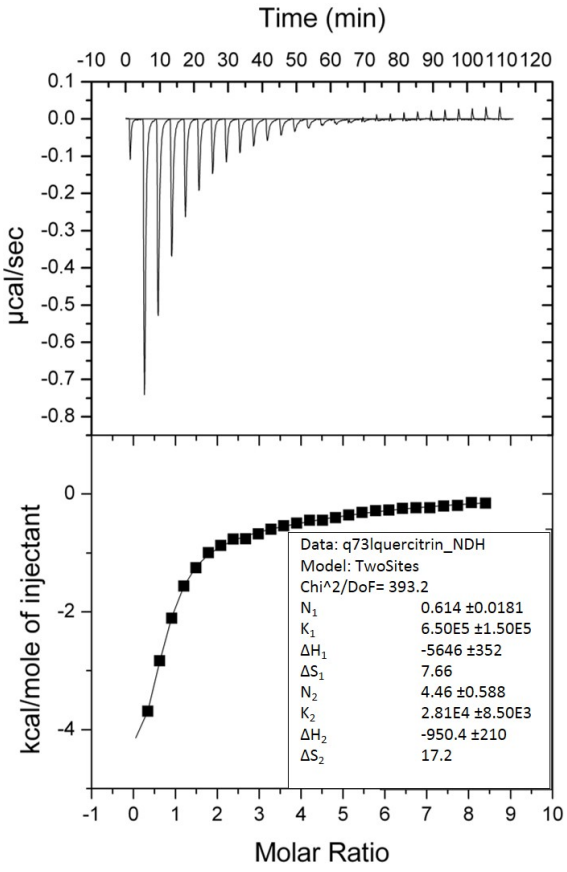


Figure 3.18. ITC of GSTF2_Q73L with quercetrin. Quercetrin (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_Q73L in buffer. The binding affinity for the first site was recorded at K_{A1}: 6.50E5 M^{-1} (K_D: 1.54 µM) whereas at the second binding site, K_D: 43.2 µM.

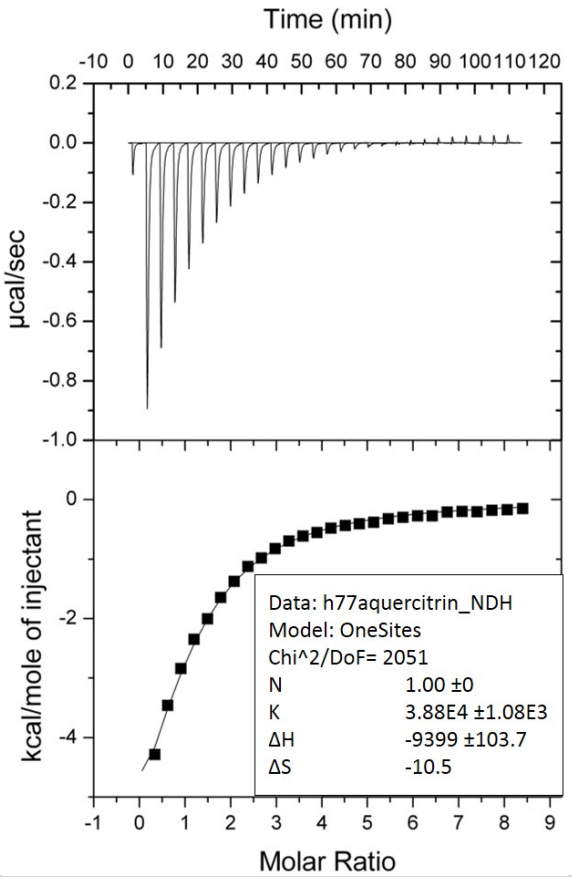


Figure 3.19. ITC of GSTF2_H77A with quercetrin. Quercetrin (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_H77A in buffer. The binding affinity for the first site was recorded at K_D: 25.77 µM.

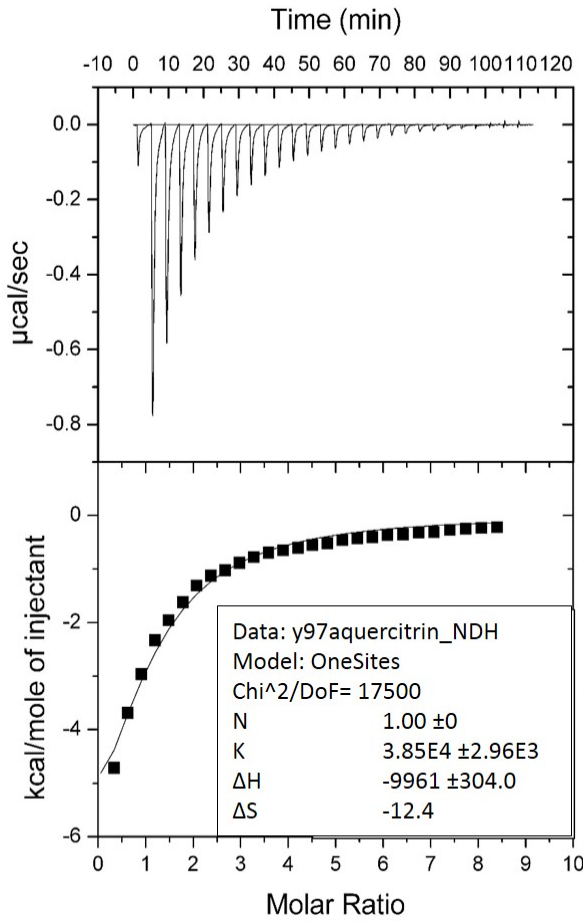


Figure 3.20. ITC of GSTF2_Y97A with quercetrin. Quercetrin (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_Y97A in buffer. The binding affinity for the first site was recorded at K_D: 25.97 µM.

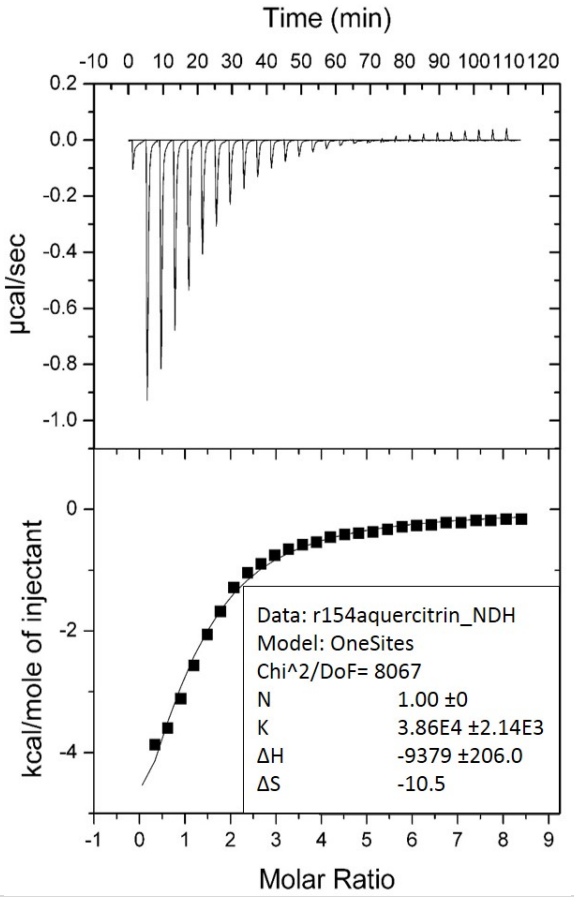


Figure 3.21. ITC of GSTF2_R154A with quercetrin. Quercetrin (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_R154A in buffer. The binding affinity for the first site was recorded K_D: 25.91 µM.

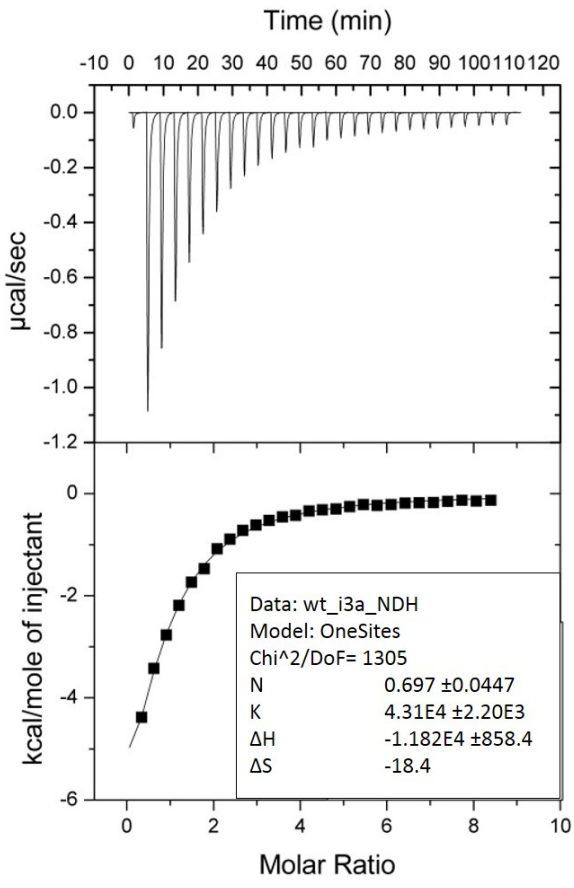


Figure 3.22. ITC of GSTF2 wild-type with indole-3-aldehyde. Indole-3-aldehyde (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2 in buffer. The binding affinity recorded was K_D: 23.20 µM.

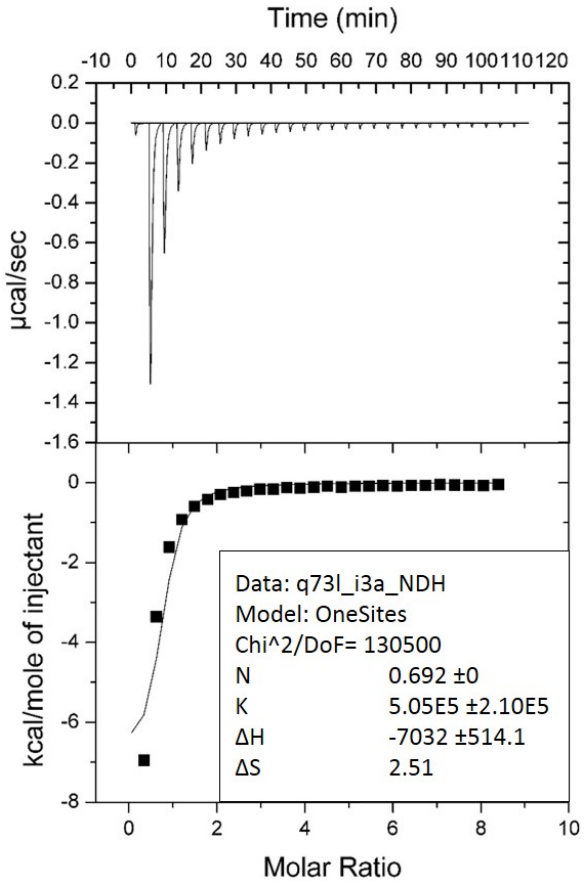


Figure 3.23. ITC of GSTF2_Q73L with indole-3-aldehyde. Indole-3-aldehyde (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_Q73L in buffer. The binding affinity recorded was K_D: 1.98 µM.

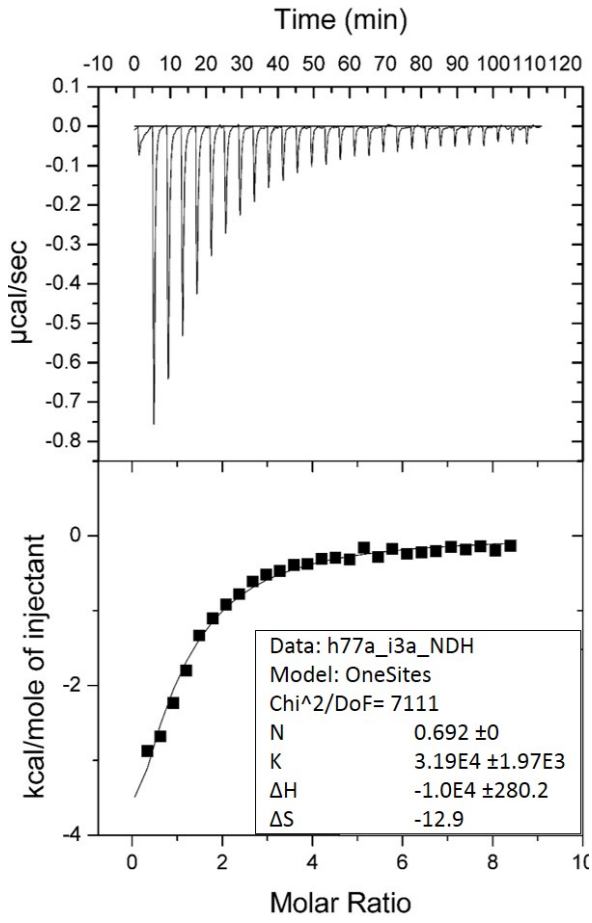


Figure 3.24. ITC of GSTF2_H77A with indole-3-aldehyde. Indole-3-aldehyde (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_H77A in buffer. The binding affinity recorded was K_D: 31.34 µM.

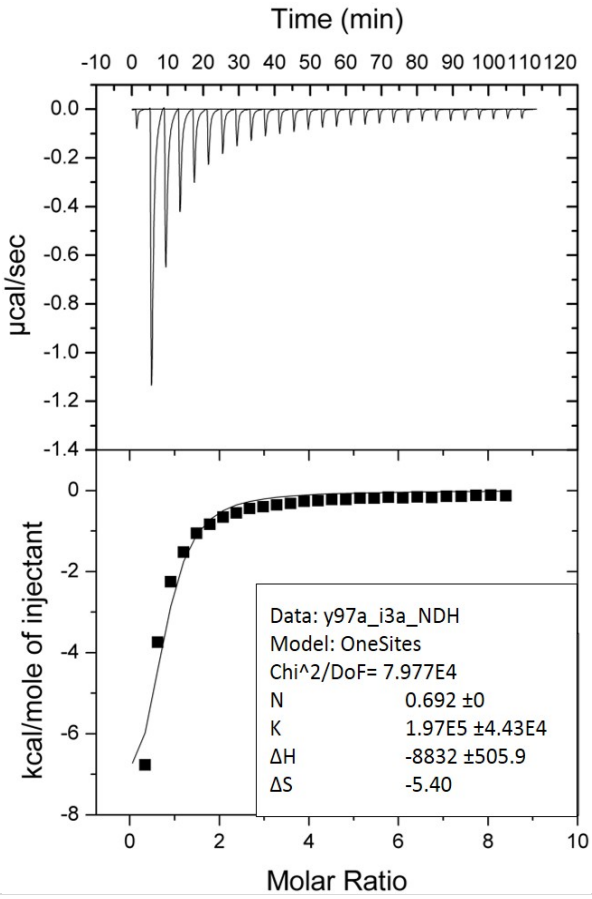


Figure 3.25. ITC of GSTF2_Y97A with indole-3-aldehyde. Indole-3-aldehyde (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_Y97A in buffer. The binding affinity recorded was K_D: 5.07 µM.

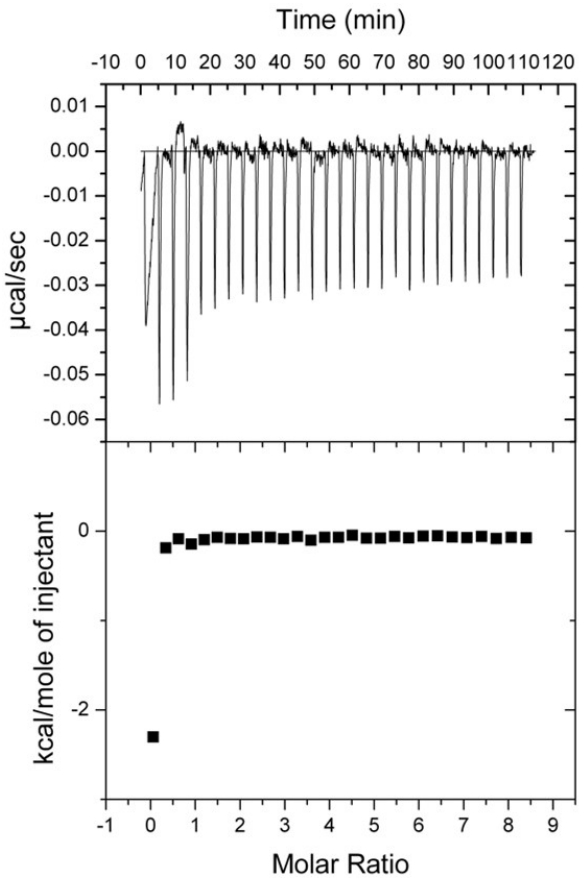


Figure 3.26. ITC of GSTF2_R154A with indole-3-aldehyde. Indole-3-aldehyde (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_R154A in buffer.

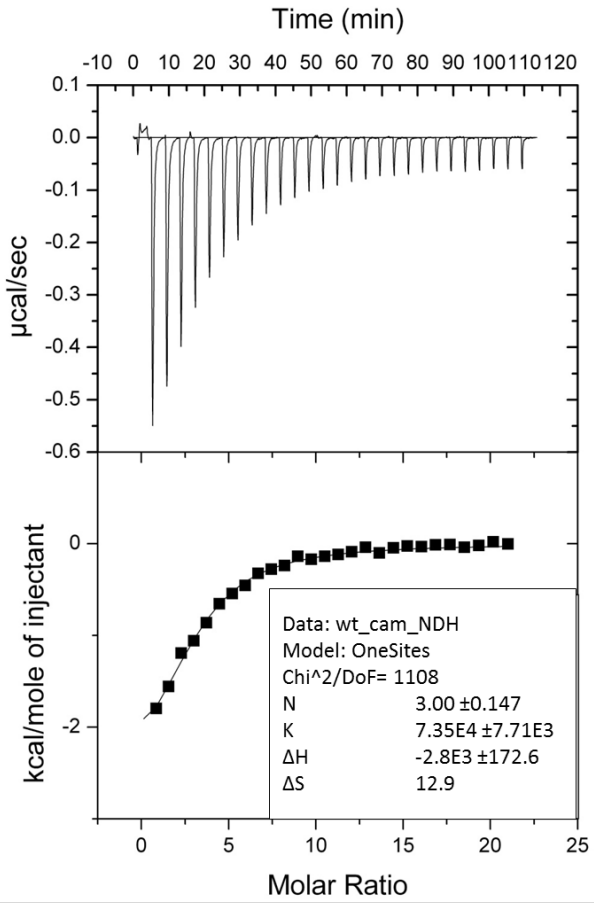


Figure 3.27. ITC of GSTF2 wild-type with camalexin. Camalexin (1 mM) in the syringe was titrated into the cell containing 0.01 mM GSTF2 in buffer. The binding affinity recorded was K_D: 30. 48 µM.

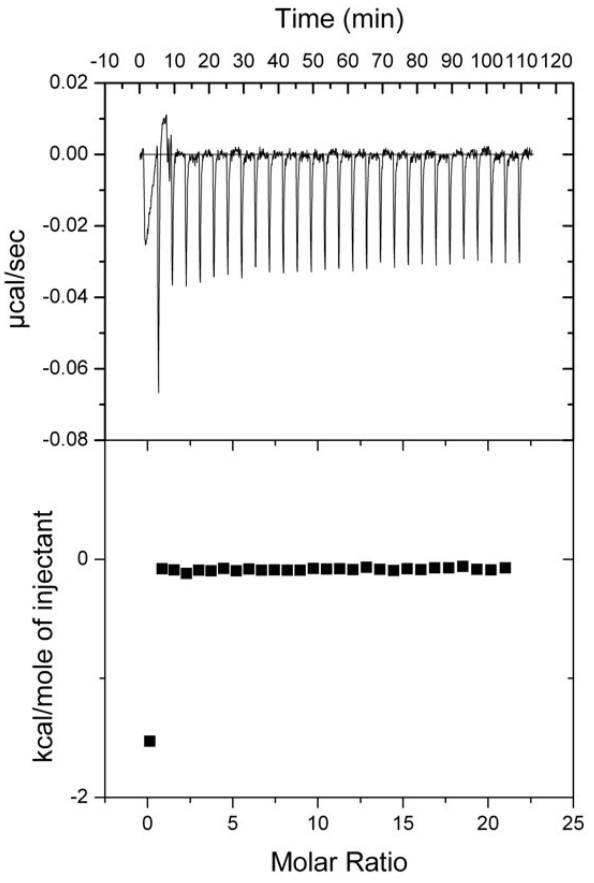


Figure 3.28. ITC of GSTF2_R154A with camalexin. Camalexin (1 mM) in the syringe was titrated into the cell containing 0.025 mM GSTF2_R154A in buffer.

3.3.8 Conformational change in GSTF2_Y97A

Based on the ITC analysis, the GSTF2_Y97A and GSTF2_Q73L mutants displayed four- and 12-fold higher binding affinities, respectively, if compared to the wild-type when reacting with indole-3-aldehyde. To examine this, purified GSTF2_Y97A was co-crystallised with indole-3-aldehyde in 0.1 mM of PCPT buffer (BIS-TRIS propane system, cacodylate and 25% PEG 1500 pH 7), with $MgCl_2$. The structure obtained was refined to 1.5 Å. The density of indole-3-aldehyde in the mutant structure was only at both **L1** sites on the dimer interface: two molecules of indole-2-aldehyde at **L1** sites in comparison to 3 molecules of the ligand at both **L1** site and one at **L2** site for the wild-type structure. The structures suggest that the mutated residue Y97 is important for the binding of indole-3-aldehyde in **L2** site. Removing the large side-chain residue of tyrosine and substituting with an alanine resulted in the inability of indole-3-aldehyde to bind at the **L2** site. The superimposed crystal structure of the GSTF2_Y97A-indole-3-aldehyde complex with the wild-type-indole-3-aldehyde complex yielded a RMS value of 0.64 Å over 414 amino acids. A slight change in the conformational structure of Y19A protein, from residue 123 to 125 of both chains can be viewed from the superimposed structure (**Figure 3.29; Figure 3.30**).

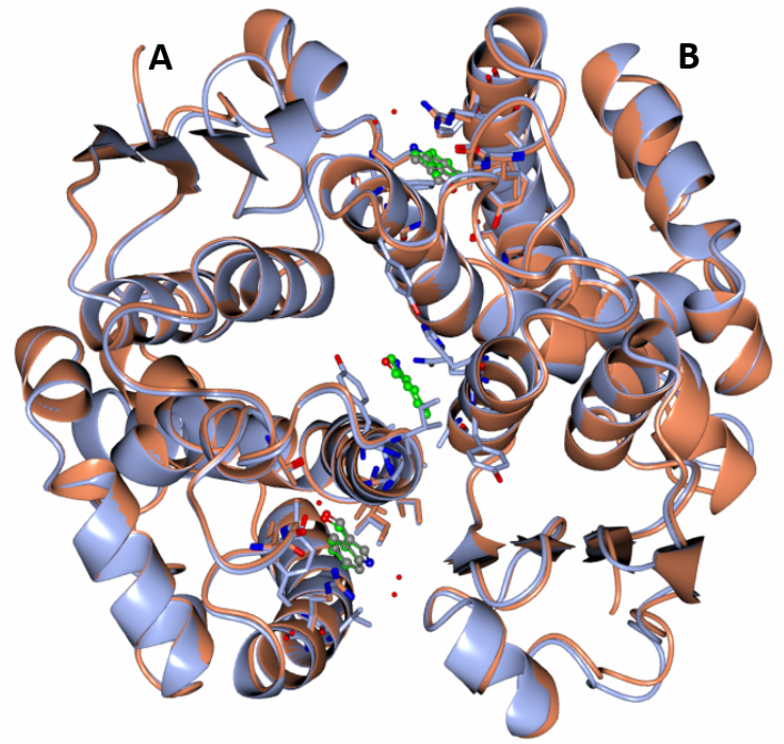


Figure 3.29. Superimposed structure of GSTF2_Y97A in orange with the wild-type in blue with indole-3-aldehyde attached. Indole-3-aldehyde molecules are shown in ball-and-stick format with carbon atom in grey (bound to GSTF2_Y97A) and green (bound to wild-type). Water molecules are shown in red.

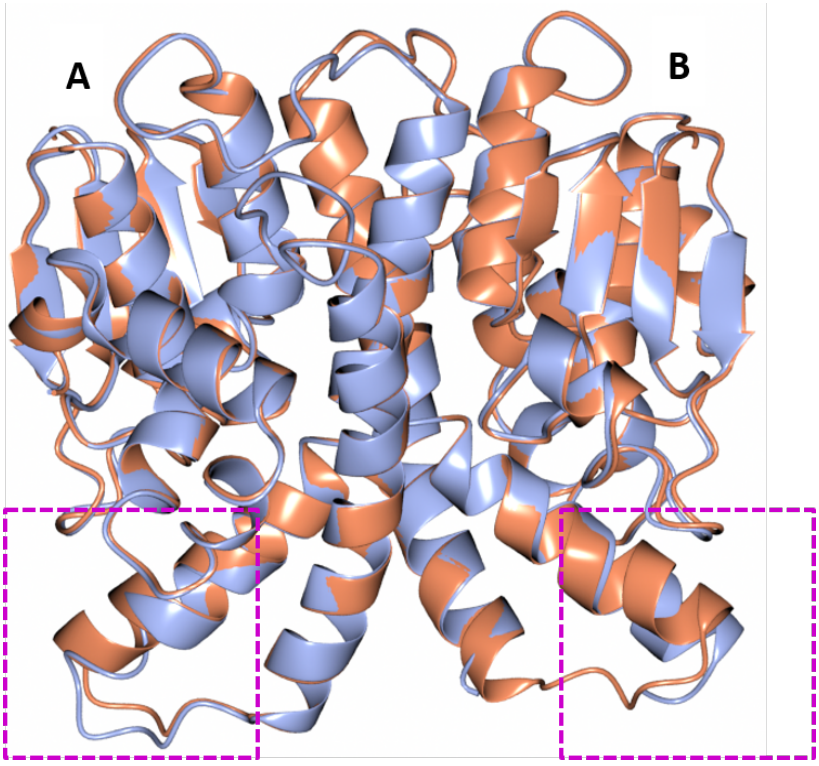


Figure 3.30. Superimposed structure of GSTF2_Y97A (orange) with the wild-type (blue) of both chains A and B. Regions with conformational change are highlighted in the magenta boxes.

3.4 DISCUSSION

Glutathione transferase is typically known to have roles in xenobiotic detoxification by catalysing the conjugation of GSH to xenobiotic compounds (Hayes *et al.*, 2005). However, several GSTs have shown the ability to interact directly with the ligand, without conjugation of GSH, rather acting as carriers or transporters; an activity known as ligandin (Sun *et al.*, 2012; Watahiki *et al.*, 1995; Smith *et al.*, 2003; Chen *et al.*, 2007). Following the detection of seven small molecules, including harmane, norharmane, lumichrome, indole-3-aldehyde, 1-acetyl-*β*-carboline, quercetrin and camalexin, binding to Arabidopsis GSTF2 by Dixon *et al.* (2011), it has become clear that GSTs, particularly GSTF2, can selectively recognise and bind a range of biologically active molecules.

This chapter presents the evidence for small molecule binding to GSTF2 through structural studies, mutagenesis and calorimetry analyses. The structure complex of GSTF2 with the ligands camalexin, quercetrin, quercetin and indole-3-aldehyde, revealed new, non-catalytic binding sites for small molecule ligands to GSTF2. These binding sites are different from the previously identified L-site that has been observed for GSTF2 in complex with *S*-hexylglutathione (PDB ID: 1GNW) (Reinemer *et al.* 1996; Prade *et al.* 1998).

The current study shows that the ligands bound at three major sites, two of the **L1** sites and one **L2** site, at the dimer interface. The GSTs site of interaction at a dimer interface with other ligands has been identified previously in other organisms. The distinct location of L-sites from the catalytic G-site and H-site have been recorded in GST structure from fluke worm *Schistosoma japonica* (*Sj*GST) (PDB: 1GTB), and the Sigma class GST from squid (PDB: 2GSQ) (Mctigue *et al.*, 1995; Ji *et al.*, 1996). The *Sj*GST structure indicated that the ligand antischistosomal drug praziquantel (PZQ) bound to the groove at the dimer interface of the protein and interacted with residues Q67, G97, L100, D101, Y104 and R108 side chains of one subunit and Y104 of the second subunit (Mctigue *et al.*, 1995). The structure of GST from squid showed the binding of two molecules of *S*-3-iodobenzyl-GSH conjugate (GSBz1) at two different sites. The first GSBz1 molecule bound at the G-site and the second GSBz1 molecule bound at the subunit interface and interacted with the hydrophobic residues of F8, L10, R13, V102 and F106 (Ji *et al.*, 1996). The superimposed structure of all three proteins revealed that the PZQ molecule is embedded deeper in the cavity between the subunits, while quercetrin and GSBz1 are located near the surface of the dimer **Figure 3.31**. Binding sites located at the surface of the protein perhaps signify more access to bulk solvent which provides protection to the bound ligand.

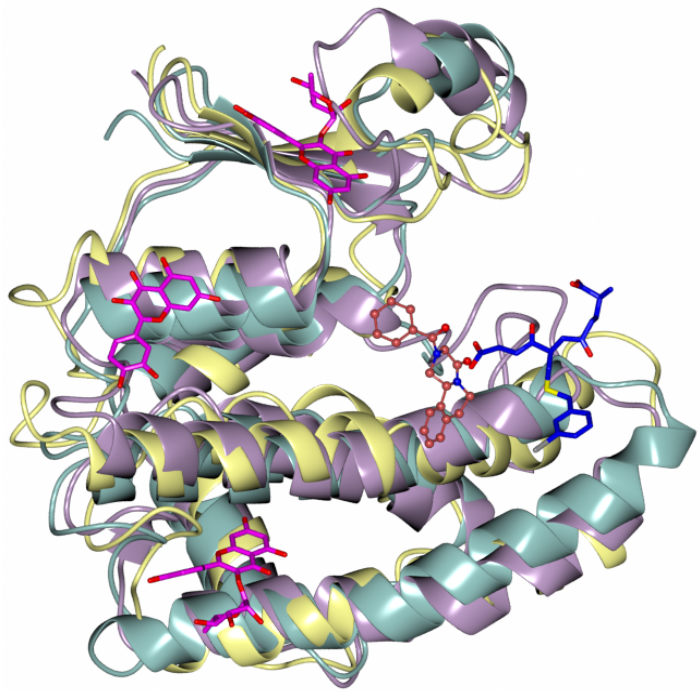


Figure 3.31. Cross section of a GSTF2 monomer in complex with quercetrin (PDB: 5A4W), superimposed with *Sj*GST with PZQ (PDB ID: 1GTB) and sigma class GST from squid complex with GSBz1 (PDB ID: 2GSQ). GSTF2 in yellow with magenta quercetrin, *Sj*GST in lilac with red PZQ and squid GST in green with blue GSBz1. The superimposition was done with GSTF2 complex fixed, with a RMS value of 2.35 Å over 171 residues for *Sj*GST and 2.50 Å over 169 residues for squid GST.

The **L1** and **L2** sites are dominated mainly by π-stacking, hydrophobic interactions and hydrogen bonds. The π-stacking particularly, is essential in small molecule substrate binding as it has been found abundantly in DNA-protein interactions (Wilson *et al.*, 2014). Similarly to the interface binding sites of *Sj*GST and squid GST, the mainly involved residues were the hydrophobic residues. In human GSTP1-1, although not located at the dimer interface, a hydrophobic surface pocket was also involved in the interaction between a ligand compound and a GST. The compound was a drug known as sulfasalazine; lodged in the hydrophobic area between the side chain of P9, V10, P202 and G205 of the GSTP1-1 (Hayeshi *et al.*, 2006).

The ITC studies allow an evaluation of the ligand-protein binding activity, by analysing the heat variations produced during the binding. The goal of ITC analysis is to model the experimental data using the simplest model and one that make sense in terms of what is already known about the reaction; the protein and ligand concentration. Firstly, a one-site binding model was employed for all data. The analysis was a curve fitting process by which a nonlinear regression protocol was used to fit a model to the data. The model, that neither

failed to fit the curve nor delivered higher error values than the parameter values, was further analysed using a more complex protocol i.e. two-sites binding model analysis. This was the case for quercetrin binding to GSTF2_Q73L that yielded a curve which could not be fitted using the one-site binding model. The data were then processed with a two-site model, which resulted in a fitted curve with credible parameter values. To gain confidence in this analysis method, and also because crystallographic data showed more than one binding site for each ligand, the two-sites model analysis was also employed on the data that were already fitted using the one-site model. The two-site analyses resulted in high error values for these data, although some of the curves could be fitted properly. The high error values are not an indication that the experimental data from the ITC analyses are compromised, but rather derive from the fact that the binding for two-sites analysis was beyond the detectable limit of the given experimental resources. For example, the maximum ligand concentration that could be used in the ITC experiment was 1 mM, but more concentrated (5-10 mM) ligands were used in the crystallography studies. This was because quercetrin, indole-3-aldehyde and camalexin have poor solubility in the buffer used for ITC analysis. Therefore, when one-site analysis was applied, the results were critically indicated for one of the sites; the site that has the detectable heat changes by ITC.

The stoichiometry values from the ITC analysis were different from the stoichiometry data derived from the crystallographic structure. Stoichiometry values derived from ITC are known to be affected by several factors, including the quality of the prepared proteins (particularly the amount of active protein and the non-functioning or misfolded proteins) and ligands. Different sample preparation for protein and ligands, as well as the buffer, could also cause the discrepancy between ITC stoichiometry and the crystallography stoichiometry. The analysis was further complicated by the poor solubility of the ligands used in this experiment.

The binding affinity of the ligands to GSTF2 was consistent with the chemical characteristics of the ligands. The largest and the most soluble ligand in this study, quercetrin, exhibited the highest affinity toward GSTF2, followed by indole-3-aldehyde and then camalexin. The lower affinity of ligand-mutants binding agrees with the hypothesis that the mutated residues are important for the interaction, except for GSTF2_Q73L and GSTF2_Y97A where a higher binding affinity was observed with indole-3-aldehyde. The complex structure of GSTF2_Y97A with indole-3-aldehyde showed that the mutant lacks one binding site at **L2**, where Y97 is located. The superimposed structure with wild-type protein in complex with indole-3-aldehyde however, revealed a different conformational change although not

significant as indicated by the RMS value of 0.64 Å over 414 atoms. The slight changes in the conformation of the mutants could potentially contribute to better interaction with the ligand and therefore resulting in higher binding affinity.

This study complemented the non-catalytic activity analyses of GSTF2 that have been reported (Dixon *et al.*, 2011; Smith *et al.*, 2003). This non-catalytic binding activity is not specific for Arabidopsis. Hydrophobic ligandin sites for two anthraquinone dyes VBAR (Vilmafix Blue A-R) and CB3GA (Gibacron Blue 3GA) have been characterised in maize *Zm*GST1 (Axarli *et al.*, 2004) and anthocyanin transport has been identified in *Petunia hybrid* GST*,* AN9 (Mueller *et al.*, 2008). In rats and humans, direct interaction with bilirubin, steroids, carcinogens and several organic anions have been determined (Litwack *et al.*, 1971; Simons & Jagt 1980). The work in this chapter reveals how GSTF2 interacts with different types of ligands. This interaction confirms that GSTs are more than GSH-conjugating enzymes and more investigations could be done to further understand this function.

4 Crystal structure of GSTU25

This research aims to investigate the ability of GSTU25 to catalyse the formation of TNT-GSH conjugates, specifically conjugate **3,** by elucidating the structure of GSTU25 in complex with the TNT-GSH conjugate. The results will be used to complement previous site-directed mutagenesis analysis carried out by Tzafestas (2016), who has investigated the key residues involved in the formation of the different conjugate profiles by GSTU25 and GSTU24. Here, the crystal structure of GSTU25 was determined in complex with disulphide glutathione (GSSG) at a resolution of 1.99 Å. One GSSG molecule bound to each GSTU25 subunit at the known catalytic binding site of GSTs (G-site). The structure of GSTU25 represents the first structure of Arabidopsis Tau class GSTs. The GSTU25 structure also validates the modelled structure by Tzafestas (2016) and most importantly, the GSTU25-GSSG complex provides the insight of GSTU25 folding upon substrate binding.

4.1 INTRODUCTION

4.1.1 The Arabidopsis GSTU25

A member of the Tau (U) class of GSTs, the gene encoding GSTU25 (*At*1g17180) is located from 5872145 to 5873086 bp on the forward orientation of *Arabidopsis thaliana* Columbia ecotype 0 (Col-0) chromosome 1 (TAIR, www.arabidopsis.org). In the genome, *GSTU25* is flanked by *GSTU24*, located 1.3 kb upstream and *GSTU26* 2.5 kb downstream; a pattern indicative of gene duplication. The *GSTU25* comprises of a 315 bp and a 351 bp exon which are linked together by an 85 base pair intron.

At the amino acid level, GSTU25 encodes a protein containing 221 amino acids with a predicted molecular mass of 26 kDa. Based on polypeptide similarity, GSTU25 is clustered in the third sub clade of the Tau GST along with GSTU19 to GSTU24 and GSTU26 to GSTU28, which all share more than 60% identity with GSTU25 (Dixon & Edwards 2010) (**Figure 1.10**).

4.1.2 The characterised role of GSTU25

Microarray data available through Genevestigator (Hruz *et al.*, 2008) indicate that *GSTU25* is highly expressed in roots and especially upregulated during seedling, seed germination and in the young rosetta developmental stage (**Figure 4.1**). The expression of *GSTU25* is

induced by hypoxia, iron deficiency, the hormones salicylic acid and 12-oxophytodienoic acid (OPDA), chemicals including fenclorim, hydrogen peroxide (H_2O_2) and infection by *Sclerotinia sclerotiorum*.

GSTU25 has high activity towards the standard GST substrate CDNB and also high GPOX activity towards cumene hydroperoxide but lacks activity towards lipid hydroperoxides. The protein has also exhibited preference to bind shorter chain length fatty acid-GSH adducts and hydroxylated fatty acids when expressed in *E. coli* (Dixon & Edwards 2009).

More recently GSTU25 was found to demonstrate enhanced tolerance and ability to detoxify the explosive compound 2,4,6-trinitrotoluene (TNT), a property that could be applied for phytoremediation of soils contaminated with TNT (Gunning *et al.*, 2014).

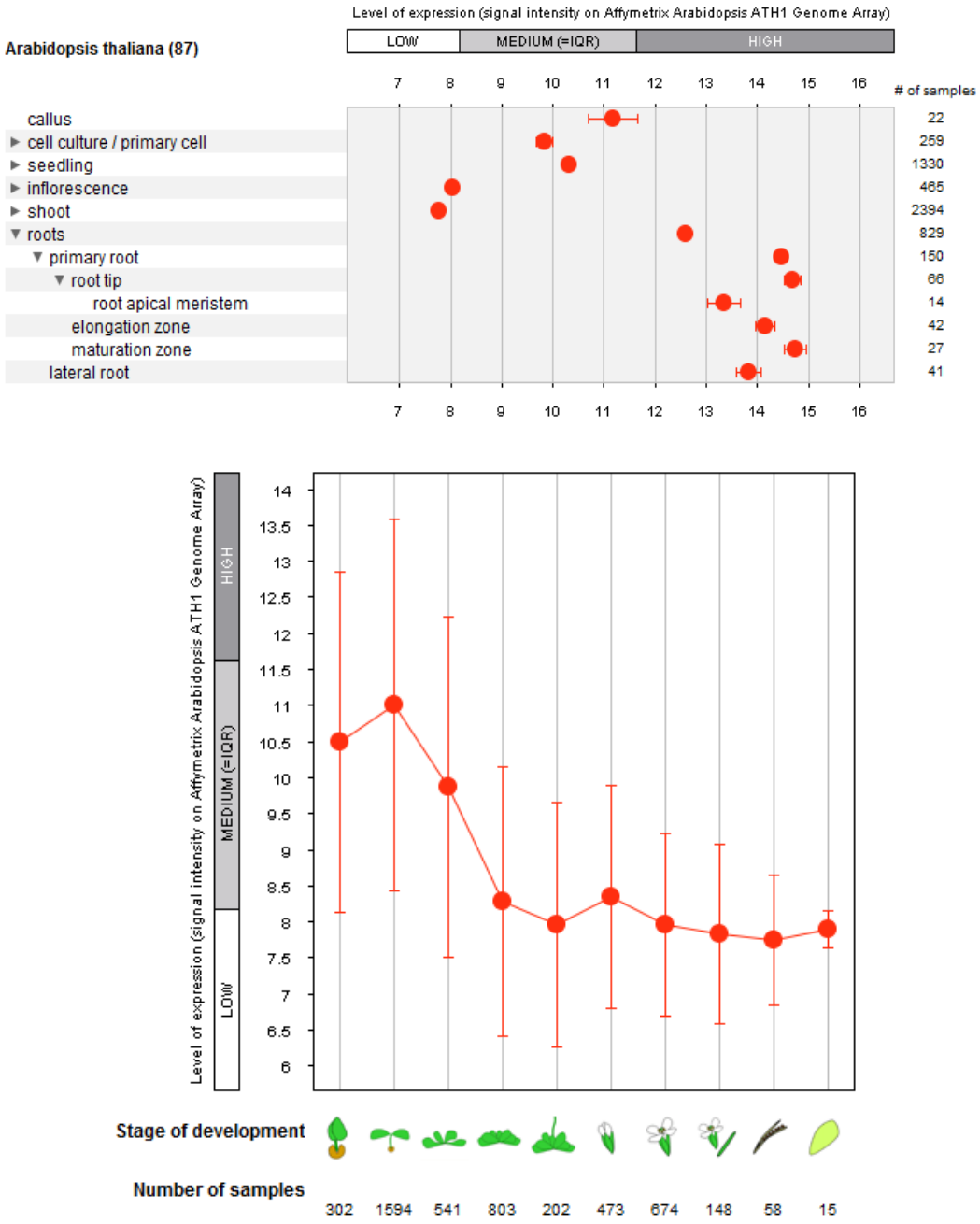


Figure 4.1. Expression of *GSTU25* gene in Arabidopsis. **Top panel**: *GSTU25* expression in different Arabidopsis compartment and; **Bottom panel**: *GSTU25* expression throughout different stage of Arabidopsis development. Level of expression in log2 scale. The stage of development is defined as; (from left) germinated seed, seedling, young rosetta, developed rosetta, bolting, young flower, developed flower, flowers and siliques and senescence. Source: Genevestigator (Hruz *et al.*, 2008), accessed 10th January 2017.

4.1.3 2,4,6-Trinitrotoluene (TNT)

4.1.3.1 An explosive and environmental pollutant

A yellow coloured compound, 2,4,6-trinitrotoluene (TNT) was invented by Joseph Wilbrand in 1863 and originally used as a yellow dye. Due to its stability and durability with reduced risk of accidental detonation compared to more sensitive explosives such as nitroglycerin, TNT was widely used in World War I and II. However, TNT is particularly resistant to degradation in the environment and as a result, high concentrations of TNT contamination has been reported at many military ranges worldwide (Amaral *et al.*, 2009; Lewis *et al.*, 2004; Zheng *et al.*, 2009). The contamination in training ranges can reach concentrations of 87,000 mg kg^{-1} with averages of 100 to 1,000 mg kg^{-1} detected in surface soils (Clark & Boopathy 2007; Jenkins *et al.*, 2006).

4.1.3.2 The chemical structure of TNT

The electron-withdrawing properties of the nitrite groups of TNT delocalise the electrons from the aromatic ring: turning it from electron rich to electron deficient (**Figure 4.2**). This makes oxidative attack on the aromatic ring difficult, preventing ring cleavage and hindering mineralisation of TNT. Thus, remediation of this molecule in the environment is extremely challenging.

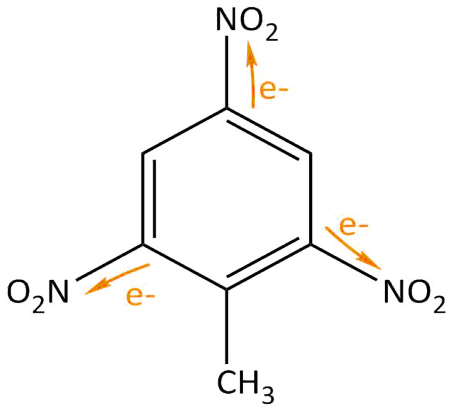


Figure 4.2. The chemical structure of TNT. The arrows indicate the electron-withdrawing properties of the nitrite groups that delocalise the electrons off the aromatic ring, making the aromatic ring electron deficient.

4.1.3.3 Detoxification of TNT in soil

In TNT-contaminated soil, microbial and plants diversity has been found although in a reduced amount to uncontaminated soil. In soil, the TNT binds tightly to the organic components of soil and is extremely recalcitrant to degradation (Rylott *et al.*, 2011). TNT has a high octanol-water partition coefficient [log K_{OW} values: 1.86 (Spain 1995)] and is most likely to enter the plant by passive diffusion across the plasma membrane, predominantly localising in root tissues (Brentner *et al.*, 2010). Enzymatic degradation of TNT has not yet been reported but several findings suggest that the TNT can be mineralised, although at low rates by bacteria in soil including *Pseudomonas* sp. and *Burkholderia* sp. (Nishino & Paoli, 2000; Spanggor *et al.*, 1991) and, species of fungi including *Phanerochaete chrysosporium* (Bumpus & Tatarko, 1994; Kim & Song, 2000; Rieble *et al.*, 1994).

In soil, bacteria has been found to reduce TNT molecules to hydroxylaminodinitrotoluenes (HADNTs), aminodinitrotoluenes (ADNTs) and diaminonitrotoluenes (DANTs) (Subramanian *et al.*, 2006; Wang *et al.*, 2003). Bacteria including *Enterobacter cloacae* and *Pseudomonas fluorescens* carrying flavoproteins such as pentaerythritol tetranitrate reductase and xenobiotic reductase XenB have the ability to transform TNT to monohydride-Meisenheimer (H-TNT) or dihydride Meisenheimer (2H-TNT) complexes (Williams *et al.*, 2004; Symons *et al.*, 2006). This mechanism results in a release of nitrite by the addition of hydride to the TNT aromatic ring. The 2H-TNT compound was found to form stable secondary diarylamines by a condensation reaction with hydroxylamines. The production of diarylamines was only detectable in bacteria.

A study in Arabidopsis indicated that oxophytodienoic acid reductases (OPRs) transformed TNT to nitroreduced TNT derivatives, HADNTs and ADNTs. Subsequently UDP-glycosyltransferase (UGT) formed conjugates with HADNT and ADNT, at either the 2-isomer or 4-isomer position to form *C*-glucosidic or *O*-glucosidic bonds (Beynon *et al.*, 2009; Gandia-Herrero *et al.*, 2008). These conjugates are most likely bound to macromolecular structures such as the cell wall components (Rylott *et al.*, 2015). The OPRs was also shown to produce both H-TNT and 2H-TNT complexes through reductive transformation of TNT (Beynon *et al.*, 2009).

The pathway for TNT detoxification in plants and bacteria is shown in **Figure 4.3**.

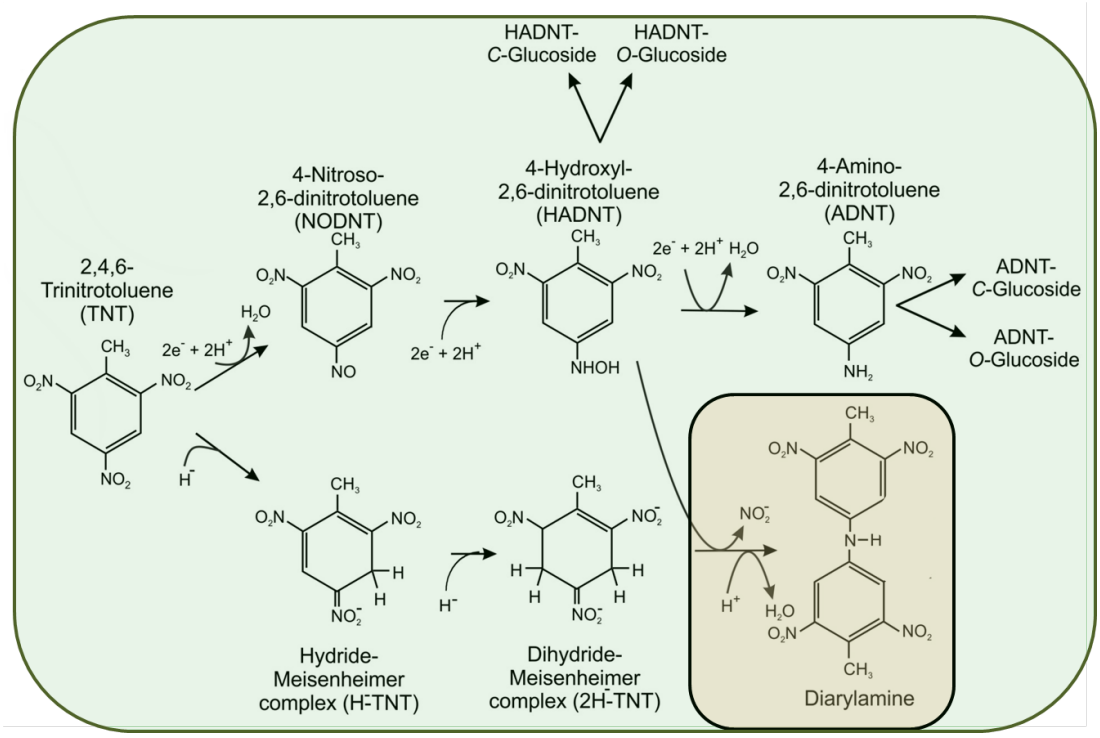


Figure 4.3. Biochemical pathway for the detoxification of TNT in plant and bacteria. The compound in the orange box was only detectable in bacteria. Figure adapted from Rylott 2011.

4.1.3.4 Role of GSTU25 in TNT detoxification

To understand more about the genes involved in the detoxification of TNT, Dr. Astrid Lorenz and co-workers in Prof. Bruce's group conducted microarray analysis (Lorenz 2007; Gandia-Herrero *et al.* 2008). Their preliminary experiments identified optimal conditions in which i) TNT could be uniformly applied, ii) the plants were recorded as having relatively low levels of stress, and iii) expression was limited to genes directly affected by TNT. To meet these criteria, 14 day-old Arabidopsis seedlings were grown in liquid culture followed by treatment with 60 µM TNT for 6 h and subject to microarray analysis to determine the responsive genes. The results of the microarray analysis revealed the upregulation of 633 genes by 2- to 204-fold. Among the upregulated genes were UGTs and OPRs which were subsequently shown to contribute towards the detoxification of TNT (Beynon *et al.* 2009), cytochromes P450, protein kinases and GSTs. The GSTs recorded from the analysis were almost exclusively members of the Tau class GSTs including *GSTU1 - U4, U7 - U9, U11, U12, U22, U24* and *U25,* with *GSTU25* and *GSTU24* the most highly upregulated, by 46 and 37-fold, respectively.

Similar studies in Arabidopsis by Ekman *et al.* (2003) reported that transcripts of *GSTU22*, *GSTU24*, *GSTU1* were upregulated following TNT treatment, and Mezzari *et al.* (2005)

showed that *GSTU1* and *GSTU24* were significantly induced by TNT. In Prof. Neil Bruce's group, the GSTUs identified from the microarray analysis were cloned, overexpressed in *E. coli* and tested for activity towards TNT by Dr. Helen Sparrow using the Griess assay (Griess 1879). This colorimetric assay analyses the concentration of released nitrite in a reaction. The released nitrite binds to sulphanilamide to form a diazonium cation which then couples to the aromatic amine *N*-(1-napthhyl)ethylenediamine (NED), producing a red-violet coloured product, with a maximum absorption at ~540 nm. Among the GSTUs tested (U1, U3, U4, U7, U12, U22, U24 and U25), U24 and U25 showed the highest levels of nitrite release (5% and 20% respectively), after incubation with TNT for 24 h (Sparrow 2010).

Subsequently, the mechanism of TNT detoxification was shown to involve the removal of nitrite from the aromatic ring of TNT (Gunning *et al.*, 2014; Lorenz 2007; Sparrow 2010). Purified GSTU24 and GSTU25 catalyse a TNT-GSH conjugating activity; K_m: 1.6 mM and V_{max}: 369.3 nkat mg^{-1} for U24 and K_m: 1.2 mM and V_{max}: 393.6 nkat mg^{-1} for GSTU25 (Gunning *et al.*, 2014). HPLC analysis demonstrated that these proteins produced three different TNT-GSH conjugates as shown in **Figure 4.4**.

Although GSTU24 and GSTU25 share 79% protein identity, they produce different profiles of TNT-GSH conjugates, depending on the pH and temperature of the reaction condition. At pH 6.5 and 9.5 with temperature 20°C, GSTU25 produced more conjugates **1** and **3** while GSTU24 predominantly produced conjugate **2** (**Figure 4.5**). The GSTU24 showed the highest rate of TNT conjugation at temperatures between 30°C to 37°C, whereas GSTU25 showed an increase of TNT-conjugating activity even at 50°C, indicating high stability of GSTU25. At temperatures above 60°C, both GSTs were rapidly inactivated (Gunning *et al.*, 2014).

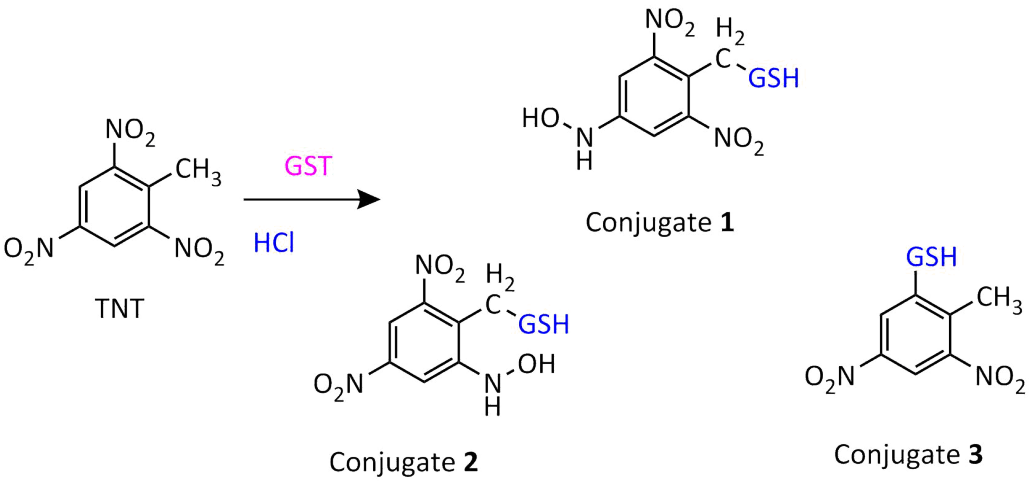


Figure 4.4. Chemical structure of TNT and three different TNT-GSH conjugates. The GSH conjugation occurs at the methyl group for conjugate **1** and **2** whereas for conjugate **3**, the GSH conjugation occurs by the substitution of a nitrite group at the 2/6 position, eliminating one nitrite group. Figure adapted from Gunning *et al.* (2014).

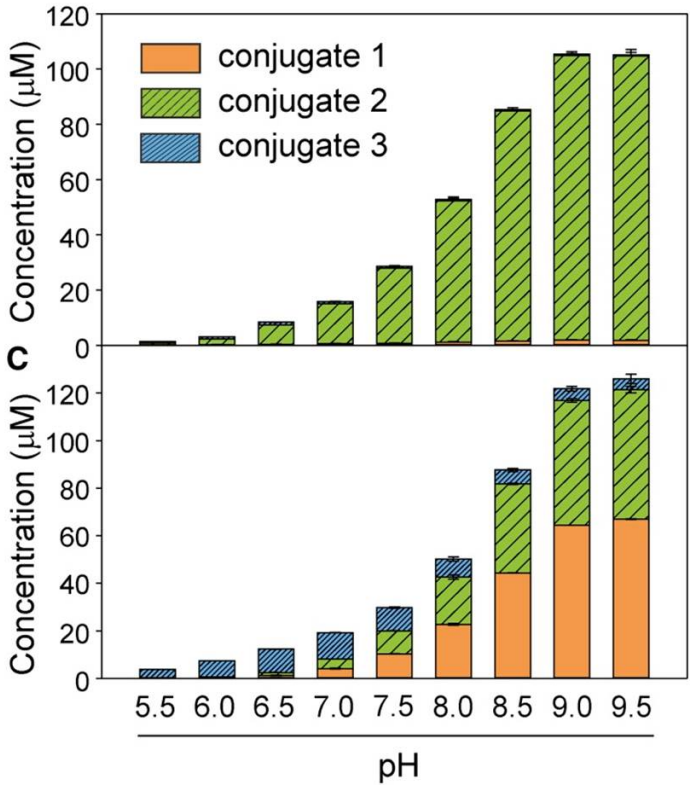


Figure 4.5. Types of GSH-TNT conjugate produced by GSTU24 (top panel) and GSTU25 (bottom panel) after 1 h of incubation with the TNT in increasing pH environment. The conjugation products were analysed in a reaction performed at 20°C in 100 mM phosphate with 150 µg of GST, 200 µM TNT and 5 µM GSH by HPLC and Griess assays. Results are the average of three biological replicates ± SD. Figure reproduced from Gunning *et al.* (2014).

Conjugate **3** is formed by the replacement of a nitrite on TNT ring with sulfur from the GSH, catalysed mainly by the GSTU25. This denitration destabilises the overall aromatic structure and promotes the opportunity for subsequent degradation and mineralisation of the TNT.

Alongside the study presented here, Dr. Kyriakos Tzafestas in Prof. Neil Bruce's group, carried out a mutagenesis study to investigate the residues responsible for TNT-conjugating activities of GSTU24 and GSTU25 (Tzafestas 2016). These residues were identified by sequence alignment and homology modelling with structurally-characterised Tau class GSTs, *Triticum aesvitum* (wheat) *Ta*GSTU4-4 and *Glycine max* (soybean) *Gm*GSTU4-4. The structure of *Ta*GSTU4-4 was in complex with *S*-hexylglutathione (Thom *et al.*, 2002) and the *Gm*GSTU4-4 structure included the conjugate of *S*-(*p*-nitrobenzyl)-GSH (Axarli *et al.*, 2009). Sequence alignment analysis in **Figure 4.6** showed the residues important for substrate specificity of the Tau class GSTs (highlighted in red), and reveal the difference in residues at position 107 between GSTU25 and GSTU24.

Figure 4.6. Sequence alignment of GSTU24, GSTU25, *Ta*GSTU4-4 and *Gm*GSTU4-4 by ClustalW indicated conservation for both GSTU24 and GSTU25 at all residues important for substrate specificity (red boxes) in Tau class GSTs, except for residue at position 107 (as marked by the yellow star). Figure reproduced from Tzafestas (2016).

Subsequent homology modelling of GSTU25 using *Gm*GSTU4-4 that is 60% identical to the sequence of GSTU24 and GSTU25 showed that residues at position 12, 115, 209 and 212 of GSTU25 were oriented towards the centre of the active site, suggesting their role in the binding of the substrate. Thus, these five residues; P12, Y107, G115, V209 and L212 of GSTU25 were targeted for site-directed mutagenesis (Tzafestas 2016).

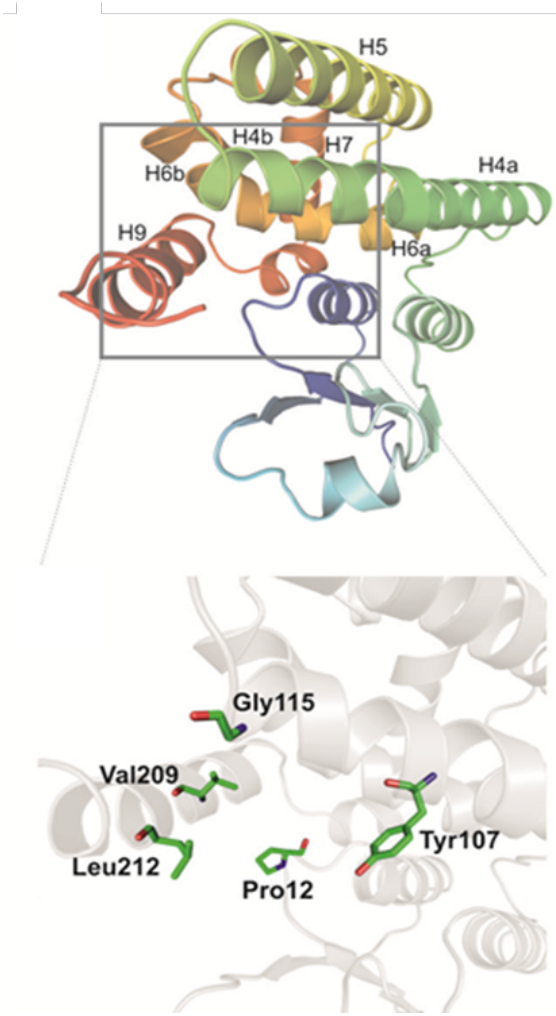


Figure 4.7. Model of GSTU25 built by homology modelling using *Glycine max* GSTU4-4 (PDB ID: 2VO4) as template. The five residues included for mutagenesis are displayed in the bottom panel. Figure reproduced from Tzafestas (2016).

Tzafestas (2016) showed that by switching the residues at position 12 and 107 for GSTU24 and GSTU25, that the conjugate produced was also switched (**Figure 4.8**). This finding suggests that both residues are critically important in the production of specific TNT-GSH conjugate for each GST.

Enzyme	Mutation identifier	Substitution
GSTU24	A	A12P
	B	N107Y
	C	A115G
	D	I208V
	E	R212L
GSTU25	F	P12A
	G	Y107N
	H	G115A
	I	V208I
	J	L212R

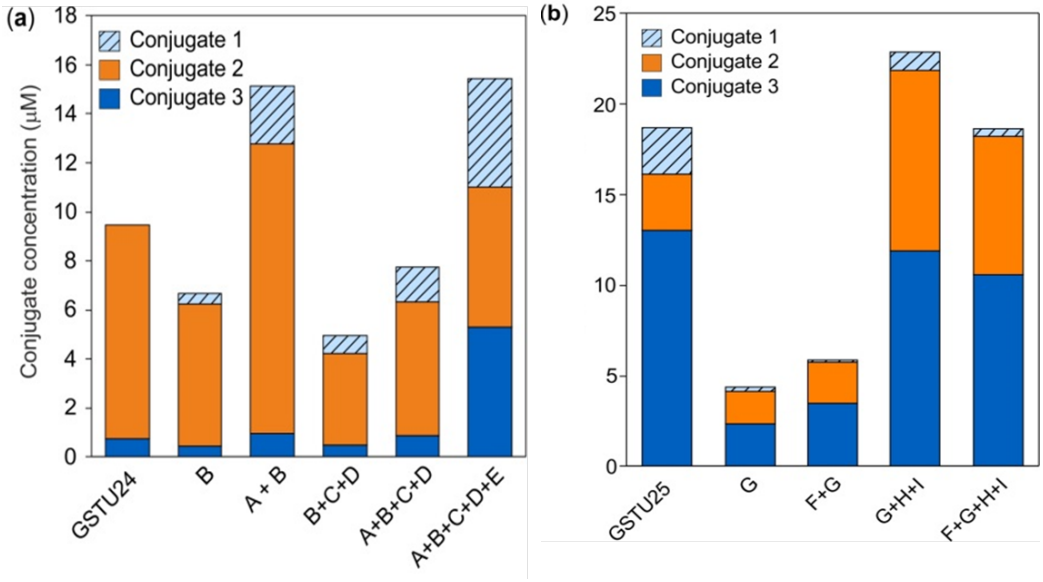


Figure 4.8. The residues involved in the mutagenesis study. TNT-GSH conjugate profiles produced by purified wild-type and mutant of GSTU24 and GSTU25 (a) GSTU24 (b) GSTU25 after 1 h with 200 µM TNT at 20°C. The results are means of three technical replicates. Figure reproduced from Tzafestas (2016).

4.2 MATERIAL AND METHODS

4.2.1 The preparation of GSTU25 protein for crystallisation

The pET22b-GSTU25 construct, without a tag, was generated using the In-fusion cloning method (Clonetech Laboratories, USA) with primers listed in **Table 4.1**. The GSTU25 sequence was amplified in a reaction mix containing 1 x CloneAmp HiFi PCR Premix (supplied by the manufacturer), 300 nM reverse primer (GSTU25_R), 300 nM forward primer (GSTU25_F) and 1 µL DNA template (GSTU25 cloned in pET-YSBLIC3C vector), prepared in a total volume of 25 µL. The amplification reaction was 10 s at 98°C, 15 s at 55°C and 7 s at 72°C for 33 cycles. The amplified products were analysed on an agarose gel and the DNA band for GSTU25 excised and purified. Following purification, the product was mixed in an In-Fusion cloning reaction containing 2 µL of 5X In-Fusion Premix, 100 ng of purified product and 100 ng of linearized pET22b (Merck, UK) in a total reaction of 5 µL. The reaction mix was incubated for 15 min at 50°C for ligation, and transformed into XL-10 Gold Ultra-competent *E. coli* cells (Strategene, UK) followed by plasmid purification and DNA sequencing. The constructs were then transformed into *E. coli* Tunetta cells (Edward group stock) as described in **Section 2.4.9**.

Table 4.1. Primers used for PCR amplification of GSTU25.

Primer	Nucleotide (5'-3')	TM (°C)	% GC
GSTU25_F	GGAGATATACATATGGCAGACGAGGTGATTCTTCTT	70	42
GSTU25_R	GTGGTGGTGCTCGAGTTATTCGATTTCGATCCCAAGTTT	78	46

4.2.2 GSTU25 protein expression

Protein expression processes were performed according to the method described for GSTF2 in **Section 3.3.2**.

4.2.3 GSTU25 protein purification

4.2.3.1.1 Purification of His-tagged GSTU25

The soluble supernatant was loaded onto 1 mL His-Trap HP column (GE Healthcare). The protein was eluted using gradient of 0-300 mM imidazole in 50 mM Tris HCl pH 7.5 containing 500 mM NaCl. The purified enzyme was desalted using 10 kDa cutoff Centricon filters (50 mL) and washed twice with 50 mM Tris HCl pH 7.5 to remove imidazole and NaCl.

4.2.3.1.2 Purification using GSH chromatography and size exclusion chromatography (SEC)

Affinity chromatography using GSH column and size exclusion chromatography using HiLoad 16/60 Superdex 75 or Superdex 200 PrepGrade (GE Healthcare, UK), was performed according to method described for GSTF2, **Section 3.3.3**. Protein yield of 17 mg mL^{-1} from 1 L of the starting culture was obtained and used in this study.

4.2.4 GSTU25 protein crystallisation

The ligand-protein complexes were subjected to crystallisation trials using a Mosquito® ROBOT (TTP LabTech) and commercially available crystallisation screens including PACT premier™ HT-96 (Molecular Dimension), Index HT (Hampton Research) and Crystal Strategy™ Screen 1 and 2 (Hampton Research). The screening was carried out in a 96-well plate in a sitting drop format, with each drop containing 150 nL protein and 150 nL of screening solution. To get the crystals of GSTU25, the method by Aleku *et al.* (2016) which recommended incubation of ligand in the screening buffer before the addition of protein was employed. In this experiment, 2 mM TNT and 2 mM GSH were incubated for 1 hour in 54 µL of crystal screening buffer. The 2 mM TNT was dissolved in dimethyl sulfoxide (DMSO) and 2 mM GSH was prepared in 20 mM Tris-HCl buffer pH 7.5. The pure GSTU25 was then subjected to crystallisation trials using a Mosquito® ROBOT (TTP LabTech).

Larger crystals for diffraction analysis were obtained from the optimised buffer in the hanging drop vapour diffusion method in 24-well Linbro dishes, in which 2 µL drops containing 1:1 ratio of reservoir solution to protein solution were pipetted onto premade siliconized cover slips which were then sealed onto the wells with high-vacuum grease. Again, in the hanging drop format, 2 mM of GSH and 2 mM of TNT were resuspended in the screening buffer and incubated for 1 h before adding GSTU25. Prior to analysis with the in-house X-ray equipment, the crystals were washed with the screening solution added with 20% (v/v) ethylene glycol as cryoprotectant, followed by flash-cooling in liquid nitrogen. Crystals were tested for diffraction using Rigaku Micromax-007HF robot fitted with Osmic multilayer optics and a MARRESEARCH MAR345 imaging plate detector by Sam Hart in York Structural Biology Laboratory, University of York. Crystals that diffracted to a resolution of equal to or better than 3 Å resolution were retained for dataset collection.

4.2.5 Data collection, structure solution, model building and refinement

The complete dataset described in this report was collected at the Diamond Light Source, Didcot, Oxfordshire, UK on beamline I02. The data were processed and integrated using the

X-ray Detector Software (XDS) (Kabsch 2010) and scaled using SCALA program (Evans 2006) included in the Xia2 processing system (Winter 2010). The structure of GSTU25 was obtained in space group $P2_12_12_1$, with four molecules in the asymmetric unit, constituting two dimers. The structure was solved by Prof. Gideon Grogan, using the Molecular Replacement program (MOLREP) (Vagin & Teplyakov 1997), with the structure of the Tau GST from *G. max* (PDB ID: 4TOP; 65% sequence identity to GSTU25) as template. The solvent content in the crystals was 51%. Structures were built and refined using the iterative cycles of Crystallographic Object-Oriented Toolkit (Coot) (Emsley & Cowtan 2004) and REFMAC5 molecular refinement program (Murshudov *et al.*, 2011), by employing local Non-Crystallographic Symmetry (NCS) restraints in the refinement cycles. After building and refinement of the protein and addition of water molecules, clear residual density was observed in the omit maps at the GSH binding site which was modelled as glutathione disulphide (GSSG). The final structure exhibited % R-cryst and R-free values of 20.5 and 21.7 respectively. Data collection and refinement statistics for GSTU25-GSSG complex are listed on **Table 4.2**. All structures were validated and checked using PDB validation software upon deposition. The Ramachandran plot for GSTU25-GSSG showed 98.4% of residues to be situated in the most favoured regions, 1.1% in additional allowed and 0.5% residues in the outlier regions. The coordinates have been deposited in the Protein Data Bank (PDB) with the accession code of 5G5A.

Table 4.2. Data collection and the refinement statistics for GSTU25-GSSG complex. Numbers in brackets refer to data for highest resolution shells.

	GSTU25 with GSSG
Beamline	Diamond I02
Wavelength (Å)	0.97949
Resolution (Å)	48.54-1.95 (1.99-1.95)
Space Group	$P2_12_12_1$
Unit cell (Å)	a = 87.83; b = 107.67; c = 108.75' a = b = g = 90°
No. of molecules in the asymmetric unit	4
Unique reflections	75638 (4446)
Completeness (%)	99.8 (100.0)
R_{merge} (%)	0.07 (0.54)
$R_{p.i.m.}$	0.05 (0.36)
Multiplicity	6.4 (6.2)
$<I/\sigma(I)>$	12.7 (3.0)
Overall *B* factor from Wilson plot (Å^2)	25
R_{cryst}/ R_{free} (%)	20.5/21.7
r.m.s.d 1-2 bonds (Å)	0.02
r.m.s.d 1-3 angles (^{o})	1.94
Average main chain B (Å^2)	31
Average side chain B (Å^2)	35
Average water B (Å^2)	42
Average ligand B (Å^2)	40

4.3 RESULTS

4.3.1 Expression and purification of GSTU25

The gene encoding full–length of GSTU25 from *Arabidopsis thaliana* incorporating a 6x His-tag, had previously been cloned in the pET-YSBLIC3C expression vector (Gunning *et al.*, 2014). Purification of GSTU25 using a HisTrap HP column (GE Healthcare, UK) did not yield a pure single band, as viewed on an SDS-PAGE gel (**Figure 4.9a**). An additional step of purification using size exclusion chromatography (SEC) to separate the expected band of GSTU25 from the larger-size proteins visible on SDS-PAGE gel did not produce homogeneous protein; however, the fractions were collected and stored in -80°C for crystallography screening. The identity of the protein band (shown in red box) was determined using Matrix Assisted Laser Desorption/Ionisation Time-of-Flight mass spectrometry (MALDI-TOF) and confirmed as GSTU25. When this fraction was analysed on a 10% native gel, the protein was observed as regularly-spaced, with the multiple bands indicating possible protein denaturation and aggregation in comparison to GSTF2, which was purified about five months earlier (**Figure 4.9c**).

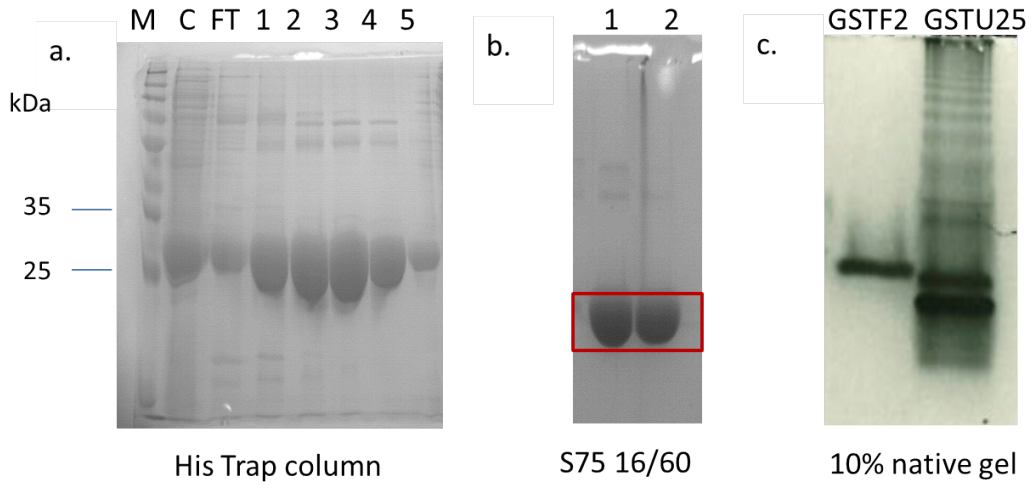


Figure 4.9. The protein purification profiles of GSTU25 cloned in the pET-YSBLIC3C expressed by *E. coli* Tunetta. a) Purification profile from his-tag affinity chromatography using HisTrap column (GE Healthcare). M: prestained protein marker, C: crude lysate, FT: unbound fractions in flow-through, 1-5: purified fractions, b) size exclusion chromatography (SEC) using HiLoad 16/60 Superdex S75 PrepGrade gel filtration column (GE Healthcare, UK), excised and submitted for protein identity analysis. 1 and 2: purified protein fractions on 12% SDS-PAGE gel, c) purified protein fractions after SEC were analysed on 10% native gel.

4.3.2 The GSTU25 from pET-YSBLIC3C construct failed to crystallise

Despite the detection of contaminating bands in purified GSTU25 fractions, the protein fractions that eluted from the SEC column were concentrated to 18 mg mL^{-1} and used for crystal screens. The crystal screening with commercial screening buffers (as mentioned in **Section 3.3.5**) was performed using a Mosquito Crystal robot (TTP Labtech). In each well of the 96-well dishes, 54 µL of screening buffer was dispensed. Co-crystallisation was performed by incubation of 18 mg mL^{-1} GSTU25 with 2 mM quercetrin. Quercetrin was used for two reasons: to replicate the success of GSTF2 crystallisation and because quercetrin was the most soluble ligand used successfully for the crystallisation of GSTF2. The mixture of GSTU25 and quercetrin was incubated in drops of the crystal screening buffers, in a sitting drop format, each at a total volume of 300 nL (150 nL protein and 150 nL crystal screening solutions).

Within a week, crystals were observed in several wells. The biggest single crystal was observed in the ammonium sulfate drop containing 3.2 M ammonium sulfate and 0.1 M N-N-bis(2-hydroxyethyl)glycine (BICINE) at pH 9. The size of the crystals was improved by optimisation in 1 µL drop in a hanging drop format. Upon analysis using in-house X-ray equipment, a Rigaku Micromax-007HF in YSBL, the crystals were found to be salt (**Figure 4.10**).

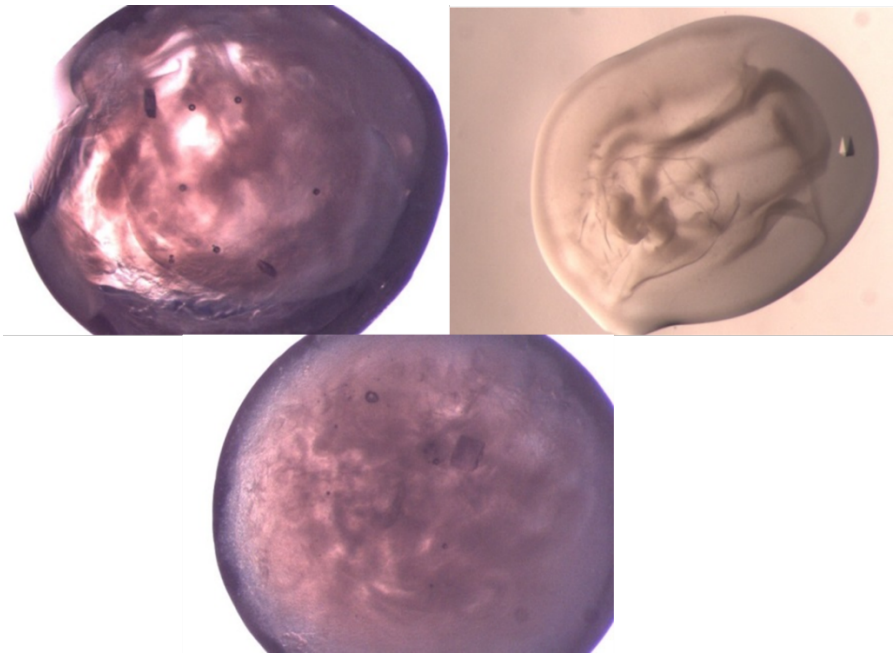


Figure 4.10. The crystals observed in ammonium sulfate buffer. The X-ray diffraction confirmed that they were salt crystals.

4.3.3 Purification of GSTU25 after sub-cloning into pET22b

The failure to obtain protein crystals from purified protein derived from the GSTU25-pET-YSBLIC3C construct was believed to be due to the presence of protein impurities in the eluted fractions. Therefore, the expression vector was switched to a different vector; pET22b, designed to contain a stop codon before the 6x-his codons to enable the expression of GSTU25 protein without the affinity tag. The recombinant vector was overexpressed in *E. coli* Tunetta cells and purified using GSH affinity chromatography and further purified by SEC using HiLoad 16/60 Superdex 200 PrepGrade gel filtration column GE Healthcare, UK). Purified protein was successfully obtained from the GSH chromatography with some degradation observed below the expected protein band for GSTU25 (**Figure 4.11a**). The purified fractions were pooled and loaded on an SEC S200 column, finally yielding a cleaner single band when viewed on an SDS-PAGE gel (**Figure 4.11b**). The fractions from the SEC column were then pooled, concentrated to 17 mg mL^{-1} and stored at -80°C until crystallisation screening.

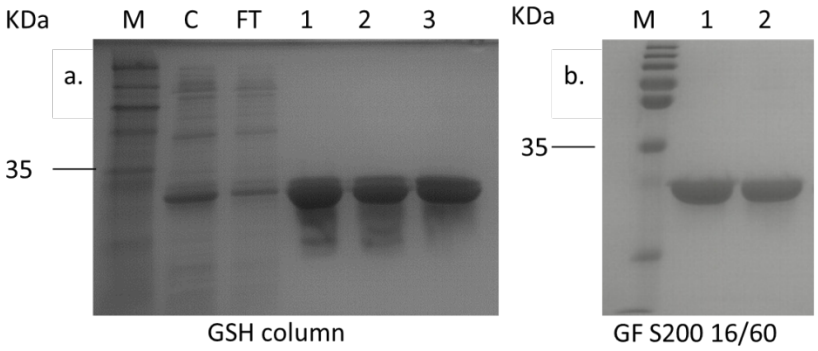


Figure 4.11. Purification profile for GSTU25 re-cloned into pET22b vector and overexpressed in *E. coli* Tunetta cells. a) The GSH affinity chromatography using GSTrap column (GE Healthcare). M: prestained protein marker, C: crude lysate, FT: unbound fractions in flow-through, lane 1-3: purified fractions and b) GSTU25 band after an SEC S200 (16/60) chromatography. Lane 1 and 2: GSTU25 fractions eluted from an SEC column.

4.3.4 Crystallisation of the apo GSTU25-pET22b

Screening for the apo form of GSTU25 crystals, without ligands or GSH, was carried out using commercial buffers available in the lab. Crystals were observed in three different drops i) 0.02 M phosphate buffer (Na/KPO_4) with 18% (w/v) PEG 3350, ii) 0.15 M D,L-malic acid pH 7 with 20% (w/v) PEG 3350 and, iii) 0.02 M sodium sulfate with 20% (w/v) PEG 3350. The largest single crystal was observed in 0.02 M sodium sulfate with 20% PEG 3350 solution. Optimisation was carried out in a hanging drop format at a larger scale: 1 µL drop of 1:1 ratio of protein to buffer. Crystals were observed within two weeks and the largest crystals for X-ray diffraction were collected for in-house X-ray robot testing. The data yielded a diffraction pattern close to the beam stop indicating diffraction of protein crystals, at low resolution (**Figure 4.12**).

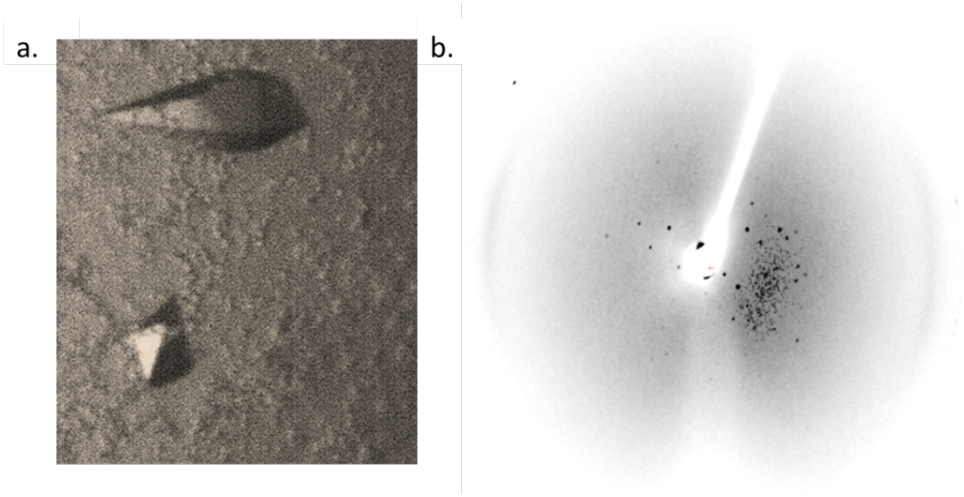


Figure 4.12. a) Crystals of apo-GSTU25 observed in 0.02 M sodium sulfate with 20% (w/v) of PEG 3350. b) Diffraction pattern produced by the in-house X-ray equipment.

4.3.5 Co-crystallisation of GSTU25 with TNT

Following confirmation that the crystals were protein, and not salt crystals, attempts to obtain the structure of GSTU25 in complex with GSH-TNT conjugate were carried out. The purified protein was incubated with a high concentration of TNT (5 and 10 mM TNT) for 1 h on ice.

After two weeks, crystals were observed in four different drops i) 0.2 M sodium thiocyanate pH 6.9 with 20% (w/v) PEG 3350, ii) 0.15 M D,L-malate pH 7.0 with 20% (w/v) PEG 3350, iii) 0.2 M potassium nitrate pH 6.9 with 20% (w/v) PEG 3350 and, iv) 0.17 M ammonium sulfate with 25.5% (w/v) PEG 4000 and 15% (v/v) glycerol. The largest crystal was observed in 0.2 M potassium nitrate pH 6.9 with 20% PEG 3350 and therefore optimisation using the hanging drop method was performed. However, routine testing using in-house X-ray robot did not result in diffraction.

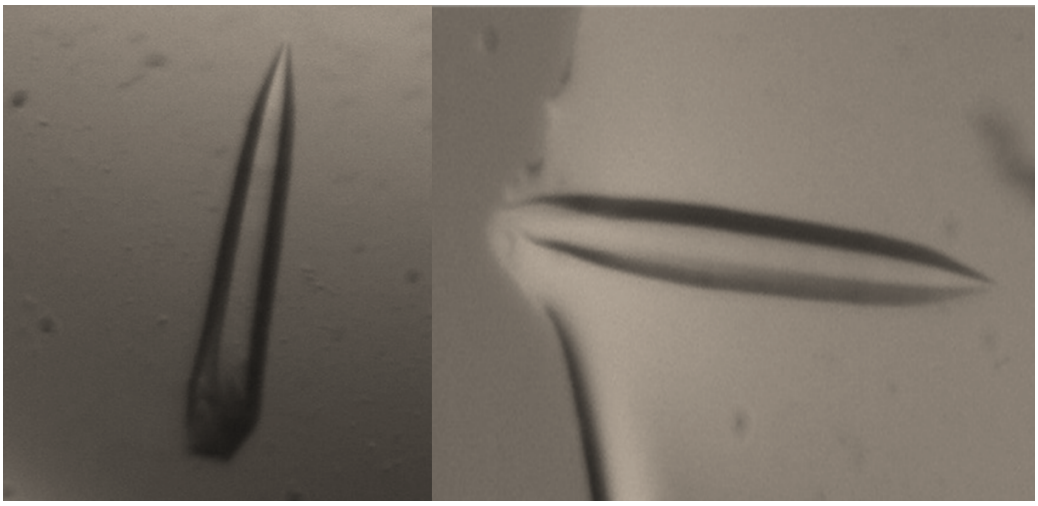


Figure 4.13. The crystals of GSTU25 observed in the drops. The protein was first incubated with 5 to 10 mM of TNT prior incubation in the commercial crystal screening buffers.

4.3.6 Crystallisation of GSTU25 with TNT and GSH in the screening buffer

To further improve the conditions for crystallisation, the method described by Aleku *et al.* (2016) was employed (refer **Section 4.3.4**). After a week of incubation in the screening buffers, pink coloured crystals were observed in three different drops, i) 0.2 M ammonium acetate with 0.1 M of BIS-TRIS pH 5.5 and 25% (w/v) PEG 3350; ii) 0.2 M of $MgCl_2$ with 0.1 M HEPES pH 7.5 and 25% (w/v) PEG 3350 and, iii) 0.2 M sodium formate 20% (w/v) PEG 3350. All drops contained 2 mM of TNT and 2 mM of GSH. The pink crystals were a promising indicator as this is the colour of TNT in solution, and thus indicative of TNT being integrated into the crystals.

The largest crystals were observed from 0.2 M ammonium acetate with 0.1 M BIS-TRIS pH 5.5 and 25% (w/v) PEG 3350 in the presence of 2 mM GSH and 2 mM TNT. This condition was used for further optimisation using hanging-drop vapour diffusion method.

After the optimisation, the largest crystals were collected from drops containing 0.2 M ammonium acetate with 0.1 M BIS-TRIS pH 5.5, 23% (w/v) PEG 3350, 2 mM of GSH and 2 mM TNT. The crystals collected were diffracted to 3 Å using the in-house X-ray equipment and improved to a higher resolution of 1.99 Å at Diamond Synchrotron, UK.

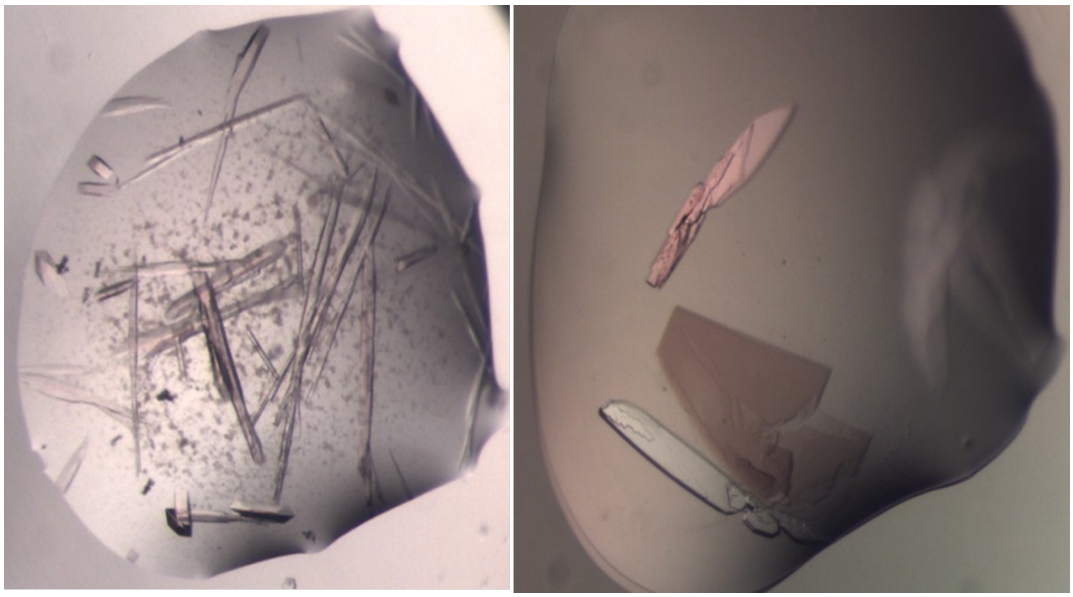


Figure 4.14. Crystals of GSTU25 observed in 0.1 M bis-tris pH 5.5, 23% (w/v) PEG 3350, 2 mM GSH and 2 mM TNT.

4.3.7 Two GSSG molecules bound to GSTU25 dimer

The structure of GSTU25 was solved by Prof. Gideon Grogan using molecular replacement at a resolution of 1.99 Å using *Gm*GSTU4-4 as template (PDB ID: 2VO4). Analysis of the protein structure using the DALI server (Holm & Rosenström 2010), indeed revealed that the monomer was more similar to the structure of a Tau class GST mutant from *G. max,* called Sh14 (PDB ID: 5AGY) (Axarli *et al.*, 2016). Both structures were 68% identical with RMS value of 1.2 Å over 219 residues.

The structure of GSTU25 has four monomers in the asymmetric unit, represented by two dimers. Each monomer has four *β*-strands and nine *α*-helices adopting the canonical GST fold. The first 77 residues at the N-terminus folded into a thioredoxin-like domain followed by an *α*-helical domain at the C-terminus starting from position 89 to 216. The two domains were linked together by a short linker.

Although the crystals were incubated with TNT and GSH, no binding of TNT was detected after the inspection of the electron density map. Multiple rounds of structure refinement cycles using REFMAC5 program revealed two GSH molecules covalently linked by a disulphide bond, showing the structure of GSTU25 was in complex with the oxidised form of GSH, known as glutathione disulphide (GSSG) (**Figure 4.15** to **Figure 4.17**).

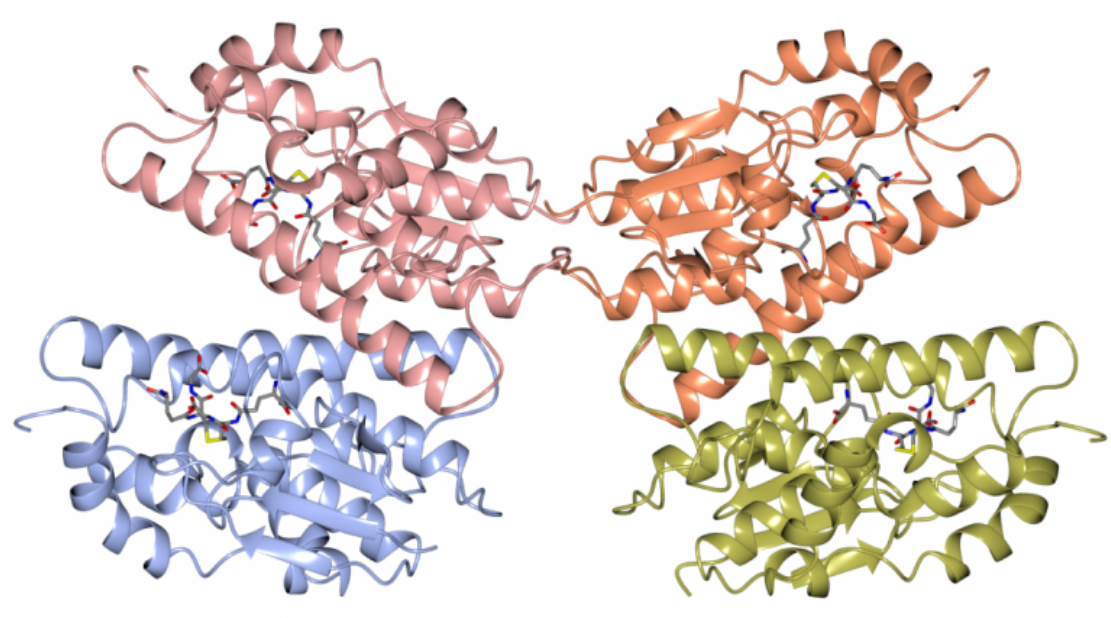


Figure 4.15. Two dimers of GSTU25 in complex with four molecules of disulphide GSH (GSSG).

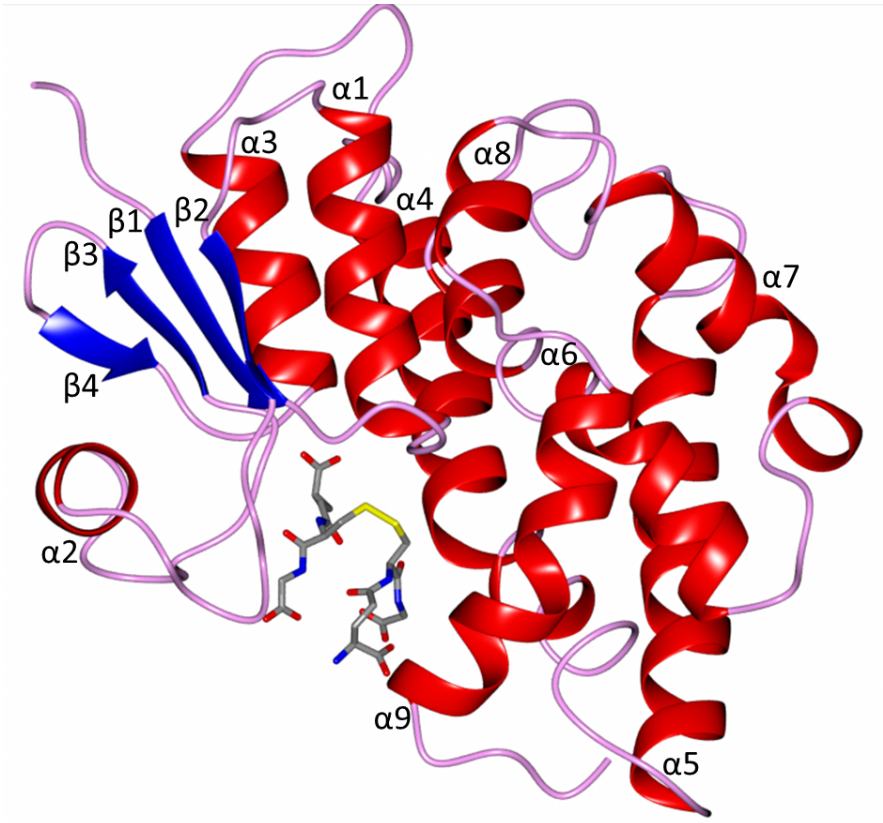


Figure 4.16. Close up view of GSTU25 monomer in ribbon conformation. The helices are coloured in red, strands in blue and turns in pink. The bound GSSG molecule is represented as cylinder and coloured in grey is the carbon, yellow is the sulfur, blue is the nitrogen and red is the oxygen atom.

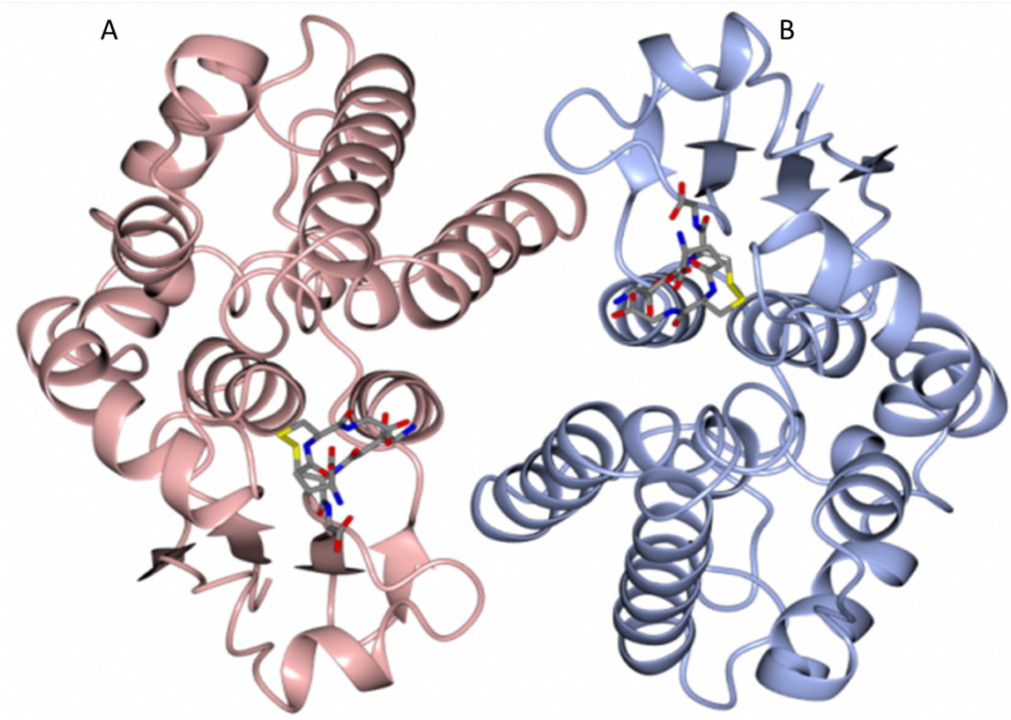


Figure 4.17. GSTU25 dimer labelled with subunit A and B, bound with two GSSG molecules.

At the binding site, the GSSG subunits: GSH-**1** and GSH-**2**, were located in a binding pocket surrounded by polar, nonpolar and charged amino acids (**Figure 4.18**). This pocket was similar to the active site identified for most GSTs, where the hydroxyl group of S13 and Y107 has been shown to contribute to the ionisation of GSH sulfhydryl group (Brock *et al.*, 2013). Similar locations were observed for the same S and Y residues of *Gm*GSTU4-4 in complex with *S*-(*p*-nitrobenzyl)-glutathione (PDB ID: 2VO4) (Axarli *et al.*, 2009). The S residue was found to stabilise the thiolate anion of GSH and enhance its nucleophilicity while the Y residue was important in regulating the enzyme catalytic function.

The GSTU25-GSSG structure showed that the terminal carboxylate group of GSH-**1** *ϒ*-glutamyl moiety formed a hydrogen bond at 3.0 Å to the nitrogen atom of guanidinium group of R111. The glycine moiety of GSH-**1** protruded towards the GSTU25 *α*4 chain. The GSH-**2** molecule, at the carboxylate terminal of the glycine moiety, formed a 2.6 Å hydrogen bond to the oxygen atom of K4 and the *ϒ*-glutamyl moiety was located in between the *α*1 chain and *α*3 chain.

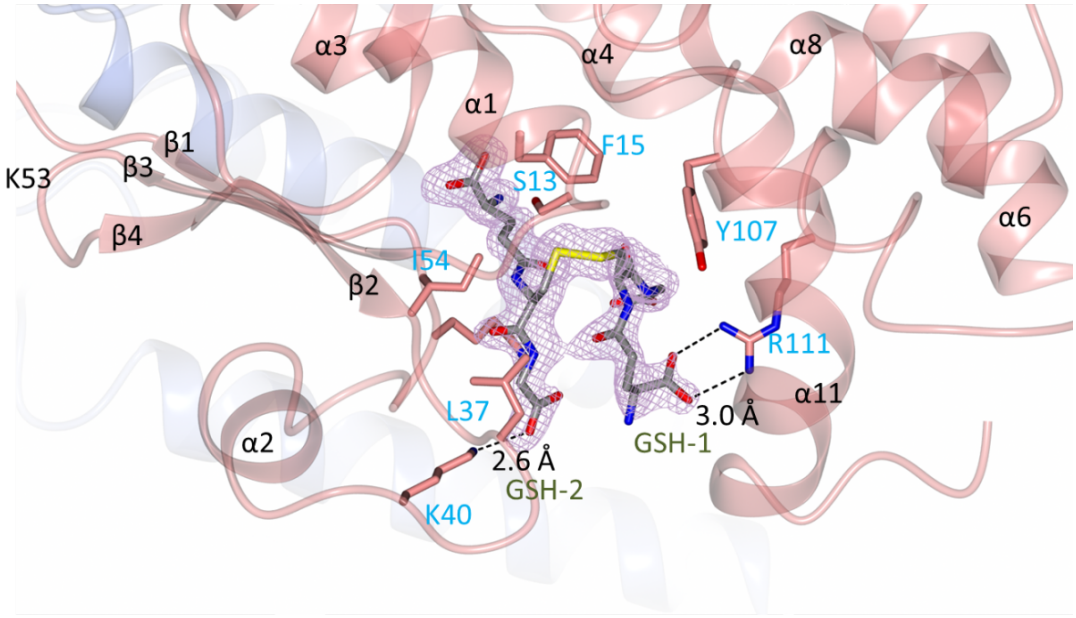


Figure 4.18. Close up view of GSTU25-GSSG binding site. GSSG is depicted in cylinder format with the disulphide bridge in yellow, the nitrogen atom in blue and the oxygen atom in red. The hydrogen bond between the ligands and the interacting residue are indicated as dashed lines. The electron density is the F_o-F_c omit map contoured at 3 σ, which was obtained prior to the refinement of the ligand atoms. The atoms have been added from the refined ligand complex for clarity.

4.4 DISCUSSION

Mutagenesis studies performed by Tzafestas (2016) revealed the importance of five residues; P12, Y107, G115, V209 and L212, of GSTU25 that are important in the formation of different TNT-GSH conjugates. His study revealed that the specificity of GSTU24 could be switched to that seen for GSTU25 when the residues at position 12 and 107 were mutated to proline (P) and tyrosine (Y) respectively. The GSTU24 and GSTU25 each produce a different set of TNT-GSH conjugates; GSTU24 predominantly produces conjugate **2** while GSTU25 produces conjugate **3,** with conjugate **3** being potentially more susceptible to further biodegradation processes. This finding highlights the importance of characterising the structure of the protein; since GSTU25 has the unique ability to produce the more desirable conjugate **3**, the detailed structural characterisation of GSTU25 in complex with TNT-GSH conjugate is, therefore, highly desirable.

Obtaining the structural complex of GSTU25 with conjugate **3** was hindered by the lack of a suitable TNT-GSH adduct. Instead, the electron density map revealed unambiguously one disulphide glutathione (GSSG) per GSTU25 subunit (**Figure 4.15**). The structure of GSTU25 was determined by molecular replacement using the *G. max* GST, PDB ID: 4TOP as a model. Structural comparison using a network service for comparing 3D protein structures, the DALI server (ekhidna.biocenter.helsinki.fi/dali_server), revealed that the monomer of GSTU25 was indeed 68% identical with the structure of a GST mutant from *G. max*, Sh14 (PDB ID: 5AGY), which was determined in complex with *S*-(p-nitrobenzyl-GSH) (Axarli *et al.*, 2016). The Sh14 is a chimeric protein, generated from segments of sequences and point mutations of the parental sequences of *Gm*GSTU2-2, *Gm*GSTU4-4 and *Gm*GSTU10-10. This mutant GST was expressed in *E. coli* and exhibited an enhanced activity towards CDNB and fluorodifen but showed reduced activity with GSH (Axarli *et al.*, 2016).

From the structure of GSTU25, S13 was strategically located at the pocket of the GSSG binding site, this residue is conserved in the other members of the Tau class, and those in the Phi, Theta and Zeta class GSTs. This conservation is essential for the stabilisation of the GSH thiolate anion and enhances GSH nucleophilicity (Axarli *et al.*, 2009). Mutation of the serine to alanine in *Gm*GSTU4-4 reduced the ability of the mutant to bind GSH. The alanine mutant showed an increase in the K_m towards GSH to 159 µM in comparison to 66 µM for the wild-type. However, the same mutant when tested for the binding activity towards herbicide fluorodifen, showed only a mild increase in the K_m: 186 µM compared to 116 µM

for the wild-type, suggesting a binding specificity towards GSH (Axarli *et al.*, 2009). In other classes of GSTs such as DHAR and Lambda, instead of a serine, the role for GSH thiolate stabilisation is provided by a cysteine residue (Dixon & Edwards 2010).

Figure 4.19 compares the structure of GSTU25 modelled by Tzafestas (2016) through homology modelling tools, Modeller software (Šali *et al.*, 1995) with the GSTU25-GSSG complex structure obtained in this study. The superimposed structure clearly shows that P12 and Y107 were located in the similar position, towards the binding pocket and making direct interaction with the GSSG.

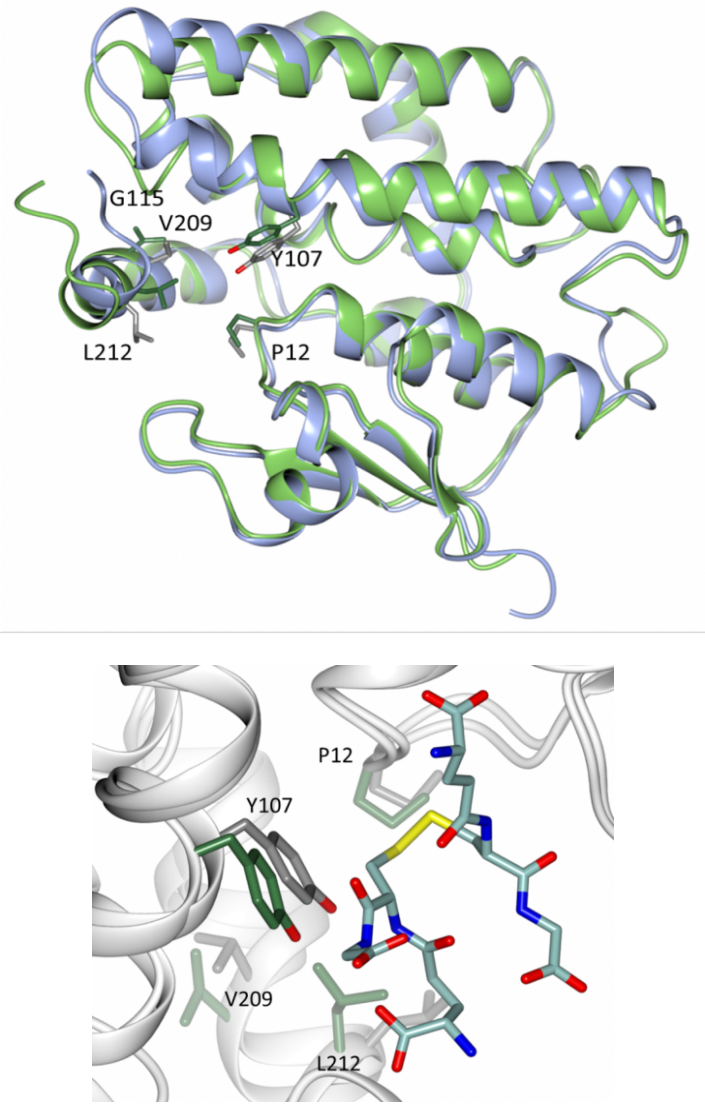


Figure 4.19. Superimposed structure of GSTU25 obtained in this study with the GSTU25 model generated by Tzafestas (2016). The GSTU25 in this study is coloured in green and the GSTU25 model is in blue (both in ribbon conformation). **Top panel:** The superimposed structures yield an RMS value of 1.22 Å over 207 residue, showing the side-chains of residues involved in mutagenesis study by Tzafestas (2016). **Bottom panel:** The superimposed structures with the addition of bound GSSG determined in this study.

In the GSTU25-GSSG structure, the GSH-**1** moiety is stabilised by the side chain of arginine (R111) while the GSH-**2** moiety is located at the known binding site for GSH in other GSTs (Axarli *et al.*, 2016, Axarli *et al.*, 2009, Skopelitou *et al.*, 2015). While only reduced GSH was added during the co-crystallisation process, the crystals took seven days to form, and in the aerobic environment present, significant levels of GSH would be expected to oxidise. A stable reducing agent, such as tris(2-carboxyethyl)-phosphine (TCEP), can assist crystallisation of a compound that is highly sensitive to oxidation, as observed in the case of *Ɣ*-glutamylcysteine synthetase, a GSH synthetase from *Streptococcus agalactiae* (Nakashima *et al.*, 2009). Additionally, protein purification and crystallisation of ferrodoxin II from strict anaerobic sulfate-reducing bacterium *Desulfovibrio gigas* in an anaerobic chamber has been carried out to achieve crystallisation in a reduced state (Hsieh *et al.*, 2005). Although this method successfully yielded a high quality crystal for diffraction, it does not guarantee to mediate the production of TNT-GSH conjugate by GSTU25 in the crystallisation reaction. Therefore, the most reasonable way to obtain the desired structure is to synthesise a suitable TNT-GSH adduct chemically prior the crystallisation screening.

In regards to the structure obtained in this study, the binding of GST to GSSG is not unique to GSTU25. The fungal specific GST from wood fungus, *Phanerochaete chrysosporium*, *Pc*Ure2pA, was found to bind two GSSG molecules per dimer (PDB ID: 4F0B) (Roret *et al.*, 2015) and the *Pc*Ure2pA GST homologues in *E. coli*, *Ec*YghU and *Ec*YfcG displayed a binding property with two GSSG molecules at the active sites (PDB ID: 3C8E and 3GX0 respectively). The *Ec*YghU and *Ec*YfcG structural complexes with GSSG were obtained in the presence of 20 mM GSH in the screening buffer. No additional GSH was loaded in the crystallisation of *Pc*Ure2pA, but the resulting structure suggests that the protein possibly interacts with GSH during protein production (Thuillier *et al.*, 2013; Stourman *et al.*, 2011; Wadington *et al.*, 2009). *Ec*YghU and *Ec*YfcG were previously found to be as efficient as glutaredoxin and thioredoxin from *E. coli* when reducing disulphide bond in 2-hyroxyethyl-disulphide (HED) (Stourman *et al.* 2011).

A recently determined structure of a bacterial glutaredoxin from *Clostridium oremlandii* (cGrx2; PDB ID: 4TR0) indicated the binding of two GSSG molecules per dimer when crystallised in the presence of 10 mM GSH (Lee *et al.*, 2014). When the structure of GSTU25 was superimposed on to cGRX2 (Lee *et al.*, 2014), significant similarity could be observed between the two proteins at the core of the thioredoxin fold where four *β*-strands and *α*-

helices, can be observed (**Figure 4.20**). The cGRX2 has cysteine as the GSH thiol stabiliser, while GSTU25 has serine at the same position, interacting with the GSSG molecule.

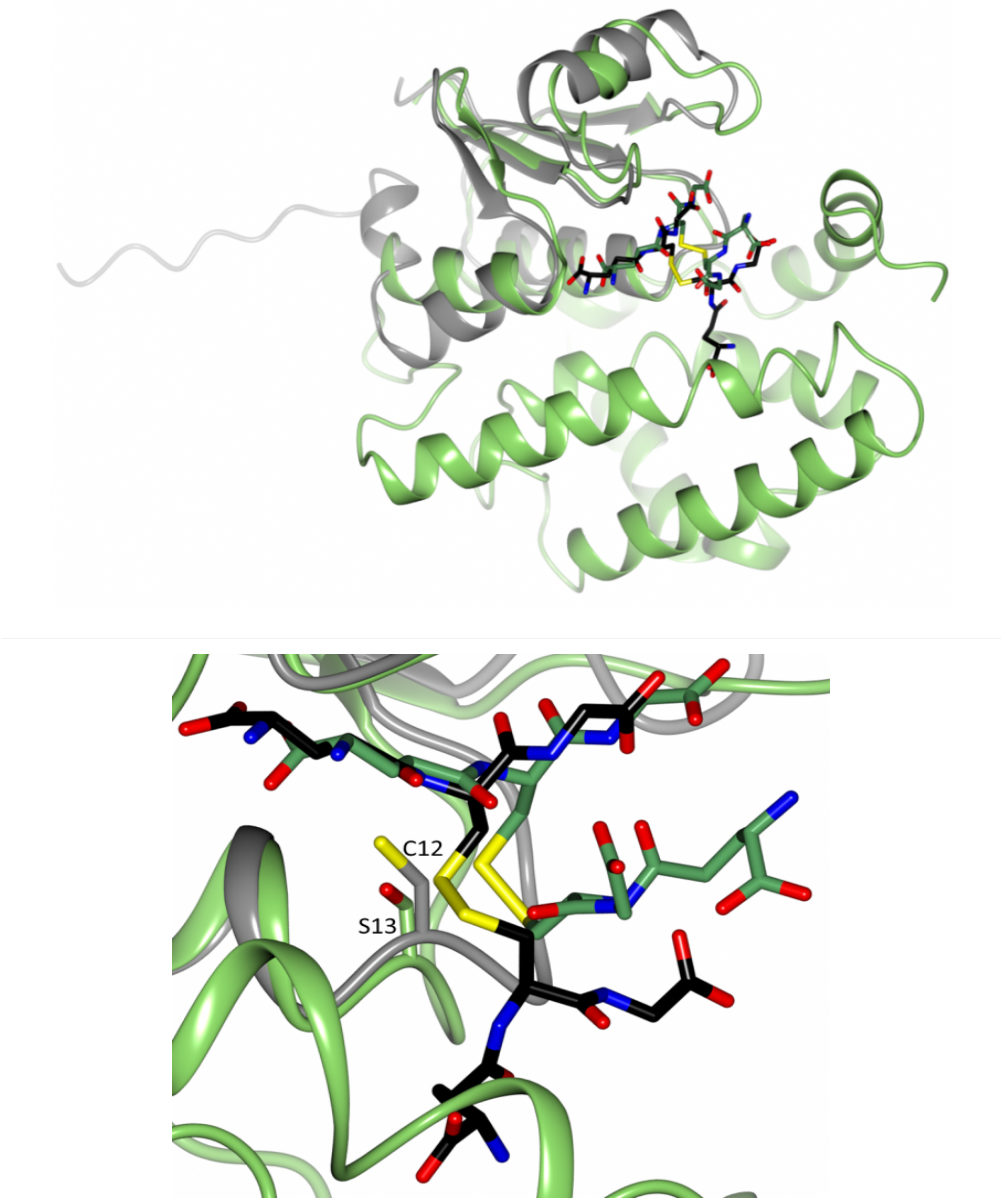


Figure 4.20. Superimposed structures of the glutaredoxin subunit from *C. oremlandii* in complex with GSSG (PDB ID: 4TR0) and the GSTU25 subunit in complex with GSSG in ribbon conformation. GSTU25 is coloured in green ribbon and the GSSG bound is displayed in green carbon chain. Glutaredoxin is displayed in grey with black carbon chain of GSSG. **Top panel:** the RMS value for the superimposed structures was 2.3 Å over 73 residues. **Bottom panel:** The active residue for GSH thiol stabilisation serine 13 of GSTU25 and cysteine 12 of cGRX2.

Glutaredoxin, also known as thiol transferase, is an enzyme that uses GSH and NADPH as cofactors along with glutathione reductase to reduce disulphide bonds (Fernandes & Holmgren, 2004; Prinz *et al.*, 1997). The interchangeable functions of GSTs and glutaredoxins have been reported. Yeast (*Saccharomyces cerevisiae)* glutaredoxins (GRX1 and GRX2) are able to catalyse the conjugation of CDNB to GSH, suggesting that glutaredoxin displays GST-like activities (Collinson & Grant 2003). Meanwhile, an Omega class GST from a parasitic worm (*Schistosoma mansoni)* could participate in the reduction of mixed disulphides, including hydroxyethyl disulphide (HEDS), which is a typical glutaredoxin activity (Girardini *et al.*, 2002). Since the binding of GSSG to GSTU25 is evidently documented in this structure, the disulphide reduction activity of GSTU25 could perhaps be investigated in the future to understand the possibility of a glutaredoxin-like activity of GSTU25, or the interaction of GSTU25 with the disulphide bridge compounds.

Although the structure of GSTU25 with a TNT-GSH conjugate was not successfully obtained, the present structure of GSTU25-GSSG complex provides the insight of GSTU25 folding upon substrate binding. In the future, the crystallisation of GSTU25 could be done in the presence of the synthesized TNT-GSH adduct, conjugate **3.** The production of TNT-GSH conjugates has been previously demonstrated by Gunning *et al.* (2014) by incubation of purified GSTU25 in 100 mM potassium phosphate buffer pH 6.5 at 50°C with 2 mM TNT, 25 mM GSH and 50 µL DMSO in a reaction volume of 500 µL for 6 h. The conjugate can then be purified from the reaction using HPLC and stored in -80°C until crystallisation screening. Using the current GSTU25-GSSG structure, the interaction of GSTU25 with TNT-GSH conjugate can be predicted by *in silico* analysis of ligand docking software such as AutoDOCK by Trott & Olson (2010).

5 Knockout of Tau class members using CRISPR/Cas9

As a continuation of the structural characterisation of GSTU25 in the previous chapter and as a complement to the previously published response of GSTU24 and GSTU25 to TNT (Gunning *et al.*, 2014), a gene knockout study of *GSTU25* and related GSTs using CRISPR/Cas9 was conducted. The aims of this study were to use the CRIPSR/Cas9 method of Xing *et al.*, (2014) to produce and characterise; 1) knockout lines of *GSTU24* and *GSTU25*, and 2) knockout lines of *GSTU25* and the three most related genes sharing the same subclade: *GSTU24*, *GSTU21* and *GSTU19*. Here, only *GSTU25* was successfully knockout and showed to have no significant difference in the root length between the mutant and the wild type line at each tested TNT concentration; 7, 15 and 30 µM (T-test analysis, P≥0.05). As more improved constructs and methods have been developed to enhance the efficiency of CRISPR/Cas9 technology, the results here acts as starter for future studies to knock out multiple members of GST family.

5.1 INTRODUCTION

5.1.1 CRISPR/Cas

The CRISPR acronym stands for Clustered Regularly Interspaced Short Palindromic Repeats. The CRISPR locus composes an operon encoding the Cas protein and a series of repetitive sequences, separated by short non-repetitive sequences called spacers. The spacers correspond to segments of phage sequences that were inserted into the bacterial genome upon infection. The repeat-spacers array transcription produces precursor CRISPR RNA (pre-crRNA) molecules that require maturation, becoming short CRISPR RNA (crRNA), complimentary to the unique invader DNA sequence. The individual crRNA guides Cas protein(s) to cleave the specific invading nucleic acids and prohibit the phage infection (Krzysztof *et al.*, 2013).

The CRISPR/Cas systems studied to date have been classified into two main groups, Class I and Class II. These groups contain five major types and 16 subtypes which are classified according to features within the Cas nuclease (Makarova & Koonin 2015; Makarova *et al.*, 2011; Makarova *et al.*, 2015). Class I activity is defined by multiple effector protein

complexes, and encompasses Types I, III and IV. Class II is described as having a single effector protein, and contains Types II and V systems (Barrangou *et al.*, 2015).

5.1.1.1 Type II CRISPR/Cas9 system

The first discovered and most studied CRISPR/Cas is the Class II, Type II with Cas9 nuclease, which works in a CRISPR/Cas9 system in *Streptococcus pyogenes*. This system only requires three components: Cas9, crRNA and trans-activating CRISPR RNA (tracrRNA) makes this system the simplest, most versatile and precise method for genome engineering (Ding *et al.*, 2016).

The Cas9 is a 163.6 kDa protein responsible for the cleavage of DNA by the activity of two nuclease domains: a Ruv-C like nuclease domain at the amino-terminus and a HNH-like nuclease domain in the mid region of the protein (Sapranauskas *et al.*, 2011). These nuclease domains cut both DNA strands, at sites defined by a 20 nucleotide guide sequence within the crRNA transcript.

The highlight in CRISPR activation is the maturation of active crRNAs from the CRISPR precursor transcript (pre-crRNA). The pre-crRNAs contains the nuclease guide sequences (spacers) interspaced by direct repeats. The activation of pre-crRNA involves the trans-activating CRISPR RNA (tracrRNA) to direct the maturation of pre-crRNAs into crRNA by the activities of the host's RNase III and CRISPR-associated nucleases (Deltcheva *et al.*, 2011).

In CRIPSR/Cas tools, the mature crRNA and tracrRNA can be replaced by one RNA gene, called single guide RNA (sgRNA) which is engineered to contain a hairpin that mimic the crRNA-tracrRNA complex. The binding specificity of Cas9 nuclease to the DNA is specifically guided by the pairing of sgRNA to the complementary DNA and a protospacer-adjacent motif (PAM: with sequence NGG), immediately downstream of the target region. The Cas9 nuclease domain (HNH-like and Ruv-C like nuclease domain) cut at the respective DNA three nucleotide distance away from the PAM motif, resulting double strand break (DSB) (**Figure 5.1**).

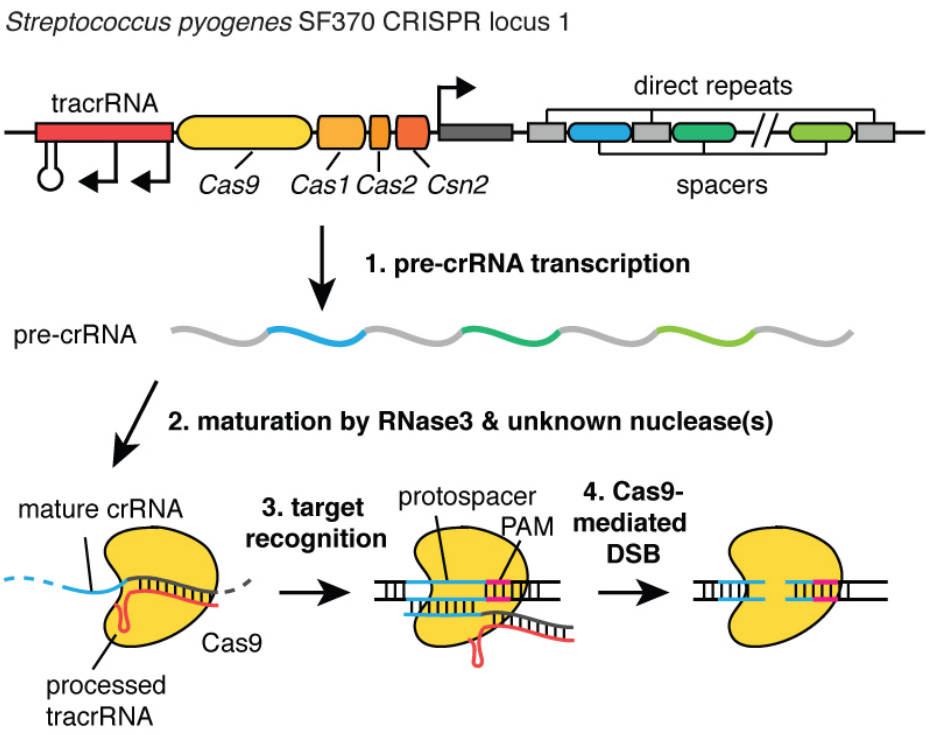


Figure 5.1. Overview of CRISPR/Cas9 type II system in *Streptococcus pyogenes*. 1. The transcription tracrRNA and crRNA from the CRIPSR locus. 2. Maturation of crRNA involved the hybridisation of tracrRNA. 3. The mature crRNA-tracrRNA complex directs Cas9 to the target DNA nucleotides, adjacent to PAM site. 4. Cas9 mediates double strand break (DSB) at the target site. Figure from CRISPR Resources, Zhang Lab (Broad Institute).

5.2 DOUBLE STRAND BREAK (DSB) REPAIR

Following the DSB, the host cell utilises an endogenous mechanism to repair the broken genomic DNA, either by the error-prone non-homologous end-joining (NHEJ) process or by the relatively error free homology-directed repair (HDR) (**Figure 5.2**). The NHEJ is useful for sequence-specific gene knockout as it can lead to the production of nucleotide insertions, deletions and substitutions. The resulting frame shifts will lead to the production of a premature stop codon that will truncate the protein. In contrast to the NHEJ process, HDR repair depends on the presence of an exogenous DNA template on another cloning vector inserted at the same time of CRISPR/Cas9 introduction. The DNA template contains the desired modification, flanked by segments of homologous DNA sequence to the blunt ends of the cleaved DNA (Chu *et al.*, 2015). The preference between these pathways depends on the phase of the cell cycle. The NHEJ is predominantly used in the G1 phase of cell cycle while the HDR is primarily used in the S and G2 phase of cell cycle (Symington & Gautier 2011).

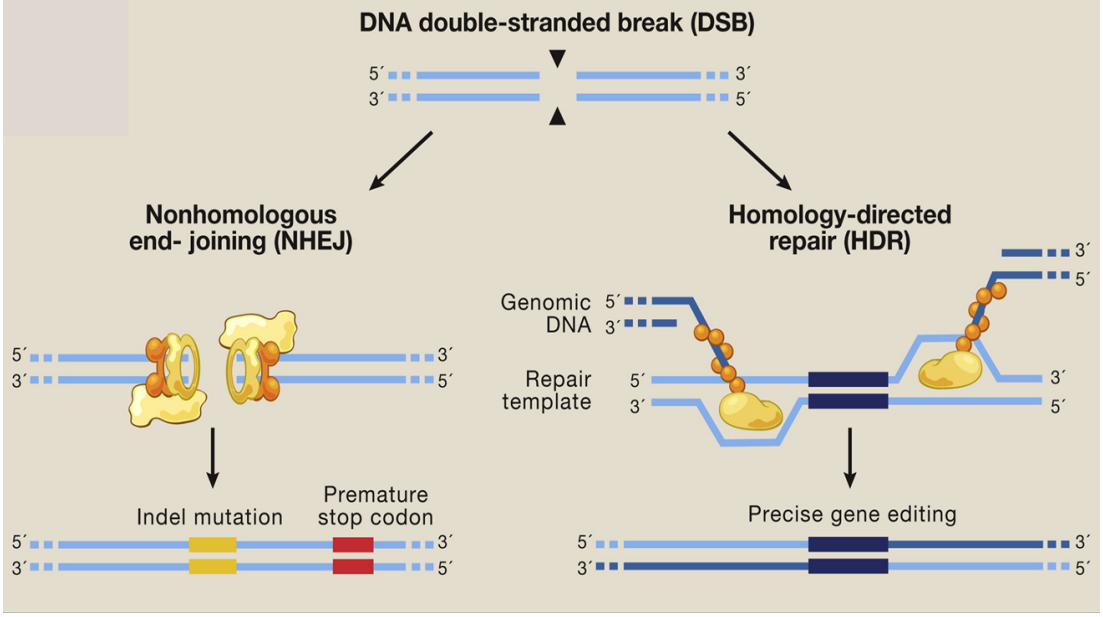


Figure 5.2. The double strand break (DSB) introduced by Cas9 can be repaired by nonhomologous end-joining (NHEJ) or homology-directed repair (HDR). The NHEJ repairs and rejoins the DNA which can result in random insertion or deletions (Indel) at the site of junction. The indel mutation can result in frameshifts and creation of premature stop codon, leading to gene knockout. The alternative repair pathway is carried out by HDR repair which allows precise and high-fidelity editing, provided by the repair template. Figure from Hsu *et al.* (2014).

5.2.1 Plants genome editing using CRISPR/Cas9

The CRISPR/Cas9 system has been applied to model and crop plants including Arabidopsis (Mao *et al.*, 2013; Jiang *et al.* 2013; Li *et al.* 2013), *Zea mays* (maize) (Svitashev *et al.*, 2015), *Solanum lycopersicum* (tomato) (Ito *et al.*, 2015), *Triticum aestivum* (bread wheat) (Wang *et al.*, 2014), *Nicotiana benthamiana* (tobacco) (Jiang *et al.*, 2013; Lowder *et al.*, 2015), *Oryza sativa* (rice) (Jiang *et al.*, 2013; Woo *et al.*, 2015), *Hordeum vulgare* (barley) (Lawrenson *et al.*, 2015), *Solanum tuberosum* (potato) (Butler *et al.*, 2015), *Brassica oleracea* (cabbage) (Lawrenson *et al.*, 2015) and *Populus tomentosa* (poplar) (Fan *et al.*, 2015). The targets include magnesium-chelatase subunit I (CHLI 1) (*At*4g18480) and CHLI2 (*At*5g45930) in Arabidopsis (Mao *et al.*, 2013) and the phytoene desaturase gene 8 (*PtoPDS*) in poplar (Fan *et al.*, 2015). Both studies were carried out to test the efficiency of CRISPR/Cas9 system by using albino as the observable phenotype.

When using CRISPR/Cas technology in plants, the expression of sgRNAs are generally directed by ubiquitin 3 (U3) or ubiquitin 6 (U6) RNA gene promoters and are transcribed by endogenous RNA polymerase III (Nekrasov *et al.*, 2013; Jiang *et al.*, 2013). The sgRNA expression cassette carrying the U3/U6 promoter can be generated by target-adaptor ligation or overlapping PCR (Li *et al.*, 2013). The expression of *Cas9* is driven by the Cauliflower Mosaic Virus 35S promoter (CaMV 35S) (Xie & Yang 2013). In common with many constructs for CRISPR/Cas, pCambia-like binary vector and pGreen-like binary vectors can accommodate several different sgRNAs to be transcribed and incorporated within a single CRISPR/Cas9 complex to simultaneously target multiple sites in the genome (Zhang *et al.* 2014; Xing *et al.* 2014). The presence of the CRISPR/Cas9 is followed by the addition of a marker gene, in the case of pCambia (Addgene id: pHSE401, *Supplementary 1*) that is used in this study; it is the hygromycin phosphotransferase (*hptII*) gene, which encodes resistance to the antibiotic hygromycin.

To create a transgenic Arabidopsis plant, the *Agrobacterium tumefaciens* (Agrobacterium)-mediated 'floral dip' technique is used to target the germ-line tissues with DNA constructs encoding the CRISPR/Cas9 system (Ye *et al.*, 1999; Clough & Bent 1998). In the T1 plants, DSBs followed by host-cell repair, introduces mutations throughout the cells (Feng *et al.*, 2013; Mao *et al.*, 2013; Jiang *et al.*, 2014; Xing *et al.*, 2014; Fauser *et al.*, 2014; Feng *et al.*, 2014). Research by Liu *et al.* (2015) and Kumar & Jain (2015), used sgRNA targeted the chlorophyll pathway gene, the magnesium-chelatase subunit genes *chli1* and c*hli2*. Thus DSB and mis-repairs in the targeted gene resulted in sectors of white tissue; an observable phenotype, in an otherwise green plant (**Figure 5.3**).

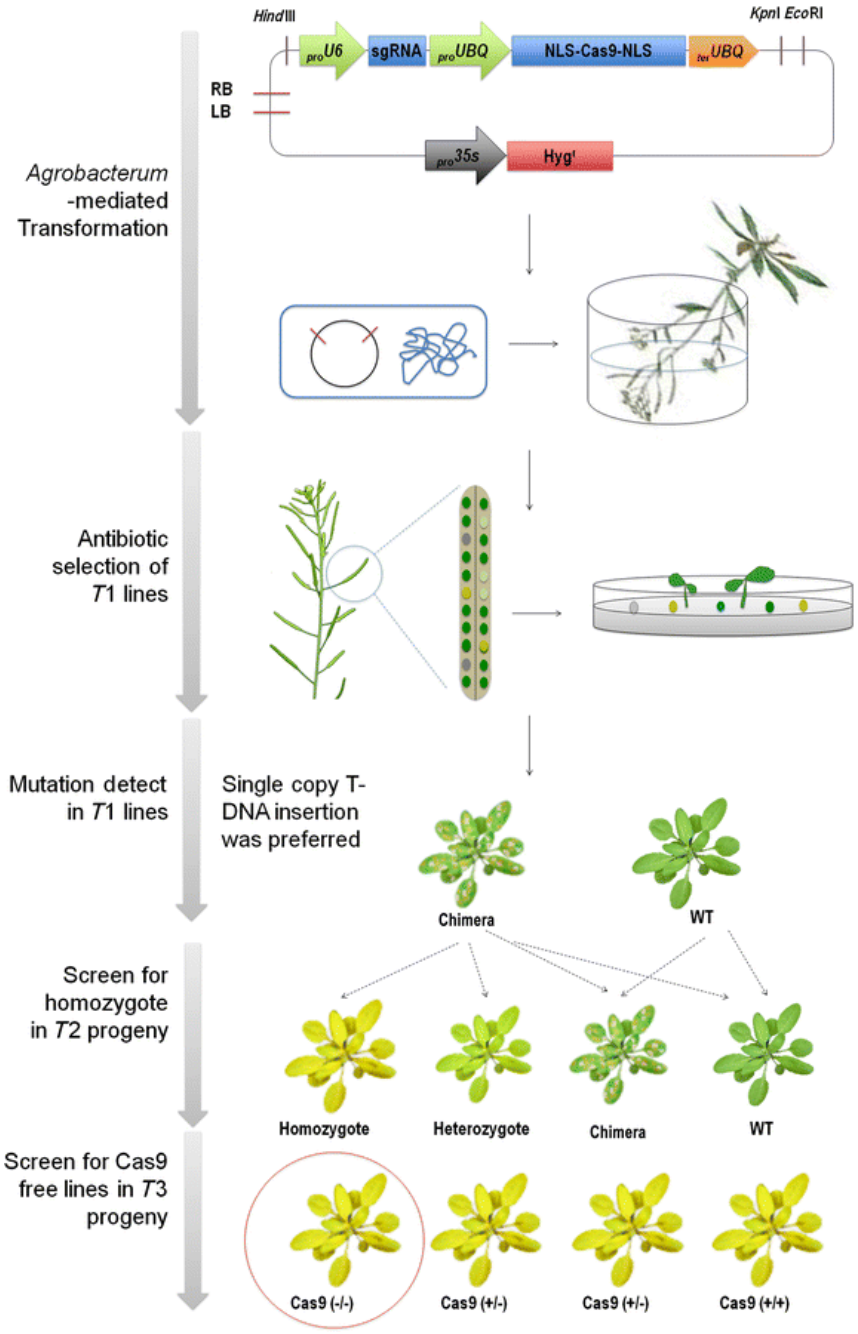


Figure 5.3. The workflow of generating the CRISPR/Cas9 mutant in Arabidopsis. The Agrobacterium-mediated flower dipping method transforms CRISPR/Cas9 vector to Arabidopsis. Selective ½ MS agar of the sowed T0 seeds allows the selection of transgenic plants. Often that all T1 generations will be chimeras which after self-pollination will segregate to homozygous, heterozygous or biallelic plants in T2 generation. Segregation of the T-DNA carrying Cas9 in T3 generation allows the identification of mutant free plants. Figure from (Liu *et al.*, 2015).

If mutations occur in the germ line, these will be passed on, either/or both in the egg or pollen, to the T2 generation upon self-pollination. Thus, if the germline yielding both egg and pollen have been mutated, it is possible to obtain homozygous mutants in the T2 generation. Although in practice, many T2 plants will be heterozygous; the result of a mutation is in either the egg or pollen-yielding germ line. Screening for homozygous mutants is possible in the T2 population by sequencing the 23bp region in DNA from leaf tissue, but to ensure the mutation is stable, the inserted Cas9 T-DNA needs to be segregated away from the plants containing the mutation. In the case of the system used by Xing *et al.* (2014), the pCambia-Cas9 free mutants will lack the *hptII* gene and can be identified by counter-selective screening for sensitivity of seedlings to hygromycin.

5.2.2 Targeting Multigene families using CRIPSR/Cas9 in plants

The CRISPR/Cas9 system provides an excellent method to understand the function encoded by individual members of a multigene family with apparent redundancy. This technique is also useful for polyploid plants, where disruption of a multihomologous gene array is essential to observe clear changes on the phenotype (Endo *et al.*, 2015). Such a multiplex gene editing strategy has been used to target members of multigene families, replacing or deleting multiple genomic regions to improve the genetics of multiple traits in crop breeding (Cao *et al.*, 2016; Ma *et al.*, 2016).

Using CRISPR/Cas9, effective editing of multiple genes can be accomplished by combining multiple single guide RNAs (gRNAs), each controlled by a promoter upstream of the gRNA, into a single construct to edit multiple targets in plants (Wang *et al.*, 2016). Several strategies to assemble multiple sgRNA expression cassettes into single CRISPR/Cas9 binary constructs have been developed. One strategy uses sequential rounds of a more traditional, restriction digestion-based, cloning to insert sgRNAs into the construct; however this traditional cloning approach, although effective is time-consuming, negating its use in high-throughput systems. More advanced techniques employ Gibson Assembly (Ma *et al.*, 2015) or Golden Gate technology (Wang *et al.*, 2016). Gibson Assembly uses the simultaneous actions of 5' exonucleases to generate 3' single stranded overhangs to produce 'sticky ends'. This activity is similar to restriction enzymes but has a greater length of complementarity and has been used in the genome modification of rice and Arabidopsis with success (Ma *et al.*, 2015).

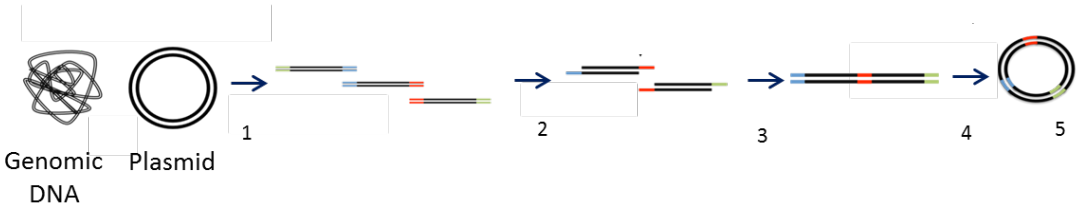


Figure 5.4. The Gibson Assembly method to create a circular DNA plasmid. The basic methodology is as follow: 1: The production of DNA fragments with overlapping ends by PCR digestion or restriction enzymes digestion, 2: The 5' exonuclease digestion of DNA fragments to yield 'sticky' ends, 3, 4: Overlapping ends of the complementary base pairing allows fragments to form circular plasmid, 5: T4 DNA ligase seals the nicks in the DNA backbone and the plasmid is ready to be transformed into *E. coli*. Figure from Sinfield (2014).

Golden Gate uses numerous restriction enzymes that yield sequential palindromic sticky ends to assemble multiple sgRNAs into one binary vector. The restriction enzymes cleave at specific sites creating overhangs, which then facilitate the correct annealing aided by ligation enzymes. The entire cloning step, digestion and ligation can be carried out in single reaction; the use of different restriction enzymes within the reaction produces specific overhangs preserving the directionality of the cloning reaction.

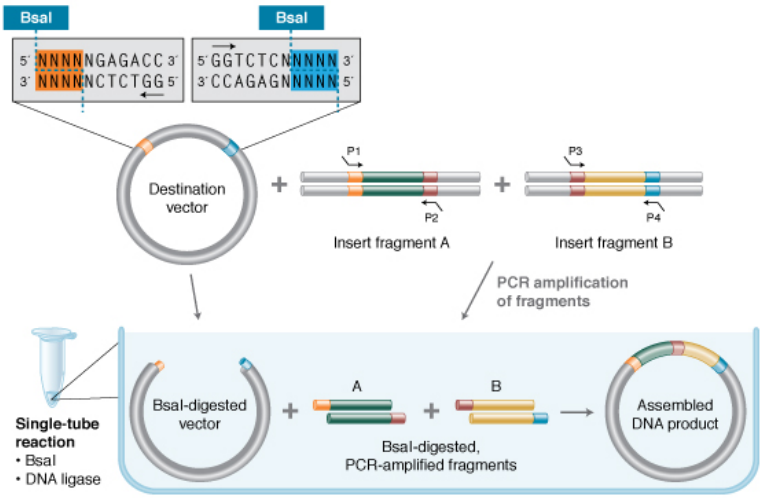


Figure 5.5. The Golden Gate assembly to insert multiple DNA fragments into a single circular plasmid. In this case, the addition of *Bsa*I (GGTCTC) recognition site allows *Bsa*I digestion and creates non-palindromic overhangs on both the target plasmid and the fragments. The addition of T4 DNA ligase assembles the DNA into a circular plasmid. Figure from New England Biolabs (NEB) website.

The CRISPR/Cas9 technique use a single sgRNA to target multiple genes, but for success, this requires the presence of a 23 bp region of complete identity to all the target genes, and directly preceding a PAM, to be present in the genome. Studies by Xing *et al.* (2014) using

Arabidopsis has shown that a single base mismatch between the sgRNA and target DNA can reduce knockout efficiency by 10% while two mismatches cause inefficiency of about 60% and by three base mismatch, no mutation was detectable suggesting high sequence specificity is crucial. A study by Cong *et al.* (2013) in human and mouse indicated that a single-base mismatch completely prevented the DSB cleavage activity of Cas9. To enhance target sequence identity, software programs such as Genome Engineering 4.0 (Hsu *et al.*, 2013) and E-CRISP (Heigwer *et al.*, 2014) can be used to design highly specific sgRNA.

5.2.3 Targeting the *GST* Tau class using CRISPR/Cas9

In Arabidopsis, Yoon *et al.* (2007) identified that *GSTU24* was upregulated by 4.7 fold following treatment with 2,6-dinitrotoluene (DNT) and also TNT. Yoon *et al.* (2007) compared TNT uptake rates of a T-DNA mutant of *GSTU24* with wild-type based on the hypothesis that if *GSTU24* was involved in the *in planta* detoxification, the knockout transgenic line would be compromised in the ability to resist TNT toxicity. However, when the mutant was grown on soil containing 50 mg L^{-1} and 100 mg L^{-1} TNT there was no distinguishable TNT uptake rate and no statistically significant difference in root length from the wild-type. The authors concluded that this result was due to redundancy between the GST family members, which was in agreement with a later study by Gunning *et al.* (2014) which demonstrated that *GSTU25* is similarly upregulated in response to TNT, and purified GSTU25 recorded K_m and V_{max} values (1.2 mM and 393.6 nkat mg^{-1}, respectively) towards TNT, while GSTU24 has K_m of 1.6 mM and V_{max} of 369.3 nkat mg^{-1}, respectively, with TNT. It is possible that additional Tau class GSTs may also contribute to *in planta* TNT detoxification. The problem of redundancy within the GST family is not limited to TNT as the substrate or to Arabidopsis as the plant species. Resistance of the malaria-carrying mosquito, *Anopheles coluzzi* by the insecticide pyrethroid is conferred by *GSTE2* and *GSTS1_1*, with likely additional contributions from other family members (Toé *et al.*, 2015).

Using the data from Yoon *et al.* (2007) and Gunning *et al.* (2014), including the increased resistance to TNT seen in Arabidopsis plants overexpressing *GSTU25*, it seems likely that knocking out a combination of at least *GSTU24* and *GSTU25*, would reduce the remaining GST activity towards TNT enough to give a phenotype. To increase the likelihood of generating a phenotype, the two remaining Tau class GST (*GSTU19* and *GSTU21*), within the subclade shown in **Figure 1.13** were selected as CRISPR-Cas9 targets.

Neither *GSTU19* or *GSTU21* were represented in the ATH1 chip used for the microarray study previously done by Lorenz (2007), therefore it is not known whether these genes are

expressed in response to TNT and no studies have yet tested these GSTs for activity towards TNT. However, *GSTU22,* the other gene located within the subclade, was upregulated in the microarray analysis but lacked activity towards TNT (Lorenz 2007). Thus, the only clue as to whether *GSTU21* and *GSTU19* have any activity towards TNT is by comparing the protein sequence of the active residues to that are known to be important for conjugation activity in *GSTU24* and *GSTU25* (Tzafestas, 2016).

The GSTU24 and GSTU25 proteins are closely related, sharing 79% sequence identity, with GSTU19 and GSTU21 sharing 76% and 72% sequence identity respectively, with GSTU25. While these percentages are relativity high, to perform CRISPR/Cas9 gene editing using just one sgRNA requires a 23 bp identical region to be present if CRISPR/Cas9 is to successfully generate knockouts. The absence of a common, 23 bp sequence across the four GSTs can be overcome by using CRISPR/Cas9 constructs in which multiple 23 bp sequences can be cloned, with one (or more) targeting each of the GST gene sequences.

5.3 MATERIAL AND METHODS

5.3.1 The selection of sgRNA targets

To identify the regions in all four target gene sequences that could be used to design a 23 bp region for the sgRNA required for CRISPR/Cas9, multiple sequence alignment of *GSTU19*, *GSTU21*, *GSTU24* and *GSTU25* was conducted (**Figure 5.6**).

```
AtGSTU19      AAGAGTCCTTTGCTTCTCCAGATGAATCCGATTCACAAGAAGATTCCTGTTCTCATCCAC 177
AtGSTU21      AAGAGTCCCTTGCTACTAGAGATGAATCCGATTCACAAGACAATCCCGGTTCTCATCCAC 180
AtGSTU24      AAGAGCTCCTTGCTCCTCGAGATGAACCCGGTTCATAAGAAAATTCCGGTTCTCATCCAC 177
AtGSTU25      AAAAGCCCGATTCTCCTCGAGATGAATCCGGTTCATAAGAAAATACCGGTTCTCATCCAC 177
              **.**  * :* ** **. ******* ***.**** ****..** ** ***********

AtGSTU19      AATGGTAAACCGGTTAACGAATCTATCATCCAGGTTCAGTACATTGACGAGGTCTGGTCT 237
AtGSTU21      AATGGTAAACCGGTTCTTGAATCCCTTATCCAAATCCAATATATAGACGAGGTTTGGTCT 240
AtGSTU24      AATGGTAAACCGGTATGTGAATCTCTCATTCAGATCGAGTACATAGACGAGACTTGGCCC 237
AtGSTU25      AATGGTAATCCGGTATGTGAATCACTCATCCAGATCGAATACATCGACGAAGTTTGGCCT 237
              ********:*****:   ***** .* ** **..*  *.** ** *****..  *** *
```

Figure 5.6. Multiple sequence alignment of *GSTU*s selected for the knockout study using CRISPR/Cas9 system. The region selected is colour-coded by yellow for *GSTU19*, green for *GSTU21*, blue for *GSTU24* and magenta for *GSTU25* which is located at the N-terminal (~160 bp after the start codon).

5.3.2 Assembly of sgRNA targets and expression components

The Cas9-carrying template plasmids pCBC-DT1T2 (*Supplementary 2*), pCBC-DT2T3 (*Supplementary 3*) and pCBC-DT3DT4 (*Supplementary 4*), were derivatives of pCambia as published by Xing *et al.* (2014) and supplied by Addgene, UK. To incorporate the selected sgRNAs into the template plasmids, primers with overlapping sequences to the respective template plasmids and the selected sgRNA sequences were designed. The list of primers is provided in **Table 5.1**.

The production of the binary plasmid containing gRNA for two targets (*GSTU24* and *GSTU25*), hereafter known as pCambia_2T**.** Two primer pairs were added: 1 µM U25-F0, 1 µM U24-R0 and 20 µM U25-BsF, 20 µM U24-BsR in a 50 µL total reaction composed of 5x Phusion HF buffer (New England Biolabs), 10 mM dNTPs mix, 1 U Phusion DNA polymerase, 3% dimethyl sulfoxide (DMSO) and amplification was carried out from pCBC-DT1T2. The amplification reaction was performed with a cycle of 98°C for 30 s followed by 33 cycles of 98°C for 10 s, 56°C for 30 s and 72°C for 30s. The amplified product was verified on agarose gel prior purification using Wizard® SV Gel (Promega). The amplified product was used to set up restriction-ligation reactions with pCambia-Cas9 binary vector (Addgene, UK) using

*Bsa*I and T4 Ligase (New England Biolabs). The reaction was incubated in a thermocycler for 5 h at 37°C, 5 min at 50°C and 10 min at 80°C.

The production of the binary plasmid containing gRNA for four target sites (*GSTU19*, *GSTU21*, *GSTU24* and *GSTU25*), hereafter known as pCambia_4T**.** Three cassettes of single target of *GSTU24* and *GSTU25* each and a double target of *GSTU21* and *GSTU19* were generated. Primers U25-BsF, U25-F0 and DT0-BSR2 were used to produce the U25 single site cassette; U24-BsF2, U24-F0 and DT0-BSR3 were used to produce the U24 single site cassette; U21-BsF3, U21-F0, U19-R0 and U19-BsR were used to produce the U21 and U19 double target cassette. The amplification reaction used for this method was as mentioned for *GSTU24*/*GSTU25* double target cassette. All of the purified, amplified products were subjected to restriction-ligation reactions with pCambia-Cas9 binary vector as mentioned in the production of double target cassette.

The ligation products (5 µL) were transformed into *E. coli* DH5α and positive clones selected on LB agar plates containing kanamycin (50 µg mL^{-1}). The kanamycin resistant clones were validated by colony PCR, and sequencing by GATC Biotech Ltd. (UK).

Table 5.1. List of primers used to assemble multiple components including sgRNA target, U3 or U6 promoter and U3 or U6 terminator to be inserted in the expression plasmid. Golden Gate assembly using restriction enzymes (*Bsa*I) was employed for the assembly. The pink coloured bases are sgRNA for *GSTU25*, blue for *GSTU24*, green for *GSTU21* and orange for *GSTU19*. The red bases are the cleavage site of *Bsa*I.

Primer	Nucleotide (5'-3')
U25-BsF	ATATATGGTCTCGATTGTACCGGTTCTCATCCACAAGTT
U25-F0	TGTACCGGTTCTCATCCACAAGTTTTAGAGCTAGAAATAGC
U24-R0	AAC*TTGTGGATGAGAACCGGAA*CAATCTCTTAGTCGACTCTAC
U24-BsR	ATTATTGGTCTCGAAAC*TTGTGGATGAGAACCGGAA*CAA
U24-F	ATATTATTGGTCTCAAGATTGTTCCGGTTCTCATCCACAAGTT
U24-F0	TGTTCCGGTTCTCATCCACAAGTTTTAGAGCTAGAAATAGC
U21-R0	AAC*TTGTGGATGAGAACCGGGA*CAATCACTACTTCGTCTCTAACCAT
U21-BsR	ATTATTGGTCTCGAAAC*TTGTGGATGAGAACCGGGA*C
U24-BsF2	ATATTATTGGTCTCAAGATTGTTCCGGTTCTCATCCACAAGTT
U24-F0	TGTTCCGGTTCTCATCCACAAGTTTTAGAGCTAGAAATAGC
DT0-BsR3	ATATTATTGGTCTCATCACTACTTCGTCTCTAACCAT
U21-BsF3	ATATTATTGGTCTCAGTGATTGTCCGGTTCTCATCCACAAGTT
U21-F0	TGTCCGGTTCTCATCCACAAGTTTTAGAGCTAGAAATAGC
U19-R0	AAC***TTGTGGATGAGAACAGGAA***CAATCACTACTTCGACTCTAGCTGTAT
U19-BsR	ATTATTGGTCTCTAAAC***TTGTGGATGAGAACAGGAA***

5.3.3 Generation of transgenic Arabidopsis plants

The pCambia_4T and pCambia_2T were transformed separately into Agrobacterium strain GV3101 using the electroporation method: 100 µL of competent cells were incubated with 1 µL of purified plasmid on ice for 1 min then pulsed with 2.5 kV within 2 mM micropulse cuvette (EQUIBIO ECU102) in a MicroPulser TM (Bio-Rad 165-2100; setting Ec2) and immediately transferred to a clean Eppendorf tube. The cells were allowed to grow for 3 h at 30°C with gentle shaking in 1 mL of SOC medium (20 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 0.5 g L^{-1} NaCl, autoclaved, then additional 1 M $MgCl_2$, 1 M $MgSO_4$ and 2 M glucose, all filter sterilised before addition). To select transformed single colony, the cells were plated on 50 µg mL^{-1} of gentamycin and 25 µg mL^{-1} spectinomycin-LB agar for three days at 30°C.

The floral dip method was employed to transform Arabidopsis ecotype Col-0 wild-type plants with Agrobacterium carrying the binary vectors (Clough & Bent, 1998). Prior to that, a single colony of recombinant Agrobacterium was grown overnight in selective 10 mL LB broth containing 50 µg mL^{-1} of gentamycin and 25 µg mL^{-1} spectinomycin at 28°C. The overnight culture was transferred into 500 mL LB in 2 L conical flasks and incubated at 30°C, shaking at 180 rpm for 3 days. The 500 mL culture was pelleted by centrifugation (3200 x g, 10 min) and resuspended in 500 mL 5% (w/v) sucrose, 0.05% TritonX-100. The Arabidopsis plant, which had flowering buds, was dipped in this solution. Dipped Arabidopsis plants were covered within autoclaved bags for 24 h after dipping to retain moisture and transferred to the glasshouse. The transformed Arabidopsis plants were grown for 6 weeks in the glasshouse and T1 seed were collected. To identify transformed Arabidopsis, the T1 seeds were screened on selective ½ MS agar plates containing 30 µg mL^{-1} hygromycin. The seedlings that has acquired resistant to hygromycin were transferred to soil. Genomic DNA was extracted from the plants and later T2 seeds were collected. Random screening of T2 seedlings on ½ MS agar containing 0 µM, 7 µM and 15 µM TNT was carried out to select seedling with shorter root and later were rescued by transferring onto soil. The T3 seeds from the survived T2 plants were collected. The T3 seeds were sprinkled on hygromycin plates (30 µg mL^{-1}) to identify hygromycin resistant and sensitive seedlings. The overall step of producing transgenic mutants in this experiment is showed in **Figure 5.7**. Cas9 plasmid segregated T3 seedlings were screened on agar plates containing ½ MS and 7 µM, 15 µM and 30 µM of TNT.

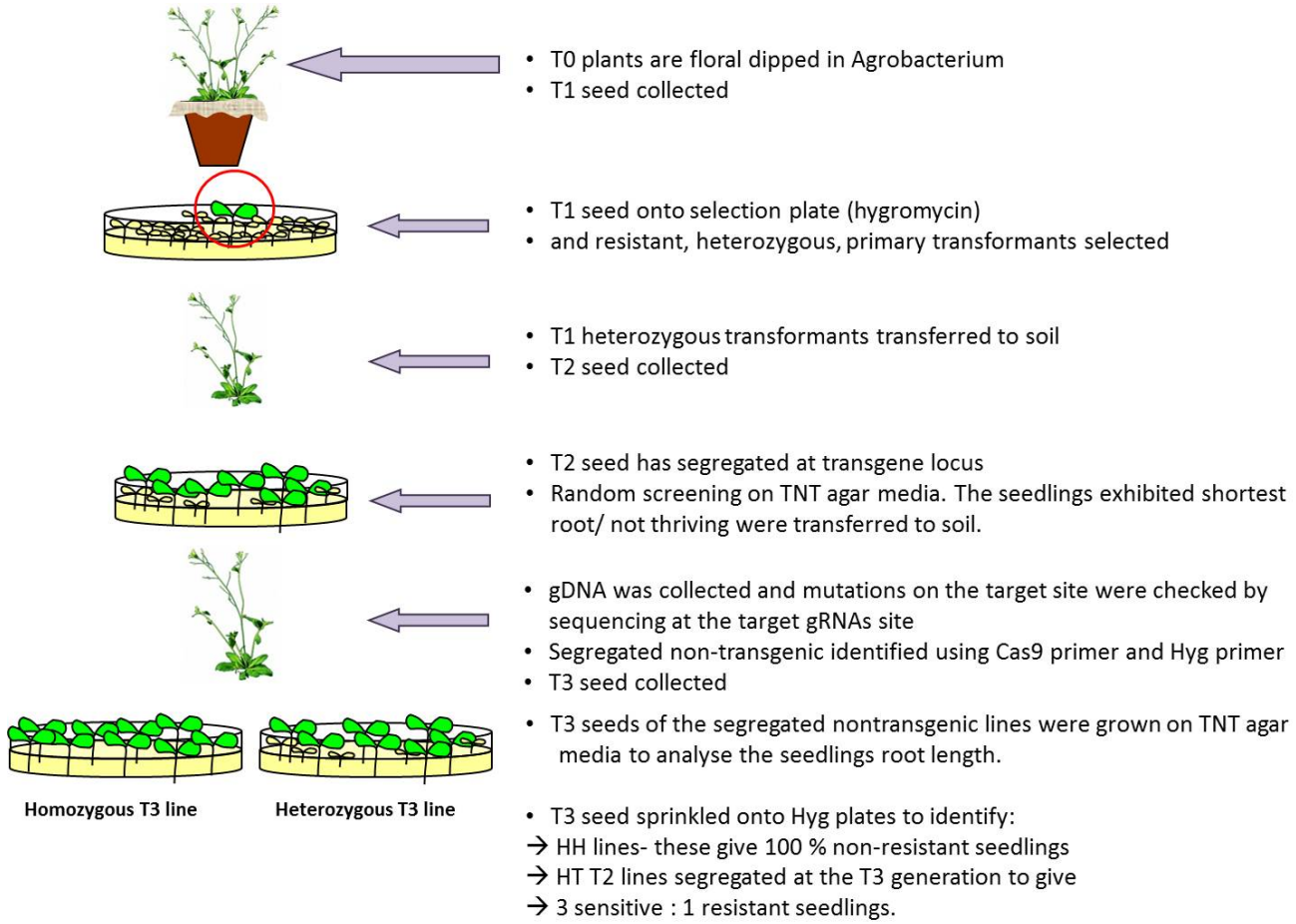


Figure 5.7. The production of transgenic mutant lines using CRISPR/Cas9 used in this experiment. The figure is kindly adapted from one by Dr. Elizabeth L. Rylott.

5.3.4 Genomic DNA extraction

Plant tissue (100 mg) was ground with a pestle within a 1.5 mL Eppendorf tube with 500 µL CTAB buffer (2% cetyl trimethylamine bromide, 1.4 M NaCl, 100 mM Tris-HCl pH 8, 20 mM NaEDTA) and incubated for 1 h at 65°C. Following incubation, 300 µL of 24:1 chloroform: iso-amyl-alcohol was added, centrifuged (13000 rpm, 10 min) and 300 µL of aqueous layer was transferred to a new 1.5 mL Eppendorf tube containing 960 µL ethanol and 40 µl 3 M NaAc pH 5.3. Genomic DNA was precipitated at RT for 40 min and collected with centrifugation (13000 rpm, 15 min, 4°C). The pellet was rinsed in 70% ethanol, dried using a speed-vac (Savant DNA Speed-Vac, high temperature setting, 10 min) and resuspended in 50 µL sterile water.

5.3.5 Validation of the mutations

Genomic DNA was extracted from T1 and T2 transgenic plants. The full open reading frames (ORFs) of the targeted genes were amplified by PCR using gene specific primers *GSTU24*, *GSTU25*, *GSTU21*, *GSTU19* (**Table 5.2**). The fragments were purified from agarose gel and were sent for sequencing at GATC Biotech, UK. The T2 plant lines that were confirmed to contain mutations in *GSTU25* and segregated from Cas9 were analysed on agar media containing TNT (7 µM, 15 µM and 30 µM of TNT). The seedlings were grown for 7 d (16 h light, 8 h dark cycle) and the root lengths of measured.

Table 5.2. The primers used for sequencing of the whole gene

Primer	Nucleotide (5'-3')
GSTU24_F	AATGATGTAATCAAGTGCGCC
GSTU24_R	GCCGTTATGGCAACCAGATTAATG
GSTU25_F	GTGTAATGGGCACGTTAATGC
GSTU25_R	GCAACGTATTACGCATGCTTAAC
GSTU21_F	GTCTACGAGATATCATAACGTTACC
GSTU21_R	GTAGTGGTTAGCAACGATGAAGATCTAAG
GSTU19_F	AGGCGAATGAAGATATGTTACG
GSTU19_R	CACAGGAATCTGCTACAGCAAAGAC

5.3.6 Root length study of transgenic Arabidopsis plants

Root-length of seven day old seedlings was measured from photographs using ImageJ (Schneider *et al.*, 2012). The pair sample t-test was conducted using Sigma PLot V.13 (Systat Software, San Jose, CA).

5.4 RESULTS

5.4.1 The selection of sgRNA targets

The protein sequence of GSTU21, GSTU19, GSTU24, GSTU25 and GSTU22, were aligned to establish if five residues, highlighted by Tzafestas (2016) as key for TNT-conjugate production were present (**Figure 5.8**). The alignment showed that GSTU19 shares all five identical active residues found in GSTU25, suggesting the ability to produce similar conjugate produced by GSTU25 whereas, GSTU21 possesses identical active residues to GSTU25 except at residue G115 of GSTU25 (labelled as: ▪) which instead was alanine (A), as found in GSTU24. The alignment also revealed that GSTU22 possesses different residues with GSTU25 at position Y107, R111, G115, V208 and L212. Despite the highly identical active residues of GSTU24 and GSTU25 with GSTU21 and GSTU19, whether GSTU21 and GSTU19 have the activity towards TNT is still unknown. Therefore, all four GSTs were included in the CRISPR/Cas9 knockout study.

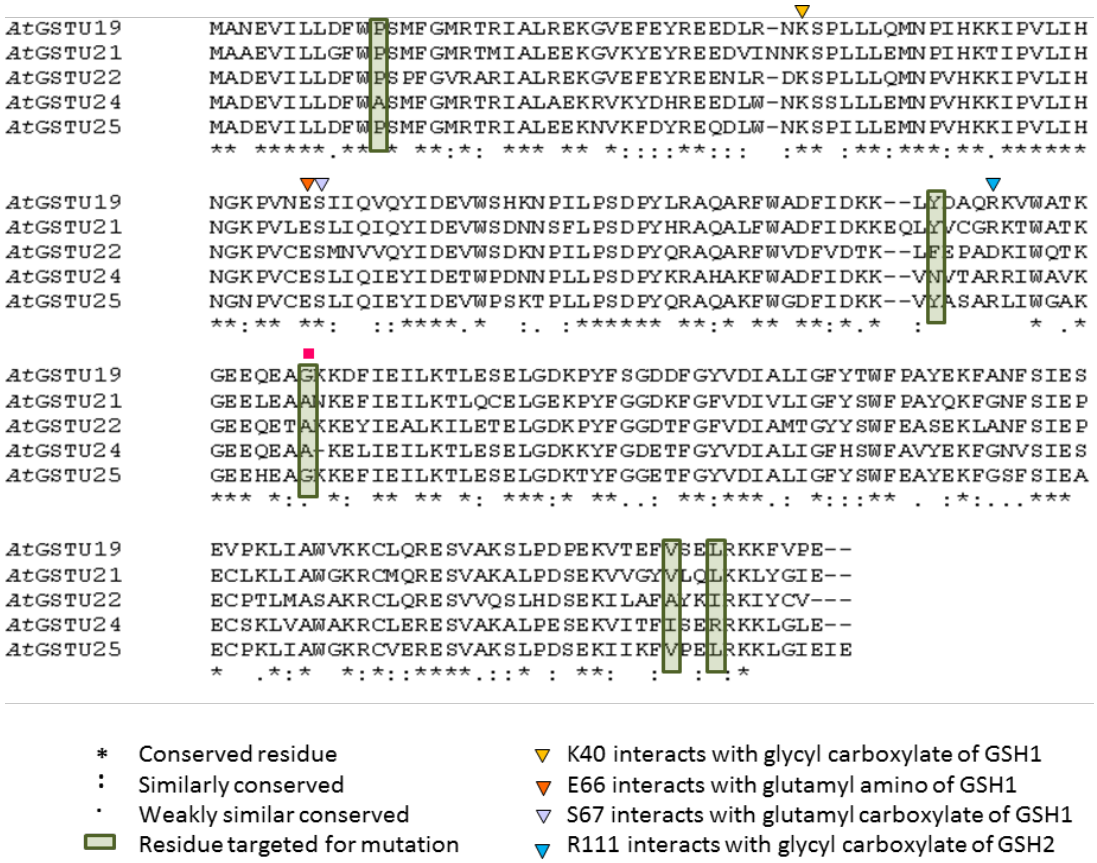


Figure 5.8. Multiple sequence alignment of GSTU19, GSTU21, GSTU24, GSTU25 and *GSTU22*. The residues highlighted were the active residues used in site-directed mutagenesis study of GSTU25 to analyse the ability of mutants to produce different types of TNT-GSH conjugates (Tzafestas 2016). Residues labelled with coloured triangles (▽) were GSTU25 residues interacted with GSSG molecule (as presented in **Chapter 4**).

The sgRNAs to target *GSTU19*, *GSTU21*, *GSTU24* and *GSTU25* were designed manually by selecting the 23 bp with lowest mismatches to each other and immediately preceding a PAM sequence (5'-NGG-3') (**Figure 5.9**).

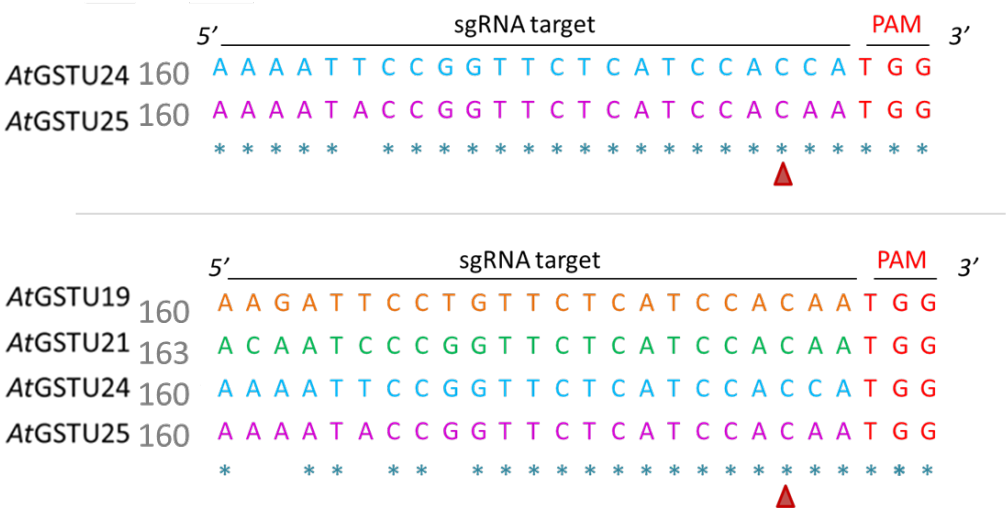


Figure 5.9. Top panel: Alignment of the sgRNA for targeting *GSTU24* and *GSTU25*. **Bottom panel:** Alignment of the sgRNA for targeting the four GSTs (*GSTU19*, *GSTU21*, *GSTU24* and *GSTU25*). The red triangles indicate the putative cleavage site (three nucleotides upstream of the Protospacer Adjacent Motif (PAM site). The asterisks (*) indicate conserved nucleotides. The target sgRNA is located ~160 bp downstream of the start codon.

5.4.2 Validation for the generation of mutants with gene mutations

5.4.2.1 Using Golden-Gate cloning method

Using the Golden Gate cloning method, PCR products containing fragments of sgRNA and binary vector carrying Cas9 (pCambia) were mixed in a reaction containing restriction enzyme *Bsa*I and T4 DNA ligase. The ligated product was transformed into *E. coli* DH5α and the positive clones were screened on kanamycin LB agar plates. The correct clones were identified by colony PCR and verified by sequencing. **Figure 5.10** shows the colony PCR of the selected *E. coli* colonies from the transformation. DNA bands of the expected size were cut out and after purification from the agarose, the bands were sent for validation by sequencing at GATC Biotech. Unfortunately, the sequencing failed. At this point, it was decided that due to the low cost, and relative speed of synthesising short DNA regions commercially, this would be a better route. Thus, the fragments shown in **Figure 5.11** were subsequently synthesised commercially by GeneArt (ThermoFisher Ltd).

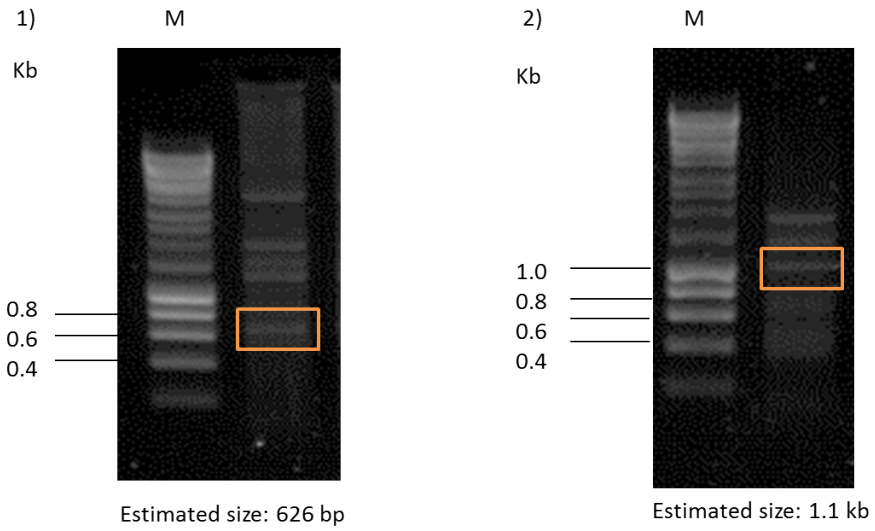


Figure 5.10. Agarose gel of 0.8% showing the product of colony PCR of *E. coli* DH5α carrying binary vector pCambia with sgRNA expression cassettes. 1) Colony PCR for *E. coli* carrying pCambia_2T yield a band at the expected size of 0.63 kb. 2) Colony PCR for *E. coli* carrying pCambia_4T yield band at 1.1 kb as expected.

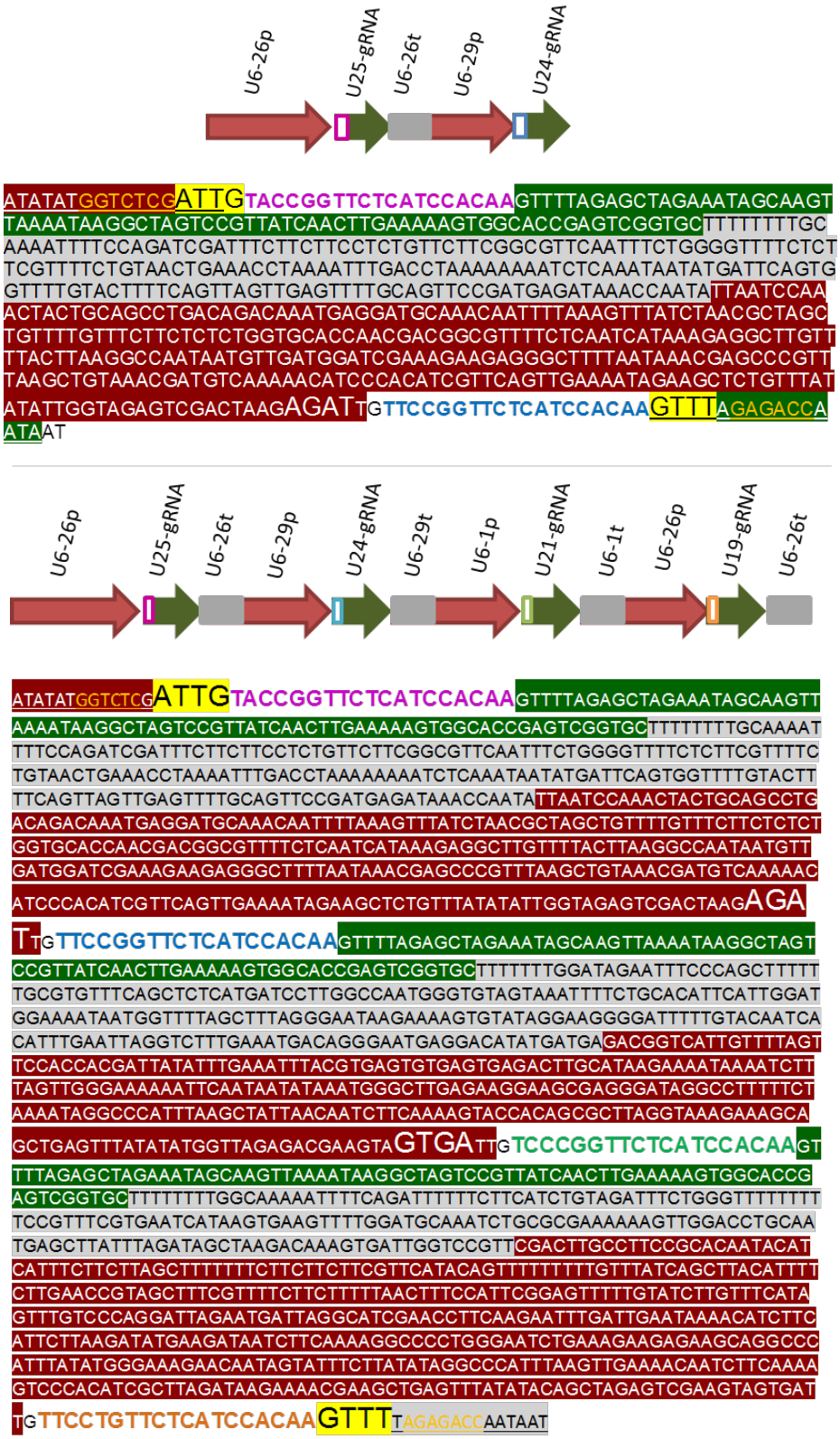


Figure 5.11. Fragment containing sgRNA scaffold with target (U25-gRNA, U24-gRNA, U21-gRNA and U19-gRNA), U6 promoters and terminators as synthesised by GeneArt. **Top panel**: fragment for pCambia_2T. **Bottom panel**: fragments for pCambia_4T. Underline letters come from pCambia vector, nucleotides highlighted in yellow: 5' protruding end produced by *Bsa*I (5'-GGTCTC-3') digestion, maroon: U6 promoters, green: gRNA scaffold, grey: U6 terminators. Coloured nucleotides represents the target sequence; pink: U25, blue: U24, green: U21, and orange: U19.

5.4.2.2 sgRNA fragments synthesised commercially by GeneArt.

The sgRNA expression cassettes containing the target site, gRNA scaffold, terminator and promoter flanked by *Bsa*I digestion sites were synthesised and cloned into pMK-RQ by GeneArt. Following digestion with *Bsa*I, the digested products were analysed on 0.8% agarose gels. Expected bands for the insert and wild-type pCambia plasmid digested with *Bsa*I were purified from agarose gels and ligated using T4 DNA ligase. The ligated products were transformed into *E. coli* DH5α. The transformants were screened on LB agar plates containing kanamycin (50 µg mL^{-1}). Colony PCR verified the insert in the colonies and plasmid purification was performed to collect the modified pCambia (pCambia_2T and pCambia_4T) (**Figure 5.12**).

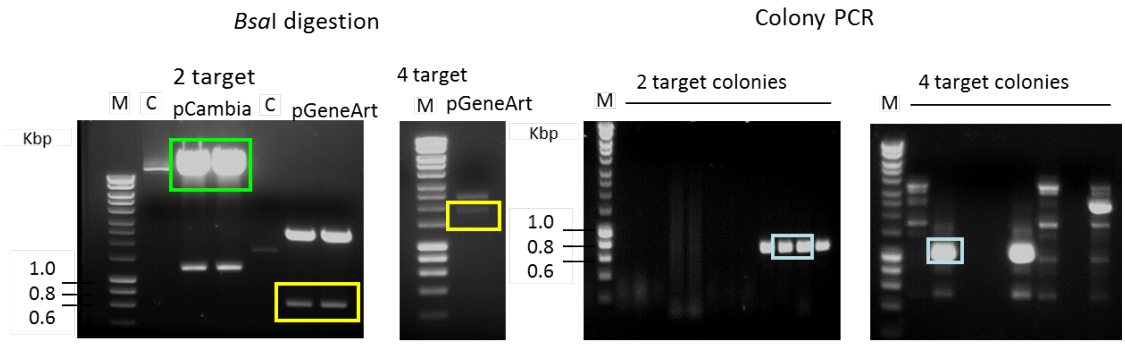


Figure 5.12. *Bsa*I digestion of plasmid pCambia and synthesised plasmid from GeneArt (pGeneArt). M indicated DNA marker Hyperladder 1 kb and C indicates undigested plasmid as control. Green box indicates digested pCambia vector while the yellow boxes indicate the sgRNA insert from pGeneArt vectors. Colony PCR confirmed the positives clones were carrying vector with sgRNA insert.

5.4.3 Transformation into Agrobacterium strain GV3101

The pCambia-2T and pCambia _4T plasmid from *E. coli* DH5α carrying two sgRNAs target and four sgRNA targets respectively, were purified and transformed into Agrobacterium by electroporation. The positive clones were identified by colony PCR (**Figure 5.13**).

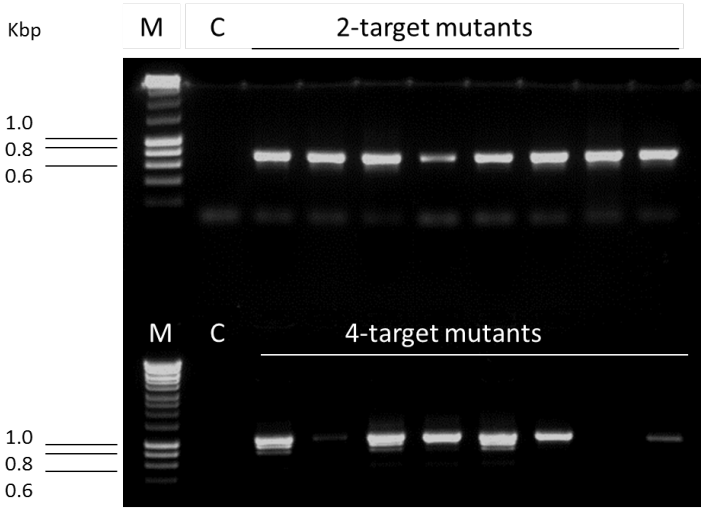


Figure 5.13. Colony PCR of Agrobacterium colonies grown on selective LB agar (50 µg mL^{-1} of gentamycin and 25 µg mL^{-1} spectinomycin). For pCambia _2T (top panel) expected size: 0.6 Kbp and pCambia_4T (bottom panel) expected size: 1,1 Kbp were observed.

5.4.4 Floral dip transformation of Arabidopsis Col-0

The floral dip transformation using Agrobacterium carrying pCambia_2T and pCambia_4T was carried out with eight pots of Arabidopsis for each type of plasmid construct. Each pot contained about 15 Arabidopsis flowering plants and pots were assigned as M2-1 to M2-8 for transformation with Agrobacterium carrying pCambia_2T while M4-1 to M4-8 were assigned for transformation of Agrobacterium carrying pCambia_4T.

After the floral dip transformation, the T1 seeds were collected and screened on ½ MS agar plates containing 30 µg mL^{-1} hygromycin.

From all eight pots transfected by Agrobacterium containing the pCambia_2T, only four seedlings, and one from each of M2-1, M2-2, M2-5 and M2-7 pots survived the screening on hygromycin agar plates. From the Arabidopsis plants transformed with pCambia_4T, only six hygromycin seedlings were obtained (one from each pot of M4-2, M4-3, M4-4, M4-6, M4-7 and M4-8).

Figure 5.14. Representative of the resistant seedlings survived on the hygromycin agar plate.

The T1 hygromycin resistant seedlings were transferred to soil and rosette leaf DNA checked for the presence of the Cas9 gene using PCR. **Figure 5.15** showed that Cas9 gene was present in all the resistant seedlings; except for line M4-7(2).

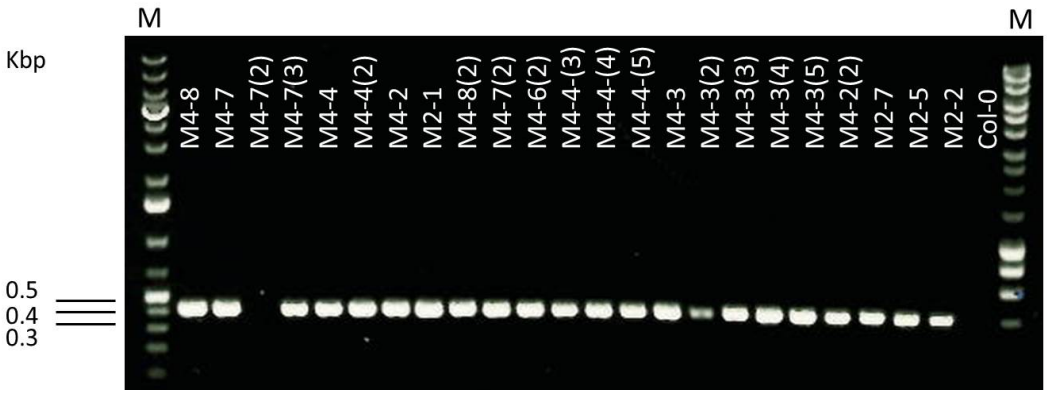


Figure 5.15. Cas9 validation of T1 plant genomic DNA by PCR. Expected size for Cas9 was 374 bp. Lane M: GeneRuler 1 kb Plus DNA ladder. M4-N: Mutant carrying four sgRNA target; M2-N: Mutant carrying two sgRNA targets. Control: DNA genomic of Arabidopsis Col-0.

5.4.5 Screening of T2 seeds on TNT agar plates.

After confirmation that the Cas9 gene was integrated into the genome of T1 transgenic Arabidopsis plants, the T2 seeds were collected and the T2 seedlings examined for altered resistance towards TNT. The seeds were germinated and seedlings were grown on ½ MS agar plates containing no TNT, 7 µM and 15 µM TNT. For the plants transformed with pCambia_2T, the seeds from T2 transgenic lines of M2-1, M2-2 and M2-5 were grown and for plants transformed with pCambia_4T, T2 seeds from M4-2, M4-3, M4-4, M4-7 and M4-8 were grown. The seedlings with TNT sensitive phenotypes, as judged by having root lengths

shorter (<5 mm) than wild-type seedlings when grown on TNT containing media, were subsequently transferred to soil (**Figure 5.16**).

For plants transformed with pCambia_2T, a total of 224 seedlings were screened on the TNT plates but only 20 seedlings from three mutant lines showed shorter root lengths, while for plants transformed with pCambia_4T a total of 77 out of 361 seedlings showed shorter root length on the TNT plate (**Table 5.3**). All of these seedlings were rescued by transferring to soil to grow and to collect the T3 seeds.

Table 5.3. Number of T2 seedlings with shorter (<5 mm) root length when grown on 0 µM, 7 µM and 15 µM of TNT for 10 d.

Mutant	No. of shorter root seedlings/Total seedlings on plate		Total number of shorter root seedlings
	7 µM	15 µM	
Progeny from plants transformed with pCambia_2T			
M2-1	3/44	4/48	7/92
M2-2	0/0	6/40	6/40
M2-5	3/44	4/48	7/92
Progeny from plants transformed with pCambia_4T			
M4-2	7/28	10/44	17/72
M4-3	5/43	14/42	19/85
M4-4	4/28	18/40	22/68
M4-7	1/17	10/47	11/64
M4-8	3/27	5/45	8/72

For the *gst*U*24*/*gst*U*25*-target plants, DNA sequencing of the 23 bp target region of 10 plants, for each mutation at the target sites, revealed a single nucleotide change was present in *gst*U*25* in three seedlings from the parent T2 transgenic line of M2-2. These were labelled as M2-2B, M2-2C and M2-2E (**Figure 5.17**). The seed for T3 generation of lines M2-2B, M2-2C and M2-2E were collected.

No mutations were observed in the *gst*U19/*gst*U21/*gst*U24/*gst*U25-target plants.

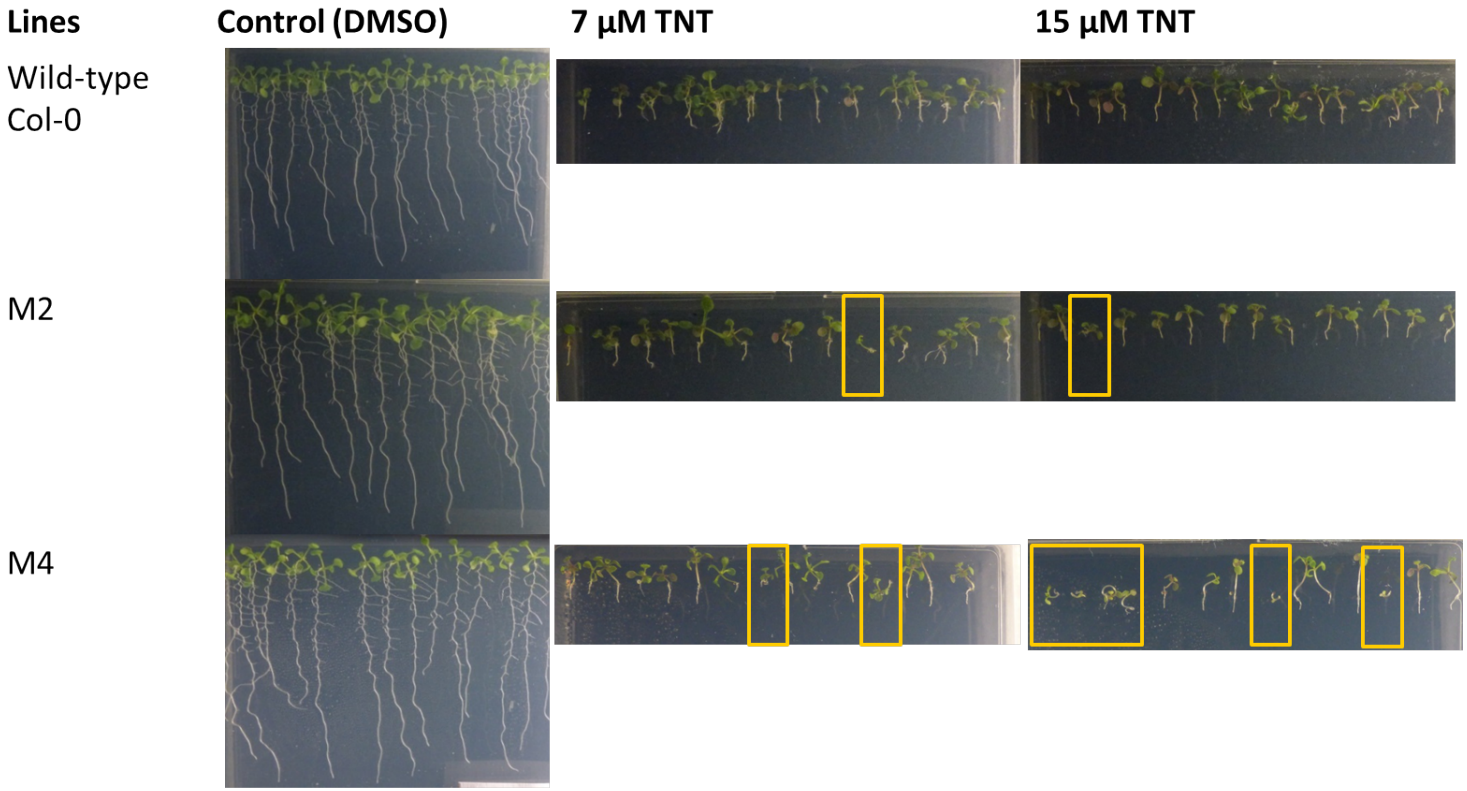


Figure 5.16. Representative plates showing the seven day old, T2 seedlings on plates containing 7 µM and 15 µM TNT. Seedlings with shortest roots (yellow boxes) were rescued by transferring to soil. The survived T2 plants were screened for mutation at the sgRNA target site by sequencing.

Wild-type Col-0	160	5 ' T A C C G G T T C T C A T - C C A - C A A T G G 3 '
M2-2B-U25	160	5 ' T A C C G G T T C T C A T C C C A - C A A T G G 3 '
M2-2C-U25	160	5 ' T A C C G G T T C T C A T - C C A T C A A T G G 3 '
M2-2E-U25	160	5 ' T A C C G G T T C T C A T C C C A - C A A T G G 3 '
		* *

Figure 5.17. Sequence alignment showing the mutation (insertion of a nucleotide base) in the 23bp target region of three seedlings from parent M2-2.

5.4.6 Segregating the mutation from Cas9 in mutant line M2-2

To ensure that the observed mutations were stable, counter selection for Cas9 was conducted on the genomic DNA T3 generation of the identified mutants: M2-2B, M2-2C and M2-2E using three primer pairs including two pairs of hygromycin-resistance gene; hygromycin-IDF/R, and hygromycin-IDF2/R2 together with zCas9-IDF/R for zCas9 and control primers for the actin gene (accession number: NC_003074), Act2-F/R for actin2. Wild-type Col-0 genomic DNA served as a negative control and genomic DNA from the T1 transgenic plants of M2-2 served as positive control. The results showed that only M2-2C had segregated from the Cas9 (**Figure 5.18**).

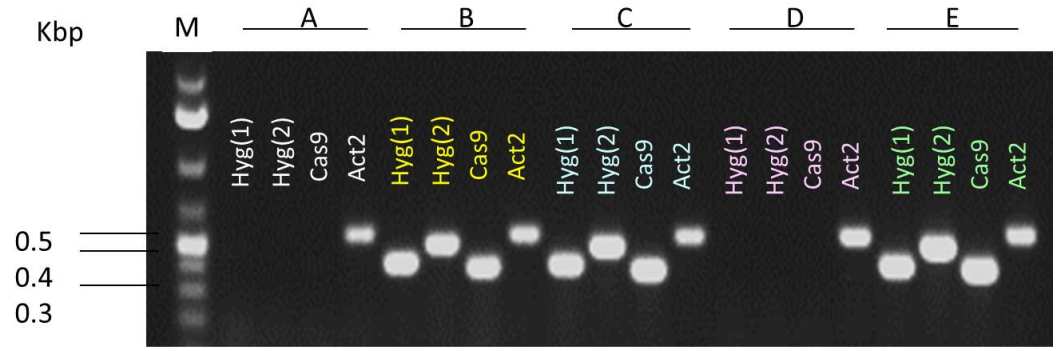


Figure 5.18. Counter selection PCR indicated that only M2-2C has segregated from Cas9. 1 kb plus DNA ladder was used as marker. A: Wild-type Col-0 (in grey), B: M2-2 (T1 line) (in yellow), C: M2-2B (T3 line) (in blue), D: M2-2C (T3 line) (in pink) and E: M2-2E (T3 line) (in green). Primers used for verification include two hygromycin primers were used (Hyg1 and Hyg2), Cas9 primers and Act2 primers.

To confirm that the *hptII* gene has been lost from the M2-2C T3 generation, seeds were germinated and grown on ½ MS agar plates containing hygromycin. The results in **Figure 5.19** demonstrated that, as expected, M2-2C T3 seedlings exhibited supressed growth similar to wild-type Col-0. Meanwhile, M2-2 from the T2 generation displayed 90% resistance to hygromycin indicative of the presence of Cas9 in the genome.

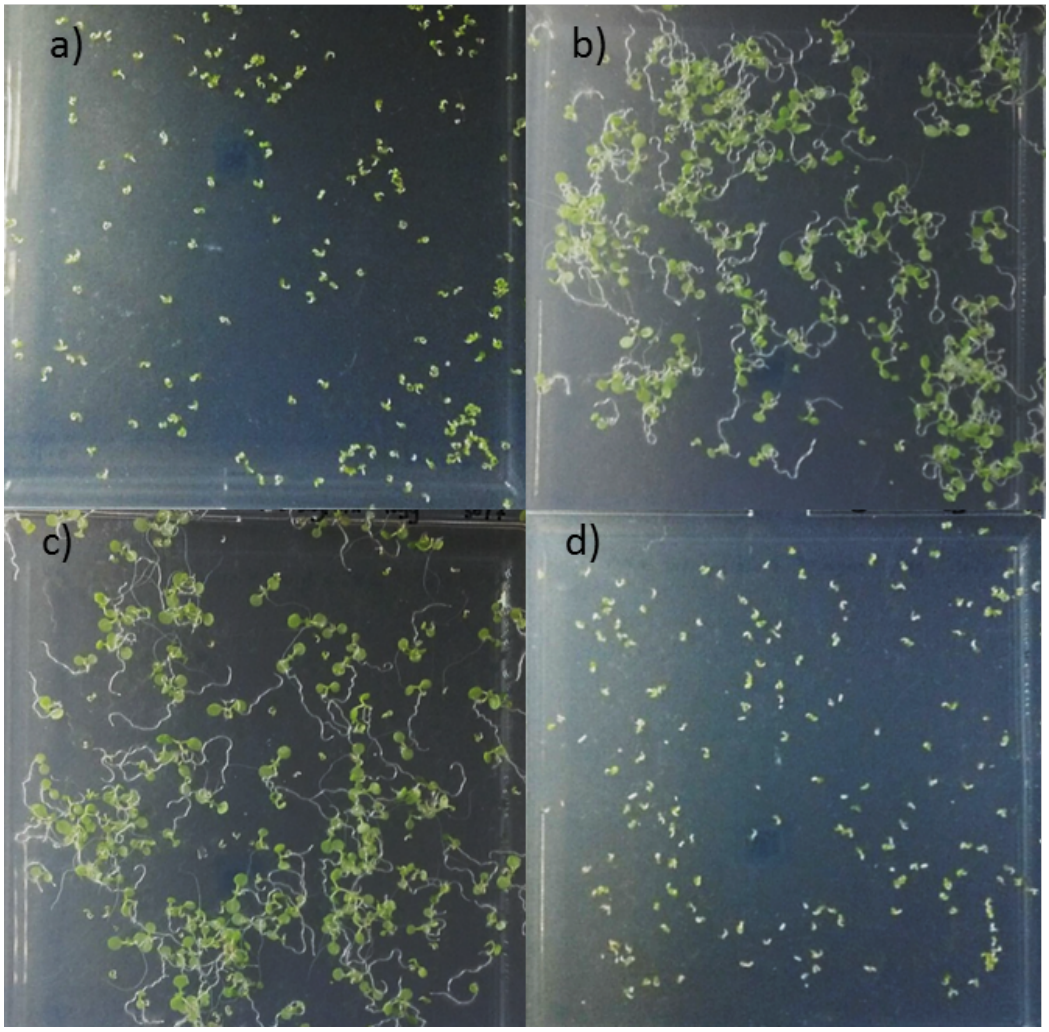


Figure 5.19. Screening of 10 d old seedlings on ½ MS agar plates containing 30 µg L^{-1} hygromycin. Panel a) Arabidopsis wild-type Col-0 (negative control), b) M2-2 T2 transgenic seedlings (positive control), c) M2-2B T3 transgenic seedlings (positive control) and d) M2-2C T3 transgenic seedlings.

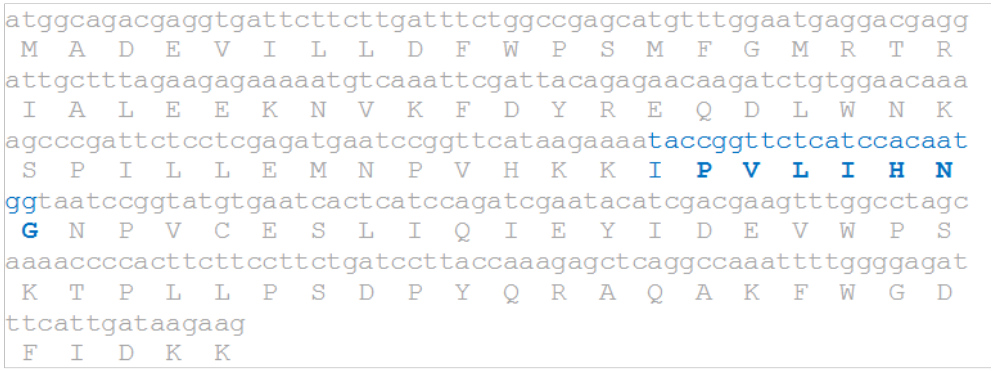
5.4.7 Sequence analysis of the *gst*U25 mutations

To understand the effect of the mutations on the translation of *gst*U25, the DNA sequences of the mutant varieties (M2-2B and M2-2C) were translated to protein sequence using the ExPASy Translate tool (Gasteiger *et al.*, 2003) and compared to the wild-type sequence, as shown in **Figure 5.20**.

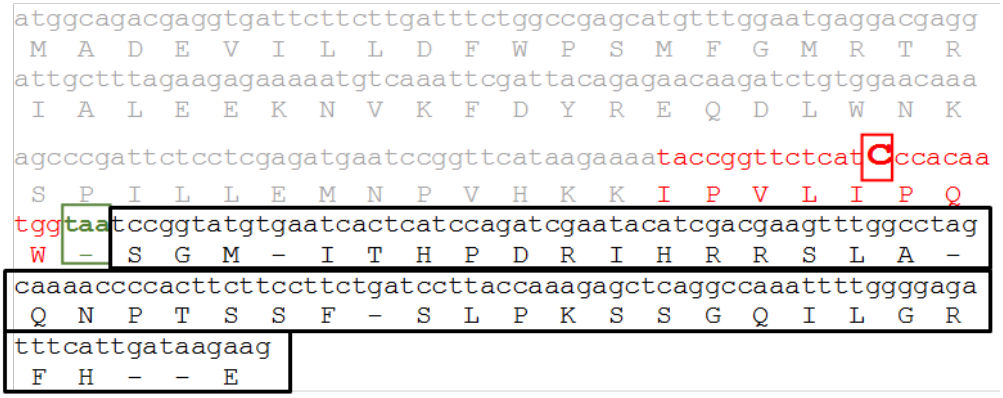
For M2-2B, the sequence analysis showed an insertion of cytosine (C) 174 base pairs from the ATG, and within the 23 bp target region as shown in **Figure 5.17**. The shifted codon frame changed the subsequent amino acid sequence to encode three non-matching amino acids, then a stop codon. This resulted in a truncated protein lacking 73% of the coding sequence from the amino terminus.

For M2-2C, the sequence analysis showed that an insertion of thymine (T) 177 base pairs from the ATG, and within the 23 bp target region as shown in **Figure 5.17**. The shifted codon frame changed the subsequent amino acid sequence to encode two non-matching amino acids, then a stop codon. As with M2-2B, this resulted in a truncated protein lacking 72% of its coding sequence at the amino terminus. Both M2-2B and M2-2C lacked the GSH binding (G-site) domain, a mutation that would most likely abolish the activity of *gst*U25 to bind to GSH and disrupt the hydrophobic binding site (H-site) found at the C-terminus of a GST protein.

A.



B.



C.

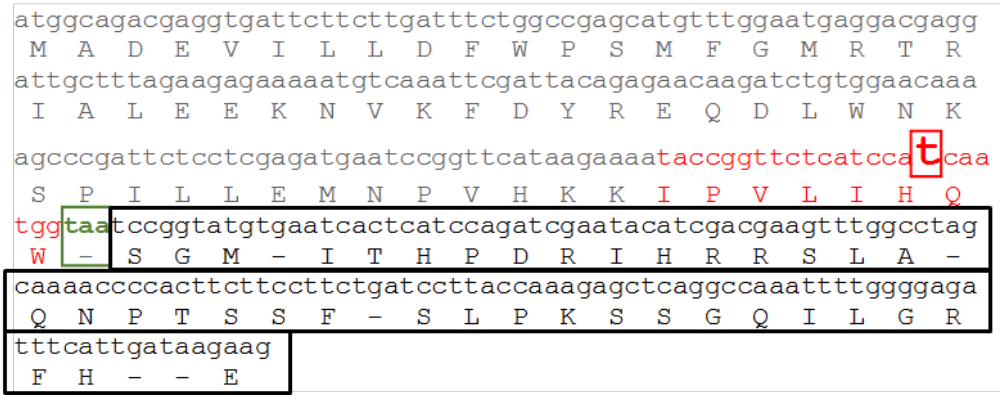


Figure 5.20. Sequence analysis of the mutant varieties showed a shifted codon after the site of insertion. The DNA sequence is translated to the respective amino acid using ExPASy Translate tool. A; Wild-type, B: M2-2B and C: M2-2C. Nucleotides in blue represent the wild-type sequence, nucleotides in red represent the mutant sequence and green nucleotides represent the stop codon. Nucleotides in red box represent the insertion and in green box represent the stop codon. Non-mutated nucleotides are in grey and changed nucleotides are in black box.

5.4.8 Testing resistance of mutant M2-2C to TNT

As segregated non-transgenic mutants carrying this mutation are fixed and no further modification from Cas9 could be happening, further analysis was carried out focusing on this segregated mutant progeny. The segregated line: M2-2C was grown on ½ MS agar plates containing TNT, root length measurements from seven day old seedlings were recorded and compared with the wild-type Col-0. The concentrations of TNT used were 0 µM, 7 µM, 15 µM and 30 µM. As shown in **Figure 5.21**, the root lengths of M2-2C seedlings grown on plates containing 0 µM, 7 µM and 15 µM of TNT concentration were indistinguishable to those of the wild-type (T-test analysis, P≥0.05). As reported by Yoon *et al.* (2007), both wild-type and M2-2C seedlings showed a decreasing trend of root length in higher concentration of TNT.

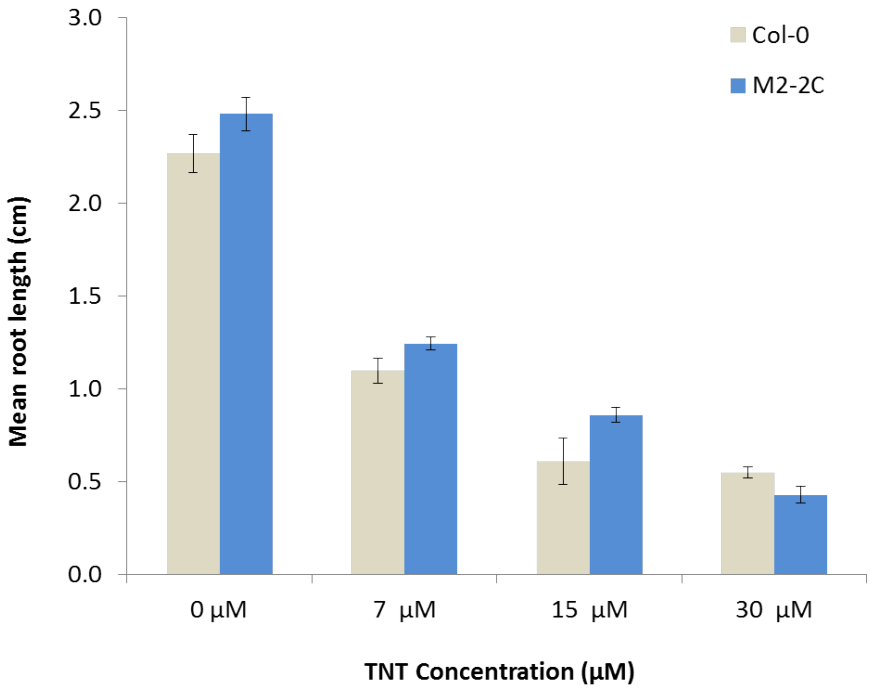


Figure 5.21. Average root length of seven day old M2-2C generation T3 Arabidopsis seedlings compared to wild-type (Col-O) (± SE of three replicates). T-test analysis indicated no significant difference at P≥0.05. Bars show the mean of ±SE of three replicate plates from about 35-47 seedlings on each plate.

As shown in **Figure 5.21**, the mean root lengths of the M2-2C are about 0.2 cm longer than the wild-type roots in the absence of TNT, although not statistically significant, (T-test analysis, P≥ 0.05). The seeds were from plants grown up alongside each other in the glasshouse, the slight differences in root length could perhaps be due to differences in the

size of the seeds or some other, unknown factor. To take this into account, the results were re-plotted relative to 100% of 0 µM of the respective lines. Similarly, no significant difference of M2-2C to the wild-type at each concentration was observed by T-test (P≥ 0.05) (**Figure 5.22**).

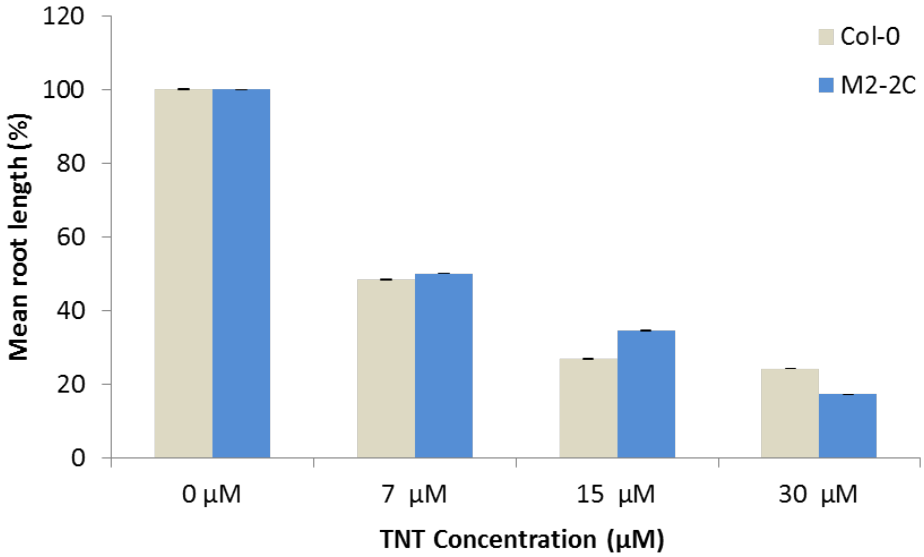


Figure 5.22. Average root length of seven day old wild-type (Col-0) seedlings and M2-2C generation T3 seedlings compared to no TNT (0 µM) for each group (± SE of three replicates). T-test analysis indicated no significant difference at P≥ 0.05. Bars show the mean of ±SE of three replicate plates from about 35-47 seedlings on each plate.

5.5 DISCUSSION

Studies by Gunning *et al.* (2014) showed that Arabidopsis plants overexpressing *GSTU24* and *GSTU25* had enhanced ability to withstand and detoxify TNT. Earlier, a transcriptomic study by Ekman *et al.* (2003) using Arabidopsis grown on TNT-containing soil recorded a series of upregulated genes including *GSTU1* and *GSTU24*. The inefficiency in studying only one gene in a multigene family, to observe any change in phenotype after exposure to TNT has encouraged the use of multiple knockouts of highly homologous genes. This study aimed to use CRISPR/Cas9 technology to knock out sub-clades of Tau class GSTs to see if a subsequent sensitivity to TNT could be elicited. Two plasmids were generated, pCambia_2T and pCambia_4T to target the knockout of *GSTU24* and *GSTU25*; and *GSTU19*, *GSTU21*, *GSTU24* and *GSTU25*, respectively.

The design of these experiments was carried out as recommended by Xing *et al.* (2014). Two specific sgRNAs were inserted, each targeting *GSTU24* and *GSTU25* for the double target mutants and for the four targets mutants, four specific sgRNAs were inserted into the expression plasmid pCambia carrying Cas9. The experiment was carried out on the basis that higher ratio of mutant frequencies would be further enhanced by the use of two or more gRNAs to target two or more genes. Xing *et al.* (2014) demonstrated that using this toolkit, higher ratios of T1 transgenic Arabidopsis seedlings were obtained (24/36 = 67% success from the T1 transformants) whereby another method, although using the same strategy of targeting two genes with two different sgRNA, expression cassette yielded a lower success rate of 23/60=38% of positive T1 transformants (Mao *et al.*, 2013).

However, the success of using multiple sgRNAs approach was severely hindered by the low number of primary transformants obtained (only four for the double *GSTU24*/*GSTU25* target and six for the quadruple *GSTU19*, *GSTU21*, *GSTU24*, *GSTU25* target). Two, albeit different, mutations in *gst*U25 were subsequently identified in the T2 progeny from only one of the four, double *GSTU24*/*GSTU25* targeted lines; with no mutations in *gst*U24. This gives a success rate of 25% for *gst*U25 and *gst*U24 respectively, and rather obviously, a 0% success rate for knocking out both genes; the main aim for this CRISPR/Cas9 approach. It is not known why the floral dip transformation rate was so low for these particular experiments, but this is a technique routinely used with high transformation frequencies by other members of the group, and a repeat experiment would be expected to yield many more primary transformants. Thus, a 25% success rate for *gst*U25 is perhaps not so bad. The *gst*U25 could be re-transformed with a construct targeting one or more alternative 23 bp sequences for *GSTU24*, or the pCambia_2T construct could be modified to include an

alternative 23 bp sequence to target *GSTU24*. Either way, it is certainly possible that a double *GSTU24*/*GSTU25* knock-out could be achieved using CRISPR/Cas technology.

In the absence of any mutations using the quadruple *GSTU19*, *GSTU21*, *GSTU24*, *GSTU25* target, and given the very low (six) number of primary transformants achieved, it is difficult to deliberate on what went wrong. As the approach taken used 23 bp taken from the same region for all four genes, it is possible that this is, for some reason, was just a 'bad' region for CRISPR/Cas9. Perhaps there is some strong secondary structure to the DNA/RNA that inhibits the binding of the Cas9.

For M2-2B, the mutation was an insertion of cytosine (C) at six nucleotides upstream of the PAM while for M2-2C, tyrosine (T) was inserted three nucleotides upstream of the PAM site. The insertion of the nucleotide by the NHEJ repair mechanism can cause the codon frameshift and introduction of a premature stop codon. The generation of stop codon is often observed as a result of insertion or deletions of a nucleotide which essentially increase the chances to form stop codon (TAG, TAA and TGA) (Ran *et al.*, 2013). The location of insertion T is known as the cleaving site of Cas9 (Belhaj *et al.*, 2015). Insertion as the type of editing has been found as the second most frequent type of genome editing by Cas9. Liang *et al.* (2016) reported that the highly efficient sgRNAs for CRISPR/Cas9 resulted about 36% of insertion and 59% of deletion at the target site. Among the three mutants, screening on agar plates containing hygromycin revealed that the mutant with insertion T was the mutant line that has segregated from the Cas9.

Of the T2 seedlings identified as having shorter roots on agar plates containing 7 and 15 µM TNT, only two were found to have mutations, and they were only in *gst*U25. For plants with pCambia_2T, a total of 20 seedlings were rescued from TNT plates and about 77 seedlings from plants with pCambia_4T. From this, only 10 healthy plants carrying pCambia_2T and pCambia_4T were sequenced. It is likely that those non-mutated seedlings with shorter root lengths were the result of natural variations within the Col-0 genome in combination with variation in seed quality; smaller seeds yielded shorter seedling root lengths.

Longer roots for the mutant line were observed when compared to the wild-type on plates without TNT (0 µM) of about 0.2 cm, although not statistically significant, (T-test analysis, $P \geq 0.05$) (**Figure 5.22**). Without excluding some unknown technical error on the wild-type plates, off-target mutation could have occurred causing mutation of another unknown gene that contributes to root growth in Arabidopsis. Although this scenario is extremely unlikely;

GSTU24 was not targeted, creating multiple, independent mutants would be required to give statistical robustness to subsequent publications on any phenotypes. When the results were plotted relatively against the lines grown in 0 µM, no significant difference was observed between the mutant line and the wild-type at each tested concentration (T-test analysis, P≥0.01). This non-observable phenotype from one gene mutation is expected because Yoon *et al.* (2007) has demonstrated that mutation of only *gst*U24 showed no significant difference from the wild-type, and later Gunning *et al.* (2014) has showed that multiple genes, *GSTU24* and *GSTU25* were responsible in the response of TNT.

Demonstrating that the CRISPR/Cas9 technique, which had not previously been performed within CNAP, has been successfully used to knock *GSTU25*, is a positive step towards knocking out sub clades of the GST family. Since this study, more improved constructs have been developed such as those used by Lawrenson *et al.* (2015) at the John Innes Centre, UK. The results here will pave the way for successful future studies knocking out multiple members of the GST family and elucidating the roles of these captivating enzymes.

6 Final discussion

The capacity of GSTs to undergo catalytic and non-catalytic interactions, either dependent or independent of GSH, has been investigated for over a decade. Despite the research, little is known of the individual roles played by GSTs, particularly with regard to non-catalytic interactions. This study builds on the knowledge gained from research on two Arabidopsis GSTs from different classes, Phi and Tau, which were shown to have two different activities. The Phi GST, GSTF2, has been shown to act as a carrier of natural metabolites and small molecules. The Tau GST, GSTU25 has direct, conjugating activity of GSH to the prominent environmental pollutant and explosive compound TNT. The results presented here revealed the crystal structure of both proteins and elucidated the site of binding between the previously identified interacting ligands and the proteins. Knockout mutagenesis of GSTU25 was used to further elaborate the significant involvement of this GST in the interaction with TNT.

Chapter 3 described the structural characterisation of GSTF2 with selection of ligands described by Dixon *et al.* (2011); quercetrin, quercetin, indole-3-aldehyde and camalexin. The structures show these non-substrate ligands present at two main binding sites on the dimer interface, defined as **L1** and **L2**. The significance of these binding sites was tested using site-directed mutagenesis and calorimetry techniques and revealed that four residues at the binding sites: Q73, Y97, H77 and R154 were important in the binding interaction.

Non-substrate binding has been postulated to be independent from the GSH-conjugation reaction. The structures obtained in this study supports this hypothesis by showing distinct binding sites between the ligands and GSH in the superimposed structure of GSTF2 in complex with *S*-hexylglutathione (PDB ID: 1GNW) and quercetrin (**Figure 6.1**). However, the independence of non-catalytic and catalytic binding sites was not found in all GSTs. For example, the crystal structure of human hGSTP1-1 with bromosulfophthalein (BS), a molecule that is relatively large (837.99 g mol^{-1}) used in dyeing industry, showed that the non-catalytic binding site is located in part of the H-site of the hGSTP1-1. Co-location results in non-competitive binding kinetics in the standard CDNB-GSH reaction assay (Oakley *et al.*, 1999).

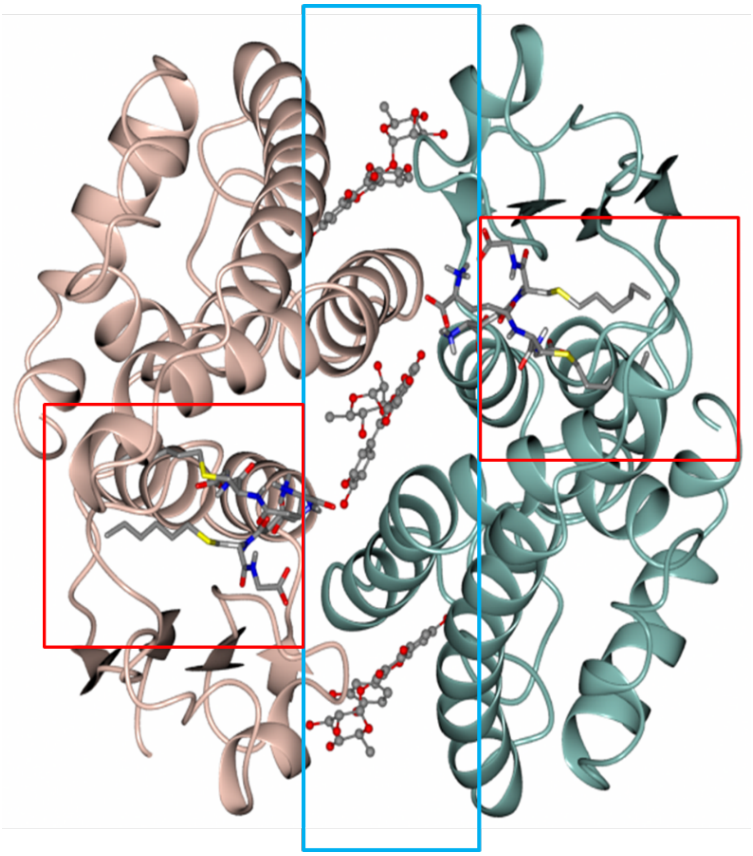


Figure 6.1. The structure of GSTF2 in complex with quercetrin superimposed with *S*-hexyl glutathione (PDB ID: 1GNW) showing distinct regions of the non-catalytic ligand binding (quercetrin) shown in blue box and the known G-site for GSH binding in red box.

The ligands investigated in this study represent the classes of biologically active plant secondary metabolites derived from indoles and polyphenols. None of these ligands have previously been described as substrates for GSTs. Moreover, most of the compounds described for non-catalytic binding to GSTs contained planar surfaces and had molecular weights below 900 g mol^{-1}. For instance, porphyrin, which has a molecular weight of 308.3 g mol^{-1}, binds to the non-catalytic site of the *Taenia solium* (tapeworm) *Ts*26GST (Plancarte *et al.*, 2014), and in plants, porphyrinogens, auxin, ethylene, phytoalexins and flavonoids which all have molecular weights below 500 g mol^{-1} contain planar aromatic rings bound to GSTs from *Z. mays* (*Zm*GSTU1 and *Zm*GSTU2) and Arabidopsis (GSTF2) (Dixon *et al.* 2008; Dixon *et al.* 2011; Smith *et al.* 2003). For a compound to diffuse passively across cell membranes, low molecular weight and relatively hydrophobic properties are essential characteristics. Furthermore, compounds can also cross the membranes by facilitated diffusion with the help of carrier proteins (Cooper 2000). The ability of GSTs to bind to these types of compounds suggests a possible role for transporting the molecules to a site of action, as demonstrated for the sequestration process of anthocyanin from the

cytoplasm into the vacuole by *Vitis vinifera* (grapevine) *Vv*GST4; Arabidopsis GSTF12, *Petunia hybrida* (petunia) AN9, *Z. mays* (maize) BZ2, *Cyclamen* spp. (cyclamen) *Ckm*GST3, *Perilla frustescens* (perilla) *Pf*GST1 and *T. aestivum* (wheat) *Ta*GSTL1 (Conn *et al.*, 2008; Kitamura *et al.*, 2004; Kitamura *et al.*, 2010; Alfenito *et al.*, 1998; Marrs *et al.*, 1995; Kitamura *et al.*, 2012; Yamazaki *et al.*, 2008; Dixon & Edwards 2010).

Heterocyclic compounds are molecules joined in five or six membered carbon rings and contain heteroatoms of nitrogen, oxygen and sulphur (IUPAC, 1997). In addition to the ligands found by Dixon *et al.* (2011) in their GSTF2 ligand binding study, the expression of *GSTF2* has also been influenced by the induction of heterocyclic herbicide safeners such as benoxacor, fenclorim and fluxofenim (DeRidder and Goldsborough 2006). Herbicide safeners are non-phytotoxic, synthetic compounds that reduce herbicide toxicity in crops by inducing the expression of GSTs to promote conjugation of the herbicide to GSH (Davies & Caseley 1999, DeRidder and Goldsborough 2006). Interestingly, more than 70% of the compounds used in the agrochemical industry carry at least one heterocyclic ring (Bartlett *et al.* 2002) which could potentially bind to GSTF2 in a similar reaction to the ligands used in this study. There are several databases such as ChEMBL (Bento *et al.*, 2014) and PubChem (Kim *et al.*, 2016) which provide structural information on biologically active chemicals that could be used for structure-based design. Furthermore, GSTF2 is a highly soluble protein, easily expressed and purified, allowing high-throughput X-ray crystallography screening to identify and investigate any interactions with other important agrochemical products.

Chapter 4 expanded the knowledge of GST enzymatic activity by elucidating the GSTU25 structure in complex with glutathione disulphide (GSSG). The structure of GSTU25 is represented by the typical GST protein architecture which consists of a dimer with mixed α-helical and β-strand domains, at the N-terminal of the protein followed by an all α-helical domain at the C-terminal. The structure complex exhibited one GSSG molecule bound to each subunit of the dimer at the glutathione binding site (G-site), a location similar to that for *S*-hexylglutahione on GSTF2 published by Reinemer *et al.* (1996) and shown in **Figure 6.1**. Although the initially desired complex of GSTU25 with GSH-TNT conjugate could not be resolved, it is intriguing to note that pinky-red-coloured crystals were observed on the crystal screening plate. The aqueous solubility limit for TNT is approximately 500 µM (at 25°C), and at this relatively low concentrations, the TNT solution is colourless. However, at 2 mM using DMSO as solvent, as used in the crystallisation process (incubated at 18°C), the solution was pinky-red. The presence of coloured crystals suggested that TNT had been

incorporated into the active site of GSTU25; however, no TNT molecules were observed in the X-ray diffraction data. Instead, GSSG molecules were seen, suggesting that the TNT-GSH conjugation involves GSH oxidation and that the reaction occurred too quickly to be captured by the crystallisation process. In agreement with this, data on the oxidation of ozone with different organic compounds, including chlorinated alkenes, phenols and ketones stated that the oxidation process occurs very rapidly, in a range of 0.01 to 10^4 M^{-1} s^{-1} (Munter 2001). While many parameters are involved in the protein crystallisation process, the crystallisation rate is much slower than the oxidation reaction.

A study by Gunning *et al.* (2014) revealed that, using TNT as substrate, GSTU25 can produce three types of GST-TNT conjugates. The profile of the conjugates varies with pH, with conjugate **3**, 2-glutathionyl-4,6-dinitrotoluene, being the dominant product at pH 6.5 – 7.0. Evidence suggests that this pH range is similar to that in the root cytosol (Irzyk & Fuerst 1993; Guddewar & Dauterman 1979), the endogenous location of GSTU25. The three opposing nitro groups in TNT give structural stability, and are a major reason for the low biodegradation rates of this compound in the environment. Therefore, in conjugate **3**, the loss of a nitro group could make it more susceptible to subsequent biodegradation either *in planta*, or by the soil microbial communities. Studies have demonstrated the detoxification pathways in Arabidopsis for TNT (Hannink *et al.*, 2001; van Dillewijn *et al.*, 2008; Rylott *et al.*, 2011). In the pathway, nitroreductases, including oxophytodienoate reductases (OPRs) (Beynon *et al.*, 2009) reduce TNT to hydroxylamino dinitrotoluenes (HADNTs) which can be further reduced to aminodinitrotoluenes (ADNTs). These transformation products are subsequently glycosylated by glucosyl transferases. A future study could investigate whether the glutathione can be catalytically cleaved from the conjugate **3**, leaving dinitrotoluene (DNT) which could be subsequently degraded *in planta* or soil bacteria. Pathways for the degradation of DNT and structurally similar compounds by microorganisms have been previously well characterised (Spanggord *et al.*, 1991; Nishino & Paoli 2000; de las Heras *et al.*, 2011). In particular, DNT degradation has been studied in detail in *Burkholderia* sp., where the process is initiated by hydroxylation of DNT (Monti *et al.*, 2005; Suen & Spain 1993; Johnson *et al.*, 2000) (**Figure 6.2**).

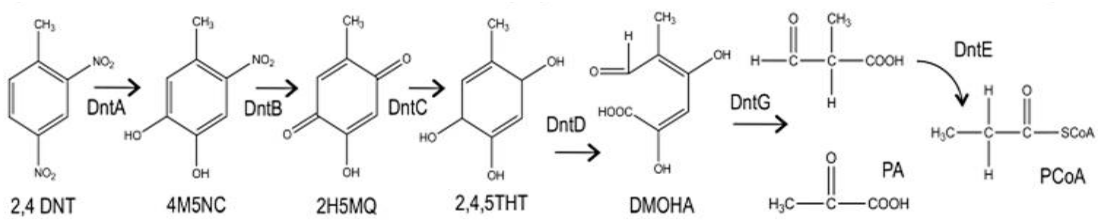


Figure 6.2 Degradation pathway of DNTs as identified in *Burkholderia* sp. strain DNT. Through oxygenase attack at the aromatic ring (Suen & Spain 1993). DntA: multicomponent DNT oxygenase, Dntb, 4M5NC monooxygenase, DntC: unidentified endogenous reductase, DntD: 2,4,5-THT oxygenase, DntG: DMOHA isomerase/ 4-hydroxy-2-keto-5-methyl-6-oxo-3-hexenoate hydrolase, DntE: a methylmalonate semialdehyde dehydrogenase. The figure and text reproduced from de las Heras *et al.*, (2011).

The initial aim of obtaining the GSTU25 structure in complex with GSH-TNT conjugates was to understand the binding site of conjugate **3** in GSTU25. As the crystallisation was conducted at pH 5.5 in the absence of a purified GSH-TNT adduct, this aim proved to be challenging. This could perhaps be overcome in the future by crystallising GSTU25 in the presence of purified GST-TNT adducts. The adduct could be synthesised from TNT and GSH using purified GSTU25. The final conjugate products can be purified by HPLC as described in Gunning *et al.* (2014). Understanding the structure of GSTU25 with TNT-GSH conjugates is important to elucidate how TNT interacts with GSTU25 and how this interaction contributes to the enhanced tolerance as evident in Arabidopsis plants overexpressing GSTU25 in TNT-treated environment, shown by Gunning *et al.* (2014).

The structure of GSTU25 in this study confirms the data obtained from GSTU25 site-directed mutagenesis studies conducted by Tzafestas (2016). He identified conserved residues by multiple sequence alignment to Tau class GST structures including from wheat, *T. aestivum* (*Ta*GSTU4-4) (Thom *et al.*, 2002), soybean, *G. max* (*Gm*GSTU4-4) (Axarli *et al.*, 2009) and a known Arabidopsis GST: GSTU24 (Gunning *et al.*, 2014), also a model of GSTU25 using *Gm*GSTU4-4 as a template. The structure of GSTU25 obtained in this study could be further analysed using *in silico* docking to look for the potential TNT-GSH conjugate binding site by the estimation of van der Waals interactions and Coulomb energy scores. Combining with enzyme kinetic data of GSTU25 mutants using TNT as the substrate, key information on the protein function in TNT detoxification can be achieved.

Chapter 5 describes a converse approach to the GSTU25 overexpression study conducted by Gunning *et al.* (2014). In this chapter, CRISPR/Cas9 technology was used in an attempt to

decrease, or knock-out, the activity of GSTU25 and closely-related GSTs. This study was designed with consideration of the previous study by Yoon *et al.* (2007) which indicated that an Arabidopsis line carrying a *gstU24* mutant did not exhibit any significant differences in the uptake rates or tolerance changes at different concentrations of TNT as compared to the wild-type plant. The authors concluded that this result was due to redundancy between the closely related GSTs, and particularly with *GSTU25*. When disruption at only one gene results in negligible changes, similar to the wild-type, simultaneous modification of multiple homologous genes are often needed to reveal the targeted phenotype.

The genes used for CRISPR/Cas9-based modification were selected according to the classification on the phylogenetic tree published by Dixon & Edwards (2010) which classified *GSTU25* within the same clade of *GSTU24*, *GSTU21* and *GSTU19* based on polypeptide similarity; along with studies by Lorenz (2007) which showed the upregulation of arrays of *GSTU*s genes in response to TNT. The *GSTU21* and *GSTU19* were not represented in the ATH1 chip used in the microarray study, however, the *GSTU22*, which was located within the same subclade in the phylogenetic tree (refer **Figure 1.13**), was recorded to be upregulated by 14-fold despite showing no TNT-GSH conjugating activity on Griess assay (Lorenz 2007). The only way to gain some information about the activity of *GSTU21* and *GSTU19* towards TNT was by comparing their sequence with *GSTU24*, *GSTU25* and the *GSTU22*. Sequence comparison carried out in this chapter suggested that the lack of GSTU22 TNT-GSH conjugating activity was due to the absence of the active residues which was not the case for GSTU21 and GSTU19 (**Figure 5.8**). Therefore, all four GSTs, *GSTU21*, *GSTU19*, *GSTU24* and *GSTU25* were selected in the design of CRIPSR/Cas9 vector construct, pCambia_4T.

Despite the technical obstacle in obtaining multiple knockout mutants, a stable *gstu25* mutant line, segregated from Cas9 was accomplished. Root-length measurements of *gstu25* seedlings on TNT-treated agar confirmed the finding that was presented by Yoon *et al.* (2007); that the *gstu25* seedling root lengths were not significantly different to those of wild-type seedling roots. Unfortunately, the question of whether *gstu25* contributes directly in TNT detoxification currently remains unanswered.

Although, only *gstu25* was mutated, there is information in the literature suggesting strategies that could be used to enhance the knock-out efficiency. It is important that the target sgRNA is specific to the gene to be knocked-out. The study presented here used the technique, and vectors as reported by Xing *et al.* (2014). The CRISPR/Cas9 technology is still rapidly advancing and further improved vectors will inevitably become available in the

future. To knock out more than one gene with this technique, it is important that the sgRNA sequence matches both genes as closely as possible. Xing *et al.* (2014) reported that a one base mismatch resulted in a 10% decrease in knock-out efficiency, two bases mismatch resulted 64% decrease in the efficiency and three bases mismatched abolished the entire knockout ability.

In the study presented here, to knock-out *GSTU24* and *GSTU25*, two sgRNAs were used, one targeting *GSTU24* and the second targeting *GSTU25*. However, both sequences targeted the same location on the respective genes. The knock-out efficiency of *GSTU24* could perhaps be improved by using either two or multiple sgRNA sequences targeting to a different DNA region of the *GSTU*24 gene. Such dual or multiple sgRNAs-directed knockout have been employed to increase the chances of gene disruption (Chen *et al.*, 2014; Song *et al.*, 2016). Caution is needed though, as a study by Chen *et al.* (2014) demonstrated that dual or multiple sgRNAs-directed approach can cause off-target deletion of related sequences (Chen *et al.*, 2014) and could be a problem in a gene family such as the GSTs where there is high sequence similarity.

A study by Ann Ran *et al.* (2013), using a double sgRNA-Cas9 nickase system, demonstrated a high specificity and increased occurrences of mutations at the target site (Ann Ran *et al.*, 2013). The system combines two Cas9 nickases with two different sgRNAs and introduces single DNA strand breaks simultaneously at two different target sites according to the designed sgRNA. The use of a pair of Cas9 nickases targeting opposite strands can generate DNA double strand break (DSB) within the target DNA. The Cas9 nickase is a modified Cas9 wild-type mutated at either one of the two nuclease domains, the HNH-like domain and RuvC-like domain at residue D10A and H840A respectively (Ann Ran *et al.*, 2013). Although two sgRNAs need to be designed, this method offers better specificity and higher chances in targeting the gene. The Arabidopsis Cas9 nickase method has been evaluated and shown to give significant increase in target specificity and knockout efficiency (Fauser *et al.*, 2014; Schiml *et al.*, 2014).

Combining information from the structure of GSTU25, the residues involved in the interaction with TNT from the mutagenesis study, and the characteristic of each residue towards TNT, plants suitable for field application can potentially be manipulated to tackle TNT contamination in soils using CRISPR/Cas9.

That Arabidopsis has many attributes making it a suitable model species for laboratory studies, it is not well-suited for phytoremediation applications, due to factors including its

small size and shallow root systems. Plant species favourable for phytoremediation generally have relatively deep, penetrating root networks, fast growth rates and high biomass production such as *Panicum virgatum* (switchgrass) (Brentner *et al.*, 2010), *Vetiveria zizaniodes* (vetiver) (Makris *et al.*, 2007), *Agropyron smithii* (western wheatgrasses), *A. fragile* (Siberian wheatgrasses) and *A. trachycaulum* (slender wheatgrasses) and tree species such as *Populus* spp. (poplar) (Brentner *et al.*, 2010) and some conifer trees including the *Picea abies* (Norway spruce) (Schoenmuth & Pestemer 2004).

Analysis of GSTU25 homology genes in poplar revealed that a *Populus trichocarpa* GST, *Pt*GSTU9 shares 67% sequence identity with GSTU25. An orthologue of GSTU24 has also been identified in *P. trichocarpa,* known as *Pt*GST173 (Brentner *et al.*, 2008). Although TNT has been found to be toxic to several species including poplar (Hannink *et al.* 2002), genetic engineering might be a way to enhance the ability of the plants to resist and detoxify TNT. Indeed, a study by Dillewijn *et al.* (2008) transformed poplar with a bacterial nitroreductase gene (*pnrA*) to enhance the ability of species in this genus in TNT detoxification. One of the most efficient and successful CRISPR/Cas9 editing was reported in the poplar tree, hybrid of *P. tremula x alba* clone 717-1B4 (Zhou *et al.* 2015). A similar approach can be applied for the remediation of TNT in soil. For instance in poplar, considering the availability of the whole genome data of *P. trichocarpa,* identification of GSTU25 orthologues in *P. trichocarpa* followed by modification of identical residues important for the production of conjugate **3** as in GSTU25, can be used to generate poplar tree that could potentially be manipulated as a 'factory' producing TNT-GSH conjugate **3**.

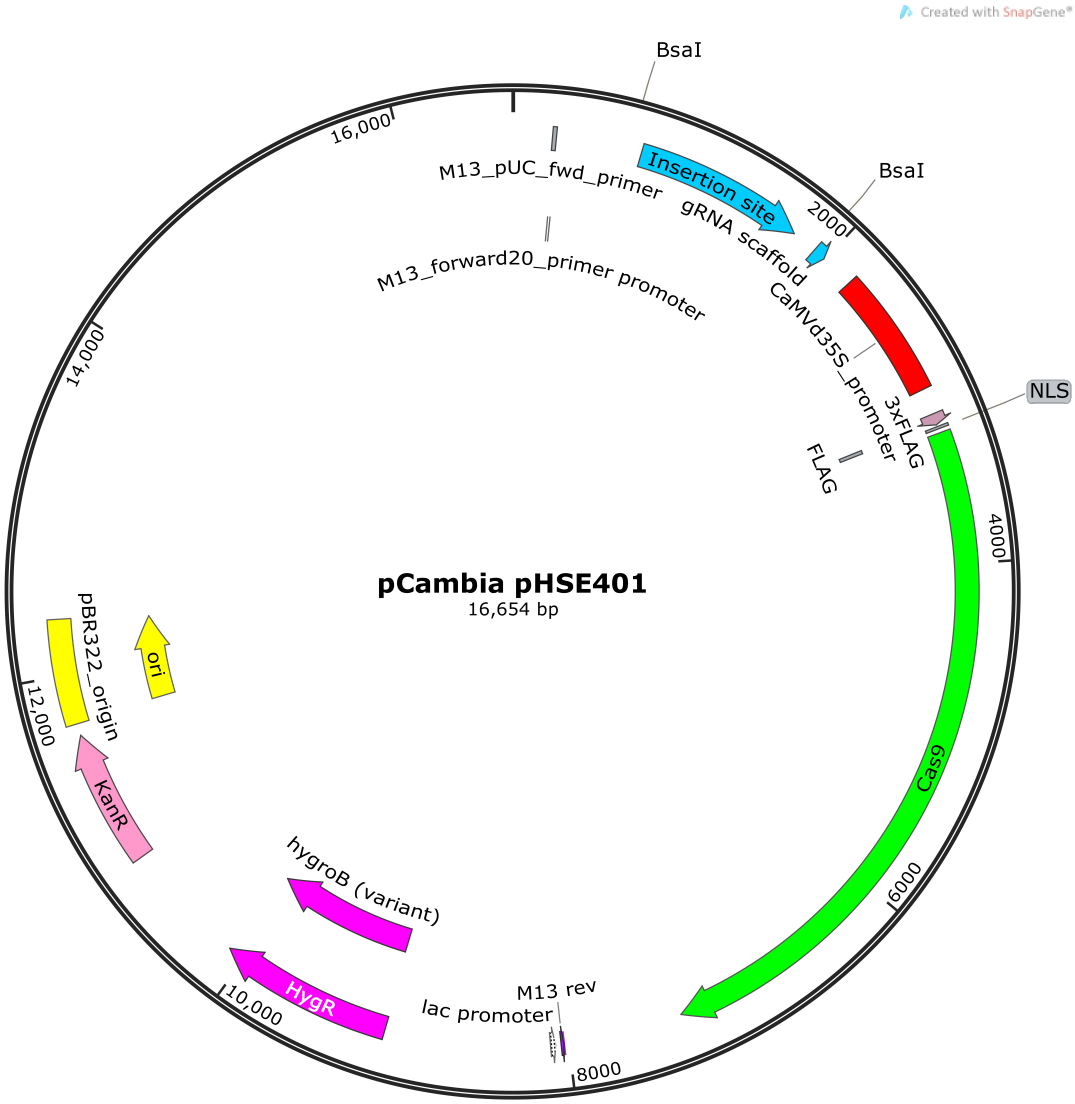
Studies have also used the rhizosphere (the area surrounding plant root) endophyte of *Acer pseudoplatanus* grown at a TNT-contaminated location, to investigate microbial TNT-transformation (Thijs *et al.*, 2014). This work characterised several *A. pseudoplatanus* associated bacteria including *Pseudomonas* spp., the leaf endophyte *Variovorax ginsengisola* and the leaf symbiont *Stenotrophomonas chelatiphaga* which were then used to inoculate *Agratis capillaris* grass (common Bent). The inoculation of these bacteria has stimulated the growth of the bent grass under TNT stress (Thijs *et al.*, 2014).

In conclusion, the work presented in this thesis has increased the knowledge of GSTs activity in interaction with non-substrate small molecule and also in regards to TNT detoxification in plant. While the structure of GSTF2 provides insights on the new binding sites of small molecules that have not been identified before, the crystal structure of GSTU25 in complex with GSSG represented the first structure of Tau class GST from

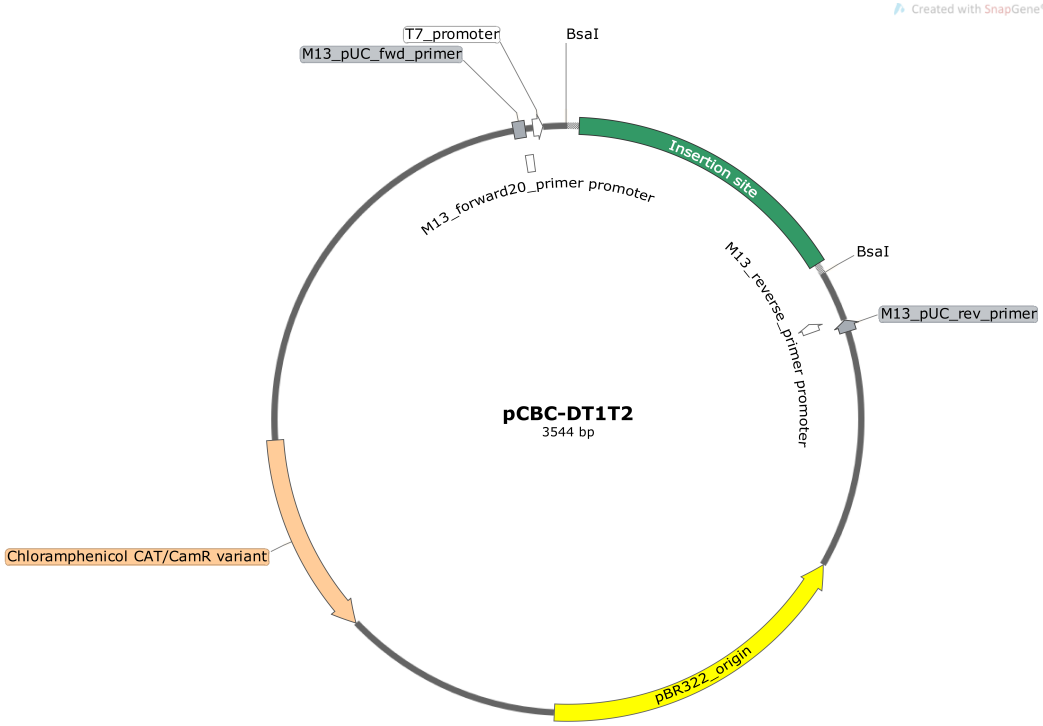
Arabidopsis and provides fundamental information in assessing the possible interaction of the GST with the TNT-GSH conjugate. Finally, the knockout study, attempted on multiple TNT-detoxification related *GSTU*s genes, although was unsuccessful, perhaps, could pave the way for a better knockout studies in the future.

Appendices

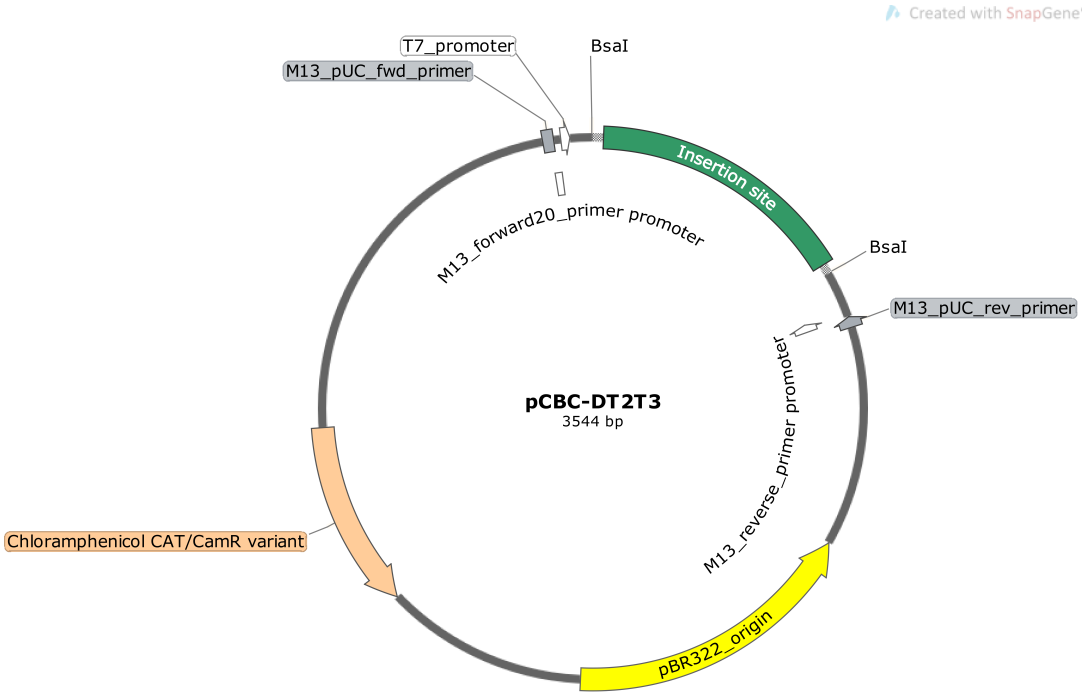
pCambia plasmid; CRISPR/Cas9 based plant genome editing and gene regulation plasmid; expresses 3×FLAGNLSzCas9NLS, gRNA scaffold, insertion site of target sequence (*At*U626 promoter), hygromycin and kanamycin resistance. Map is as generated by SnapGene from full sequence provided in Addgene by Xing *et al.*, 2014.



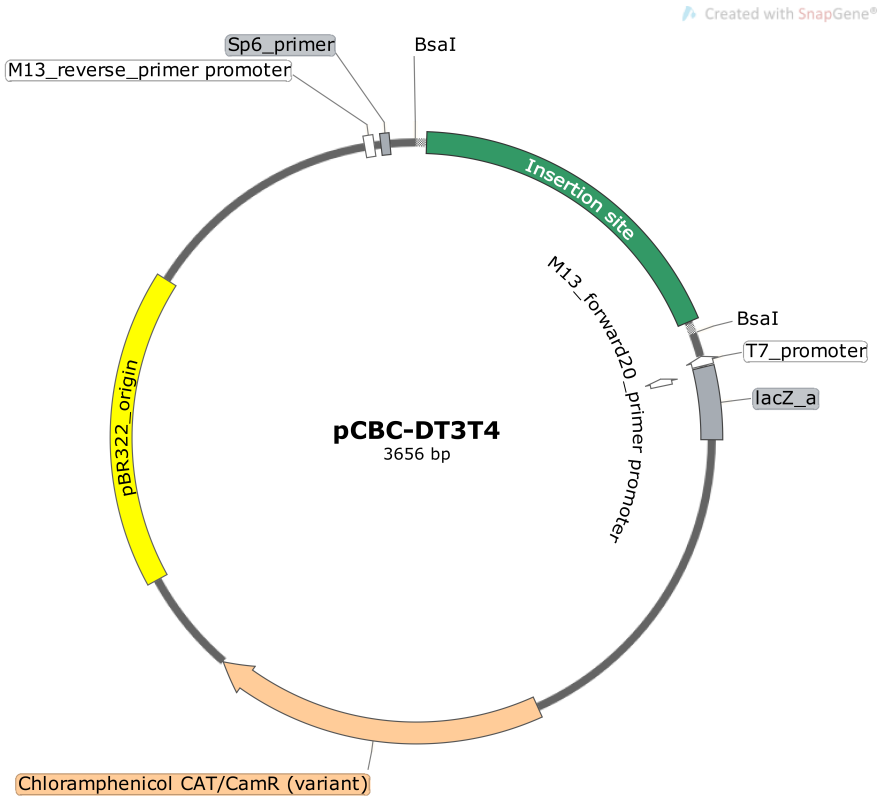
pDT1DT2 plasmid; CRISPR/Cas9 based plant genome editing and gene regulation plasmid; used as template for making expression cassette with multiple gRNA target sites; chloramphenicol resistance. Map is as generated by Addgene from full sequence provided by (Xing *et al.*, 2014).



pDT2DT3 plasmid; CRISPR/Cas9 based plant genome editing and gene regulation plasmid; used as template for making expression cassette with multiple gRNA target sites; chloramphenicol resistance. Map is as generated by Addgene from full sequence provided by (Xing *et al.*, 2014).



pDT3DT4 plasmid; CRISPR/Cas9 based plant genome editing and gene regulation plasmid; used as template for making expression cassette with multiple gRNA target sites; chloramphenicol resistance. Map is as generated by Addgene from full sequence provided by (Xing *et al.*, 2014).



Abbreviations

2H5MQ	2-hydroxy-5-methylquinone
4M5NC	4-methyl-5-nitrocatechol
ADNTs	Amino dinitrotoluenes
ATH1	Arabidopsis thaliana Homeobox 1 Gene
ATP	Adenosine triphosphate
BHA	2-tert-butyl-4-hydroxyanisole
BLASTn	Basic local alignment search tool
BSP	Sulfobromophthalein sodium
CDNB	1-chloro-2,4-dinitrobenzene
CRISPR	Clustered regularly interspaced short palindromic repeats
CHP	Cumene hydroperoxide
crRNA	Crispr RNA
CTAB	Cetyl trimethylamin bromide
CYP	Cytochrome
DCA	Dichloroacetic acid
DCNB	1,2-dichloro-4-nitrobenzene
DDT	Dichlorodiphenyltrichloroethane
DHAR	Dehydroascorbate reductase
DHCA	Dihydrocamalexic acid
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNT	Dinitrotoluenes
DSTL	Defence Science and Technology Laboratory
EDTA	Ethylenediaminetetraacetic acid
EF1BG	Elongation factor 1B gamma
ER	Endoplasmic reticulum
EPNP	1,2-epoxy-3-nitrophenoxypropane
GCS	Glutamylcysteine synthase
GPOX	Glutathione peroxidase activity
gRNAs	Guide RNA
GS	Glutathione synthetase
GSH	Glutathione
GSSG	Glutathione disulphide
GST	Glutathione transferase

HADNTs	Hydroxyl dinitrotoluenes
HCl	Hydrochloric acid
HDR	Homology direct repair
HED	*β*-hydroxyethyl disulphide
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
IAA	Indole-3-acetic acid
IAOx	Indole-3-acetaldoxime
IAN	Indole-3-acetonitrile
IDT	Integrated DNA Technology
IPTG	Isopropyl-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LB	Luria –Bertani
MAPEG	Membrane-Associated Protein in Eicosanoids and Glutathione metabolism
½ MS	½ Murashige & Skoog medium
MS/MS	Tandem mass spectrometry
NAA	Naphthalene acetic acid
NaCl	Natrium chloride
EDTA	Ethylenediaminetetraacetic acid
NBC	*p*-nitrobenzyl chloride
NCS	Non-crystallographic symmetry
NED	Amine *N*-(1-napthyl)ethylenediamine
NHEJ	Non-homologous end joining
NPA	1-N-naphthylpthalamic acid
OPRs	Oxophytodienoic acid reductases
PBO	*Trans*-4-phenyl-3-butene-2-one
PBS	Phosphate buffer saline
pCambia_2T	Plasmid Cambia with 2 targets
pCambia_4T	Plasmid Cambia with 4 targets
PCPT	Propionate-cacodylate-bistris propane buffer
PCR	Polymerase chain reactions
PDB	Protein data bank
PEG	Polyethylene glycol
pre-crRNA	Premature CRISPR RNA
PZQ	Praziquantel
RMS	Root mean square
ROS	Reactive oxygen species
SAGE	Serial analysis of gene expression

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TAIR	The Arabidopsis Information Resources
TCHQD	Tetrachlorohydroquinone dehalogenase
T-DNA	Transfer DNA
TEMED	Tetramethylethylenediamine.
TNT	2,4,6-trinitrotoluene
tracrRNA	Trans activating CRISPR RNA
UGTs	UDP-glucuronosyltransferase
XDS	X-ray Detector Software
YSBL	York Structural Biology Laboratory

Amino acid	
G	Glycine
A	Alanine
V	Valine
L	Leucine
I	Isoleucine
M	Methionine
F	Phenylalanine
W	Tryptophan
P	Proline
S	Serine
T	Threonine
C	Cysteine
Y	Tyrosine
N	Asparagine
Q	Glutamine
D	Aspartic acid
E	Glutamic acid
K	Lysine
R	Arginine
H	Histidine

Nucleotides of DNA	
A	Adenine
C	Cytosine
T	Thymine
G	Guanosine

References

Abhilash, P.C., Jamil, S. & Singh, N., 2009. Transgenic plants for enhanced biodegradation and phytoremediation of organic xenobiotics. *Biotechnology Advances*, 27(4), pp.474–488.

Aga, D.S. et al., 1996. Identification of a new sulfonic acid metabolite of metolachlor in soil. *Environmental Science & Technology*, 30(2), pp.592–597.

Aleku, G.A. et al., 2016. Stereoselectivity and structural characterization of an imine reductase (IRED) from Amycolatopsis orientalis. *ACS Catalysis*, 6, pp.3880–3889.

Alfenito, M.R. et al., 1998. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *The Plant Cell*, 10(7), pp.1135–49.

Allocati, N. et al., 2008. Cysteine 10 is critical for the activity of *Ochrobactrum anthropi* glutathione transferase and its mutation to alanine causes the preferential binding of glutathione to the H-site. *Proteins: Structure, Function and Bioinformatics*, 71(1), pp.16–23.

Allocati, N. et al., 2009. Glutathione transferases in bacteria. *FEBS Journal*, 276(1), pp.58–75.

Amaral, H.I.F. et al., 2009. Assessing TNT and DNT groundwater contamination by compound-specific isotope analysis and 3H-3He groundwater dating: A case study in Portugal. *Chemosphere*, 77(6), pp.805–812.

Arca, Pilar; Rico, Manuel; Brana, F. Alfredo; Villar, J. Claudio; Hardisson, Carlos; Suarez, E.J., 1988. Formation of an adduct between fosfomycin and glutathione: A new mechanism of antibiotic resistance in bacteria. *Antimicrobial Agents and Chemotherapy*, 32(10), pp.1552–1556.

Armstrong, R.N., 2000. Mechanistic diversity in a metalloenzyme superfamily. *Biochemistry*, 39(45), pp.13625–13632.

Atkins, W.M. et al., 1993. The catalytic mechanism of glutathione S-transferase (GST). Spectroscopic determination of the pKa of Tyr-9 in rat alpha 1-1 GST. *The Journal of Biological Chemistry*, 268(26), pp.19188–91.

Atkinson, H.J. et al., 2009. An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. R. Dunbrack, ed. *PLoS Computational Biology*, 5(10), p.e1000541.

Atkinson, H.J. & Babbitt, P.C., 2009. Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry*, 48(46), pp.11108–11116.

Axarli, I. et al., 2009. Crystallographic and functional characterization of the fluorodifen-inducible glutathione transferase from *Glycine max* reveals an active site topography suited for diphenylether herbicides and a novel L-site. *Journal of Molecular Biology*, 385(3), pp.984–1002.

Axarli, I. et al., 2016. Directed evolution of Tau class glutathione transferases reveals a site that

regulates catalytic efficiency and masks co-operativity. *Biochemical Journal*, 473, pp.559–570.

Axarli, I.A., Rigden, D.J. & Labrou, N.E., 2004. Characterization of the ligandin site of maize glutathione S-transferase I. *Biochemical Journal*, 893, pp.885–893.

Ayert, W.A. & Tewariy, J.P., 1991. The camalexin: New phytoalexins in the leaves of *Camelina sativa* (Cruciferase). *Tetrahedron*, 47(24), pp.3909–3914.

Ball, L. et al., 2004. Evidence for a direct link between glutathione biosynthesis and stress defense gene expression in Arabidopsis. *The Plant Cell*, 16(9), pp.2448–62.

Barrangou, R. et al., 2015. Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biology*, 16(1), p.247.

Bartling, D. et al., 1993. A glutathione S-transferase with glutathione-peroxidase activity from *Arabidopsis thaliana*. Molecular cloning and functional characterization. *European Journal of Biochemistry*, 216(2), pp.579–86.

Bashandy, T. et al., 2010. Interplay between the NADP-linked thioredoxin and glutathione systems in Arabidopsis auxin signaling. *The Plant Cell*, 22(2), pp.376–91.

Belhaj, K. et al., 2015. Editing plant genomes with CRISPR/Cas9. *Current Opinion in Biotechnology*, 32, pp.76–84.

Bento, A.P. et al., 2014. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42(D1), pp.D1083–D1090.

Beynon, E.R. et al., 2009. The role of oxophytodienoate reductases in the detoxification of the explosive 2,4,6-trinitrotoluene by Arabidopsis. *Plant Physiology*, 151(1), pp.253–261.

Blum, R. et al., 2007. Function of phytochelatin synthase in catabolism of glutathione-conjugates. *The Plant Journal*, 49(4), pp.740–749.

Booth, G., 2000. Nitro compounds, aromatic. In *Ullmann's Encyclopedia of Industrial Chemistry*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Bottcher, C. et al., 2009. The multifunctional enzyme CYP71B15 (PHYTOALEXIN DEFICIENT3) converts cysteine-indole-3-acetonitrile to camalexin in the indole-3-acetonitrile metabolic network of *Arabidopsis thaliana*. *Plant Cell*, 21(6), pp.1830–1845.

Böttcher, C. et al., 2014. The biosynthetic pathway of indole-3-carbaldehyde and indole-3-carboxylic acid derivatives in Arabidopsis. *Plant Physiology*, 165(2), pp.841–853.

Boyland, E. & Chasseaud, L.F., 1969. Glutathione S-aralkyltransferase. *Biochemical Journal*, 115, pp.985–991.

Boyland, E. & Williams, K., 1965. An enzyme catalysing the conjugation of epoxides with glutathione. *Biochemical Journal*, 94, pp.190–197.

Bratt, A. et al., 2016. Organelle redox autonomy during environmental stress. *Plant, Cell & Environment*, 39(9), pp.1909–1919.

Brentner, L.B. et al., 2008. Expression of glutathione S-transferases in poplar trees (*Populus trichocarpa*) exposed to 2,4,6-trinitrotoluene (TNT). *Chemosphere*, 73(5), pp.657–662.

Brentner, L.B. et al., 2010. Localization of hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) and 2,4,6-trinitrotoluene (TNT) in poplar and switchgrass plants using phosphor imager autoradiography. *Environmental Pollution*, 158(2), pp.470–475.

Brock, J., Board, P.G. & Oakley, A.J., 2013. Structural insights into Omega-class glutathione transferases: A snapshot of enzyme reduction and identification of a non-catalytic ligandin site. *PLoS ONE*, 8(4).

Bumpus, J.A. & Tatarko, M., 1994. Biodegradation of 2,4,6-trinitrotoluene by *Phanerochaete chrysosporium*: Identification of initial degradation products and the discovery of a TNT metabolite that inhibits lignin peroxidases. *Current Microbiology*, 28(3), pp.185–190.

Bushweller, J.H. et al., 1994. The nuclear magnetic resonance solution structure of the mixed disulfide between *Escherichia coli* glutaredoxin (C14S) and glutathione. *Journal of Molecular Biology*, 235(5), pp.1585–1597.

Butler, N.M. et al., 2015. Generation and inheritance of targeted mutations in potato (*Solanum tuberosum* L.) Using the CRISPR/Cas System. *PLoS ONE*, 10(12), pp.1–12.

Cairns, N.G. et al., 2006. Maturation of Arabidopsis seeds is dependent on glutathione biosynthesis within the embryo. *Plant Physiology*, 141(2), pp.446–55.

Cao, J. et al., 2016. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Research*, pp.1–10.

Chen, I.-C. et al., 2007. Glutathione S-transferase interacting with far-red insensitive 219 is involved in phytochrome A-mediated signaling in Arabidopsis. *Plant Physiology*, 143(3), pp.1189–1202.

Chen, J.H. et al., 2012. Drought and salt stress tolerance of an Arabidopsis glutathione S-transferase U17 knockout mutant are attributed to the combined effect of glutathione and abscisic acid. *Plant Physiology*, 158, pp.340–351.

Chen, X. et al., 2014. Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in *Caenorhabditis elegans*. *Nature Scientific Reports*, 4(7581), pp.1–7.

Chu, V.T. et al., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature Biotechnology*, 33(5), pp.543–548.

Clark, B. & Boopathy, R., 2007. Evaluation of bioremediation methods for the treatment of soil contaminated with explosives in Louisiana Army Ammunition Plant, Minden, Louisiana. *Journal of Hazardous Materials*, 143(3), pp.643–648.

Clough, S.J. & Bent, A.F., 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal*, 16(6), pp.735–43.

Coleman, J.O.D., Blake-Kalff, M.M.A. & Davies, T.G.E., 1997. Detoxification of xenobiotics by plants: Chemical modification and vacuolar compartmentation. *Trends in Plant Science*, 2(4), pp.144–151.

Collinson, E.J. & Grant, C.M., 2003. Role of yeast glutaredoxins as glutathione S-transferases. *The Journal of Biological Chemistry*, 278(25), pp.22492–7.

Cong, L. et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(February), pp.819–823.

Conn, S. et al., 2008. Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins. *Journal of Experimental Botany*, 59(13), pp.3621–34.

Cooper, A.J., Pinto, J.T. & Callery, P.S., 2011. Reversible and irreversible protein glutathionylation: biological and clinical aspects. *Expert Opinion on Drug Metabolism & Toxicology*, 7(7), pp.891–910.

Cooper, G.M., 2000. Transport of small molecules. In *The Cell: A Molecular Approach*. Sinauer Associates.

Dekant, W., Vamvakas, S. & Anders, M.W., 1990. Biosynthesis, bioactivation and mutagenicity of S-conjugates. *Toxicology Letters*, (53), pp.53–58.

Deltcheva, E. et al., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), pp.602–607.

van Dillewijn, P. et al., 2008. Type II hydride transferases from different microorganisms yield nitrite and diarylamines from polynitroaromatic compounds. *Applied and Environmental Microbiology*, 74(21), pp.6820–3.

Dillewijn, P. Van et al., 2008. Bioremediation of nitroreductase expressing transgenic aspen. *Environmental Science & Technology*, 42(19), pp.7405–7410.

Ding, Y. et al., 2016. Recent advances in genome editing using CRISPR/Cas9. *Frontiers in Plant Science*, 7, p.703.

Dixon, D., Cole, D.J. & Edwards, R., 1997. Characterisation of multiple glutathione transferases containing the GST I subunit with activities toward herbicide substrates in maize (*Zea mays*). *Pesticide Science*, 50(1), pp.72–82.

Dixon, D.P. et al., 2008. Binding and glutathione conjugation of porphyrinogens by plant glutathione transferases. *The Journal of Biological Chemistry*, 283(29), pp.20268–76.

Dixon, D.P. et al., 2009. Enzyme activities and subcellular localization of members of the Arabidopsis glutathione transferase superfamily. *Journal of Experimental Botany*, 60(4), pp.1207–1218.

Dixon, D.P. et al., 1998. Glutathione-mediated detoxification systems in plants. *Current Opinion in Plant Biology*, 1(3), pp.258–66.

Dixon, D.P., Cole, D.J. & Edwards, R., 2000. Characterisation of a zeta class glutathione transferase from *Arabidopsis thaliana* with a putative role in tyrosine catabolism. *Archives of Biochemistry and Biophysics*, 384(2), pp.407–412.

Dixon, D.P., Cole, D.J. & Edwards, R., 1998. Purification, regulation and cloning of a glutathione transferase (GST) from maize resembling the auxin-inducible type-III GSTs. *Plant Molecular Biology*, 36(1), pp.75–87.

Dixon, D.P., Davis, B.G. & Edwards, R., 2002. Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. *The Journal of Biological Chemistry*, 277(34), pp.30859–69.

Dixon, D.P. & Edwards, R., 2010a. Glutathione transferases. In *The Arabidopsis Book*. p. e0131.

Dixon, D.P. & Edwards, R., 2010b. Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. *The Journal of Biological Chemistry*, 285(47), pp.36322–9.

Dixon, D.P. & Edwards, R., 2009. Selective binding of glutathione conjugates of fatty acid derivatives by plant glutathione transferases. *The Journal of Biological Chemistry*, 284(32), pp.21249–56.

Dixon, D.P., Sellars, J.D. & Edwards, R., 2011. The Arabidopsis phi class glutathione transferase *At*GSTF2: binding and regulation by biologically active heterocyclic ligands. *The Biochemical Journal*, 438(1), pp.63–70.

Dixon, D.P., Steel, P.G. & Edwards, R., 2011. Roles for glutathione transferases in antioxidant recycling. *Plant Signalling & Behaviour*, 6(8), pp.1223–1227.

Droog, F., 1997. Plant glutathione S-transferases, a tale of Theta and Tau. *Journal of Plant Growth Regulation*, 16(2), pp.95–107.

Droog, F., Hooykaas, P. & Van Der Zaal, B.J., 1995. 2,4-dichlorophenoxyacetic acid and related chlorinated compounds inhibit two auxin-regulated Type-III tobacco glutathione S-transferases. *Plant Physiology*, 107(4), pp.1139–1146.

Edwards, R. & Dixon, D.P., 2005. Plant glutathione transferases. *Methods in Enzymology*, 401(5), pp.169–86.

Edwards, R., Dixon, D.P. & Walbot, V., 2000. Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends in Plant Science*, 5(5), pp.193–8.

Ekman, D.R. et al., 2003. SAGE analysis of transcriptome responses in Arabidopsis roots exposed to 2,4,6-trinitrotoluene. *Plant Physiology*, 133(3), pp.1397–406.

Emsley, P. & Cowtan, K., 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), pp.2126–2132.

Endo, M., Mikami, M. & Toki, S., 2015. Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. *Plant and Cell Physiology*, 56(1), pp.41–47.

Epp, O., Ladenstein, R. & Wendel, A., 1983. The refined structure of the selenoenzyme glutathione peroxidase at 0.2 nm resolution. *European Journal of Biochemistry*, 133(1), pp.51–69.

Evans, P., 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography*, 62(1), pp.72–82.

Fan, D. et al., 2015. Efficient CRISPR/Cas9-mediated targeted mutagenesis in Populus in the first generation. *Nature Scientific Reports*, 5, p.12217.

Fauser, F., Schiml, S. & Puchta, H., 2014. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant Journal*, 79(2), pp.348–359.

Feng, Z. et al., 2013. Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research*, 23(10), pp.1229–1232.

Feng, Z. et al., 2014. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(12), pp.4632–7.

Fernandes, A.P. & Holmgren, A., 2004. Glutaredoxin: Glutathione-dependent redox system with functions far beyond a simple thioredoxin backup system. *Antioxidant & Redox Signalling*, 6(1), pp.63–74.

Fratelli, M. et al., 2005. Gene expression profiling reveals a signaling role of glutathione in redox regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), pp.13998–4003.

Frear, D.S. & Swanson, H.R., 1970. Biosynthesis of S-(4-ethylamino-6-isopropylamino-2s-triazino) glutathione: Partial purification and properties of a glutathione-s-transferases from corn. *Phytochemistry*, 9(1969), pp.2123–2132.

Frendo, P. et al., 2001. A *Medicago truncatula* homoglutathione synthetase is derived from glutathione synthetase by gene duplication. *Plant Physiology*, 126(4), pp.1706–15.

Frova, C., 2003. The plant glutathione transferase gene family: Genomic structure, functions, expression and evolution. *Physiologia Plantarum*, 119(4), pp.469–479.

Gandia-Herrero, F. et al., 2008. Detoxification of the explosive 2,4,6-trinitrotoluene in Arabidopsis:

Discovery of bifunctional O- and C-glucosyltransferases. *Plant Journal*, 56(6), pp.963–974.

Gasteiger, E. et al., 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), pp.3784–8.

Geu-Flores, F. et al., 2011. Cytosolic γ-glutamyl peptidases process glutathione conjugates in the biosynthesis of glucosinolates and camalexin in Arabidopsis. *The Plant Cell*, 23(June), pp.2456–2469.

Gipp, J.J., Bailey, H.H. & Mulcahy, R.T., 1995. Cloning and sequencing of the cDNA for the light subunit of human liver γ-glutamylcysteine synthetase and relative RNA levels for heavy and light subunits in human normal tissues. *Biochemical and Biophysical Research Communications*, 206(2), pp.584–589.

Girardini, J. et al., 2002. Characterization of an omega-class glutathione S-transferase from *Schistosoma mansoni* with glutaredoxin-like dehydroascorbate reductase and thiol transferase activities. *European Journal of Biochemistry*, 269(22), pp.5512–21.

Griess, P., 1879. Bemerkungen zu der abhandlung der HH. Weselsky und Benedikt Ueber einige Azoverbindungen. *Berichte der deutschen chemischen Gesellschaft*, 12(1), pp.426–428.

Griffith, W., 1979. Glutathione synthetase. *Journal of Biological Chemistry*, 254(12), pp.5184–5190.

Gronwald, J.W. & Plaisance, K.L., 1998. Isolation and characterization of glutathione S-transferase isozymes from sorghum. *Plant Physiology*, 117(3), pp.877–92.

Grzam, A. et al., 2007. Gamma-Glutamyl transpeptidase GGT4 initiates vacuolar degradation of glutathione S-conjugates in Arabidopsis. *FEBS Letters*, 581(17), pp.3131–8.

Grzam, A. et al., 2006. Vacuolar sequestration of glutathione S-conjugates outcompetes a possible degradation of the glutathione moiety by phytochelatin synthase. *FEBS Letters*, 580(27), pp.6384–6390.

Guddewar, M.B. & Dauterman, W.C., 1979. Purification and properties of a glutathione-S-transferase from corn which conjugates s-triazine herbicides. *Phytochemistry*, 18(5), pp.735–740.

Gunning, V. et al., 2014. Arabidopsis glutathione transferases U24 and U25 exhibit a range of detoxification activities with the environmental pollutant and explosive, 2,4,6-trinitrotoluene. *Plant Physiology*, 165(2), pp.854–865.

Habig, W.-H. et al., 1974. The identity of glutathione S-transferase B with ligandin, a major binding protein of liver. *Proceedings of the National Academy of Sciences*, 71(10), pp.3879–3882.

Habig, W.H., Pabst, M.J. & Jakoby, W.B., 1974. Glutathione S-transferase: The first enzymatic step in mercapturic acid formation. *The Journal of Biological Chemistry*, 249, pp.7130–7139.

Hannink, N. et al., 2001. Phytodetoxification of TNT by transgenic plants expressing a bacterial

nitroreductase. *Nature Biotechnology*, 19(12), pp.1168–1172.

Hannink, N.K. et al., 2007. Enhanced transformation of TNT by tobacco plants expressing a bacterial nitroreductase. *International Journal of Phytoremediation*, 9(5), pp.385–401.

Hannink, N.K., Rosser, S.J. & Bruce, N.C., 2002. Phytoremediation of explosives. *Critical Reviews in Plant Sciences*, 21(5), pp.511–538.

Hao, G.-F. et al., 2011. Protoporphyrinogen oxidase inhibitor: An ideal target for herbicide discovery. *Chemistry in China*, 65(12), pp.961–969.

Hayes, J.D., Flanagan, J.U. & Jowsey, I.R., 2005. Glutathione transferases. *Annual Review of Pharmacology and Toxicology*, 45, pp.51–88.

Hayes, J.D. & Pulford, D.J., 1995. The glutathione S-transferase supergene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Critical Reviews in Biochemistry and Molecular Biology*, 30(6), pp.445–520.

Hayeshi, R., Chinyanga, F., Chengedza, S., Mukanganyama, S., 2006. Inhibition of human glutathione transferases by multidrug resistance chemomodulators in vitro. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 21(5), pp.581–587.

Heigwer, F., Kerr, G. & Boutros, M., 2014. E-CRISP: Fast CRISPR target site identification. *Nature Methods*, 11(2), pp.122–123.

Henderson, C.J., McLaren, A.W. & Wolf, C.R., 2014. In vivo regulation of human glutathione transferase GSTP by chemopreventive agents. *Cancer Research*, 74(16), pp.4378–4387.

Holm, L. & Rosenström, P., 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38, pp.W545-9.

Honaker, M.T. et al., 2011. Ensemble perspective for catalytic promiscuity: Calorimetric analysis of the active site conformational landscape of a detoxification enzyme. *The Journal of Biological Chemistry*, 286(49), pp.42770–6.

Hruz, T. et al., 2008. Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. *Advances in bioinformatics*, 2008, p.420747.

Hsieh, Y.-C. et al., 2005. Anaerobic purification and crystallization to improve the crystal quality: Ferredoxin II from *Desulfovibrio gigas*. *Acta Crystallographica Section D: Biological Crystallography*, 61, pp.780–783.

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), pp.827–32.

Hsu, P.D., Lander, E.S. & Zhang, F., 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6), pp.1262–1278.

Huang, C.S. et al., 1995. Amino acid sequence of rat kidney glutathione synthetase. *Biochemistry*, 92, pp.1232–1236.

Ibarra, C.A. et al., 2003. The anomalous pKa of Tyr-9 in glutathione S-transferase A1-1 catalyzes product release. *The Journal of biological chemistry*, 278(21), pp.19257–65.

Ibraheem, F., Gaffoor, I. & Chopra, S., 2010. Flavonoid phytoalexin-dependent resistance to anthracnose leaf blight requires a functional yellow seed1 in *Sorghum bicolor*. *Genetics*, 184(4), pp.915–926.

Irzyk, G.P. & Fuerst, E.P., 1993. Purification and characterization of a glutathione S-transferase from benoxacor-treated maize (*Zea mays*). *Plant Physiology*, 102(3), pp.803–10.

Ito, Y. et al., 2015. CRISPR/Cas9-mediated mutagenesis of the RIN locus that regulates tomato fruit ripening. *Biochemical and Biophysical Research Communications*, 467(1), pp.76–82.

Jablonkai, I. & Ko, T., 1999. Chemical catalysis of the isomerisation of peroxidising herbicidal thiadiazolidines. *IUPAC Congress*, 675(February), pp.1998–1999.

Jakobsson, P. et al., 2000. Membrane-associated proteins in eicosanoid and glutathione metabolism (MAPEG). *Critical Care Medicine*, 886(21), pp.S20–S24.

Jakobsson, P.-J. et al., 1999. Common structural features of MAPEG - A widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. *Protein Science*, 8(3), pp.689–692.

Jeandet, P. et al., 2014. Deciphering the role of phytoalexins in plant-microorganism interactions and human health. *Molecules*, 19(11), pp.18033–18056.

Jenkins, T.F. et al., 2006. Identity and distribution of residues of energetic compounds at army live-fire training ranges. *Chemosphere*, 63(8), pp.1280–1290.

Ji, X. et al., 1996. Location of a potential transport binding site in a sigma class glutathione transferase by X-ray crystallography. *Proceedings of the National Academy of Sciences*, 93(16), pp.8208–8213.

Ji, X. et al., 1992. The three-dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-A resolution. *Biochemistry*, 31(42), pp.10169–10184.

Jiang, H.W. et al., 2010. A glutathione S-transferase regulated by light and hormones participates in the modulation of Arabidopsis seedling development. *Plant Physiology*, 154(4), pp.1646–58.

Jiang, W. et al., 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Research*, 41(20), pp.1–12.

Jiang, W. et al., 2014. Efficient CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana* and

inheritance of modified genes in the T2 and T3 generations F. Börnke, ed. *PLoS ONE*, 9(6), p.e99225.

Johnson, G.R., Jain, R.K. & Spain, J.C., 2000. Properties of the trihydroxytoluene oxygenase from *Burkholderia cepacia* R34: An extradiol dioxygenase from the 2,4-dinitrotoluene pathway. *Archives of Microbiology*, 173(2), pp.86–90.

Johnson, Z.L. & Chen, J., 2017. Structural basis of substrate recognition by the Multidrug Resistance Protein MRP1. *Cell*, 168, pp.1075–1085.

Kabsch, W., 2010. XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.125–132.

Ketterer, B., Coles, B. & Meyer, D.J., 1983. The role of glutathione in detoxication by glutathione. *Environmental Health*, 49, pp.59–69.

Ketterman, A.J., Saisawang, C. & Wongsantichon, J., 2011. Insect glutathione transferases. *Drug Metabolism Reviews*, 43(2), pp.253–65.

Kim, H.Y. & Song, H.-G., 2000. Comparison of 2,4,6-trinitrotoluene degradation by seven strains of white rot fungi. *Current Microbiology*, 41(5), pp.317–320.

Kim, S. et al., 2016. PubChem substance and compound databases. *Nucleic Acids Research*, 44(D1), pp.D1202–D1213.

Kitamura, S. et al., 2010. Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* flavonoid accumulation mutants. *The Plant Journal*, 62(4), pp.549–559.

Kitamura, S. et al., 2012. Molecular characterization of an anthocyanin-related glutathione S-transferase gene in cyclamen. *Journal of Plant Physiology*, 169(6), pp.636–642.

Kitamura, S., Shikazono, N. & Tanaka, A., 2004. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *The Plant Journal*, 37(1), pp.104–114.

Kiyosue, T., Yamaguchi-Shinozaki, K. & Shinozaki, K., 1993. Characterization of two cDNAs (ERD11 and ERD13) for dehydration-inducible genes that encode putative glutathione S-transferases in *Arabidopsis thaliana* L. *FEBS Letters*, 335(2), pp.189–192.

Klapheck, S., 1988. Homoglutathione: Isolation, quantification and occurrence in legumes. *Physiologia Plantarum*, 74(4), pp.727–732.

Koprivova, A., Mugford, S.T. & Kopriva, S., 2010. Arabidopsis root growth dependence on glutathione is linked to auxin transport. *Plant Cell Reports*, 29(10), pp.1157–1167.

Krzysztof, C., Le, R.A. & Emmanuelle, C., 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas

immunity systems. *RNA Biology*, 10(5), pp.726–737.

Kumar, S. & Pandey, A.K., 2013. Chemistry and biological activities of flavonoids: An overview. *The Scientific World Journal*, 2013, p.162750.

Kumar, V. & Jain, M., 2015. The CRISPR-Cas system for plant genome editing: Advances and opportunities. *Journal of Experimental Botany*, 66(1), pp.47–57.

Labrou, N.E., Mello, L. V & Clonis, Y.D., 2001. Functional and structural roles of the glutathione-binding residues in maize (*Zea mays*) glutathione S-transferase I. *The Biochemical Journal*, 358(Pt 1), pp.101–110.

Lallement, P.A., Meux, E., et al., 2014. Structural and enzymatic insights into Lambda glutathione transferases from *Populus trichocarpa*, monomeric enzymes constituting an early divergent class specific to terrestrial plants. *The Biochemical Journal*, 462(1), pp.39–52.

Lallement, P.A., Brouwer, B., et al., 2014. The still mysterious roles of cysteine-containing glutathione transferases in plants. *Frontiers in Pharmacology*, 5, p.192.

Lamoureux, G.L. & Rusness, D.G., 1980. In vitro metabolism of pentachloronitrobenzene to pentachloromethylthiobenzene by onion: Characterization of glutathione S-transferase, cysteine C-S lyase, and S-adenosylmethionine methyl transferase activities. *Pesticide Biochemistry and Physiology*, 14(1), pp.50–61.

Lamoureux, G.L., Rusness, D.G. & Schroder, P., 1993. Metabolism of a diphenylether herbicide to a volatile thioanisole metabolite in spruce (*Picea*). *Pesticide Biochemistry and Physiology*, 47(1), pp.8–20.

Lan, T. et al., 2009. Extensive functional diversification of the *Populus* glutathione S-transferase supergene family. *The Plant Cell*, 21(12), pp.3749–66.

de las Heras, A., Chavarría, M. & de Lorenzo, V., 2011. Association of dnt genes of *Burkholderia* sp. DNT with the substrate-blind regulator DntR draws the evolutionary itinerary of 2,4-dinitrotoluene biodegradation. *Molecular Microbiology*, 82(2), pp.287–299.

Laskowski, R. a. et al., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, pp.283–291.

Lawrenson, T. et al., 2015. Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome Biology*, 16(1), p.258.

Lederer, B. & Böger, P., 2003. Binding and protection of porphyrins by glutathione S-transferases of *Zea mays* L. *Biochimica et Biophysica Acta*, 1621(2), pp.226–233.

Lee, E.H., Kim, H.Y. & Hwang, K.Y., 2014. The GSH- and GSSG-bound structures of glutaredoxin from *Clostridium oremlandii*. *Archives of Biochemistry and Biophysics*, 564, pp.20–25.

Lemarié, S. et al., 2015. Camalexin contributes to the partial resistance of *Arabidopsis thaliana* to the biotrophic soilborne protist Plasmodiophora brassicae. *Frontiers in Plant Science*, 6, p.539.

Lewis, T.A., Newcombe, D.A. & Crawford, R.L., 2004. Bioremediation of soils contaminated with explosives. *Journal of Environmental Management*, 70(4), pp.291–307.

Li, J.-F. et al., 2013. Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnology*, 31(8), pp.688–691.

Liang, G. et al., 2016. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Scientific Reports*, 6, p.21451.

Licciardello, C. et al., 2014. Characterization of the glutathione S-transferase gene family through ESTs and expression analyses within common and pigmented cultivars of *Citrus sinensis* (L.) Osbeck. *BMC Plant Biology*, 14(1), p.39.

Litwack, G., Ketterer, B. & Arias, I.. I.M., 1971. Ligandin: A hepatic protein which binds steroids, bilirubin, carcinogens and a number of exogenous organic anions. *Nature*, 234(5330), pp.466–467.

Liu, W. et al., 2015. A detailed procedure for CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana*. *Science Bulletin*, 60(15), pp.1332–1347.

Liu, Y.-J. et al., 2013. Functional divergence of the glutathione S-transferase supergene family in *Physcomitrella patens* reveals complex patterns of large gene family evolution in land plants. *Plant Physiology*, 161(2), pp.773–86.

Lorenz, A., 2007. *Bioengineering transgenic plants to detoxify nitroaromatic explosive compounds*. University of York.

Lowder, L.G. et al., 2015. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiology*, 169(2), pp.971–85.

Ma, X. et al., 2015. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant*, 8, pp.1274–1284.

Ma, X. et al., 2016. CRISPR/Cas9 platforms for genome editing in plants: Developments and applications CRISPR/Cas9 platforms for genome editing in plants. *Molecular Plant*, 9, pp.961–974.

Makarova, K.S. et al., 2015. An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, 13(11), pp.722–736.

Makarova, K.S. et al., 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct*, 6(1), p.38.

Makarova, K.S. & Koonin, E. V., 2015. Annotation and classification of CRISPR-Cas systems. In *Methods in Molecualr Biology*. pp. 47–75.

Makris, K.C. et al., 2007. High uptake of 2,4,6-trinitrotoluene by vetiver grass – Potential for phytoremediation? *Environmental Pollution*, 146(1), pp.1–4.

Mao, Y. et al., 2013. Application of the CRISPR-Cas system for efficient genome engineering in plants. *Molecular Plant*, 6(6), pp.2008–2011.

Marrs, K.A. et al., 1995. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2. Nature*, 375(6530), pp.397–400.

Martin, J.L., Bardwell, J.C.A. & Kurlyan, J., 1993. Crystal structure of the DsbA protein required for disulphide bond formation in vivo. *Nature*, 365, pp.464–468.

May, M.J. & Leaver, C.J., 1994. *Arabidopsis thaliana* gamma-glutamylcysteine synthetase is structurally unrelated to mammalian, yeast, and *Escherichia coli* homologs. *Proceedings of the National Academy of Sciences of the United States of America*, 91(21), pp.10059–63.

McGonigle, B. et al., 2000. A genomics approach to the comprehensive analysis of the glutathione-S-transferase gene family in soybean and maize. *Plant Physiology*, 124(3), pp.1105–20.

Mctigue, M.A. et al., 1995. Crystal structures of a schistosomal drug and vaccine target: Glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *Journal of Molecular Biology*, 246, pp.21–27.

Meyer, A.J. et al., 2007. Redox-sensitive GFP in *Arabidopsis thaliana* is a quantitative biosensor for the redox potential of the cellular glutathione redox buffer. *Plant Journal*, 52(5), pp.973–986.

Mezzari, M.P. et al., 2005. Gene expression and microscopic analysis of *Arabidopsis* exposed to chloroacetanilide herbicides and explosive compounds. A phytoremediation approach. *Plant Physiology*, 138(2), pp.858–869.

Mhamdi, A. et al., 2010. Cytosolic NADP-dependent isocitrate dehydrogenase contributes to redox homeostasis and the regulation of pathogen responses in *Arabidopsis* leaves. *Plant, Cell and Environment*, 33(7), pp.1112–1123.

Middleton Jr., E. & Kandaswami, C., 1998. Effect of flavonoids on immune and inflammatory cell function. In J. Manthey & B. Buslig, eds. *Biochemical Pharmacology*. Advances in Experimental Medicine and Biology. Springer US, pp. 175–182.

Møldrup, M.E. et al., 2013. De novo genetic engineering of the camalexin biosynthetic pathway. *Journal of Biotechnology*, 167(3), pp.296–301.

Monti, M.R. et al., 2005. Engineering *Pseudomonas fluorescens* for biodegradation of 2,4-dinitrotoluene. *Applied and Environmental Microbiology*, 71(12), pp.8864–72.

Morel, M. et al., 2013. Xenomic networks variability and adaptation traits in wood decaying fungi. *Microbial Biotechnology*, 6(3), pp.248–263.

Mueller, L. a et al., 2000. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiology*, 123(4), pp.1561–70.

Mueller, S. et al., 2008. General detoxification and stress responses are mediated by oxidized lipids through TGA transcription factors in Arabidopsis. *The Plant Cell*, 20, pp.768–785.

Munter, R., 2001. Advanced oxidation processes - current status and prospect. *Proceedings of Estonian Academy of Sciences. Chemistry*, 50(2), pp.59–80.

Munyampundu, J., Xu, Y. & Cai, X., 2016. Phi class of glutathione S-transferase gene superfamily widely exists in nonplant taxonomic groups. *Evolutionary Bioinformatics*, 12, pp.59–71.

Murphy, A.S. et al., 2002. Identification, purification and molecular cloning of N-1-naphthylphthalmic acid-binding plasma membrane -associated aminopeptidases from Arabidopsis. *Plant Physiology*, 128, pp.935–950.

Murshudov, G.N. et al., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp.355–367.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography*, 53(3), pp.240–255.

Nakashima, Y. et al., 2009. Crystallization and preliminary crystallographic analysis of bifunctional gamma-glutamylcysteine synthetase-glutatione synthetase from *Streptococcus agalactiae*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 65(Pt 7), pp.678–80.

Nekrasov, V. et al., 2013. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology*, 31(8), pp.691–693.

Nishida, M. et al., 1994. Molecular cloning and site-directed mutagenesis of glutathione S-transferase from *Escherichia coli*. *The Journal of Biological Chemistry*, 269(51), pp.32536–32541.

Nishino, S.F. & Paoli, G.C., 2000. Aerobic degradation of dinitrotoluenes and pathway for bacterial degradation of aerobic degradation of dinitrotoluenes and pathway for bacterial degradation of 2 , 6-dinitrotoluene. *Applied and Environmental Microbiology*, 66(5), pp.2139–2147.

Noctor, G. et al., 2012. Glutathione in plants: An integrated overview. *Plant, Cell and Environment*, 35, pp.454–484.

Oakley, A., 2011. Glutathione transferases: A structural perspective. *Drug Metabolism Reviews*, 43(2), pp.138–151.

Oakley, a J. et al., 1999. The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site). *Journal of Molecular Biology*, 291(4), pp.913–26.

Ohkama-Ohtsu, N. et al., 2007. Glutathione conjugates in the vacuole are degraded by γ-glutamyl transpeptidase GGT3 in Arabidopsis. *Plant Journal*, 49(5), pp.878–888.

Pasternak, M. et al., 2007. Restricting glutathione biosynthesis to the cytosol is sufficient for normal plant development. *The Plant Journal*, 53(6), pp.999–1012.

Pedras, M.S.C. et al., 2000. Phytoalexins from crucifers: Synthesis, biosynthesis, and biotransformation. *Phytochemistry*, 53(2), pp.161–176.

Perito, B. et al., 1996. Molecular cloning and overexpression of a glutathione transferase gene from *Proteus mirabilis*. *Proteus*, 318, pp.157–162.

Pettigrew, N.E. & Colman, R.F., 2001. Heterodimers of glutathione S-transferase can form between isoenzyme classes Pi and Mu. *Archives of Biochemistry and Biophysics*, 396(2), pp.225–230.

Peuke, A.D. & Rennenberg, H., 2005. Phytoremediation. *EMBO reports*, 6(6), pp.497–501.

Plancarte, A. et al., 2014. Evaluation of the non-catalytic binding function of Ts26GST a glutathione transferase isoform of *Taenia solium*. *Experimental Parasitology*, 138, pp.63–70.

Prade, L., Huber, R. & Bieseler, B., 1998. Structures of herbicides in complex with their detoxifying enzyme glutathione S-transferase - explanations for the selectivity of the enzyme in plants. *Structure*, 6(11), pp.1445–52.

Prinz, W. a et al., 1997. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *The Journal of Biological Chemistry*, 272(25), pp.15661–15667.

Queval, G. et al., 2011. Increased intracellular H_2O_2 availability preferentially drives glutathione accumulation in vacuoles and chloroplasts. *Plant, Cell and Environment*, 34(1), pp.21–32.

Ran, F.A. et al., 2013. Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–1389.

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), pp.2281–308.

Rausch, T. et al., 2007. Novel Insight into the regulation of GSH biosynthesis in higher plants. *Plant Biology*, 9(5), pp.565–572.

Reinemer, P. et al., 1991. The three-dimensional structure of class pi glutathione S-transferase in complex with glutathione sulfonate at 2.3 A resolution. *The EMBO Journal*, 10(8), pp.1997–2005.

Reinemer, P. et al., 1996. Three-dimensional structure of glutathione S-transferase from *Arabidopsis thaliana* at 2.2 A resolution: structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. *Journal of Molecular Biology*, 255(2), pp.289–309.

Rennenberg, H., 1982. Glutathione metabolism and possible biological roles in higher plants. *Phytochemistry*, 21(12), pp.2771–2781.

Rieble, S., Joshi, D.K. & Gold, M.H., 1994. Aromatic nitroreductase from the basidiomycete *Phanerochaete chrysosporium*. *Biochemical and Biophysical Research Communications*, 205(1), pp.298–304.

Rigsby, R.E. et al., 2005. Fosfomycin resistance proteins: A nexus of glutathione transferases and epoxide hydrolases in a metalloenzyme superfamily. *Methods in Enzymology*, 401(5), pp.367–379.

Roret, T. et al., 2015. Evolutionary divergence of Ure2pA glutathione transferases in wood degrading fungi. *Fungal Genetics and Biology*, 83, pp.103–112.

Rossjohn, J. et al., 1998. A mixed disulfide bond in bacterial glutathione transferase: Functional and evolutionary implications. *Structure*, 6(6), pp.721–734.

Rouhier, N. et al., 2001. Isolation and characterization of a new peroxiredoxin from poplar sieve tubes that uses either glutaredoxin or thioredoxin as a proton donor. *Plant Physiology*, 127(3), pp.1299–309.

Rowe, H.C. et al., 2010. Deficiencies in jasmonate-mediated plant defense reveal quantitative variation in *Botrytis cinerea* pathogenesis. *PLoS Pathogen*, 6(4), p.e1000861.

Russell Thom, ⊥ et al., 2002. Structure of a Tau class glutathione S-transferase from wheat active in herbicide detoxification. *Biochemistry*, 41, pp.7008–7020.

Rylott, E.L. et al., 2015. Phytodetoxification of the environmental pollutant and explosive 2,4,6-trinitrotoluene. *Plant Signaling & Behavior*, 2324, pp.e977714-1-e977714-4.

Rylott, E.L., Johnston, E.J. & Bruce, N.C., 2015. Harnessing microbial gene pools to remediate persistent organic pollutants using genetically modified plants - A viable technology? *Journal of Experimental Botany*, 66(21), pp.6519–6533.

Rylott, E.L., Lorenz, A. & Bruce, N.C., 2011. Biodegradation and biotransformation of explosives. *Current Opinion in Biotechnology*, 22(3), pp.434–440.

Šali, A. et al., 1995. Evaluation of comparative protein modeling by MODELLER. *Proteins: Structure, Function, and Genetics*, 23(3), pp.318–326.

Santos, P.M. et al., 2002. Occurrence and properties of glutathione S-transferases in phenol-degrading *Pseudomonas* strains. *Research in Microbiology*, 153, pp.89–98.

Sappl, P.G. et al., 2009. The Arabidopsis glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress. *The Plant Journal*, 58(1), pp.53–68.

Sapranauskas, R. et al., 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research*, 39(21), pp.9275–82.

Schiml, S., Fauser, F. & Puchta, H., 2014. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny. *Plant Journal*, 80(6), pp.1139–1150.

Schoenmuth, B.W. & Pestemer, W., 2004. Dendroremediation of trinitrotoluene (TNT). *Environmental Science and Pollution Research*, 11, pp.331–339.

Schröder, P. et al., 2007. How plants cope with foreign compounds. Translocation of xenobiotic glutathione conjugates in roots of barley (*Hordeum vulgare*). *Environmental science and pollution research international*, 14(2), pp.114–22.

Schuhegger, R. et al., 2006. CYP71B15 (PAD3) catalyzes the final step in camalexin biosynthesis. *Plant Physiology*, 141(4), pp.1248–1254.

Schüttelkopf, A.W. & Van Aalten, D.M.F., 2004. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, 60(8), pp.1355–1363.

Sens, C. et al., 1998. Distribution of 14C-TNT and derivatives in different biochemical compartments of *Phaseolus vulgaris*. *Environmental science and pollution research international*, 5(4), pp.202–8.

Sheehan, D. et al., 2001. Structure, function and evolution of glutathione transferases: Implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*, 360(Pt 1), pp.1–16.

Shelly C. Lu, 2014. Glutathione synthesis. *Biochimica Biophysica Acta*, 1830(5), pp.3143–3153.

Sies, H., 1999. Glutathione and its role in cellular functions. *Free Radical Biology and Medicine*, 27(9), pp.916–921.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), p.539.

Simons, P.C. & Jagt, D.L. Vander, 1980. Bilirubin binding to human liver ligandin (Glutathione S-transferase). *The Journal of Biological Chemistry*, 255(10), pp.4740–4744.

Skopelitou, K. et al., 2012. A glutathione transferase from *Agrobacterium tumefaciens* reveals a novel class of bacterial GST superfamily. *PLoS ONE*, 7(4), pp.1–10.

Smith, A.P. et al., 2003. Arabidopsis *At*GSTF2 is regulated by ethylene and auxin, and encodes a glutathione S-transferase that interacts with flavonoids. *The Plant Journal*, 36(4), pp.433–442.

Snyder, B. a & Nicholson, R.L., 1990. Synthesis of phytoalexins in sorghum as a site-specific response to fungal ingress. *Science*, 248(4963), pp.1637–1639.

Song, Y. et al., 2016. Efficient dual sgRNA-directed large gene deletion in rabbit with CRISPR/Cas9 system. *Cellular and Molecular Life Sciences*, 73(15), pp.2959–2968.

Spain, J.C., 1995. Biodegradation of notroaromatic compounds. *Annual Review of Microbiology*, 49, pp.523–555.

Spanggord, R.J. et al., 1991. Biodegradation of 2,4-dinitrotoluene by a *Pseudomonas* sp. *Applied and Environmental Microbiology*, 57(11), pp.3200–5.

Sparrow, H., 2010. *The role of glutathione transferases in the detoxification of TNT*. University of York.

Stourman, N. V. et al., 2011. Structure and function of YghU, a Nu class glutathione transferase related to YfcG from *Escherichia coli*. *Biochemistry*, 50(7), pp.1274–1281.

Su, T. et al., 2011. Glutathione-indole-3-acetonitrile is required for camalexin biosynthesis in *Arabidopsis thaliana*. *The Plant Cell*, 23(1), pp.364–380.

Subramanian, M., Oliver, D.J. & Shanks, J.V., 2006. TNT phytotransformation pathway characteristics in *Arabidopsis*: Role of aromatic hydroxylamines. *Biotechnology Progress*, 22(1), pp.208–216.

Suen, W.C. & Spain, J.C., 1993. Cloning and characterization of *Pseudomonas* sp. strain DNT genes for 2,4-dinitrotoluene degradation. *Journal of Bacteriology*, 175(6), pp.1831–7.

Sun, Y., Li, H. & Huang, J.-R., 2012. *Arabidopsis* TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Molecular Plant*, 5(2), pp.387–400.

Svitashev, S. et al., 2015. Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiology*, 169(2), pp.931–45.

Symington, L.S. & Gautier, J., 2011. Double-strand break end resection and repair pathway choice. *Annual Review of Genetics*, 45(1), pp.247–271.

Symons, Z.C. et al., 2006. Bacterial pathways for degradation of nitroaromatics. *Natural Product Reports*, 23(6), p.845.

Takahashi, Y., Ishida, S. & Nagata, T., 1995. Auxin-regulated genes. *Plant Cell Physiology*, 36(3), pp.383–390.

Takusagawa, F., 2013. Microsomal prostaglandin E synthase type 2 (mPGES2) is a glutathione-dependent heme protein, and dithiothreitol dissociates the bound heme to produce active prostaglandin E2 synthase in vitro. *Journal of Biological Chemistry*, 288(14), pp.10166–10175.

Tao, S. et al., 2008. Characterization of Ser73 in *Arabidopsis thaliana* glutathione S-transferase Zeta class. *Journal of Genetics and Genomics*, 35(8), pp.507–512.

Taylor, V.L. et al., 2013. Protective responses induced by herbicide safeners in wheat. *Environmental and Experimental Botany*, 88, pp.93–99.

Taylor, V.L., 2012. *The activities of herbicide safeners in wheat (Triticum aestivum L).* University of Durham.

Tepperman, J.M. et al., 2001. Multiple transcription-factor genes are early targets of phytochrome A signaling. *Proceedings of the National Academy of Sciences (PNAS)*, 98(16), pp.9437–9442.

Thijs, S. et al., 2014. Exploring the rhizospheric and endophytic bacterial communities of *Acer pseudoplatanus* growing on a TNT-contaminated soil: towards the development of a rhizocompetent TNT-detoxifying plant growth promoting consortium. *Plant and Soil*, 385(1–2), pp.15–36.

Thom, R. et al., 2001. The structure of a zeta class glutathione S-transferase from *Arabidopsis thaliana*: Characterisation of a GST with novel active-site architecture and a putative role in tyrosine catabolism. *Journal of Molecular Biology*, 308(5), pp.949–62.

Thomma, B.P.H.J. et al., 1999. Deficiency in phytoalexin production causes enhanced susceptibility of *Arabidopsis thaliana* to the fungus *Alternaria brassicicola*. *Plant Journal*, 19(2), pp.163–171.

Thuillier, A. et al., 2013. Atypical features of a Ure2p glutathione transferase from *Phanerochaete chrysosporium*. *FEBS Letters*, 587(14), pp.2125–2130.

Toé, K.H. et al., 2015. The recent escalation in strength of pyrethroid resistance in *Anopheles coluzzi* in West Africa is linked to increased expression of multiple gene families. *BMC Genomics*, 16(1), p.146.

Travensolo, R.F. et al., 2008. Cloning, expression, purification and characterization of recombinant glutathione-S-transferase from *Xylella fastidiosa*. *Protein Expression and Purification*, 59, pp.153–160.

Trott, O. & Olson, A.J., 2010. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, 31, pp.455–461.

Tzafestas, K., 2016. *Investigating the role of glutathione transferase in the phytodetoxification of explosives*. University of York.

Ullmann, P. et al., 1996. Cloning of *Arabidopsis thaliana* glutathione synthetase (GSH2) by functional complementation of a yeast GSH2 mutant. *European Journal of Biochemistry*, 236(2), pp.662–669.

Urano, J. et al., 2000. Molecular cloning and characterization of a rice dehydroascorbate reductase.

FEBS Letters, 466(1), pp.107–11.

Vagin, A. & Teplyakov, A., 1997. MOLREP : An automated program for molecular replacement. *Journal of Applied Crystallography*, 30(6), pp.1022–1025.

Wachter, A. et al., 2004. Differential targeting of GSH1 and GSH2 is achieved by multiple transcription initiation: Implications for the compartmentation of glutathione biosynthesis in the *Brassicaceae*. *The Plant Journal*, 41(1), pp.15–30.

Wadington, M.C. et al., 2009. Analysis of the structure and function of YfcG from *Escherichia coli* reveals an efficient and unique disulfide bond reductase. *Biochemistry*, 48(28), pp.6559–6561.

Wagner, U. et al., 2002. Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family. *Plant Molecular Biology*, 49(5), pp.515–32.

Wang, C. et al., 2003. Role of hydroxylamine intermediates in the phytotransformation of 2,4,6-trinitrotoluene by *Myriophyllum aquaticum*. *Environmental Science & Technology*, 37(16), pp.3595–3600.

Wang, W. et al., 2016. Optimizing multiplex CRISPR/CAS9Cas9 system for wheat genome editing. *Cold Spring Harbor Laboratory*, pp.15–35.

Wang, Y. et al., 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology*, Advance on(July), pp.1–6.

Wangwattana, B. et al., 2008. Characterization of PAP1-upregulated glutathione S-transferase genes in *Arabidopsis thaliana*. *Plant Biotechnology*, 25(2), pp.191–196.

Watahiki, M.K., Mori, H. & Yamamoto, K.T., 1995. Inhibitory effects of auxins and related substances on the activity of an *Arabidopsis* glutathione S-transferase isozyme expressed in Escherichia coli. *Physiologia Plantarum*, 94(4), pp.566–574.

Wiktelius, E. & Stenberg, G., 2007. Novel class of glutathione transferases from cyanobacteria exhibit high catalytic activities towards naturally occurring isothiocyanates. *The Biochemical Journal*, 406(1), pp.115–123.

Williams, R.E. et al., 2004. Biotransformation of explosives by the old yellow enzyme family of flavoproteins. *Applied and environmental microbiology*, 70(6), pp.3566–74.

Wilson, K.A., Kellie, J.L. & Wetmore, S.D., 2014. DNA-protein π-interactions in nature: Abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Research*, 42(10), pp.6726–6741.

Winkel-Shirley, B., 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology and biotechnology. *Plant Physiology*, 126(2), pp.485–493.

Winter, G., 2010. Xia2: An expert system for macromolecular crystallography data reduction. *Journal*

of Applied Crystallography, 43(1), pp.186–190.

Wolf, A.E., Dietz, K.J. & Schröder, P., 1996. Degradation of glutathione S-conjugates by a carboxypeptidase in the plant vacuole. *FEBS Letters*, 384(1), pp.31–34.

Woo, J.W. et al., 2015. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology*, 33(11), pp.1162–1164.

Wu, J., Cramer, C.L. & Hatzios, K.K., 1999. Characterization of two cDNAs encoding glutathione S-transferases in rice and induction of their transcripts by the herbicide safener fenclorim. *Physiologia Plantarum*, 105(1), pp.102–108.

Xie, K. & Yang, Y., 2013. RNA-guided genome editing in plants using a CRISPR–Cas system. *Molecular Plant*, 6(6), pp.1975–1983.

Xing, H.L. et al., 2014. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, 14(1), p.327.

Yamazaki, M. et al., 2008. Differential gene expression profiles of red and green forms of *Perilla frutescens* leading to comprehensive identification of anthocyanin biosynthetic genes. *FEBS Journal*, 275(13), pp.3494–3502.

Ye, G.N. et al., 1999. *Arabidopsis* ovule is the target for *Agrobacterium* in planta vacuum infiltration transformation. *The Plant Journal*, 19(3), pp.249–57.

Yoon, J.M., Oliver, D.J. & Shanks, J. V., 2007. Phytotoxicity and phytoremediation of 2,6-dinitrotoluene using a model plant, *Arabidopsis thaliana*. *Chemosphere*, 68(6), pp.1050–1057.

Zechmann, B. et al., 2008. Subcellular immunocytochemical analysis detects the highest concentrations of glutathione in mitochondria and not in plastids. *Journal of Experimental Botany*, 59(14), pp.4017–4027.

Zechmann, B. & Müller, M., 2010. Subcellular compartmentation of glutathione in dicotyledonous plants. *Protoplasma*, 246(1–4), pp.15–24.

Zettl, R., Schell, J. & Palme, K., 1994. Photoaffinity labeling of *Arabidopsis thaliana* plasma membrane vesicles by 5-azido-[7-3H]indole-3-acetic acid: Identification of a glutathione S-transferase. *Proceedings of the National Academy of Sciences (PNAS)*, 91, pp.689–693.

Zhang, C. hua et al., 2013. Purification and identification of glutathione S-transferase in rice root under cadmium stress. *Rice Science*, 20(3), pp.173–178.

Zhang, H. et al., 2014. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnology Journal*, 12(6), pp.797–807.

Zhao, J. & Dixon, R. a, 2010. The “ins” and “outs” of flavonoid transport. *Trends in Plant Science*, 15(2), pp.72–80.

Zheng, W. et al., 2009. Fate and transport of TNT, RDX, and HMX in streambed sediments: Implications for riverbank filtration. *Chemosphere*, 76(9), pp.1167–1177.

Zhou, N., Tootle, T.L. & Glazebrook, J., 1999. Arabidopsis PAD3, a gene required for camalexin biosynthesis, encodes a putative cytochrome P450 monooxygenase. *The Plant cell*, 11(12), pp.2419–28.

Zhou, X. et al., 2015. Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate: CoA ligase specificity and redundancy. *New Phytologist*, 208(2), pp.298–301.

Zhou, Y. et al., 2014. Genetic analysis of tissue glutathione concentrations and redox balance. *Free Radical Biology and Medicine*, 71, pp.157–164.

Zybailov, B. et al., 2008. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, 3(4), p.e1994.