Durham E-Theses

Structural and Functional Characterisation of Glutathione-S-Transferases to Combat Multiple Herbicide Resistance in Black Grass (Alopecurus myosuroides)

ENO, REBECCA,FLORENCE,MARY

Structural and Functional Characterisation of Glutathione-S-Transferases

to Combat Multiple Herbicide Resistance in Black Grass (*Alopecurus myosuroides*)

Rebecca Florence Mary Eno

PhD Thesis

Durham University

Department of Chemistry

2019

Abstract

Black grass (*Alopecurus myosuroides*) is one of the most common weeds in Western Europe. With the increasing use of herbicides, in response to pressure to produce higher crop yields, incidences of multiple herbicide resistance have been widely reported. Previous work in this area has identified a phi class glutathione-s-transferase, *Am*GSTF1, as playing a causative role in this multiple herbicide resistance. Two classes of inhibitor have previously been identified: the first CNBF, a multi drug resistance drug; and the second a group of flavone compounds identified from ligand fishing experiments. These have been shown to have activity *in vitro* inhibiting the action of *Am*GSTF1 and *in vivo* against multiple herbicide resistant black grass. However, little was known about their mode of action. This work has focussed on investigating the interactions between these inhibitors and *Am*GSTF1 using crystallographic, biochemical and complementary biophysical techniques.

Apo *Am*GSTF1 has successfully been crystallised, with the structure solved to 1.5 Å. In addition a structure has been solved to 2.0 Å with CNBF covalently modifying the Cys120 residue. For both these structures the crystal packing results in loops in the active site region being disordered as well as preventing small molecule binding within the active site. In order to determine the complete structure, a series of mutants were designed to alter crystal packing. The structure of these were determined, and they adopt a different packing arrangement which results in the previously disordered loops being ordered, as well as exposing the binding site. The structure of the F122T mutant was used for *in silico* modelling to determine the likely binding site for flavonoid ligands.

The mutants were subsequently used for seeding and soaking experiments which allowed for a complete structure of wild type *Am*GSTF1 to be determined in complex with a glutathione conjugate of CNBF in the active site. This structure gives a significantly better understanding of the mode of action of these inhibitors, in addition to allowing for the possibility of future development of the inhibitors using structure based design.

Table of Contents

Table of Figures

Abbreviations

4-HPPD	4-hydroxyphenyl-pyruvatedioxygenase
ABC	ATP-binding cassette
ACCase	Acetyl CoA Carboxylase
ALS	Acetolactate synthase Inhbitors
CDNB	1-chloro-2,4-dinitrobenzene
Chlor	Chloramphenicol
CNBF	4-chloro-7-nitrobenzofurazan
CoA	coenzyme A
CYP450	Cytochrome P450 monooxygenases
DHAR	Dihydroascorbate reductase
DLS	Diamond light source
DMSO	dimethylsulfoxide
DOXP synthase	1-deoxy-D-xyulose 5-phosphate synthatase
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPSP	enolypyruvl-shikimate-3-phosphate
EPSPS	5-enolypyruvl-shikimate-3-phosphate-synthase
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
FPLC	Fast protein liquid chromatography
GPOX	Glutathione Peroxidase
GS-DNB	Dinitrophenyl thioether
GSH	Glutathione
GST	Glutathione-S-Transferase
GT	Glycosyl transferase
HRAC	Herbicide resistance action committee
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JNK1	c-Jun N-terminal kinases
Kan	Kanamycin
LB	Lysogeny broth
LC-MS	Liquid chromatography mass spectrometry

MDR	Multiple drug resistance
MeCN	Acetonitrile
MHR	Multiple Herbicide Resistance
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NBF-GS	7-nitrobenzofurazan glutathione conjugate
PAGE	polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PDS	Phytoene desaturase
PEG	Polyethylene glycol
PEP	phosphoenolpyruvate
PLP	Piecewise Linear Potential
QToF	Quadropole time of flight
rmsd	Root mean square deviations
RT-PCR	Real time PCR
S3P	shikimate-3-phosphate
SAR	Structure activity relationships
SDS	Sodium dodecyl sulfate
SOC	Super optimal broth with catabolite repression
TCEP	tris (2-carboxyethyl)phosphine
T_m	Melting temperature
TSA	Thermal shift assays
UDP-Glc	Uridine diphosphate glucose
WSSA	Weed science society of America
WT	Wild type

Acknowledgments

I would like to take this opportunity to thank some people without whom this work would not have been possible.

Firstly my sincerest thanks to my supervisor Ehmke Pohl for giving me the opportunity to undertake this PhD. For his continual support, invaluable advice extending far beyond just my research and for making this PhD an overwhelmingly positive experience.

To my second supervisor Patrick Steel for his advice, guidance and enthusiasm throughout the project.

Thanks go to those at Syngenta, my industrial sponsor. To my industrial supervisor Glynn Mitchell for his help throughout the project, to Jenny Moore for hosting me during my placement and for her invaluable advice and guidance, and to Nathan Kidley for his help with docking and computational modelling. Also to all of the members of the Structural Biology and Computational Chemistry teams for making me so welcome during my placement.

Thanks to Ian Edwards for his ability to fix anything and everything in the lab, his protein care packages while I was at Syngenta and for his encyclopedic knowledge of free food opportunities!

To all the members past and present of labs CG229, CG234 and CG001 for their help and friendship over the years. In particular to Charlie, Dan and Katy for countless laughs and help in the lab and Stef for endless GST knowledge and crystal mounting expertise.

Finally, to my friends and family. To my family for their support, and for pretending they are interested in my work even when they don't understand it. To Naomi for being a voice of reason and source of comfort, to Vicki and Niamh for lunchtime entertainment and for eating cake until we felt sick on Food Fridays, to Beth for chats over hot chocolate and to Durham University Sailing Club and Durham Ladies Netball Club for reminding me that there was a world outside of the lab.

Declaration

The work described in this thesis was carried out in the Department of Chemistry, Durham University and at Syngenta International Research Centre, Jealott's Hill, Berkshire between October 2015 and February 2019. All work is the author's own, unless otherwise stated. This work has not been previously submitted for a degree at this or any other institution.

Statement of Copyright

1 Introduction

1.1 Herbicides

With the need to feed the growing population, and with the added pressure of decreasing land space, there has been a rising pressure to increase crop yields. This has led to the use of chemical control in the form of herbicides, fungicides and insecticides. Herbicides control weeds which compete with crops for light and nutrients. Their use has grown rapidly since the first herbicide, 2,4-dichlorophenoxyacetic acid- originally a biological weapon, was used in 1945.[1]

Herbicides can be either non-selective, where they kill all plants or selective, where they target the differences between monocots and dicots, killing weeds while sparing crops. It is possible to enhance selectivity using herbicide safeners which reduce herbicide toxicity to crops whilst maintaining activity against weeds.[2] Herbicides can be applied either pre-emergence of weeds, where the herbicide works to stop weed germination, or post emergence where they act to kill existing weeds. [3]

Herbicides are further classified into groups based on their mode of action, they target only a few proteins and processes within the cell (Figure 1-1).
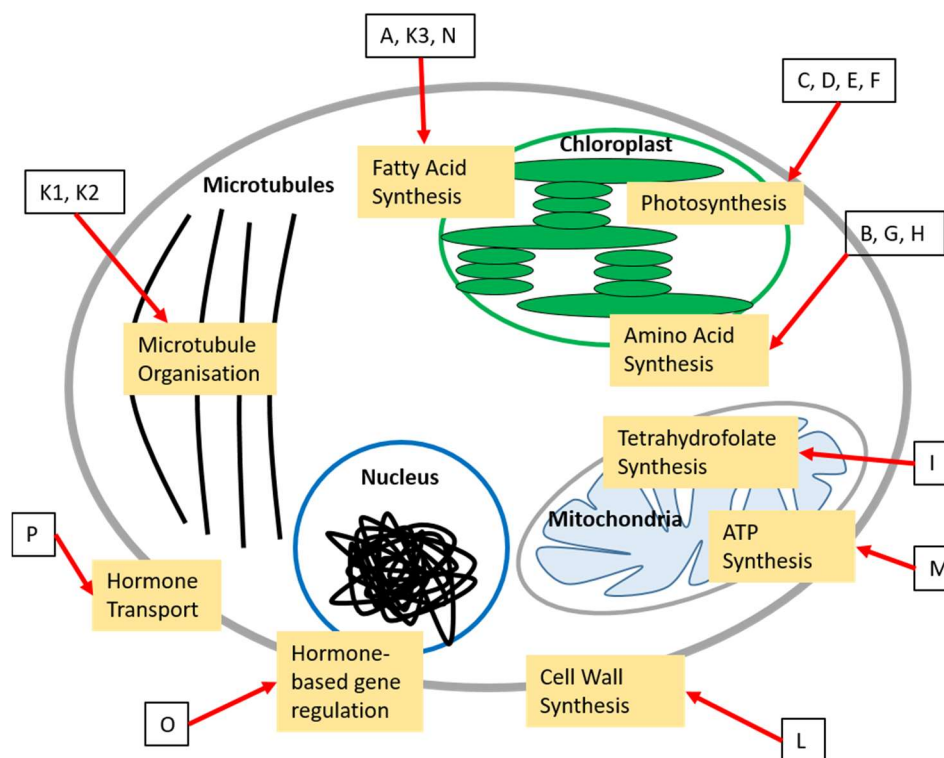
Figure 1-1: Schematic view of herbicide sites of action. Letters represent mode of action of herbicide according to Herbicide Resistance Action Committee classification.[4]

Herbicide groups are assigned either a number (Weed Science Society of America (WSSA)), or a letter (Herbicide Resistance Action Committee (HRAC)) (Table 1-1) based on their mode of action.

Table 1-1: Classification of herbicides according to their mode of action. Classification given according to WSSA classifications (numbers) or HRAC classification (capital letters in brackets)[5]

Group	Site of Action	Examples	Chemical Family
1 (A)	Acetyl CoA Carboxylase (ACCase) Inhibitor	Pinoxaden	Phenylpyrazoline ('DEN')
		Diclofop	Aryloxyphenoxy-propionate ('FOPs')
2 (B)	Acetolactate Synthase (ALS) Inhibitors	Chlorsulfuron	Sulfonylurea
		Flumetsulam	Triazolopyrimidine
3 (K1)	Inhibitor of microtubule assembly	Pendimethalin	Dinitroanaline
4 (O)	Synthetic auxin	2,4-D	Phenoxy-carboxylic acid
5 (C1)	Inhibitor of photosystem II site A	Atrazine	Triazine
6 (C3)	Inhibitor of photosystem II site B	Bromoxynil	Nitrile

7 (C2)	Inhibitor of photosystem II site A (Different behaviour from group 5)	Chlortoluron	Urea
8 (N)	Inhibitor of lipid synthesis; not ACCase inhibition	Molinate	Thiocarbamate
9 (G)	Inhibitor of 5-enolypyruvyl-shikimate-3-phosphate synthase (EPSPS)	Glyphosate	Glycine
10 (H)	Inhibitor of glutamine synthetase	Glufosinate	Phosphinic acid
11 (F3)	Inhibitor of carotenoid biosynthesis (unknown target)	Aclonifen	Diphenylether
12 (F1)	Inhibitor of phytoene desaturase (PDS)	Diflufenican	Pyridinecarboxamide
13 (F3)	Inhibitor of 1-deoxy-D-xyulose 5-phosphate synthatase (DOXP synthase)	Clomazone	Isoxazolidinone
14 (E)	Inhibitor of protoporphyrinogen oxidase (Protox, PPO)	Acifluorfen	Diphenylether
15 (K3)	Mitosis Inhibitor	Alachlor	Chloroacetamide
		Flufenacet	Oxyacetemide
18 (I)	Inhibitor of 7,8-dihydro-preroate synthetase (DHP)	Asulam	Carbamate
19 (P)	Inhibitor of indoleacetic acid transport	Naptalam	Phthalamate Semicarbazone
20 (L)	Inhibitor of cell wall synthesis site A	Dichlobenil	Nitrile
21 (L)	Inhibitor of cell wall synthesis site B	Isoxaben	Benzamide
22 (D)	Photosystem I electron diverter	Paraquat	Bipyridylium
23 (K2)	Inhibitor of mitosis	Carbetamide	Carbamate
24 (M)	Membrane disruptor	Dinoterb	Dinitrophenol
26 (Z)	Unknown site of action	Difenzoquat	Pyrazolium
27 (F2)	Inhibitor of 4-hydroxyphenyl-pyruvatedioxygenase (4-HPPD)	Mesotrione	Triketone
28 (Z)	Unknown site of action	DMSA	Organoarsenical
29 (L)	Inhibitor of cell wall synthesis site C	Indaziflam	-
30 (R)	Tyrosine aminotransferase	Methiozalin	

1.2 Herbicide Metabolism

The metabolism of herbicides in plants occurs in a process similar to that of xenobiotics in humans, in a three stage process: activation, conjugation and export (discussed further below) (Figure 1-2). The notable difference is that while in humans the process occurs in the liver and

waste is excreted in urine there is no such excretory system in plants and as a result compounds undergo internal compartmentation in the vacuole (soluble conjugates) and cell wall (insoluble conjugates).[6]
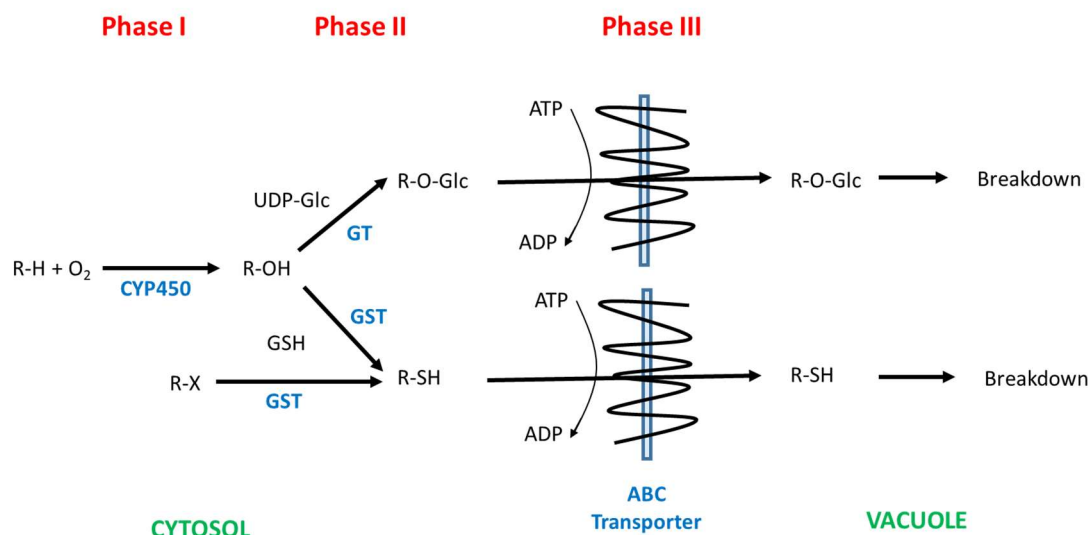
Figure 1-2: Metabolism of herbicides in plants. Showing enzymes used (blue), and locations of the plant where the differing phases occur (green).[7]

In phase I metabolism, compounds are activated to allow for conjugation in phase II. This most commonly involves oxidation by cytochrome p450 monooxygenases (CYP450s). CYP450s are heme dependent proteins that typically use NADPH and/ or NADH to cleave atmospheric dioxygen and produce a functionalised organic substrate and a molecule of water.[8] They are capable of carrying out a wide range of functionalisations most often producing hydroxylated derivatives but they are also capable of aromatic hydroxylation, epoxidation, N, S and O dealkylations, oxidative deamination, sulfoxide formation, N-oxidation, N-hydroxylation oxidative dehalogenation and reductive dehalogenation (Figure 1-3)[9] Most CYP450s are anchored by a single N-terminal transmembrane helix in the endoplasmic reticulum (ER) with the remainder of the protein in the cytosol. They are localised alongside NADPH dependent P450 reductases which transfer electrons from NADPH into the buried catalytic site of the CYP450.[8]
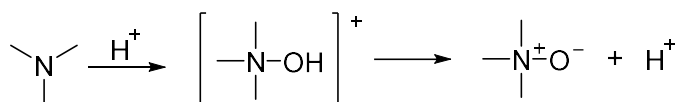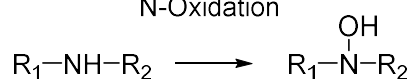
R−CH$_3$ ⟶ R−CH$_2$·OH

Aliphatic Hydroxylation

R−⟨benzene⟩ ⟶ [R−⟨epoxide⟩] ⟶ R−⟨phenol⟩−OH

Aromatic Hydroxylation

R−NH−CH$_3$ ⟶ [R−NH−CH$_2$OH] ⟶ R−NH$_2$ + CH$_2$O

N-Dealkylation

R−O−CH$_3$ ⟶ [R−O−CH$_2$OH] ⟶ R−OH + CH$_2$O

O-Dealkylation

R−S−CH$_3$ ⟶ [R−S−CH$_2$OH] ⟶ R−SH + CH$_2$O

S-Dealkylation

R−CH−CH$_3$ (NH$_2$) ⟶ R−C(OH)−CH$_3$ (NH$_2$) ⟶ R−C(=O)−CH$_3$ + NH3

Oxidative Deamination

R$_1$−S−R$_2$ —H$^+$→ [R$_1$−S(OH)−R$_2$]$^+$ ⟶ R$_1$−S(=O)−R$_2$ + H$^+$

Sulfoxide Formation

⟩N⟨ —H$^+$→ [−N(−OH)]$^+$ ⟶ −N$^\pm$O$^-$ + H$^+$

N-Oxidation

R$_1$−NH−R$_2$ ⟶ R$_1$−N(OH)−R$_2$

N-Hydroxylation

R$_1$−C(H)(R$_2$)−X ⟶ R$_1$−C(OH)(R$_2$)−X ⟶ R$_1$−C(R$_2$)=O + HX

Oxidative Dehalogenation

R$_1$−C(R$_3$)(R$_2$)−X —e$^-$→ R$_1$−C·(R$_3$)(R$_2$)−X$^-$ —H$^+$→ R$_1$−CH(R$_3$)(R$_2$) + HX

Reductive Dehalogenation

Figure 1-3: Reactions carried out by cytochrome P450 enzymes[9]

Phase II involves the conjugation of a small molecule to the compound. This acts to increase water solubility, as well as targeting the molecule for transport. It involves the use of either glutathione transferases (GST) or glycosyltransferases (GT).[10] GTs catalyse the addition of a sugar molecule to a functional group, using uridine diphosphate glucose (UDP-Glc) as a sugar donor. O-GTs add to –OH or –COOH groups, while N-GTs add to NH_2 groups.[11] GSTs catalyse the addition of a glutathione (GSH) molecule to an electrophilic centre. Unlike glycosylation this does not always require initial functionalisation by a phase I enzyme as many herbicides contain an electrophilic centre.

Phase II metabolism detoxifies compounds, but it is possible they may have another biological activity and therefore must be removed from the cytosol. For example, herbicide synergist tridiphane is converted to a GSH conjugate in phase II metabolism which is a powerful suppressant of GST-mediated atrazine metabolism.[12] As a result metabolites are either deposited as bound residues in the extracellular matrix, or stored as water soluble metabolites in the vacuole.[10] The uptake of GSH conjugates into the vacuole is ATP dependent, and uses specific ATP-binding casette (ABC) transporters.[13]

1.3   Herbicide Resistance

With increasing use of herbicides, resistance has become an increasing problem. Resistance has been reported across all modes of herbicide action, and across virtually every weed species. Since 1975 a rapid growth has been observed in the number of resistance cases that have been reported, with 495 unique cases being reported up to 2017, although an increase in surveillance and better reporting has probably partially factored in this rapid incline.[14]
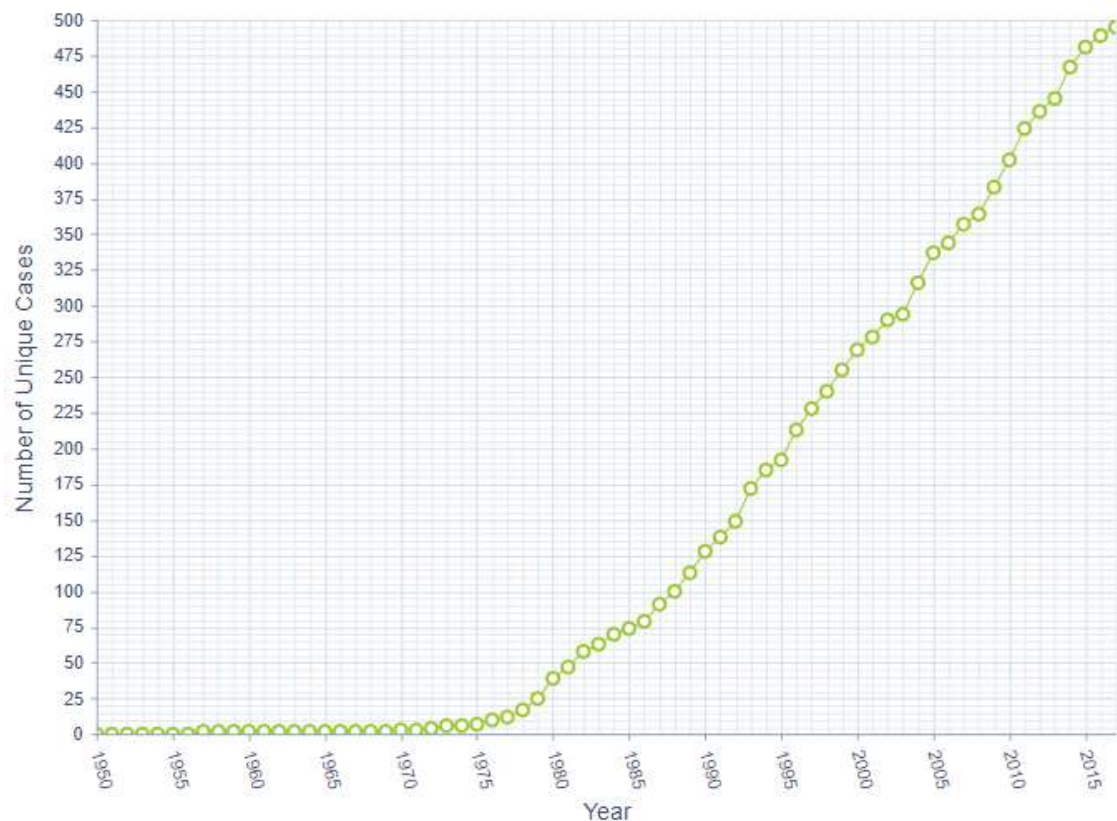
Figure 1-4: The chronological increase in herbicide resistance cases observed globally since 1950 reported in the international survey of herbicide resistant weeds.[14]

Herbicide resistance can take two forms: target site and non-target site. Target site comes from mutations in the protein which the herbicide targets causing the herbicide to become less efficient while non-target site occurs from other complementary methods.

Herbicide action is divided into 3 steps: penetration into the plant; translocation and accumulation at the target site; and binding to the biological target causing a phenotypic response and resulting in plant death. Resistance mechanisms can occur at each and every point of this pathway.[4]

1.3.1 Target-site resistance

Target site resistance occurs most commonly via a gene mutation conferring an amino acid change which prevents herbicide binding while allowing substrate binding. It can also occur by overexpression of the target enzyme.[15] This can result in a fitness cost if the mutation impairs enzyme functionality or plant performance.[16] This section will explore examples of target site resistance occurring against a range of herbicide mode of actions.

1.3.1.1   ALS Inhibiting Herbicides

Acetolactate synthase (ALS) is the first enzyme in the biosynthetic pathway of branched amino acids (Val, Leu and Ile) (Figure 1-5). It is an essential protein, making it an excellent herbicide target. A huge number of resistance endowing mutations of ALS have been identified. By 2009 22 had been observed across seven different amino acid sites. Most remarkably at Pro197 11 different amino acid substitutions were observed which resulted in herbicide resistance.[15] A Pro197Ser mutation is most common as this does not confer a fitness cost and requires only one nucleotide substitution. Crystal structures demonstrate that ALS inhibiting herbicides do not bind to the active site, instead binding across the channel that gives access to the active site for substrate binding.[17] As a result resistance is common as different mutations are able to be made without affecting native substrate binding.[18]

Figure 1-5: Synthesis of branched chain amino acids showing function of ALS and action of ALS inhibiting herbicides

1.3.1.2   ACCase Inhibiting Herbicides

Acetyl CoA carboxylase (ACCase) is a key enzyme in lipid biosynthesis, responsible for catalysing the formation of malonyl CoA from the carboxylation of acetyl CoA (Figure 1-6). This blocks fatty acid biosynthesis altering membrane integrity, resulting in plant death.
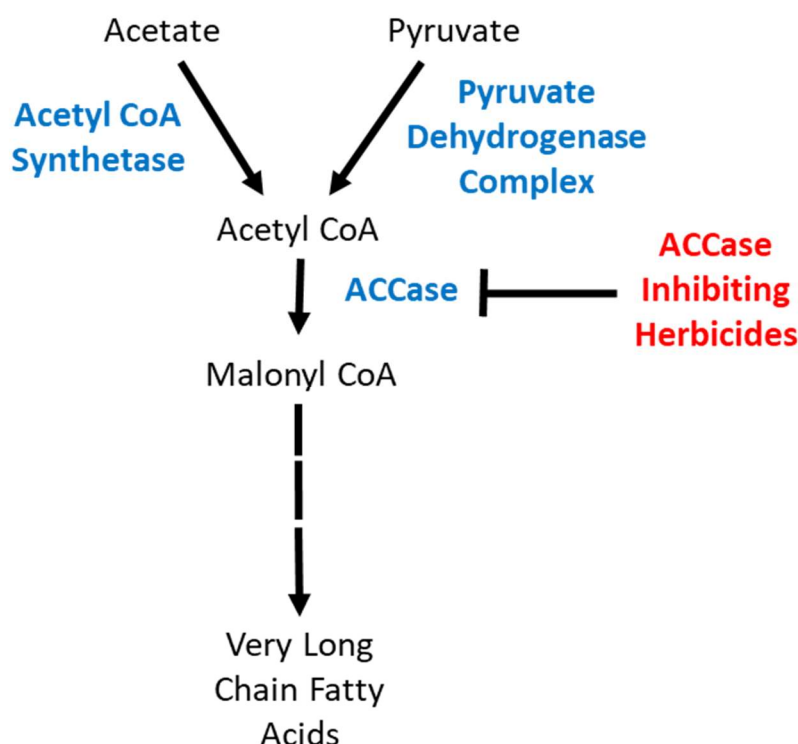
Figure 1-6: Lipid biosynthesis pathway showing enzymes in blue and action of ACCase inhibiting herbicides[19]

48 species of grass weeds have been identified as having developed resistance (2018).[14] The Leu1781Ile mutation is the most common resistance substitution but seven others have also been identified. Crystal structures have demonstrated that herbicides bind to a domain close to and partially overlapping the catalytic site.[20] This means some amino acid substitutions can be made which prevent herbicide binding while not disrupting acetyl CoA binding, while others do disrupt acetyl CoA binding, resulting in a fitness cost.[21]

1.3.1.3   Glyphosate resistance: EPSPS Inhibitors

Glyphosate is the world's most widely used herbicide, popular as it controls a very broad spectrum of weeds.[15] Glyphosate is a potent inhibitor of EPSPS which catalyses the reaction of

shikimate-3-phosphate (S3P) and phosphoenolpyruvate (PEP) to form enolpyruvylshikimate-3-phosphate (EPSP). Glyphosate inhibits this activity disrupting the shikimate pathway and inhibiting aromatic amino acid production.
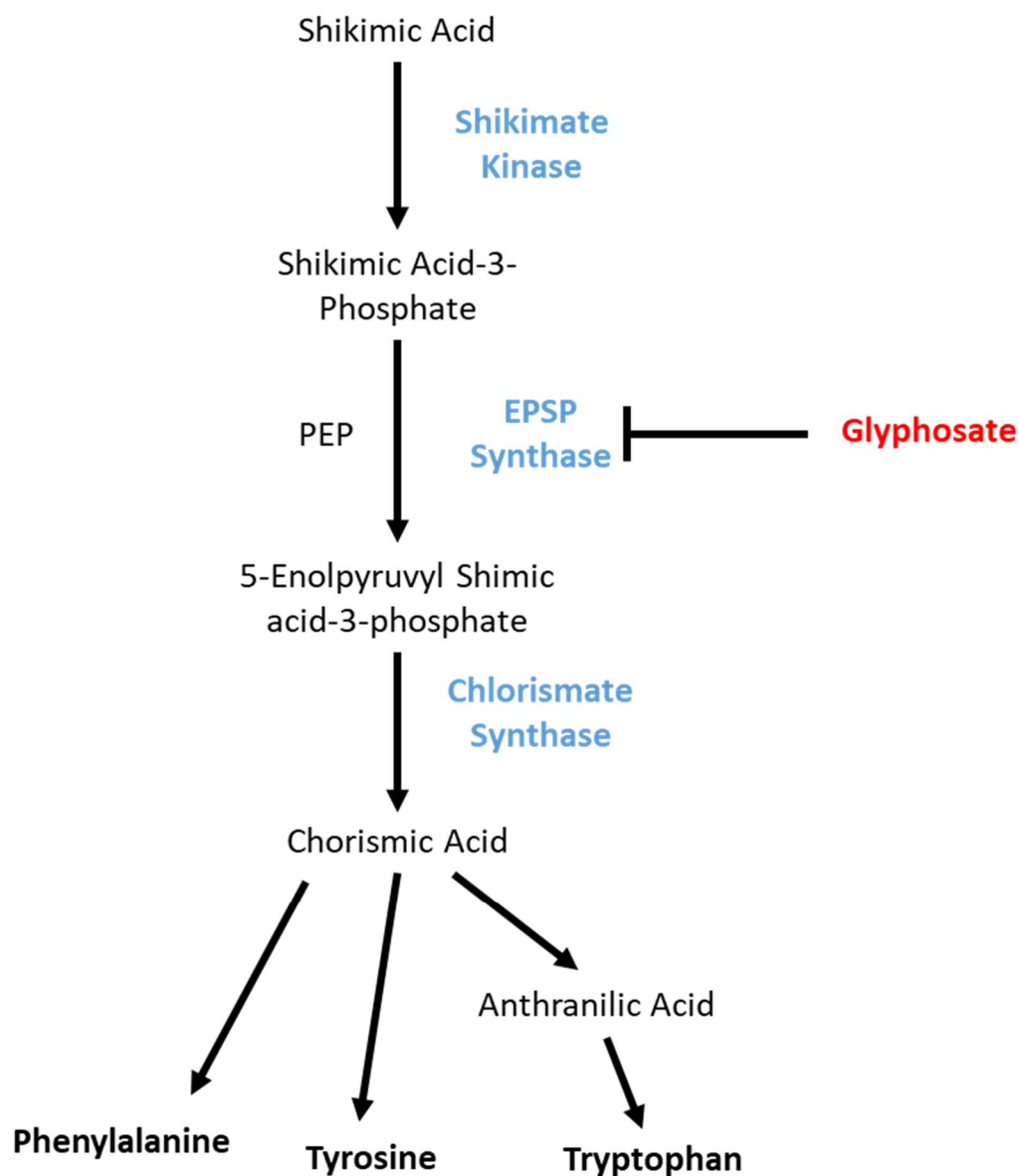
Figure 1-7: Synthesis pathway of aromatic amino acids showing inhibitory action of glyphosate[22]

A single mutation has been observed conferring resistance where Pro106 is mutated to a number of different amino acids. While Pro106 is not involved in herbicide binding, mutating it results in a structural change in the active site reducing available space, evidenced by cyrstallography.[23,24] No other resistance conferring mutations have been observed- it appears to be challenging to prevent herbicide binding while maintaining enzyme function.

Target site resistance has also been observed in glyphosate resistant crops where a large overexpression of EPSPS is observed. A 100 fold gene amplification resulting in a 40 fold overexpression of EPSPS has been observed in *Amaranthus palmeri.*[25]

1.3.2   Non-target site resistance

Non-target site resistance can occur via a number of different mechanisms. Weeds have been observed which show reduced penetration to herbicides, as well as those with altered translocation of herbicides. For example, plants have developed resistance to glyphosate by reducing its translocation through the plant and to its roots.[26] Reduced translocation has also been observed in response to paraquat, a herbicide that targets the electron transfer chain. Reduced translocation inherits as a single nuclear semi-dominant gene meaning it is easily spread to future populations.[27]

Perhaps the most common methods of non-target site resistance involve the metabolic enzymes CYP450s and GSTs. CYP450s are known to be involved in herbicide resistance although this is poorly understood. Non-target site resistance occurs when CYP450s are upregulated. Crops are able to CYP450 metabolise many herbicides so their use on weeds acts as a strong selection pressure for weeds with the same ability. It is a very threatening form of resistance as it is able to confer resistance to many different types of herbicides (Cross resistance discussed in more detail in 1.3.3.1).[15]

GSTs are also known to be involved in resistance to herbicides. Maize is able to detoxify triazine herbicides because of GST activity. A mutation has been reported in *A. theophrasti* which leads to an improved catalytic activity of a GST resulting in atrazine resistance.[28] The role of GSTs in resistance will be discussed in more detail later as part of their role in multiple herbicide resistance (MHR). (See section 1.3.3.1.2)

1.3.3   Cross Resistance and Multiple Resistance

Cross resistance and multiple resistance are both methods by which weeds can become resistant to a number of different herbicides. In cross resistance plants have a single mechanism which gives resistance to two or more chemical families targeting one site of action. In multiple resistance plants possess multiple target site resistance traits that give resistance to several different sites of action.[29]

1.3.3.1   Cross resistance

Cross resistance is a genetically endowed mechanism conferring the ability to withstand herbicides from different chemical classes. It can occur as a result of either target site or non-target site methods.[29]

1.3.3.1.1   Target site cross resistance

Target site cross resistance occurs where a change at the biochemical site of action of a herbicide confers resistance to herbicides from different chemical classes that inhibit the same site of action. It does not necessarily give resistance to all herbicide classes with similar mode of action or even all herbicides within a chemical class.[29] For example, cross resistance is observed in response to ALS inhibiting herbicides. The Asp376Glu mutant is resistant to imidazolinone, pyrimidinylthiobenzoate, sulfonylaminocarbonyltriazoline, sulfonylurea and triazolopyrimidine herbicides.[30] Increased amplification of a target gene and overexpression of a protein will also result in resistance to all chemical classes targeting a particular mode of action.[25]

1.3.3.1.2   Non-target site cross resistance

Non target site resistance weeds often have a central defence mechanism which gives resistance to many different sites of action. The most studied form of these is multiple herbicide resistance. This is seen in P450s, GSTs, GTs and ABC transporters.[4]

CYP450 activity has been seen to correlate with herbicide resistance in weeds. CYP450 involvement in black grass resistance to chlortoluron was identified through exogenous application of a P450 enzyme inhibitor and by analysing metabolites that accumulated after herbicide treatment.[31]

GSTs were first implicated in herbicide resistance in 1970s with their involvement in atrazine resistance in grasses. It has been seen to occur both as a result of increased activity of GSTs[32] and in "Peldon" populations of black grass in response to an increase in gene expression.[33,34] This role in peldon populations will be discussed further with regards to the role of *Am*GSTF1 in multiple herbicide resistance in black grass (see section 2.2).

1.3.3.2   Multiple resistance

Multiple resistance occurs when a weed develops resistance to multiple different sites of action through a number of different genetic changes. Resistance is able to spread and accumulate quickly particularly with cross pollination.[15] Some weed populations have been seen to show resistance to up to seven different modes of action (Figure 1-8).[14]

©2019 WeedScience.org, Dr. Ian Heap 01/05/2019
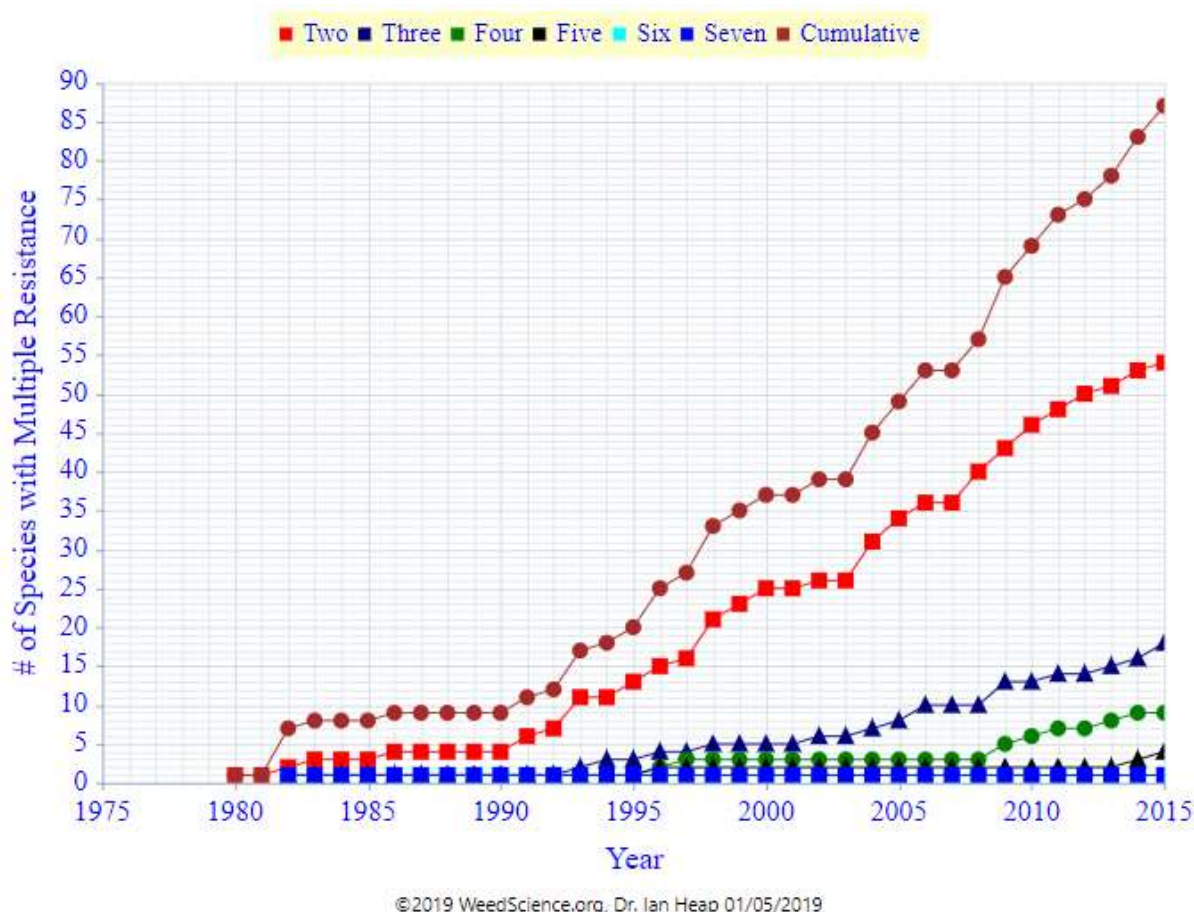
Figure 1-8: Weed populations resistant to herbicides of different sites of action. Showing the number of individual weed populations which have been identified as showing resistance to multiple sites of action and the number of sites of action and the number of weed populations identified in each case.

1.4 Glutathione-S-Transferases

GSTs are a class of enzyme primarily associated with their role in catalysing the nucleophilic addition of reduced glutathione (GSH) to electrophilic centres in compounds to produce a less reactive compound and increase solubility.[35]

They are divided into a number of classes based on sequence similarity of which six classes are present in plants (Table 1-2). Phi, tau, lambda and dihydroascorbate reductase (DHAR) are plant specific while theta and zeta classes are also found in mammals. They are ubiquitous, with 53 members encoded for in the *Arabidopsis* genome.[36] They are soluble proteins,[37] with phi and tau classes most represented.[38]

Table 1-2: Classes of plant GST and their single letter codes

GST Class	Single Letter Code	Kingdom
Phi	F	Plant
Tau	U	Plant
Theta	T	Plant and Mammal
Zeta	Z	Plant and Mammal
Dihydroascorbate	DHAR	Plant
Lambda	L	Plant

There is specific nomenclature used to describe and identify these GSTs, eg. *Am*GSTF1-1. The italicised letters denote a two letter code for the organism where the GST is found, usually based on binomial nomenclature, in this case *Alopecurus myosuroides*. The letter following the GST represents the class, in this case phi. This is then followed by a number which represents the dimer subunits. In this case F1-1 represents a homodimer of the F1 subunit.

1.4.1 The multiple roles of GSTs

While GSTs are most commonly associated with, and were initially discovered for their ability to detoxify metabolites, they have also been found to participate in a number of other cellular activities. These roles will be discussed in this section.

GSTs are able to catalyse the quantitative conjugation of GSH, which is the major phase II metabolism reaction in many species. GSTs catalyse nucleophilic aromatic substitutions, Michael additions to α,β unsaturated ketones and epoxide ring-opening reactions, all of which result in the formation of GSH conjugates.[39]

1. Nucleophilic Aromatic Substitution
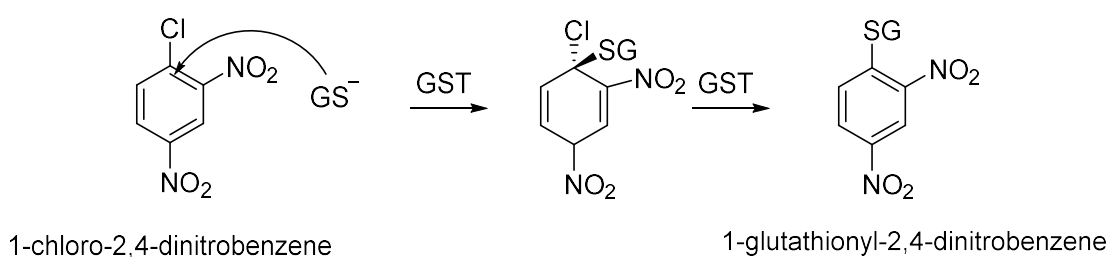
1-chloro-2,4-dinitrobenzene   1-glutathionyl-2,4-dinitrobenzene

2. Michael Addition to α, β-unsaturated ketones

Ethacrynic Acid   EA-GSH Complex

3. Epoxide ring opening reactions

Phenanthrone 9,10-dioxide
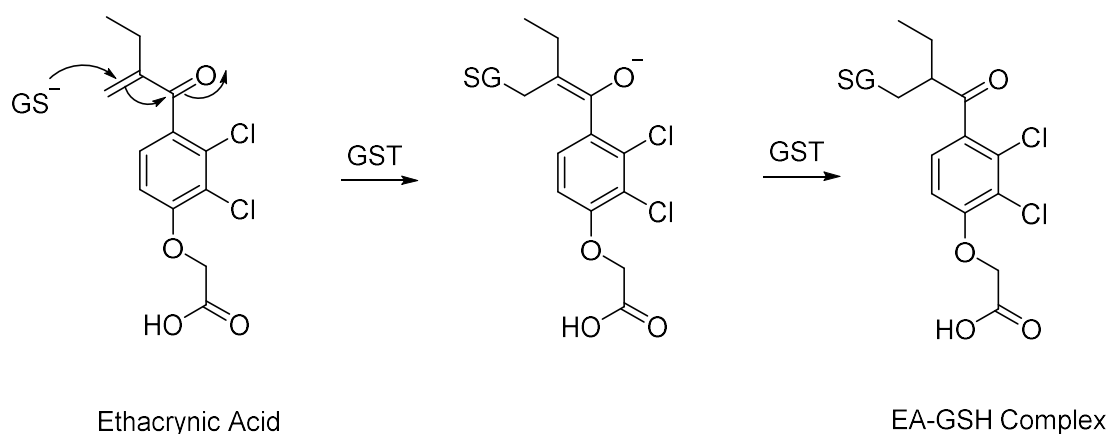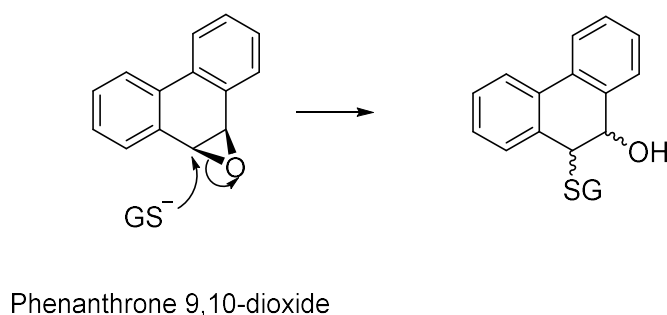
Figure 1-9: Examples of reactions carried out by Glutathione Transferases.[39] GS⁻ represents a glutathione anion.

1.4.1.2 Other Catalytic Activity

A number of plant GSTs have been identified as showing glutathione peroxidase activities, including proteins present in *Arabidopsis,*[40,41] wheat[42] and black grass.[33]

Some GSTs have been identified as having isomerase activity. They use GSH in a catalytic role, with the end product not including a GSH conjugate. For example, the zeta class GST in *Arabidopsis* adds GSH across a cis-double bond in maleylacetoacetate, allowing bond rotation followed by elimination of GSH to yield fumarylacetoacetate.[43]

1.4.1.3 Targeting transmembrane transport

Secondary metabolites are often toxic, and require glutathione conjugation to mark them for transport to the vacuole. For example, in anthocyanin biosynthesis if the pigment is retained in the cytoplasm it prevents pigment production and is toxic to cells.[44]

1.4.1.4 Protection from oxidative damage

Oxidative damage caused by hydroxyl radicals is cytotoxic. GSTs are able to conjugate GSH with endogenously produced electrophiles to detoxify them.[40]

During pathogen attack, some plants protect themselves using a process called systemic acquired resistance.[45,46] This process involves production of salicylic acid, which prevents catalase activity, stopping breakdown of H_2O_2. This pathway leads to GST genes being induced to protect the plant from the damaging effect of the reactive oxygen species.[47]

1.4.1.5 Ligandins: Non-enzymatic binding and intracellular transport

In addition to their catalytic function, GSTs are able to function as carrier proteins. In mammals they have been identified to be involved in transport of steroids, heme, bilirubin, bile salts, carcinogens and some drugs.[48]

In plants some GSTs serve as carriers of natural auxins,[49] while others have been identified to bind flavonoids in their non-catalytic L-site.[50]

1.4.1.6 Other functions in Mammals

GSTs have been studied more extensively in mammals than plants, and a number of other function identified. GSTs are responsible for catalysing the penultimate step in the catabolism of Phe and Tyr amino acids.[51] They are also involved in steroid synthesis of testosterone and progesterone.[52] They contribute to the biosynthesis of pharmacologically important metabolites of arachidonic acid.[53] Finally GSTs are involved in modulating the signalling pathways of cyclopentenone prostaglandins.[54,55]

1.4.2   Structure of GSTs

GST structure in plants is well understood thanks to crystal structures being available for all cytosolic classes except the theta class for which mammalian structures are available (Table 1-3). Sequence similarity is low and substrate specificity varies, however 3D structure is highly conserved, especially within the glutathione binding site.[35]

Table 1-3: Representative crystal structures of plant GSTs available in the Protein Data Bank (PDB). For Theta class no plant structure is available however the human structure is well characterised. Rmsd values are given for each structure compared to the Phi class Maize structure (1axd), and are calculated using Pymol.

GST Class	Organism	GST code	PDB Code	Ligand	Reference	rmsd/ Å
Phi	*Arabidopsis thaliana*	*At*GSTF2	5a4w	Quercetrin	[50]	0.92
	Zea mays	*Zm*GSTF1	1axd	Lactoylglutathione	[56]	-
Tau	*Mangifera indica*	*Mi*GSTU1	5kej	S-hexyl glutathione	[57]	3.223
	Arabidopsis thaliana	*At*GSTU23	6ep6	-	[58]	3.17
Theta	*Homo sapiens*	*Hs*GSTT1	2c3t	-	[59]	2.72
Zeta	*Arabidopsis thaliana*	*At*GSTZ1	1e6b	-	[43]	6.36
DHAR	*Oryza sativa*	*Os*DHAR1	5d9v	-	[60]	3.95
	Arabidopsis thaliana	*At*DHAR2	5lol	Glutathione	[61]	3.79
Lambda	*Populus trichocarpa*	*Pt*GSTL1	4pqh	Glutathione	[62]	2.52
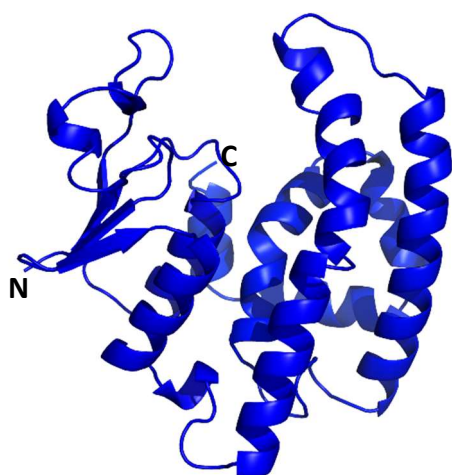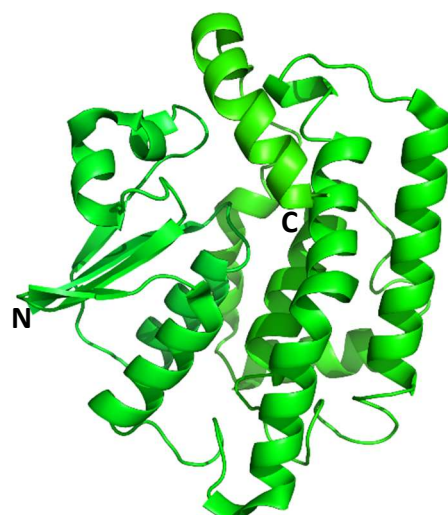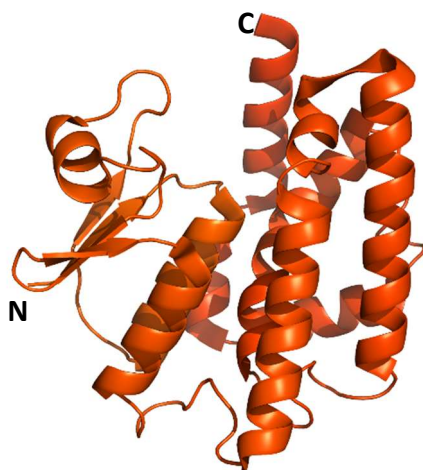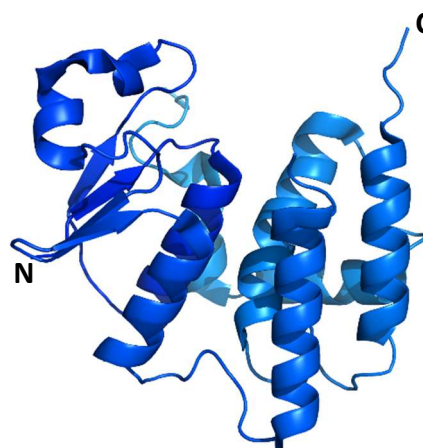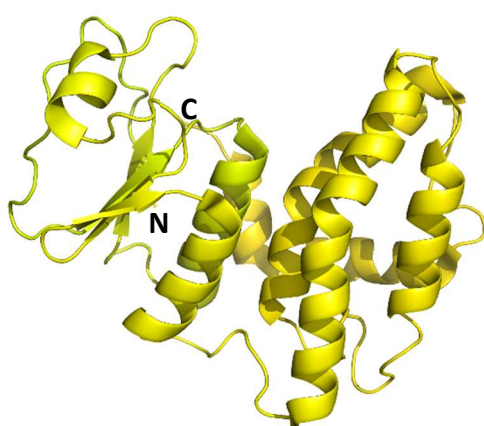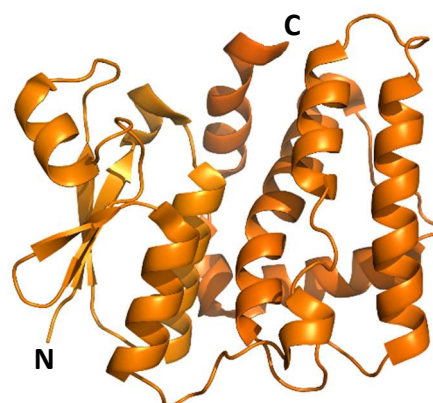
Figure 1-10: Representative structures for each of the classes of plant GST. PDB codes are given in brackets. Structures are all orientated in the same manner with domain I on the left and domain II on the right. N and C termini are marked for each structure. Figures produced in pymol.

Most classes have a dimeric composition (except DHAR and lambda classes) with the enzyme active as a homo or heterodimer of subunits from the same gene class of about 25 kDa in size. Each subunit is composed of a catalytic site with two components (domain I and domain II) joined by a short linker region (Figure 1-11).
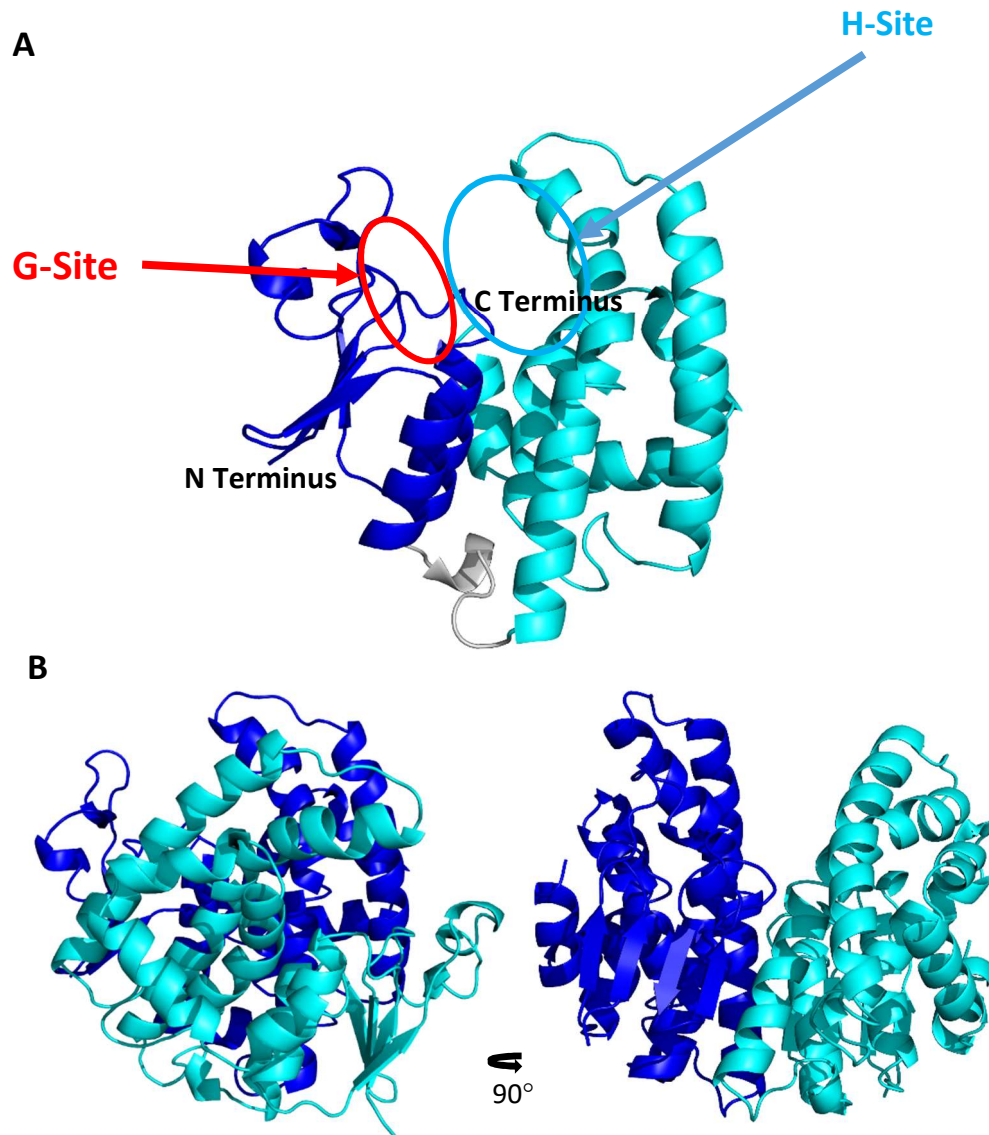
Figure 1-11 A: Structure of a Phi Class GST (PDB:1AXD) showing location of domain I and II, position of G and H sites are circled, and location of N and C terminus marked. Domain I is coloured in blue, with domian II in cyan, and the linker region is coloured in grey. The protein is orientated with the dimer interface facing out. B: Dimer form of a Phi Class GST (PDB 1AXD) with the monomer form in the same orientation as for figure A (left) and rotated by 90° (right). Each half of the dimer is coloured either cyan or blue. Figure produced in pymol.
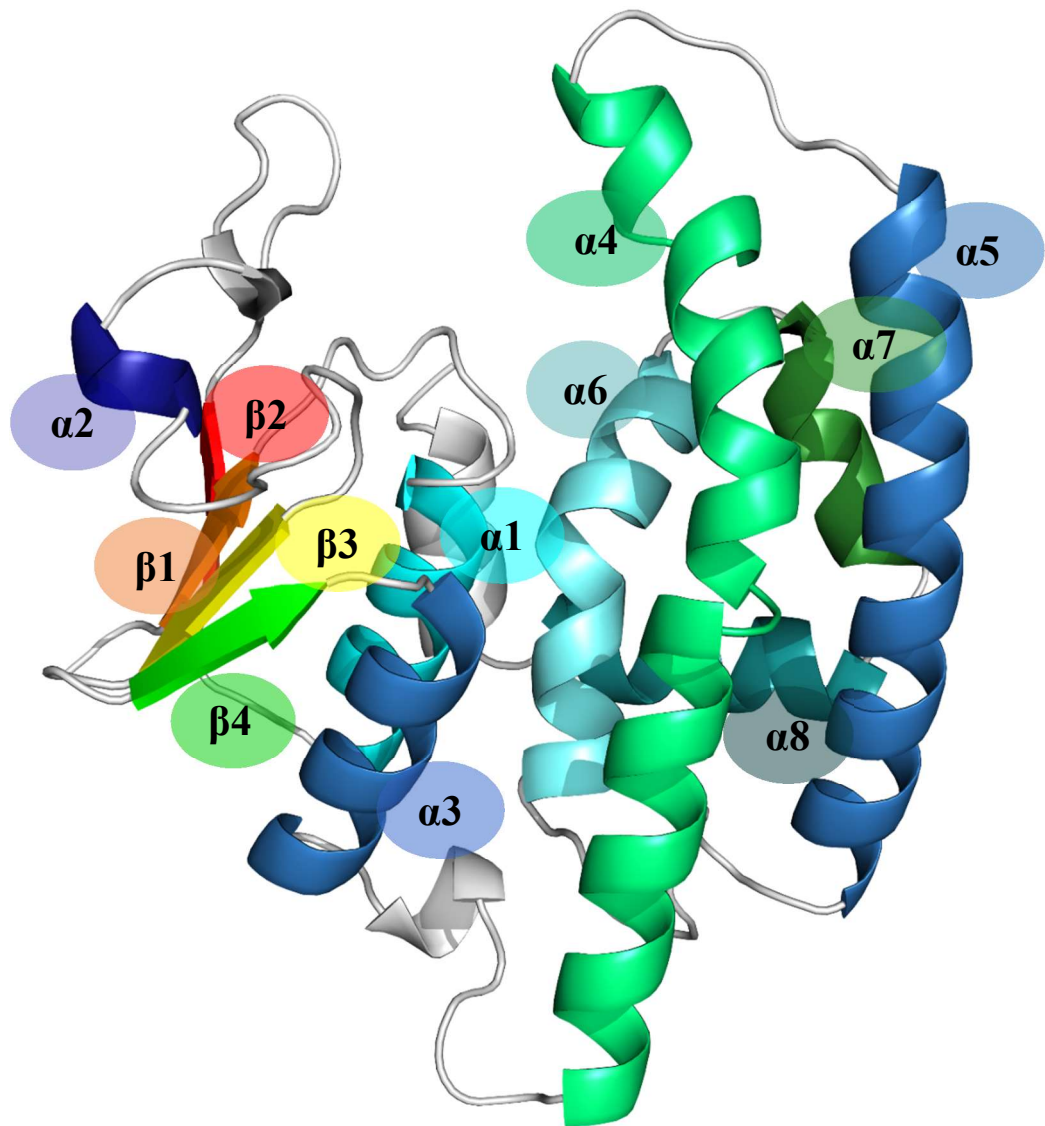
Figure 1-12: Structure of a Phi Class GST (PDB 1AXD) showing conventions for numbering secondary structure elements. Labels for secondary structure elements are shown in the same colour as the secondary structure element. Figure produced in Pymol.

Domain I is composed of residues from the C-terminus and provides the residues for GSH binding and catalytic activity.[63] It contains an α/β core structure composed of a four stranded β-sheet (β1-4) flanked by two α-helices (α1 and 3) (Figure 1-12). It resembles the thioredoxin fold observed in crystal structures for *E. coli* thioredoxin,[64] bacteriophage T4 glutaredoxin,[65] and glutathione peroxidase.[66]

Domain II is an all α structure composed of 5 α helices (α4-8).[63] It provides the residues for hydrophobic substrate binding (H-site) and as a result shows greater sequence diversity than domain I to provide substrate specificity.[35]

Dimers are stable with no spontaneous exchange of subunits occurring. Class specific heterodimers are seen, but dimers cannot include monomers from differing classes. Dimerisation contacts are primarily hydrophobic and occur between domain I of one subunit and domain II of the neighbouring subunit.[63] While dimerisation is essential for activity, and it is essential to form a full G-site, it is unclear why it is essential for GSTs to form dimers. Positive cooperativity has never been demonstrated,[38] however in Pi class GSTs it has been demonstrated that dimerisation stabilises the tertiary structure of the subunit.[67]

In plants Phi, Tau, Theta and Zeta classes have an active site serine residue (tyrosine in mammals) which stabilises the formation of a thiolate anion of GSH, enabling glutathionylation of substrates in the H-site. This residue has been shown to be catalytically essential via site directed mutagenesis across a range of organisms.[43,59,63,68] While the serine residue in plants is not located at the equivalent amino acid position to the tyrosine, Ser is at the N-terminus of α1 while Tyr is at the C-terminus of β1, their hydroxyl groups are positioned such that they occupy a similar location in the active site. By contrast Lambda and DHAR have a cysteine in their active sites. They do not catalyse the S-glutathionylation of substrates present in the H-site and instead are more likely to catalyse redox reactions.[69]

The G-site is made of a network of specific polar interactions between the glutathione tripeptide and protein moieties in domain I, and one interaction with a residue in domain II of the other subunit of the dimer. Glutathione binds in an extended conformation at one end of the four strand β-sheet in domain I.[70] The orientation is important to catalysis, and is governed by the selectivity of the binding site.[71] Surprisingly, given the number of interactions formed, and the specificity observed for glutathione the residues are predominantly conservatively replaced between classes and members of classes rather than conserved.[63] Glutathione binding occurs via an induced fit mechanism with the loop over the G-site displaying a large conformational change on binding.[56]
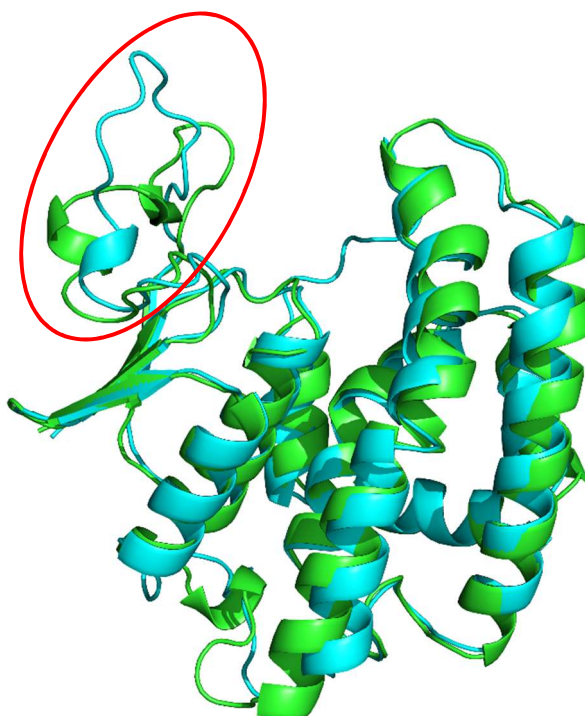
Figure 1-13: Structures demonstrating the induced fit mechanism of loop over the G-site. GSH bound (PDB: 1AXD; green) and apo (PDB: 1AW9; cyan). On GSH binding the loop is seen to adopt an alternative structure, allowing it to form additional interactions with the GSH molecule. Figure produced in Pymol.

The molecular details of the H-site differ more than the G-site both within and between classes. It is composed of clusters of hydrophobic amino acid side chains which provides a highly hydrophobic surface. Structural elements arise from both domains I and II including the active site loop between $\beta1$ and $\alpha1$, the C-terminal region of $\alpha4$ and the C-terminal segment of the polypeptide chain. The variation observed in H-site residues explains the variation observed in xenobiotic specificities by GSTs.[63]

In a number of GSTs a third non-catalytic binding site has been identified. The L-site is found in the dimer interface (Figure 1-14), and is involved predominantly in transport.[50,72]
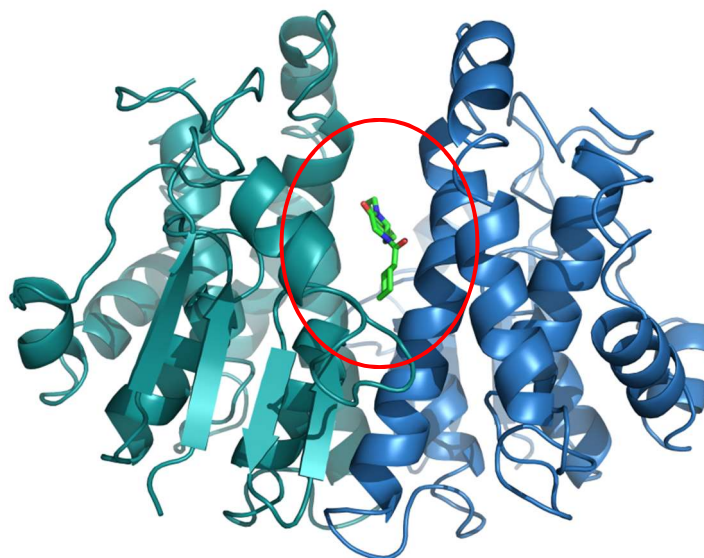
Figure 1-14: GST from squid showing location of L-site.[72] Each half of the dimer is shown in a different colour, with the drug Praziquantel (shown in green) bound in the L-site. Figure produced in Pymol.

1.4.3   Mechanism of GSTs

In plants nucleophilic substitution is most common. This forms a conjugate that is less reactive than the parent and increases solubility. During the reaction a Meisenheimer complex is formed, and this has been seen in an enzyme bound GSH anion with 1,3,5-trinitrobenzene.[73] The catalytic mechanism lowers the pKa of the GSH thiol group from around 9 in aqueous conditions to between 6 and 7 when bound to the protein. Kinetics suggests this would occur via a random sequential two substrate, two product reaction. However, in cellular conditions GSH addition occurs first because GSH concentration is 1-10 mM (about 3 orders of magnitude higher than the GSH dissociation constant from the enzyme).[35,74]

In an *in vitro* system, it appears that σ complex formation is rate limiting. If the Cl group is substituted for a more electrophilic group the rate constant increases approximately 50 fold. If σ complex decomposition was rate limiting we would expect the opposite. However, viscosity experiments suggest that *in vivo* there is most likely a non-chemical rate limiting step, and that K_{cat} is limited by product release which is limited by a diffusional barrier.[75]

1.5   Summary

This chapter has discussed herbicides and the mechanisms of resistance that have developed in response to their use. It has also looked at glutathione-s-transferases and their role in herbicide metabolism. The next chapter will focus on previous work which has been done on this project investigating multiple herbicide resistance in black grass, and the role GSTs play.

2 Previous Work and Aims

2.1 Black Grass

Black grass (*Alopecurus myosuroides)* is an annual rye grass that is found on cultivated land, and is a major weed of cereal crops. It produces a large quantity of seeds, and occurs at very high densities competing with the crop and resulting in major yield losses.[76] Whilst wild type black grass can be well controlled through herbicide application, incidences of herbicide resistance have led to a loss of control. A recent survey determined that 80% of black grass collected from 132 fields was highly resistant to all herbicides that can be used for selective black grass control in a wheat crop.[77]

Figure 2-1: Black Grass[78]

MHR was first reported in the UK in Peldon in 1982. It has now been reported in independent outbreaks all over the UK. MHR has been linked to an increase in metabolic enzymes and occurs through resistance to several herbicides.[56] The work in this project has focussed on MHR observed in this Peldon population.

2.2 Identification of *Am*GSTF1

GSTs are known to play a large role in non-target resistance. In velvetleaf (*Abutilon theophrasti)*, a resistant maize weed, GSTs have been implicated in atrazine resistance,[79] and GSTs have also

been identified as responsible for herbicide detoxification in maize crops. Work in the Edwards group has focussed on identifying the role of GST activity in herbicide resistance in black grass.[33,34,42] Work has focussed on a population from Peldon, identified to be particularly herbicide resistance, and has also used a herbicide susceptible population from Rothamsted. The Peldon population is seen to have higher GST activity (1.9 ×) than the herbicide susceptible Rothamsted population.[42] Using anti-wheat GST sera for western blotting, it was possible to identify three populations of GST that were up-regulated in the herbicide resistant populations (25, 27 and 28 kDa). In order to allow for detailed characterisation, cDNAs encoding the different types of GST subunits in black grass were cloned from Peldon seedlings and the recombinant proteins overexpressed.[33] It was concluded that the 25 kDa protein observed in western blots was *Am*GSTU1. However, exhaustive screening was unable to identify peptides of 27 and 28 kDa, instead identifying *Am*GSTF1a and *Am*GSTF1c of nearly identical mass (24.8-24.9 kDa). When these peptides were analysed with western blotting they were seen to migrate with the 28 kDa polypeptide. The 27 kDa polypeptide was determined to be a partial degradation product, with an increase observed after prolonged dialysis. Both *Am*GSTF1 and *Am*GSTU1 were identified to be active as their respective homodimers by gel filtration chromatography.[33]

RNA transcripts were analysed from susceptible Rothamsted and resistant Peldon populations. For both *Am*GSTF1 and *Am*GSTU1 it was clear that the enhancement of GSTs observed in resistant black grass was due to increased gene expression.[33]

2.3   Role of *Am*GSTF1 in Multiple Herbicide Resistance

Activity assays were undertaken to identify the possible roles of the GSTs.[33] *Am*GSTU1 was seen to be highly active in conjugating the herbicides fluorodifen and fenoxaprop-ethyl. Significantly however, *Am*GSTF1 was seen to have low activity towards these herbicides, suggesting the protective role played by these proteins was unlikely to be as a result of accelerating herbicide metabolism. The proteins were then assayed for glutathione peroxidase (GPOX) activity. While *Am*GSTU1 had no activity as a GPOX, *Am*GSTF1 was seen to be very active catalysing the reduction of both cumene hydroperoxide and linoleic acid hydroperoxide. This suggests that *Am*GSTF1 may be functioning to protect the black grass from toxic organic hydroperoxides formed as a result of herbicide injury.

Further work was undertaken to investigate the role that *Am*GSTF1 plays in multiple herbicide resistant black grass using transgenic *Arabidopsis*.[34] *Arabidopsis* plants expressing either high or low levels of *Am*GSTF1 were shown to have increased resistance to a range of herbicides, chlortoluron, alachlor and atrazine, in both spray and germination trials (Figure 2-2). Both transgenic lines showed an enhancement in GPOX and CDNB conjugating activities in line with the activity profile of the enzyme. Interestingly, the transgenic plants also showed increased conjugative activity towards atrazine, which is not a substrate of *Am*GSTF1. These activity increases are therefore likely to arise as a result of increased expression of endogenous *Arabidopsis* enzymes.[34] Furthermore, an enhancement in unrelated glycosylating activities towards the xenobiotic 2,4,5-trichlorophenol were also observed. Thus, two independent routes of bioconjugation of atrazine were both enhanced in the transgenic plants.
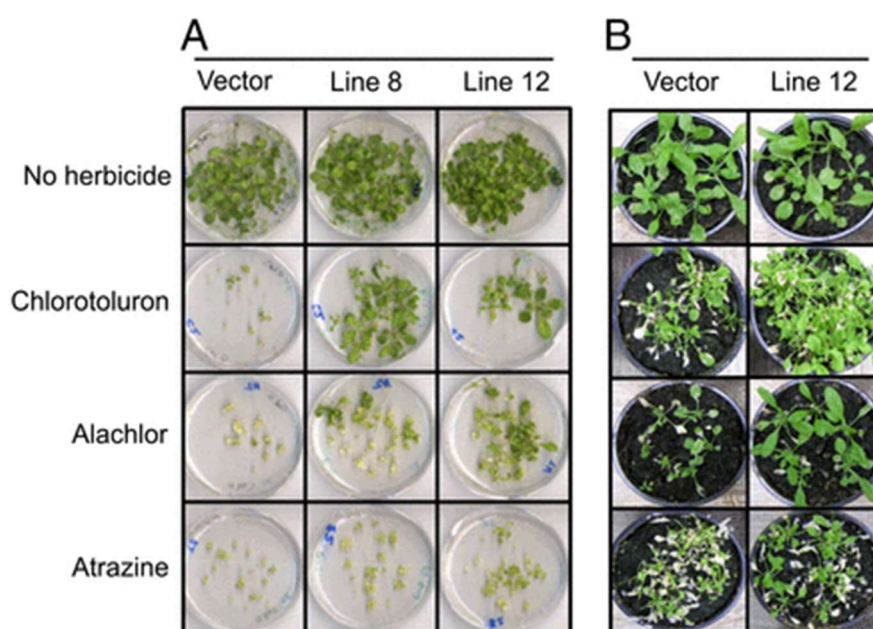
Figure 2-2: Transgenic Arabidopsis expressing AmGSTF1 in intermediate (line 8) and high (line 12) amounts. Plants were grown on agar containing herbicide (A) or sprayed after germination (B).[34] Plants expressing *Am*GSTF1 are observed to show resistance to the herbicides.

In addition, the transgenic plants were analysed using liquid chromatography-mass spectrometry (LC-MS), and a number of metabolites were identified which accumulated at higher levels in these plants than the vector only controls. The compounds were identified as conjugates of the flavonoid kaempferol and the anthocyanin cyanidin. The transcriptome of the transgenics was compared with wild type plants, and no consistent changes in gene expression were determined. This suggests that the changes in biochemistry observed were not regulated at the level of transcription.[34]

2.4 Chemical Inhibition of *Am*GSTF1

Having determined using transgenic *Arabidopsis* that *Am*GSTF1 played a causative role in MHR, a chemical strategy was sought which would restore herbicide activity. Multi-drug resistance (MDR) in humans shows parallels to MHR, as it involves a drug detoxifying GST, *Hs*GSTP1. *Hs*GSTP1 has been a successful target for inhibitors in medicinal chemistry programs.[80] These inhibitors can be subdivided into GSH conjugates and peptidomimetics which bind the G-site and compounds which act on the larger H-site.[81] These simpler H-site compounds were tested for their ability to inhibit *Am*GSTF1 and restore herbicide efficacy. Four compounds were tested: 4-chloro-7-nitrobenzofurazan (CNBF), ethacrynic acid, cyanuric acid and bromoenol lactone (Figure 2-3a). Ethacrynic acid and cyanuric acid were found to be phytotoxic, and bromoenol lactone showed no ability to enhance herbicide activity so these were discontinued from further exploration. However, CNBF was seen both to inhibit *Am*GSTF1 *in vitro* and to enhance the herbicidal activity of chlortoluron, fenoxaprop-p-ethyl and clodinafop-propargyl when pre-applied to Peldon plants (Figure 2-3).
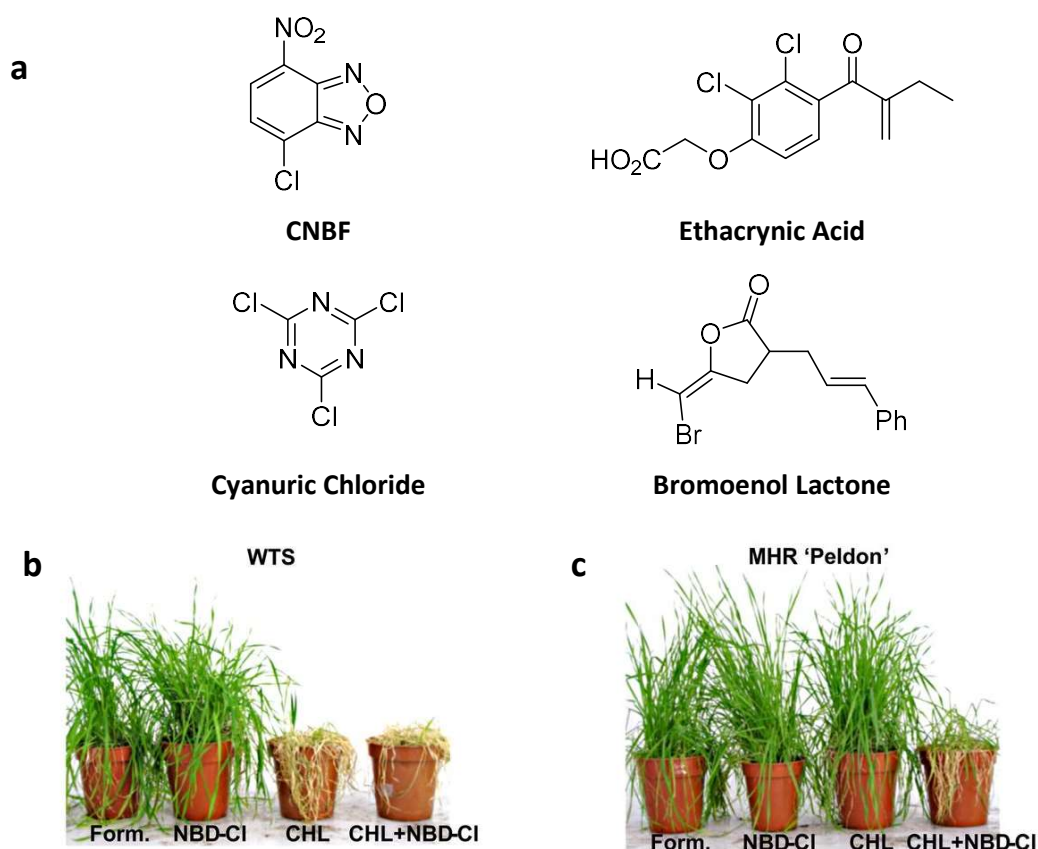
Figure 2-3: (a)GST Inhibitors tested for activity against MHR black grass (b and c) effect of CNBF (written as NBD-Cl) against WT and MHR black grass using Chlortoluron (CHL) as a herbicide. While CNBF alone is seen to have no toxic effect on plants, in combination with Chlortoluron this is seen to lead to plant death.

Work was carried out to identify the mechanism by which CNBF was having an inhibitory effect on *Am*GSTF1. In *Hs*GSTP1 CNBF forms a covalent interaction with Cys47.[82] Cys47 is thought to play a regulatory role, modulating activity with a c-Jun N-terminal kinase (JNK1), which is involved in the apoptotic pathway.[83,84] Mass spectrometry was used to confirm that in *Am*GSTF1 CNBF was also forming a covalent interaction, with it being found to modify the Cys120 residue.[34]

2.5   Identification of novel inhibitors

Attempts to modify the CNBF inhibitor with groups similar to those used in human MDR drugs proved ineffective. While several maintained activity *in vitro* against *Am*GSTF1 no activity was seen against MHR plants, probably due to problems in uptake by the leaves.[34] As a result work focused on identifying a novel class of inhibitors.

A ligand fishing experiment was used to identify possible compounds that bind to *Am*GSTF1.[85] Strep tagged *Am*GSTF1 was immobilised on a streptactin column, and a cocktail of compounds washed over it. Liquid Chromatography-Mass Spectrometry (LC-MS) was then used to identify which compounds had bound to the *Am*GSTF1 (Figure 2-4).
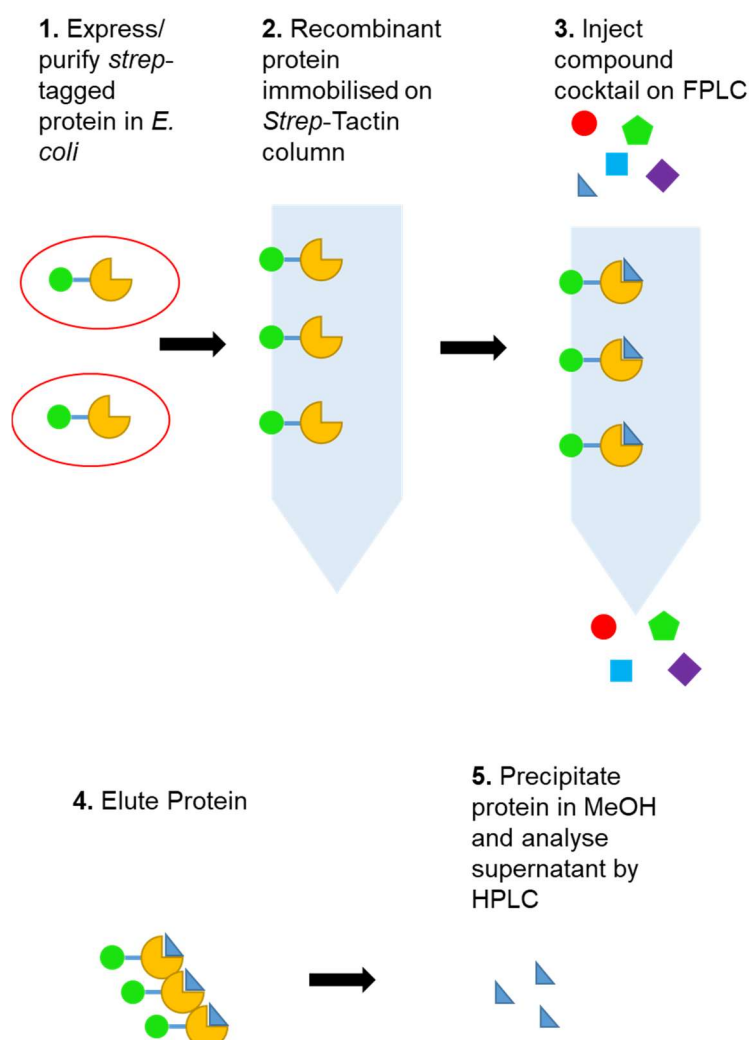
Figure 2-4: Shematic showing ligand fishing method used to identify new inhibitors. *E. coli* expressed strep-tagged *Am*GSTF1 enzyme is immobilised on a strep column. A cocktail of inhibitors are passed over the column and those that bind identified by HPLC.

This experiment identified three main compounds that bound to *Am*GSTF1: the 4-Glutathione conjugate of CNBF; a purine-CNBF hybrid and apigenin. The most significant binding was observed with apigenin. A subsequent ligand fishing experiment with a selection of commercial flavonoids identified that *Am*GSTF1 was capable of binding to a range of flavonoids. These commercial flavonoids along with a selection from the Syngenta compound library were tested for their ability to inhibit *Am*GSTF1 activity in the CDNB assay. This identified compound SYG8,

which gave 100% inhibition at 10 μM, as a new lead compound. *In vivo* tests demonstrated that it was also capable of acting as a synergist, restoring the activity of pinoxaden against MHR black grass.
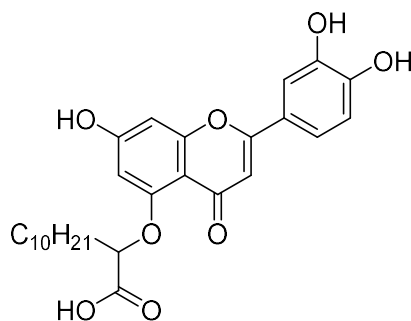
SYG8

Figure 2-5: Initial lead compound identified from screening of Syngenta library

2.6 Flavone Inhibitors

Recent efforts in the Steel group have focussed on the synthesis of derivatives of this lead compound SYG8.[85,86] Work to improve activity afforded compound MS-1-134 with a structure very similar to SYG8 and a similar level of inhibition (71% at 1 μM). The structure activity relationships (SAR) observed from these compounds will be discussed in more detail in (Section 4.3.1.2).
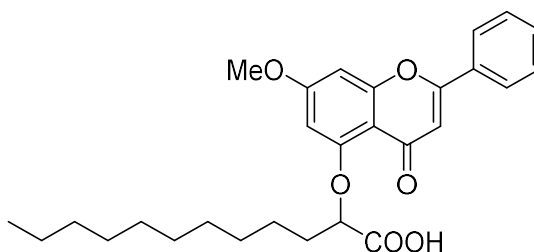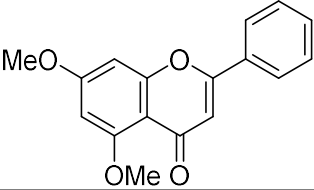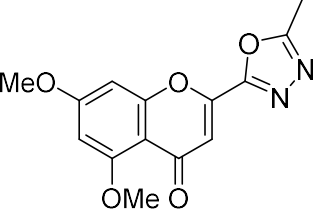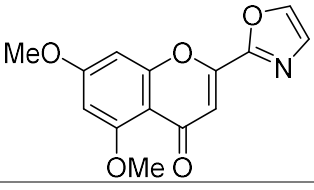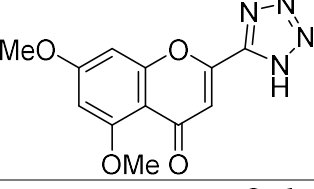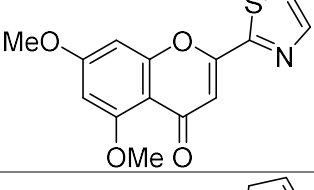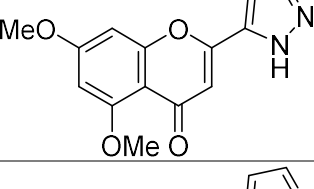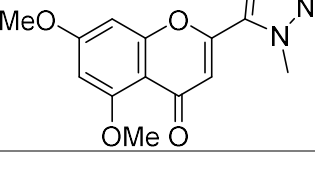
MS-1-134
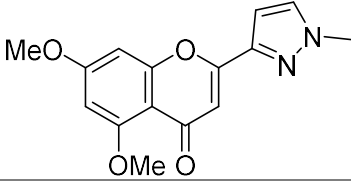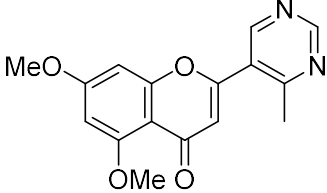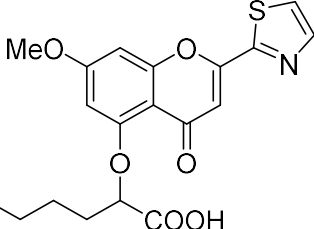
Figure 2-6: Lead compound obtained from initial SAR studies

Further work focussed on increasing the solubility of compound MS-1-134.[86] While the long alkyl chain significantly increases activity it also results in the compound being highly insoluble, which

was thought to be limiting activity *in planta* as a synergist. The effects of adding different groups on solubility and activity were investigated and the results shown in Table 2-1.

Table 2-1: Compounds designed to increase solubility, giving aqueous solubility measured at Syngenta Jealott's Hill, and inhibition against *Am*GSTF1 measured in the CDNB assay

Compound	Aqueous solubility/ ppm	Inhibition at 10 µM
	25.0	20%
	70.1	8%
	88.3	11%
	15.9	13%
	467.7	23%
	16.7	6%
	16.9	9%

Structure		
	21.4	19%
	148.6	8%
	N.D.	95%

A thiazole group was observed to significantly increase solubility, whilst maintaining activity, resulting in this group being incorporated into compounds with the long chain to give compound MS-4-48 as a more soluble lead compound (Figure 2-7). This compound was observed to show activity both *in vitro* against *Am*GSTF1 and *in vivo* acting as a synergist to restore pinoxaden sensitivity in MHR black grass.[86]
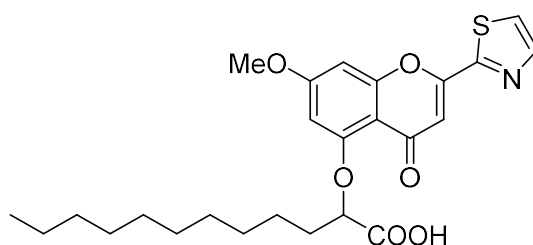
MS-4-48

Figure 2-7: Lead compound produced from work to increase solubility

2.7   Project Aims

The principal aim of this project was to gain a molecular understanding of the interactions and mechanisms of inhibitors for *Am*GSTF1. Work would primarily focus on the flavonoid inhibitors synthesised in the Steel group, with the aim of using this information to guide synthesis decisions. However, CNBF was also investigated to try to understand the mode of action of this inhibitor.

Previously, an apo crystal structure of *Am*GSTF1 had been solved at a resolution of 1.95 Å,[87] although this was incomplete, and with two disordered loops including one in the active site area. This work aimed to increase the resolution and completeness of this structure. In addition, it was aimed to crystallise protein-ligand complexes to identify ligand binding sites and allow for rational inhibitor design. An *in silico* approach would also be used to identify and evaluate possible binding sites for the inhibitors. Furthermore, thermal shift assays would be used to assess the binding of compounds to *Am*GSTF1 and to act as a complementary method for assessing them alongside the previously used CDNB activity assays.

Biophysical methods would be used to improve understanding of the mode of action of inhibitors, and the mechanisms by which they are able to restore herbicide sensitivity to MHR black grass.

Finally, it was aimed to correlate the results of spray tests, biochemical and structural studies, in order to better understand the inhibition of MHR in black grass.

3 Materials and Methods

3.1 Materials

All chemicals were purchased from Sigma-Aldrich unless otherwise specified.

All protein concentrations were calculated using A_{280} on the DeNovix DS-11+ spectrophotometer, using extinction coefficients calculated using ExPASy ProtParam.[88]

Sonication was carried out using the Bandelin Sonopuls HD2200 sonicator.

3.2 Protein Expression and Purification

3.2.1 Transformation of *E. coli*

Plasmid (2 μL) was added to competent *E. coli* cells (New England Biolabs, 50 μL, see 3.2.3 and 3.2.4) and incubated on ice for 30 minutes. Bacteria were heated at 42 °C for 45 secs and immediately incubated on ice for 2 min. SOC medium pre warmed to 37 °C was added and the cells incubated with shaking (1h, 37 °C). 100 μL of culture was spread onto a lysogeny broth (LB) agar plate (inoculated with appropriate antibiotics), the remaining cells spun down and resuspended in LB (1mL) and 100 μL spread on an LB agar plate (with appropriate antibiotics). Plates were grown overnight at 37 °C. One culture was taken from the plate and grown overnight in LB (with appropriate antibiotics). 500 μL of culture was combined with 500 μL 50% glycerol to make glycerol stocks which were stored at -80 °C until needed.

3.2.2 Plasmid Purification

A colony from an LB agar bacterial plate was added to LB inoculated with appropriate antibiotics (see 3.2.3 and 3.2.4). The culture was incubated overnight with agitation at 37 °C. The plasmid was purified from the overnight culture using GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific) according to the manufacturer's instructions.

3.2.3   Strep tagged *Am*GSTF1 and Strep tagged *At*GSTF8

Strep tagged proteins were in pETStrpIII plasmids,[89] with resistance to kanamycin (Kan), and were expressed in *E. coli* Rosetta II cells. For all steps requiring antibiotics, Kan was used at a concentration of 50 µg/ mL and Chloramphenicol (Chlor) at a concentration of 34 µg/ mL.

3.2.3.1   Expression of Strep tagged protein

Buffer A: HEPES free acid (20 mM), NaCl (150 mM) pH 7.6

LB (15 mL, Kan/Chlor) was inoculated with cells from a glycerol stock. The culture was grown overnight (37 °C, 200 rpm). The overnight culture was added to 1 L LB (Kan/ Chlor) and grown at 37 °C with shaking at 200rpm until OD_{600}= 0.6. Cultures were induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG; 100 µM) and incubated at 37 °C with shaking at 200 rpm for 16 h. Cells were collected by centrifugation (4000 g, 30 min), the media decanted off, and pellets resusupended in Buffer A (10 mL/ L *E. coli* expression) with cOmplete ULTRA Tablets, Mini, EDTA free (1 per 15 mL), and dithiothreitol (DTT, 1 mM). Pellets were frozen at -20 °C until needed.

3.2.3.2   Purification of Strep tagged protein

Pellets were removed from the freezer defrosted on ice, and lysed by sonication (3 × 30 sec, 30%) and centrifuged (20,000 rpm, 50 min). Purification was carried out using a 5 mL Strep-tactin sepharose high performance column (GE Healthcare) attached to a peristaltic pump. The column was pre-equilibrated with buffer A. The cell lysate was filtered through a 0.2 µm filter and loaded onto the column at a rate of 1 mL/min. Unbound protein was washed off the column using buffer A. Once all unbound protein was removed from the column according to the A_{280} reading the protein was eluted using desthiobiotin (5 mL, 2.5 mM in buffer A). Protein was dialysed against buffer A overnight, and stored at -80 °C until needed when it was defrosted on ice.

3.2.3.3 *At*GSTF8 Protein

Strep tagged *At*GSTF8 was expressed in *E. coli*, from a plasmid provided by Prof. Robert Edwards (Newcastle) and purified using streptactin resin (Section 3.2.3). This produced pure protein (Figure 3-1), with no subsequent purification step required, however it was seen to express at significantly lower levels than *Am*GSTF1, with around 2 mg of purified protein per litre of recombinant *E. coli*. As only small amounts were required for assays to investigate selectivity no attempts were made to optimise this procedure.
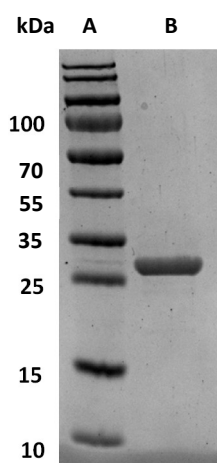
Figure 3-1: SDS-PAGE gel to show the purity of *At*GSTF8. Lane A-Molecular weight marker, B-Strep tagged *At*GSTF8. Expected molecular weight for monomer= 25941 Da

3.2.4 Untagged *Am*GSTF1 and mutants

Genes encoding untagged proteins were in a construct in pET24a plasmid with Kan resistance, and were expressed in BL21 (DE3) cells. Kan was used at a concentration of 50 µg/ mL.

3.2.4.1 Expression of untagged proteins

LB (Kan) was inoculated with cells from a glycerol stock. The culture was grown overnight (37 °C, 200 rpm). The overnight culture was added to 1 L LB (Kan) and grown at 37 °C with shaking at 200rpm until OD_{600}= 0.7. Cultures were induced with IPTG (100 µM) and incubated at 37 °C with shaking at 200rpm for 16 h. Cells were collected by centrifugation (4000g, 30 min), the media decanted off, and pellets resusupended in Buffer A with cOmplete™ ULTRA Tablets, Mini, EDTA free (1 per 15 mL), and DTT (1 mM). Pellets were frozen at -20 °C until needed.

3.2.4.2   Purification of untagged proteins

Buffer A: 20 mM HEPES free acid, 150 mM NaCl pH 7.6

Buffer B: 200 mM KCl, 20 mM TRIS, 1 mM DTT, pH 7.5

Buffer C: 5 mM GSH, 20 mM TRIS, 1mM DTT pH 7.5

Buffer D: 20 mM TRIS, 1 mM DTT, pH 7.5

Buffer E: 20 mM TRIS, 1 mM DTT, 500 mM NaCl, pH 7.5

Cell pellets suspended in Buffer A were lysed by sonication (3×30 sec, 30 % power) and re-pelleted by centrifugation (20,000 rpm, 50 min). Cell lysate was incubated (1 h) with free glutathione agarose beads (3 mL) with agitation. Using a gravity column, Buffer B was run until protein concentration was less than 0.1 mg/mL. Protein was eluted using 10 mL of buffer C, purity of fractions was confirmed by SDS-PAGE before they were pooled and dialysed against buffer D overnight. Using an Akta Explorer FPLC the protein was purified using a MonoQ column. The column was equilibrated with buffer E followed by buffer D. The sample was loaded onto the column and a gradient of 1-100 % buffer E run over 30 mins. The protein was collected in 1 mL fractions.

3.3   Cloning of mutants

Mutants (Y118S, F122T and Y118S F122T) were created from the wild type construct using PCR and the Q5® Site-Directed Mutagenesis Kit (New England BioLabs), using primers in Table 3-1 and the appropriate annealing temperatures (T_a). The Y118S F122T double mutant was created from the F122T mutant using the Y118S primers.

Table 3-1: Primers used for cloning of mutant *Am*GSTF1

Mutant	Forward Primer	Reverse Primer	T_a/ °C
Y118S	CCGATCGTGTCTCAGTGTCTG	GCTCAGTGCCGGATTATAG	62
F122T	TCAGTGTCTGACCAACCCGATGATGC	TACACGATCGGGCTCAGT	64

3.4   Labelling of *Am*GSTF1 with CNBF

DTT (5 mM) was added to *Am*GSTF1 (1 mg/mL), incubated on ice for 30 min, and desalted using a PD10 column (GE Healthcare). To 100 μL aliquots was added 11 μL of 1 mM CNBF in DMSO (100 μM final concentration). Reaction mixtures were incubated at room temperature (rt) for 20 min prior to a second desalting step (as above). Protein samples were analysed using Quad time of flight electrospray ionization mass spectrometry (QToF ESI-MS), for further details see section 3.9.

3.5   Thermal Shift Assays

3.5.1   Solvent Screens

To protein (0.5 mg/mL in 20mM HEPES, 150 mM NaCl pH 7.6) was added 20 × SYPRO orange® (4 μL/ mL protein). 10 μL of protein with dye was added to each well of a 96 well polymerase chain reaction (PCR) plate. 10 μL of solvent at varying concentrations was also added to the well. The plate was sealed with a PCR seal and centrifuged (1000 rpm, 2 min). Experiments were performed as described in 3.5.3.

3.5.2   Ligand binding

To protein (0.5 mg/mL) was added 20 × SYPRO orange (4 μL/ mL protein). 10 μL of protein (in 20mM HEPES, 150 mM NaCl pH 7.6) was added to each well of a 96 well PCR plate. 10 μL of ligand at varying concentrations (in 20 mM HEPES, 150 mM NaCl, pH 7.6 and 2% DMSO) was also added to the well. The plate was sealed with a PCR seal and centrifuged (1000 rpm, 2 min). Experiments were perfomed as described in 3.5.3.

3.5.3   Performing Thermal Shift Assays

Experiments were performed using the Applied Biosystems 7500 Fast Real-Time PCR (RT-PCR) system using Filter C with an excitation of 470 nm and an emission of 569 nm. Samples were run from 24-96 °C with a temperature increase of 1 °C/ min. Once samples had been run results were analysed using NAMI software.[90]

3.6   CDNB Activity Assay

3.6.1.1   CDNB Activity Assay

It is possible to investigate the conjugative activity of GSTs using the 1-chloro-2,4-dinitrobenzene (CDNB) assay. This is a colorimetric assay that follows the formation of dinitrophenyl thioether (GS-DNB) at 340 nm, as GS-DNB absorbs at this wavelength while CDNB does not.[91,92]
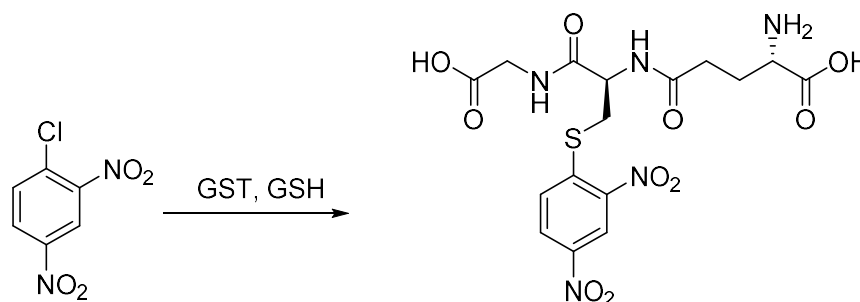
Figure 3-2: CDNB activity assay

0.1M potassium phosphate buffer (pH 6.5, 890 μL) was pre-incubated at 30 °C. 25 μL *Am*GSTF1 (1 mg/mL in HEPES buffer) was added followed by CDNB (25 μL, 40 mM in ethanol) and inhibitor (10 μL in DMSO) and a blank recorded. 50 μL GSH (100 mM in water) was added and the change in absorbance at 340 nm monitored over 30 s using a Beckman Coulter DU-530 spectrophotometer.

3.7   Protein Gels

3.7.1   SDS-PAGE gels

Sodium dodecyl sulfate- polyacrylamide gel electrophoresis (SDS-PAGE) gels were run using the Bio-Rad Mini-PROTEAN® Tetra Cell system. Gels were prepared using 12% resolving gel solution and 8% stacking gel solution. Samples were prepared by the addition of 2 × loading buffer to 10 μL of sample and heating to 95°C for 5 minutes. The gel was loaded with 10μL of Spectra™ Multicolor Broad Range Protein Ladder (Thermo Scientific) into a well and 10μL of sample into the remaining wells. The gel was run at 200V until the lowest molecular weight marker of the

ladder was 1cm from the end of the gel. The gel was removed and incubated overnight with agitation in a tray with Quick Coomassie staining solution (Generon).

3.7.2   Native PAGE gels

Mini-Protean TGX Precast Gels with 4-20% resolving gel (Bio-Rad) were run using the Bio-Rad system. Samples were made up with protein (5 μL, 1 mg/ mL), Buffer A (see 3.2.4.2, 4.5 μL) and inhibitor/DMSO (0.5 μL) and 10 μL native sample buffer (Bio-Rad). 10 μL of sample was added to each well. NativeMark™ (Thermo Fisher Scientific) was used as a molecular weight ladder. Gels were run at 200 V for 35 min. Gels were removed and stained with Quick Coomassie staining solution (Generon).

3.8   Fluorescence Anisotropy

3.8.1   Labelling of *Am*GSTF1 with HiLyte Fluor™ 488

*Am*GSTF1 (1 mg/mL in PBS buffer) was reduced with a 10-fold molar excess of tris(2-carboxyethyl)phosphine (TCEP) and incubated at rt for 2 h with 400 μM HiLyte Fluor™ 488 C2 Maleimide (AnaSpec). Unreacted dye was separated by passing through two PD10 columns (GE healthcare). Fractions containing labelled protein were analysed by absorbance at 280 nm. The contribution of HiLyte Fluor™ 488 to the absorption was estimated by the absorption value at 507 nm multiplied by 0.2 subtracted from the total absorbance at 280 nm, and protein concentration calculated using the extinction coefficient of the protein (42525 M^{-1} cm^{-1}).

3.8.2   Fluorescence anisotropy measurements

100 μL of the HiLyte Fluor™ 488 labelled protein in PBS buffer was placed in a standard black 96 well plate. 100 μL of inhibitor at various concentrations (in PBS buffer +4% DMSO) was added to each well. Fluorescence anisotropy measurements were performed on a BioTek Synergy™ H4 Hybrid Microplate Reader. Fluorescence anisotropy was measured at λ_{ex}=485 nm ± 20 nm and λ_{em}= 528 nm ± 20 nm.

3.9 Mass Spectrometry

For mass spectrometry, protein samples were buffer exchanged into water. Mass spectrometry was carried out on a QtoF Premier mass spectrometer with Acquity UPLC (Waters Ltd, UK).

3.10 Crystallisation and structure determination

3.10.1 Crystal structure analysis

Unless specified otherwise, all figures were created using PyMol.

All root-mean-square-deviation (rmsd) values were calculated using RAPIDO.[93,94]

3.10.2 Crystallisation and structure determination of wild type *Am*GSTF1 (hexagonal and CNBF labelled)

Untagged wild type *Am*GSTF1 was used for crystallographic studies. For the CNBF structure *Am*GSTF1 (37 µM) was labelled as previously described (3.4). Crystals were obtained across a range of conditions containing sodium citrate, Na/K tartrate and ammonium sulfate (Table 3-2). Crystals were cryoprotected in 25% glycerol and frozen in liquid nitrogen. Data was collected on beamline Diamond light source (DLS i04-1). Data was processed using Xia2[95] and scaled using AIMLESS/POINTLESS.[96,97] The structure was solved using Phaser[98] by molecular replacement against a lower resolution *Am*GSTF1 structure for the wild type, and against the higher resolution structure for the CNBF labelled form. Models were built in CCP4[99] using Coot[100], and refined using Refmac.[101] Ligand restraints were generated using JLigand.[102] Further experimental detail is summarised in Appendix B.

Table 3-2: Crystallisation conditions used for crystallisation of hexagonal wild type and CNBF labelled *Am*GSTF1

Well Number	pH of 0.1 M Sodium Citrate buffer	Concentration of K/ Na Tartrate/ M	Concentration of Ammonium sulfate/ M
1	5.0	0.3	2.0
2	5.5	0.3	2.0
3	6.0	0.3	2.0
4	4.5	0.2	2.0
5	5.0	0.2	2.0
6	5.5	0.2	2.0
7	6.0	0.2	2.0
8	3.5	0.1	2.0

Well Number	pH of 0.1 M Sodium Citrate buffer	Concentration of K/ Na Tartrate/ M	Concentration of Ammonium sulfate/ M
9	4.0	0.1	2.0
10	4.5	0.1	2.0
11	5.0	0.1	2.0
12	5.5	0.1	2.0
13	6.0	0.1	2.0
14	3.5	0	2.0
15	4.0	0	2.0
16	4.5	0	2.0
17	5.0	0	2.0
18	5.5	0	2.0
19	6.0	0	2.0
20	5.0	0.3	1.0
21	5.5	0.3	1.0
22	6.0	0.3	1.0
23	4.5	0.2	1.0
24	5.0	0.2	1.0

3.10.3 Crystallisation and structure determination of F122T *Am*GSTF1

Protein was crystallised using the sitting drop method in the Morpheus screen (Molecular Dimensions),[103] with crystals produced in a number of wells. Crystals were cryoprotected in 25% glycerol and frozen in liquid nitrogen. Data was collected at Diamond Light source on beamline i24. Crystals were screened and full datasets collected for those showing best diffraction. The data was processed using Xia2, and the structure solved using molecular replacement using the wild type (hexagonal) structure, and the previously disordered loops built using Buccaneer.[104] Models were built in CCP4 using Coot and refined using Refmac. Ligand restraints were generated using AceDRG.[105] Further experimental detail is summarised in Appendix B.

3.10.4 Crystallisation and structure determination of Y118S *Am*GSTF1

Protein was crystallised using the sitting drop method in the Morpheus screen,[103] with crystals produced in a number of wells. Crystals were cryoprotected in 25% glycerol and frozen in liquid nitrogen. Data was collected at Diamond Light source on beamline i04-1. The data was processed using Xia2, and the structure solved using molecular replacement using the F122T structure. Models were built in CCP4 using Coot and refined using Refmac. Ligand restraints were generated using AceDRG.[105] Further experimental detail is summarised in Appendix B.

3.10.5 Crystallisation and structure determination of wild type *Am*GSTF1 (tetragonal) and NBF-GS structure

Protein was crystallised using the sitting drop method in the Morpheus screen, using seeds of F122T protein. Seeds were created using a Hampton Seed Bead™ (Hampton Research) and associated protocol for matrix microseeding. For ligand bound structures solid NBF-GS was added directly to drops and left for 24 hours. Crystals were cryoprotected in 25% glycerol and frozen in liquid nitrogen. Data was collected at Diamond Light source on beamline i04-1. The data was processed using Xia2, and the structure solved using molecular replacement using the F122T structure. Models were built in CCP4 using Coot and refined using Refmac. Ligand restraints were generated using AceDRG.[105] Further experimental detail is summarised in Appendix B.

3.11 Molecular modelling

3.11.1 Using Chimeric GSTs (Homology models)

Models were made using Modweb from MODELLER,[106] based on 1AXD[56] and 1AW9.[56] For the 1AXD based homology model residues 30-55 and 119-139 were inserted into the black grass structure. For the 1AW9 based homology model residues 34-53 and 121-141 were inserted into the black grass structure. Geometry minimisation was carried out using Phenix. The models were used with H-atoms added, and with no waters in the binding site. Molecular modelling was carried out using GOLD[107] with ChemPLP as a fitness function. A search space of 10 Å around Ser12 was defined. 10 runs were carried out for each ligand, with an early termination setting of 3 ligands within an rmsd of 1.5 Å used.

3.11.2 Using F122T structure

The structure of the F122T mutant was used with the threonine mutant at the 122 position replaced with a phenylalanine, using Coot, to create a model of the native enzyme. The model was used with H-atoms added, and with no waters in the binding site. Molecular modelling was carried out using GOLD with ChemPLP as a fitness function. A search space of 10 Å around

Phe122 was defined. 10 runs were carried out for each ligand, with an early termination setting of 3 ligands within an rmsd of 1.5 Å used.

3.12 Synthesis of NBF-GS
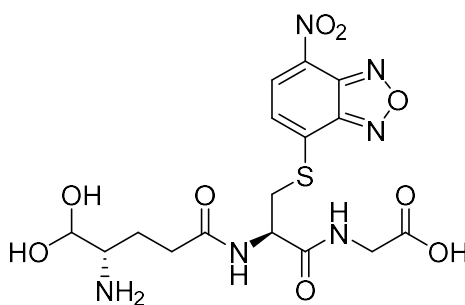
CNBF (0.5 mmol), GSH (0.5 mmol) and K_2CO_3 (0.5 mmol) were suspended in H_2O. Acetonitrile (MeCN) was slowly added, with stirring initially until all compounds were dissolved, and then in excess until a green precipitate was formed. The precipitate was separated by filtration and confirmed as NBF-GS by ESI LCMS m/z 472.2 at 97% purity.

4 Biophysical Investigations of *Am*GSTF1 with Inhibitors

In order to better understand the mode of action of both the flavonoid inhibitors identified in ligand fishing experiments, and CNBF inhibitors, a series of biophysical investigations were carried out. For the flavonoid compounds their structure-activity relationship (SAR) was investigated using binding assays, and further experiments were carried out to identify the selectivity and potential mode of action of these inhibitors.

4.1 *Am*GSTF1 Strep Production

For biophysical studies, a strep tagged construct of *Am*GSTF1 was used. The protein was over-expressed in *E. coli* and purified using a Streptactin column to produce protein of sufficient purity (Figure 4-1).[89] Using this method it was possible to produce around 20-25 mg or pure protein per litre of recombinant *E. coli*. Protein activity was confirmed using the CDNB assay. Freezing the protein at -80 °C did not significantly affect the activity, and as a result protein that had been frozen and then thawed was used for convenience for all assays.
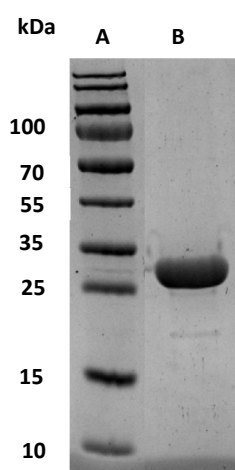
Figure 4-1: SDS-PAGE to demonstrate purity of strep tagged *Am*GSTF1. Lane A- Molecular weight marker, B- Strep tagged *Am*GSTF1. Expected molecular weight of monomer=26665 Da.

4.2   Introduction to Thermal Shift Assays

Thermal shift assays (TSA) were used to investigate compound binding. TSA uses a fluorescent dye, in this case SYPRO orange®, to determine the folding state of the protein. In aqueous solutions the fluorescence signal is quenched, however the molecule is highly fluorescent in a non-polar environment, in this case the hydrophobic sites exposed in unfolded proteins.[108] As temperature is increased, a protein will denature exposing its hydrophobic sites. The SYPRO orange® dye will bind to these sites and result in an increase in fluorescence. The unfolding of the protein can thus be followed using an RT-PCR machine (**Error! Reference source not found.**).
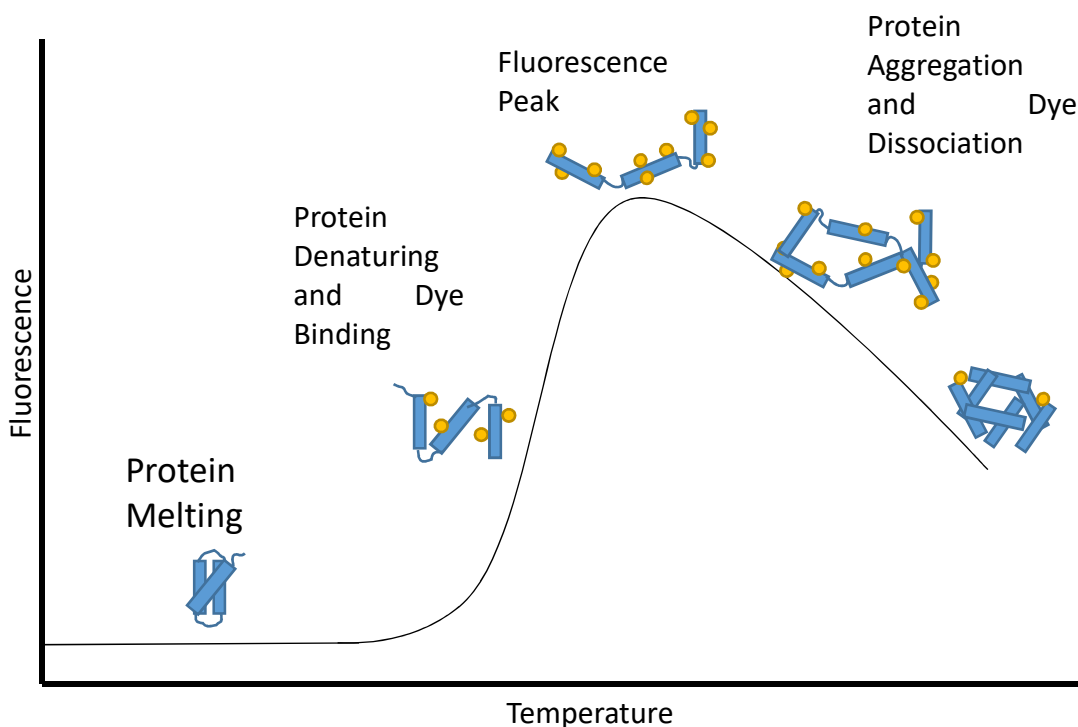
Figure 4-2: Schematic to show fluorescent probe binding in thermal shift assays, and associated curve generated. Yellow dots represent fluorescent dye binding to hydrophobic regions of the protein.

This method can be used to determine the effect of various factors on the thermal stability of a protein.[90] These can be either changes in buffer composition, where interactions are non-specific, or ligands which bind at a specific site. In the event that a ligand binds this will normally lead to a stabilisation of the folded state and an increase in melting temperature (T_m). However, occasionally destabilisation of the protein has been observed, and an associated decrease in T_m.

(**Error! Reference source not found.**) For the case of specifically binding ligands the extent of stabilisation by a ligand has been shown to correlate to the binding affinity of ligands to the protein.[109–111]
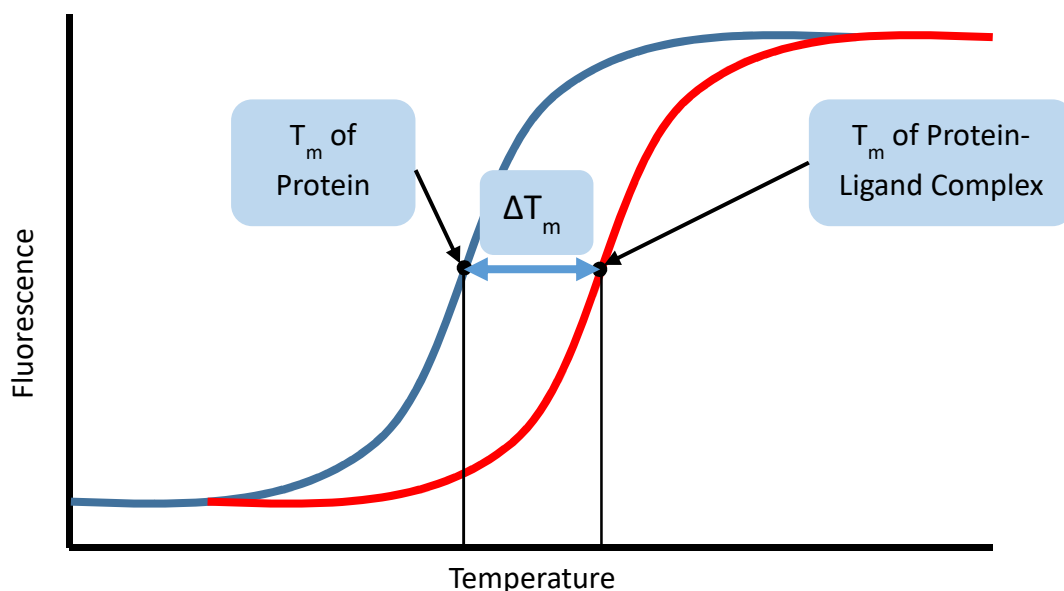
Figure 4-3: Schematic to show the effect of ligand binding on a thermal shift assay, showing a typical melting curve for protein in the absence of ligand (blue) and protein with a ligand that binds (red)

TSA allows for easy screening of proteins in a high throughput manner, as small amounts of material are needed in low concentration, and experiments can be carried out using a 96 well plate, allowing large numbers of inhibitors to be investigated in parallel. As a result this method was selected for screening of flavonoid inhibitors to identify those with highest affinity.

4.3   Flavonoid Inhibitors

4.3.1   Thermal Shift Assays

As discussed in section 2.5 a series of flavonoids had been identified as modulators of the MHR response. In studies to date the SAR of the compounds had been investigated using the CDNB assay.[85,86] As the conjugative activity of *Am*GSTF1 is low, it was a concern that this assay, which measures the inhibition of GST catalysis of GSH conjugation might not be the best way to assess compound binding. As a consequence it was decided to use thermal shift assays to directly

investigate the relative binding affinities of the flavonoids and to determine if these correlated with inhibition results from the CDNB assay.

4.3.1.1   Solvent screens

As the flavonoid inhibitors show limited aqueous solubility, it was necessary for them to be dissolved in an organic solvent for use in the assays. As a result, the stability of *Am*GSTF1 in a range of different solvents was investigated. Four water miscible organic solvents, acetonitrile, DMSO, methanol and ethanol were selected for screening. Solvents were tested at a range of final concentrations from 0.05% to 50% with the remainder of the sample composed of protein in buffer (Figure 4-4).
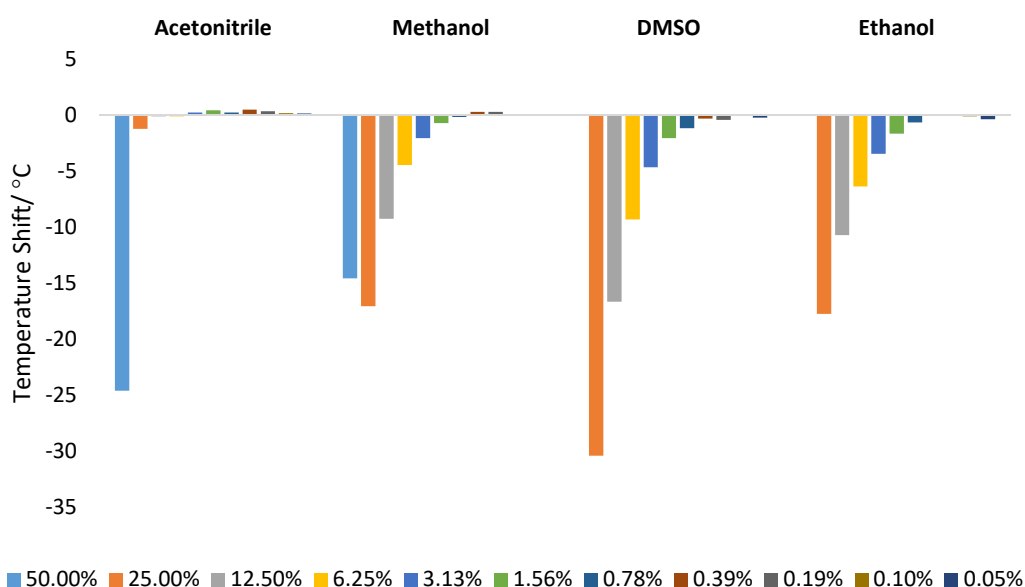
Figure 4-4: Temperature change observed on addition of various concentrations of solvent to *Am*GSTF1 in a thermal shift assay. For both DMSO and ethanol a value at 50% solvent concentration is excluded as the protein was seen to be denatured at room temperature. Results are the average of three technical replicates.

*Am*GSTF1 was seen to be most stable in acetonitrile, where only a small change in stability was detected at 25%. For the remaining solvents, protein stability was seen to decrease significantly with increasing concentration of solvent, with the greatest loss of stability observed with DMSO. However, all the solvents proved to be viable at low concentrations. Solvent selection is a compromise between protein stability and compound solubility. Although acetonitrile gave the highest degree of protein stability, the solubility of the flavonoids in this solvent proved to be an

insurmountable problem. Ultimately, as the difference in stability effect between DMSO and ethanol or methanol was minimal, and the inhibitors proved to be significantly more soluble in DMSO, this was selected for use in assays. DMSO was used at a final concentration of 1% to provide a balance between inhibitor solubility and protein stability.

4.3.1.2   Structure-activity relationship investigations

Synthetic work has focussed on modifying the core flavonoid inhibitor from an initial hit pulled from the Syngenta library by modification in a range of positions on the structure. Work focussed predominantly on investigating the effect of groups at the C5 and C7 positions, and altering the B ring (Figure 4-5).[85,86]

With thermal shift assay conditions established each flavonoid inhibitor synthesised was probed to determine relative binding affinities to *Am*GSTF1. Results from thermal shift assays were then compared with data obtained from inhibition assays from the CDNB assay, performed by Maria Schwarz, to investigate SAR.[86] In general, the results from the CDNB assays and thermal shift assays were seen to agree, with those compounds found to inhibit the enzyme most effectively also seen to show the greatest stabilisation on binding. For weaker binders, the CDNB assay was able to provide more sensitivity. While no significant change in stability could be observed using thermal shift due to the resolution of the technique, it was possible to identify a significant difference using CDNB. As the effect of a number of modifications were investigated on only a core compound, lacking an α-alkoxy carboxylate at the 5 position, this allowed these modifications to be investigated more thoroughly. For each substitution results to demonstrate the trends are shown in the section. A full table of results for all flavonoid compounds generated is available in Appendix A.
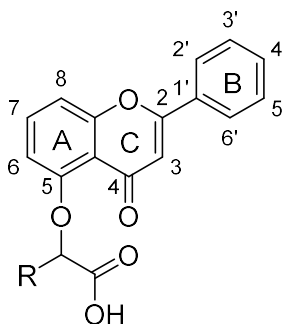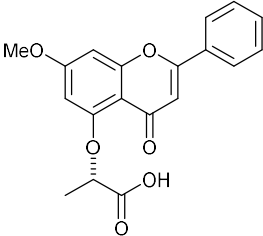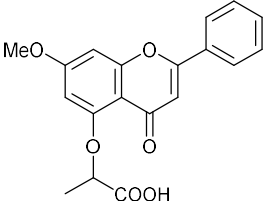
Figure 4-5: General structure of Flavonoid Inhibitors showing ring numbering and naming conventions

Substitution at the C5 Position

Table 4-1: Biophysical results obtained for a selection of flavonoid compounds. Selected to demonstrate the trends observed for substitution at the C5 position.

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-1-123		0.2	0	0.2	N.D.	25	98
MS-3-13		0.2	0.2	1.1	N.D.	14	N.D.
MS-3-21		0.2	0.4	2.3	4	69	N.D.
MS-1-134		0.2	0.5	1.2	71	99	N.D.
MS-5-34		1.2	1.3	2.1	N.D.	13	93
MS-1-139		0.1	0.3	1.1	6	39	N.D.
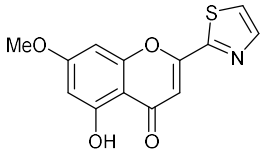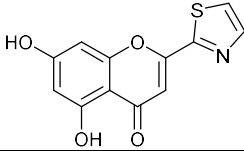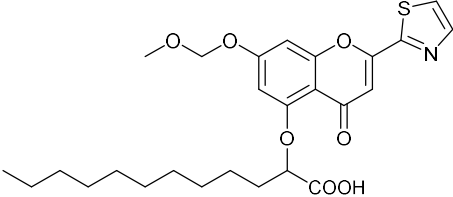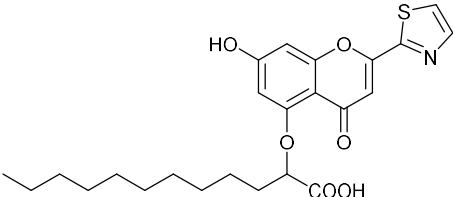
MS-3-12		0.3	0.2	0.5	N.D.	5	N.D.
MS-3-13		0.2	0.2	1.1	N.D.	14	N.D.

It was observed that the addition of an α-alkoxy carboxylate group at the C5 position resulted in a large increase in activity. This effect was increased with the addition of a long alkyl chain in place of the R group, with activity increasing with increased chain length up to 10 carbons where the effect was seen to plateau.

Compounds were synthesised where this alkyl group was replaced with a polyethyleneglycol (PEG) chain, or an amide bond inserted, to increase solubility, however these resulted in a significant loss of activity, suggesting that this tail is situated in a hydrophobic region of the protein.
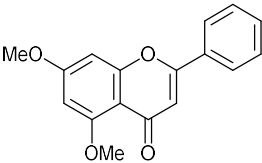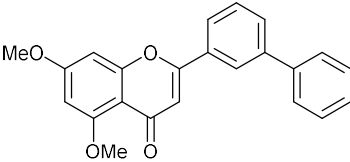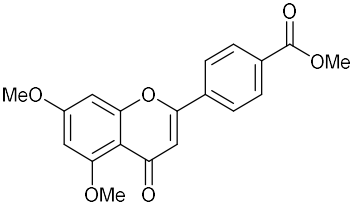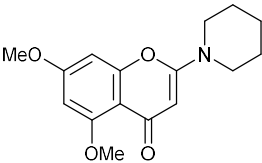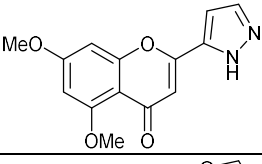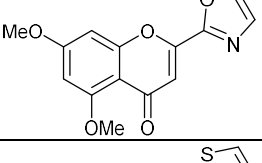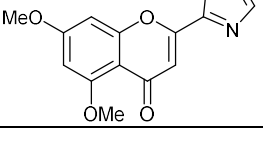
The effect of stereochemistry on the α-alkoxy carboxylate group was investigated. While both isomers were seen to be active, the S-isomer was seen to bind less well and show marginally lower activity than the racemic mixture, in compounds MS3-13 and MS3-12. For compound MS1-134 one isomer is also seen to have higher activity, indicating a possible slight preference for one isomer in the binding site.

Substitution at the C7 Position

Table 4-2: Biophysical results obtained for a selection of flavonoid compounds. Selected to demonstrate the trends observed for substitution at the C7 position.

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-4-30-2		0.4	0.4	0.1	N.D.	23	61
MS-4-50		0.3	0.4	0.9	N.D.	26	89
MS-5-6		0.7	1.6	3.8	67	98	N.D.
MS-5-7		1.3	1.7	4.2	72	100	N.D.

The majority of compounds were synthesised with either a hydroxyl group, or a methoxy group at the C7 position. Those with a hydroxyl group were observed to show better binding and better inhibition, however the methoxy variants were still seen to have good activity. Addition of a long PEG chain at this position was also observed to result in only a small loss of activity, suggesting that there is room for this group to extend into a large binding site. Previous work carried out by Hannah Straker demonstrates that the absence of an oxygen group at this position reduces activity,[85] suggesting that these groups may be involved in hydrogen bonding. Methoxy and PEG groups are able to act exclusively as hydrogen bond acceptors,[112] and in this case it would be expected that the methoxy group would be a stronger acceptor than the hydroxyl group as a result of the electron donating methyl group. It is possible that steric interactions may result in the methoxy group showing lower activity than the hydroxyl group in spite of this theoretically stronger hydrogen bond.

B Ring

Table 4-3: Biophysical results obtained for a selection of flavonoid compounds. Selected to demonstrate the trends observed for substitution of the B Ring.

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-1-51		0.2	0.3	0.3	N.D.	20	70
MS-1-53-6		0.1	0.1	-0.1	N.D.	12	87
MS-1-66-D		0.3	0.3	0.3	N.D.	20	58
MS-1-19		0.2	0.1	0.3	N.D.	12	67
MS-3-132A		-0.1	-0.1	0	N.D.	6	59
MS-3-140		-0.1	-0.2	-0.2	N.D.	11	64
MS-4-30-1		-0.4	-0.5	-0.5	N.D.	23	83

The effect of placing different functional groups on the B-ring was investigated. A range of different groups were used, with a range of properties. Hydroxyl and sulfonyl groups were added to investigate the effect of addition of hydrogen bonding groups, with no major effect to binding being observed in any case. Interestingly, addition of a large cyclic group as a substituent at the

4' position, while slightly reducing activity, did not result in total loss of activity, suggesting that the B-ring sits in a large cavity or towards the edge of the protein, allowing space for larger groups.

Secondly, in an attempt to increase solubility a number of different heterocycles were inserted in place of the B-ring. It was observed that, in all cases, insertion of a non-aromatic ring into the B-ring position resulted in a major loss of activity and binding, suggesting that this group may be involved in π-π interactions. Replacement of the B-ring with a non-cyclic functional group such as a sulfonyl or chloride group was also observed to result in loss of activity.

A range of aromatic heterocycles were added, with the thiazole showing marginally the best activity over the other rings used (oxadiazole, oxazole, tetrazole, pyrazole, methylpyrazole and methylpyridazine). As the thiazole ring also resulted in the largest increase in compound solubility, this ring was used in place of the phenyl B-ring for a further series of inhibitors.

4.3.2   Selectivity against *At*GSTF8

Ultimately it was aimed that flavonoid inhibitors would be used alongside herbicides for treatment of crops, therefore it was of interest to determine whether the compounds were acting as general or selective GST inhibitors. To determine this, their binding and inhibition effects were measured against another GST. *At*GSTF8, a protein from *Arabidopsis thaliana,* was selected as another GST which was known to bind flavonoids, and had good activity in the CDNB assay.

4.3.2.1   Selectivity results

Using a range of flavonoid inhibitors selectivity was initially assessed using the thermal shift assay to investigate binding. Each compound was observed to bind to, and inhibit both *Am*GSTF1 and *At*GSTF8 (Figure 4-6).

In general, the flavonoid compounds were seen to stabilise *At*GSTF8 to a much larger extent than *Am*GSTF1. This can be attributed to the lower T_m for *At*GSTF8 compared to *Am*GSTF1 (36.2 °C as opposed to 62.8 °C).

a

Compound	R=	Compound	R=
MS-3-13	CH_3	MS-1-134	$C_{10}H_{21}$
MS-3-21	C_5H_{11}	MS-1-139	
MS-3-22	C_7H_{15}		

b
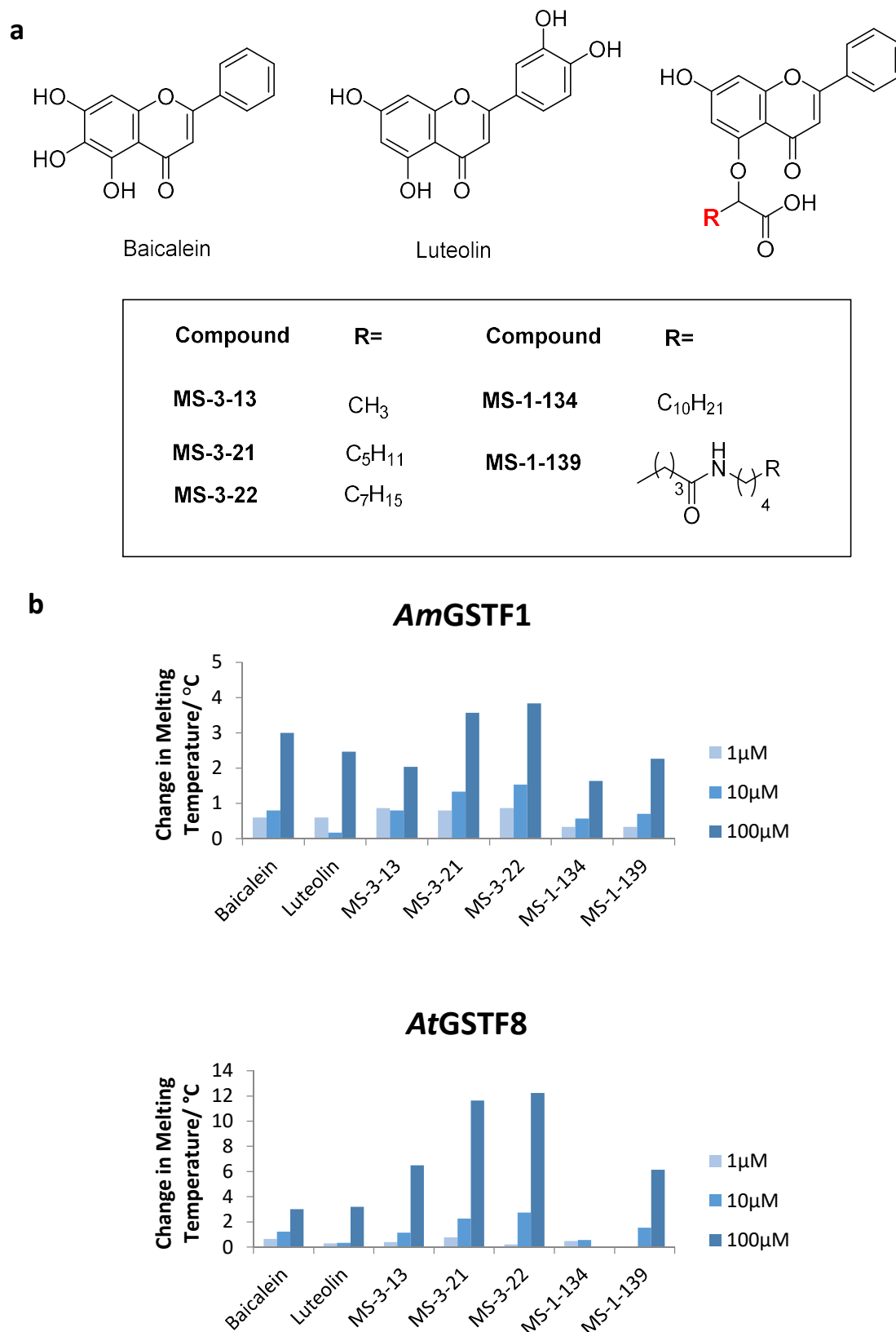
Figure 4-6: (a) Compounds used for selectivity studies. (b) Change in melting temperature observed for various compounds at 1, 10 and 100 μM using the thermal shift assay with *Am*GSTF1 and *At*GSTF8, to investigate selectivity. Results are the average of three technical replicates.

The largest differences were observed in the nature of the C5-alkylcarboxylate chain. Whilst short alkyl chains were seen to stabilise *At*GSTF8 well, and gave good inhibition, it was observed that longer alkyl chains showed a reduced effect, with MS-1-134 seen to have very minimal binding. In contrast, although showing lower changes in T_m this compound retained significant binding efficacy to *Am*GSTF1. Again, good correlation with results from the CDNB assay was obtained.[86] This further highlights the importance of this long chain, and identifies it as an area that could be further targeted for increased selectivity.

4.3.3 Investigation of Dimer Dissociation

As the long alkyl chain on the flavonoid inhibitors had been identified to have a significant effect, both in terms of binding and inhibition of enzymatic activity, and also for giving selectivity, it was of interest to identify what possible role this may be playing. *Am*GSTF1 has a large, hydrophobic dimer interface. It has a surface area of 1390.1 Å, and ePISA predicts a negative $\Delta^i G$ of -19.4 kcal/ mol indicating a highly hydrophobic surface.[113] It was hypothesised that the long alkyl chain could be binding here and disrupting dimer formation. As GSTs have been seen to be active only in their dimeric form,[38,63] this would then result in loss of activity, and it was considered possible that this was the mechanism by which these inhibitors were having an effect. This possible dimer dissociation was investigated using Native-PAGE and fluorescence anisotropy.

4.3.3.1 Native-PAGE gel

The simplest way to identify whether proteins exist in monomeric or dimeric form is using gel electrophoresis. As a result this was initially used to investigate whether the long chain flavonoid was causing dissociation of the *Am*GSTF1 dimer. In *Am*GSTF1 the dimer interface is formed by a network of hydrogen bonds and hydrophobic interfaces therefore it was not possible to use SDS-PAGE. While SDS-PAGE can be used to separate dimers held together by disulphide bridges, by using non-reducing conditions, for dimers such as that in *Am*GSTF1 the heat denaturation alone would disrupt the dimer. Unlike SDS-PAGE which uses SDS, heat and reducing agents to denature and reduce proteins, native PAGE allows for analysis in their folded, oligomeric states.
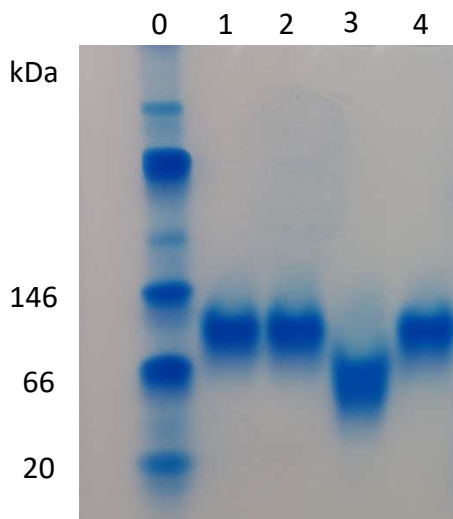
Figure 4-7: Native PAGE for *Am*GSTF1 incubated with MS-1-134 in various concentrations. Lane 0-Molecular weight marker, 1-*Am*GSTF1, 2-*Am*GSTF1+ 10% DMSO, 3- *Am*GSTF1 + 500 µM MS-1-134, 4-*Am*GSTF1 + 10 µM MS-1-134. Expected molecular weights 26665 Da (monomer), 53330 Da (dimer).

*Am*GSTF1 and *Am*GSTF1 incubated with MS-1-134 were run on a native gel in non-denaturing conditions. Preliminary experiments used a single percentage gel, however, it was not possible to get separation of low weight molecular weight markers, and as a result gradient gels were used as these achieved good separation (**Error! Reference source not found.**). *Am*GSTF1 in the absence of inhibitor was seen to run as a single band. Based on the molecular weight markers, this appeared to be higher than the expected weight of 52 kDa. In native PAGE, the isoelectric point of the protein and pH of the sample have a major effect on the separation of proteins in addition to molecular weight. Consequently, the differences observed were not a cause for concern. For protein incubated with low concentrations of MS-1-134, no difference in molecular weights could be seen. However, for *Am*GSTF1 incubated with 500 µM of MS-1-134, the protein was seen to run lower on the gel. This suggested that it was possible the protein had dissociated into its monomeric form. As it was possible that the compound may have affected the pH and therefore migration of the sample, and as no concentration dependent effect could be observed with differing inhibitor concentrations, results were not conclusive. Consequently, further investigations were carried out with complementary methods.

4.3.3.2   Fluorescence Anisotropy

To confirm the results from native PAGE, fluorescence anisotropy was used to identify if flavonoid inhibitors with long alkyl chains were resulting in dimer dissociation. Fluorescence anisotropy has previously been used to investigate the monomer-dimer equilibrium of

*Hs*GSTP1,[114] and the investigation was based on this work. A fluorescent dye, HiLyte Fluor 488 was used, which contained a reactive maleimide group, known to efficiently label cysteine residues within the protein.[115]

Fluorescence anisotropy uses polarised light to investigate the rotational diffusion of a macromolecule. Upon excitation with fluorescent light, a fluorophore will also be polarised. Anisotropy describes the extent of this polarisation, and the angular displacement that occurs between absorption and emission, caused by rotational diffusion.[116] For a small molecule, for example an unbound fluorophore, with a high tumbling state, the rotational diffusion is very high, giving an anisotropy of near zero. The maximum anisotropy that could theoretically be achieved is 0.4, which would be observed where no depolarisation occurs. As the extent of anisotropy is relative to size of molecule, it would be expected that the anisotropy of the monomer would be lower than that of the dimer. Fluorescence anisotropy was measured for two concentrations of protein (20 and 100 nM) and four different inhibitor concentrations (1, 10, 100 and 200 μM).
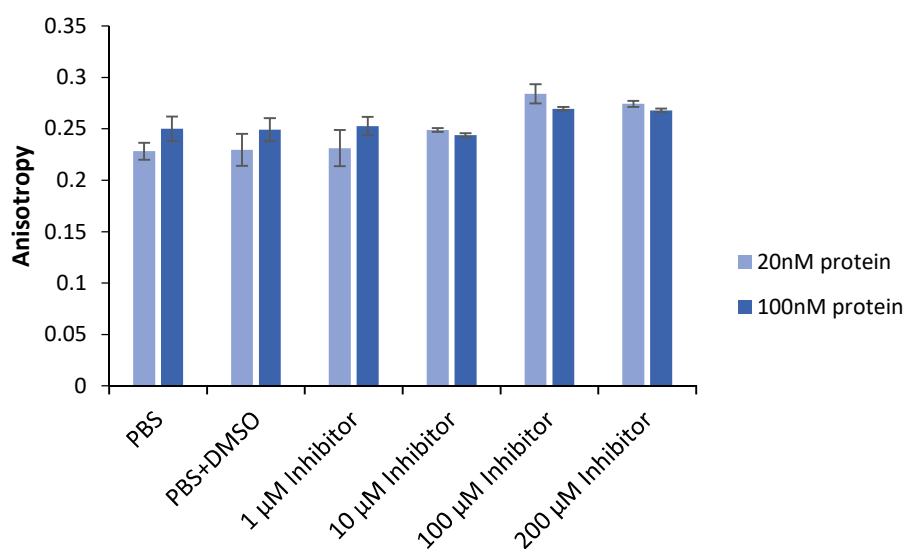
Figure 4-8: Steady state anisotropy for Hi-Lyte Fluor 488 labelled *Am*GSTF1 incubated with various concentrations of inhibitor. Controls were carried out using buffer only and buffer and DMSO, and each value is the result of eight technical replicate measurements. Error bars represent one standard deviation.

In the event that the inhibitor disrupted dimer formation a decrease in anisotropy would be observed. This is not seen suggesting that dimer disruption does not occur (Figure 4-8). Interestingly, at high inhibitor concentrations, an increase in anisotropy is observed, indicative of a possible increase in complex weight. Initially, it was suspected that at high concentrations the inhibitor might be causing protein aggregation. However, work on *Hs*GSTp1 has suggested

that MDR inhibitors may function to stabilise the dimer form, preventing protein-protein interactions and the MDR phenotype.[117] The results observed would be consistent with an increase in the quantity of dimeric protein compared to monomeric protein. Consequently, further experiments should be carried out to confirm if this may also be the case for inhibitors of *Am*GSTF1 and MHR.

4.4 Conclusions

Thermal shift assays have been used to investigate the SAR of flavonoid inhibitors. These have giving useful complementary information to that previously gained from CDNB activity assays. The results from thermal shift assays support the results from CDNB assays and demonstrate that in spite of the low conjugative activity of *Am*GSTF1 CDNB assays do represent a useful tool for testing compounds *in vitro*. From SAR data several functional groups on flavonoid inhibitors were identified as important to activity. It was found to be essential that the B-ring was aromatic, suggesting that this group is involved in π-interactions in the binding site. A hydroxyl group at the C7 position was found to increase activity. This activity increase was also seen if this is replaced with a methyl or PEG group suggesting that this is acting as a hydrogen bond acceptor. Interestingly, activity is slightly lower for either of these, in spite of their electron donating alkyl groups, suggesting that sterics may play a part at this site. In addition, the presence of an α-alkoxy carboxylate group, with a long chain alkyl group was found to dramatically increase inhibition. Introduction of hydrophilic substituents along this chain reduced activity, suggesting that it binds to a highly hydrophobic area of the protein.

Assays to investigate selectivity of the compounds were also carried out. These investigated the relative binding of compounds to *Am*GSTF1 and another flavonoid binding protein *At*GSTF8. These identified that increasing the length of the alkyl chain to C_{10} reduced activity in *At*GSTF8 whilst maintaining inhibition in *Am*GSTF1. This suggests that this alkyl chain may be an area to focus on for increasing selectivity in future.

Further preliminary investigations were also carried out to investigate the potential mode of action of the flavonoid inhibitors. Experiments investigating whether flavonoid inhibitors might be resulting in the dissociation of the dimer suggested that this was unlikely to be the case. Instead these demonstrate that flavonoids may be functioning to stabilise this dimer, functioning in a manner similar to MDR inhibitors.

5 Crystallography of Wild Type AmGSTF1

In order to unravel the molecular basis of the interactions of inhibitors with *Am*GSTF1, and their mode of action, crystal structures of the protein in complex with ligands were sought. While structures had previously been obtained by the group for Apo *Am*GSTF1 at 1.95 Å and *Am*GSTF1 with CNBF at 2.8 Å,[87] this work aimed to produce higher resolution structures, particularly of the CNBF structure. In addition to this, it was aimed to crystallise *Am*GSTF1 in the presence of a variety of the flavonoid series of inhibitors.

5.1 Production and Purification of Untagged *Am*GSTF1

Initial work in the group had found crystallisation using strep tagged *Am*GSTF1 to be unsuccessful, possibly due to the flexible tag used for purification, and as a result untagged protein was used for all crystallographic studies. Protein was expressed in *E. coli* and purified using glutathione agarose affinity chromatography in a method optimised from Cummins.[85,118,119] By streamlining the purification protocol from a four column process (Hydrophobic interactions, GSH agarose affinity, desalting and anion exchange) down to only two columns (GSH agarose affinity and anion exchange) it was possible to increase production of pure *Am*GSTF1 protein from around 2 mg/ L of recombinant *E. coli* culture to higher yields of around 10-12 mg/ L of recombinant *E. coli*. Protein purity was confirmed by SDS-PAGE as sufficient for crystallisation experiments (Figure 5-1).
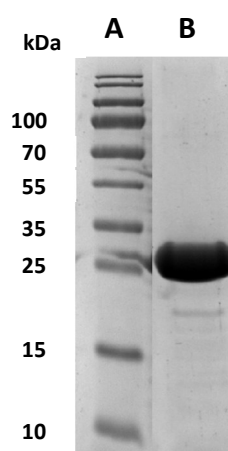
Figure 5-1: SDS-PAGE gel to show the purity of untagged AmGSTF1 protein used for crystallography. Lane A- molecular weight ladder, B-untagged *Am*GSTF1. Expected molecular weight of monomer= 24931 Da.

5.2 Crystal Structure Determination of Apo AmGSTF1

Crystals were obtained in a bipyramidal hexagonal prism morphology (Figure 5-2), which diffracted to 1.5 Å at the Diamond Light Source. Detailed information about data collection and processing are given in section 3.10.2 and Appendix B. A hexagonal modification was obtained, in space group P 6_3 2 2, with one protein chain in the asymmetric unit.
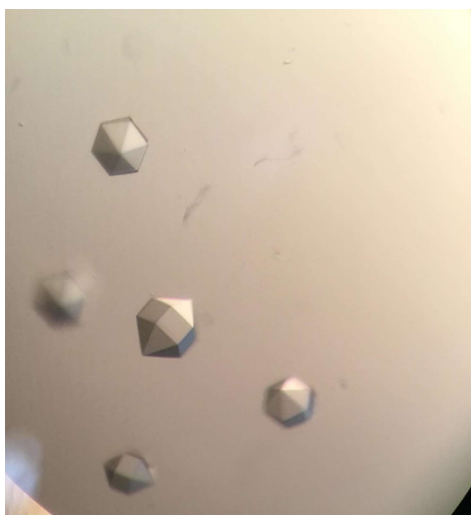
Figure 5-2: Morphology of Apo *Am*GSTF1 crystals

The structure displayed the classical GST fold, with two domains forming the two active site binding pockets. The N-terminal domain forms the G-site, lower portion of the H-site and contains the catalytic residue (Ser12). It was composed of four anti-parallel β-sheets and two α-helices (Figure 5-3a). The loop region associated with the G-site (residues 38-48) showed no electron density and as a result was not included in the model. The C-terminal domain consisted of six α-helices, the first of which (α-4) forms the upper portion of the H-site binding.[120] No density was observed for an associated loop region (residues 125-137). Based on the derived symmetry related structure (Figure 5-3b), the four residues preceding this disordered loop (Cys120-Leu121-Phe122-Asn123) were found to project into the other monomer subunit, blocking the active sites.
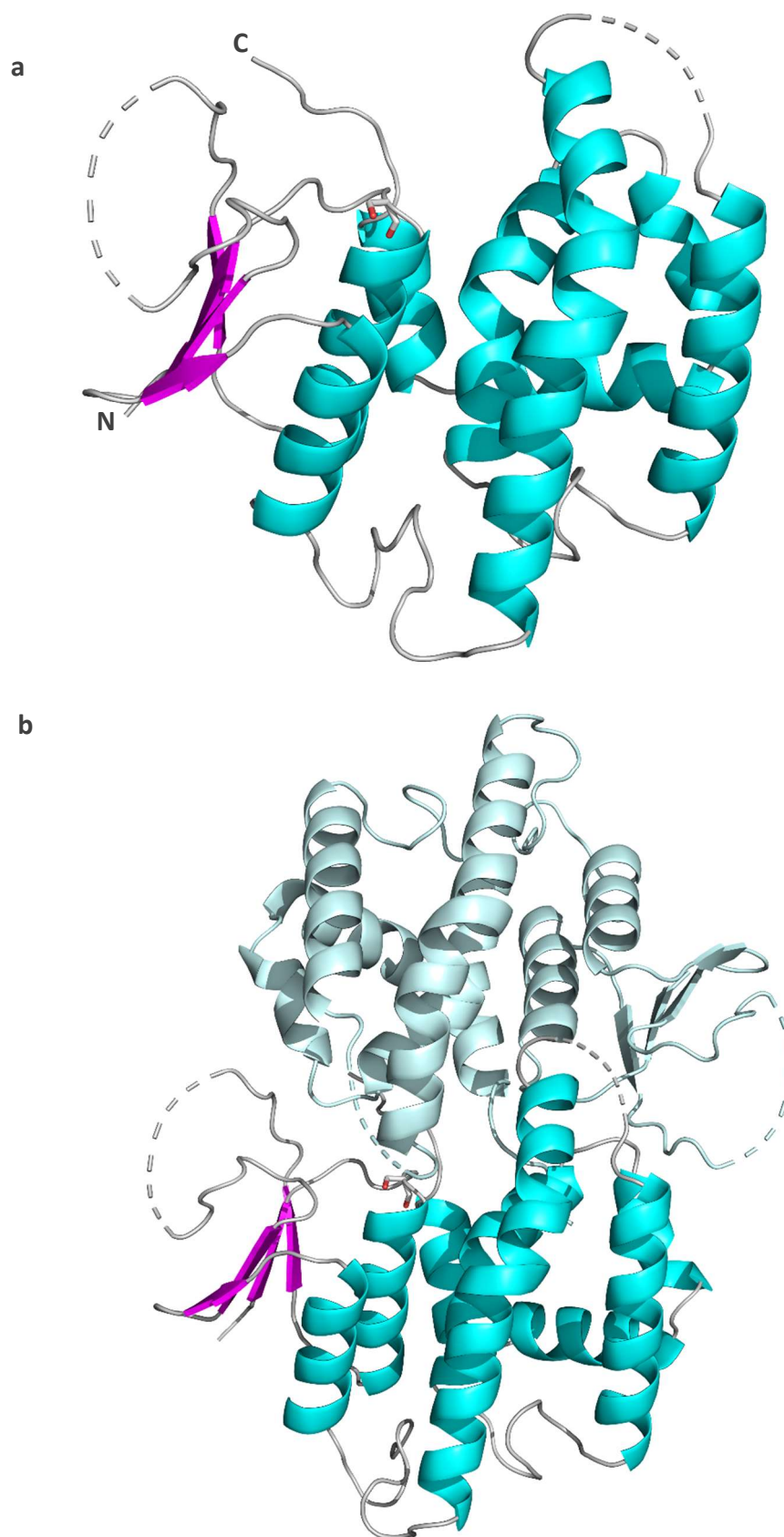
Figure 5-3: (a) Structure of *Am*GSTF1. (b) Crystal packing of *Am*GSTF1.
Α-helices are shown in cyan, β-sheets in cyan and for (b) the symmetry mate in pale cyan. Dashed lines show the location of loops for which no density is observed. The Ser12 catalytic residue is shown as sticks. Figure produced in Pymol.

Table 5-1: Details of crystallographic and active dimer inetrfaces observed in WT *Am*GSTF1 structure. Values obtained from analysis using PISA[113]

Dimer	Surface Area/ Å	$\Delta^i G$/ kcal mol^{-1}	P-value
Active	1390	-19.4	0.198
Crystallographic	1364.8	-25.1	0.048

While only one monomer is seen in each asymmetric unit, *Am*GSTF1 is active as a dimer. The interfaces for this crystal form were analysed using PISA (Table 5-1).[113] Two interfaces are observed as significant in the structure. The first (Figure 5-4) has a large area of 1390 Å^2, and is composed of mainly hydrophobic interactions in addition to 8 hydrogen bonds and 16 salt bridges. These occur between α-4 and α-5 of one monomer and the N-terminal domain of the other.[121] PISA calculates a negative $\Delta^i G$ for interface formation of -19.4 kcal/ mol, which is indicative of a hydrophobic interface. A P-value of 0.198 is indicative of a highly hydrophobic surface, very likely to be interaction specific. The second (Figure 5-3B) is also seen to have a large area of 1364.8 Å^2 and to be formed largely of hydrophobic interactions in addition to 14 hydrogen bonds. It has a more negative $\Delta^i G$ of -25.1 kcal/ mol, and a P-value of 0.048, indicative of a highly specific interface. In spite of the larger $\Delta^i G$ and smaller P-value observed for the second interface, in this case the first interface is expected to be the biological form observed in solution for two main reasons. Firstly, GSTs have been well characterised and in all cases of dimeric GSTs it is this form which has been observed. Secondly, the second interface occurs in the active site, which would prevent the protein carrying out its catalytic activity.
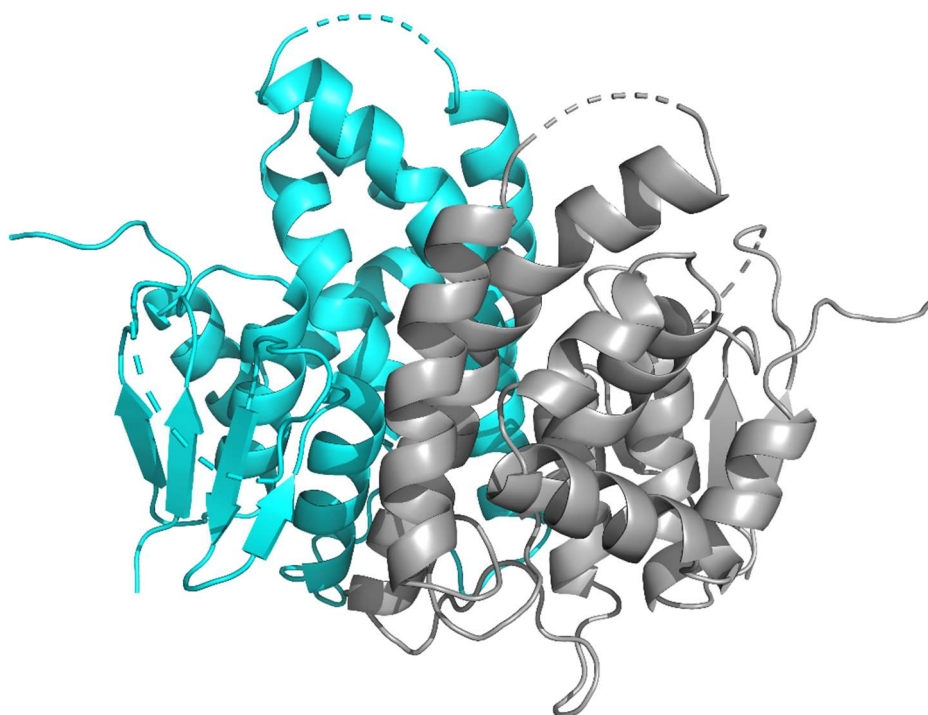
Figure 5-4: Structure of active dimer of *Am*GSTF1, across a crystallographic 2-fold axis. The two monomers are shown in different colours (cyan and grey). Figure produced in Pymol.

5.2.1 Comparison with lower resolution structure

As expected, the structure was found to be similar to the 1.95 Å resolution structure previously obtained with an RMSD of 0.21 for 189 equivalent Cα atoms.[87] The main differences occur in the regions surrounding the disordered loops. In the higher resolution structure it was possible to model more residues surrounding each of these disordered loops. For the loop over the G site (residues 38-48) one additional residue could be modelled at either end of the loop in the higher resolution structure, and for the loop in the C-terminal domain it was possible to model one additional residue at the end of the α-4 helix.

A number of changes are observed between the active sites of the two structures. Most noticeably, the side chain of the catalytic Ser12 is seen to adopt a different conformation with the oxygen pointing at almost 180° to the direction in the lower resolution structure (Figure 5-5a) Some small changes are also observed in some of the residues which form part of the H-site. Phe36, one of the final visible residues before the disordered loop, sits near to where we would expect a GSH molecule to bind and as a result could form part of the H-site.[56] In the higher resolution structure this adopts a pose 1.1 Å closer to α-4, which would have the effect of closing up the binding site slightly (Figure 5-5b). In addition, in the higher resolution structure it has been possible to model a side chain for Tyr118 and Gln119, located on α-4 which form part of the upper H-site.[120]
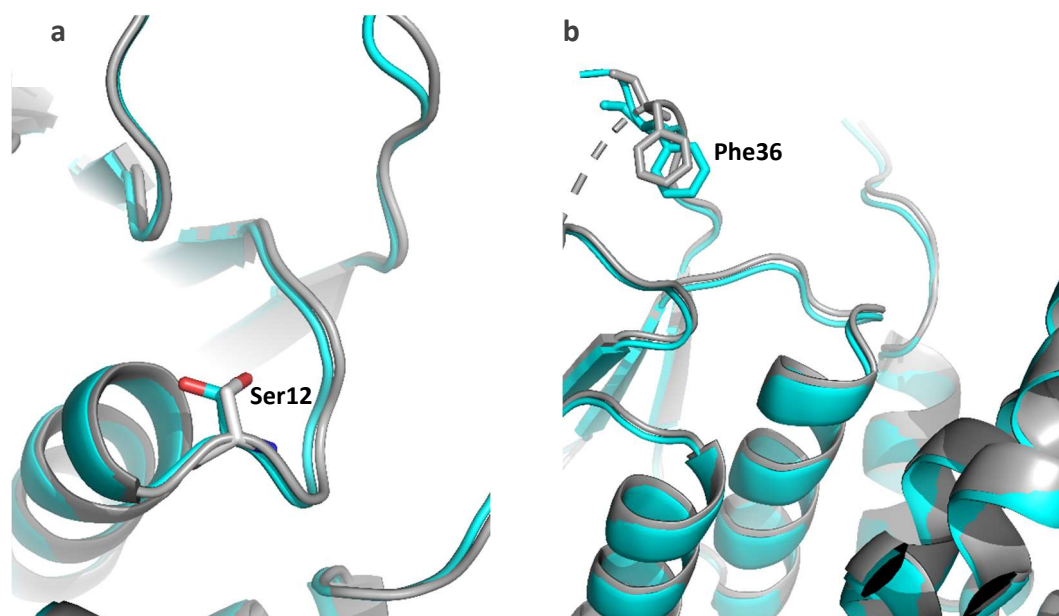
Figure 5-5: Differences between active site residues between high resolution (cyan) and low resolution (grey) structures of *Am*GSTF1. (a): Ser12 and (b): Phe36

5.2.2 Comparison with *Zm*GSTF1 structure

To gain additional insight into *Am*GSTF1 we compared it with *Zm*GSTF1 from maize. This was of interest for two main reasons. Firstly, it is the protein with a structure in the PDB which has the highest sequence similarity to *Am*GSTF1 (63%), and secondly it is known to be involved in herbicide tolerance in maize.[56,120,122] Two different structures of *Zm*GSTF1 are available in the PDB, 1AXD, a structure in complex with lactoylglutathione[56] and 1BYE, a structure in complex with an atrazine-GSH conjugate.[120] As 1AXD is the higher resolution of the two structures this comparison will focus primarily on it, however we will also look at the H-site of 1BYE, in particular the residues involved in herbicide binding.
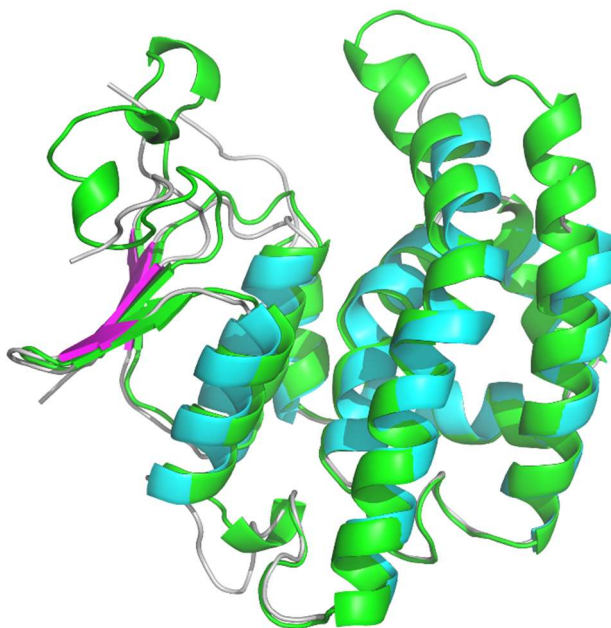
Figure 5-6: Superposition of AmGSTF1 (cyan and magenta) with ZmGSTF1 (green) PDB code 1AXD. Figure produced using Pymol.

Despite having a sequence similarity of only 63%, the overall folds of *Am*GSTF1 and *Zm*GSTF1 are very similar, with an rmsd of 1.27 Å across 182 equivalent Cα atoms (Figure 5-6). *Zm*GSTF1, like *Am*GSTF1, is composed of two domains, with the N-terminal domain being composed of four antiparallel β-sheets flanked by two α-helices, and the C terminal domain composed of six α-helices. Unlike *Am*GSTF1 it contains a short helical section in the linker between the two domains.

The largest differences between the two structures are seen to be surrounding the two loops which are disordered in the *Am*GSTF1 structure. The residues from Asp35 in the *Am*GSTF1 which are situated in the run up to the disordered loop over the G-site have a different conformation to that of the *Zm*GSTF1 structure. In *Zm*GSTF1 this loop is known to undergo a conformational change on GSH binding, resulting in an induced fit to the substrate. It is likely that this difference is due primarily to the fact that the *Am*GSTF1 structure is apo, while the *Zm*GSTF1 has a lactoylglutathione molecule bound. It would be expected that in a structure with GSH bound the loop in *Am*GSTF1 would adopt a very similar conformation to that seen in *Zm*GSTF1.

A difference is also observed in the structures in the upper portion of the α-4 helix which forms part of the H-site. In the *Am*GSTF1 structure this is seen to be bent away slightly from the active site above Asn110 compared to α-4 in the *Zm*GSTF1 structure. The degree of flexibility observed

in this area is known to be involved in determining substrate specificity of the H-site[120], and the difference observed between these two enzymes indicates that they are likely to have different substrates. Alternatively, it is possible that the crystal packing, where the upper portion of α-4 and α-5 of the crystal mate are sitting in the active site, is resulting in a slight distortion in this area.

The G-site is seen to be very well conserved between the two proteins (Figure 5-7). In 1AXD nine residues are associated with lactoylglutathione binding, which would mimic the position of a GSH molecule in the active site. Of these nine residues, eight are conserved between *Zm*GSTF1 and *Am*GSTF1 and one is a conservative substitution from Val54 in *Zm*GSTF1 to Ile55 in *Am*GSTF1. Unsurprisingly, the structure of this G-site is also very well conserved. Due to the disordered loop present in *Am*GSTF1 it is not possible to see the location of His41 or Lys42 in the *Am*GSTF1 structure, and only one sizeable difference is observed between the two structures. Phe36 in *Am*GSTF1 is seen to lie in a different orientation to Phe35 in *Zm*GSTF1. This residue is situated at the start of the disordered loop present in *Am*GSTF1. It is likely that on GSH binding this residue would move to adopt the position observed in *Zm*GSTF1. In light of the similarities between the G-sites of the two proteins it seems reasonable to expect that GSH would adopt a very similar binding pose in *Am*GSTF1 to that observed in *Zm*GSTF1.
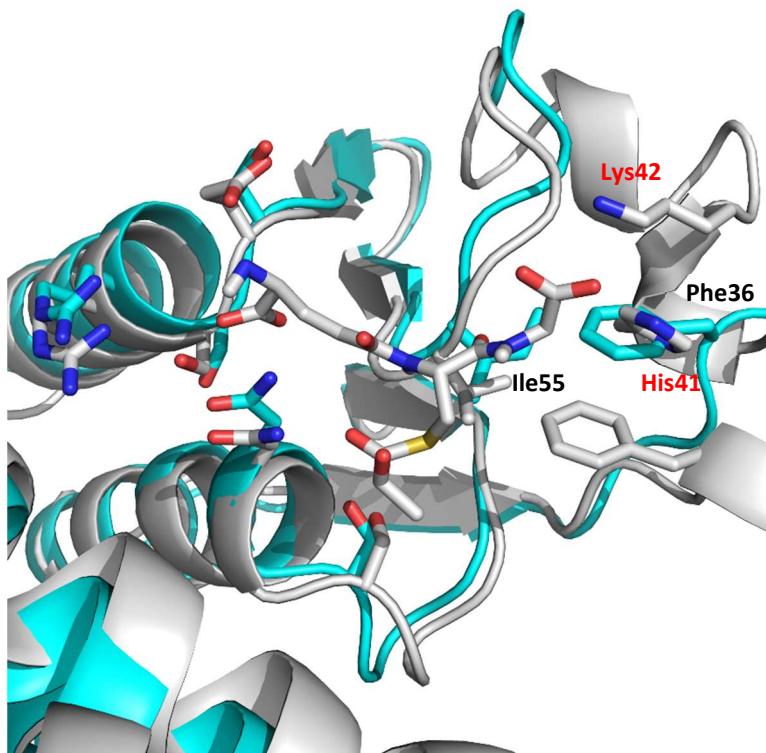
Figure 5-7: Superposition of the residues in the G-site of *Am*GSTF1 (cyan) and *Zm*GSTF1 (grey) PDB code 1AXD. Residues with differences between the structures are labelled. These refer to *Am*GSTF1 residues (black labelling) or *Zm*GSTF1 residues (red labelling). Produced using Pymol.

Next we examined the H-site from 1BYE[120], a structure of *Zm*GSTF1 in complex with a GSH-atrazine conjugate (Figure 5-8). In addition to the residues previously observed as binding GSH in the G-site, four main residues are involved in atrazine binding, with atrazine lying in a cleft between Met10, Phe35, and Trp12 and Ile118. Met10 and Phe35 are conserved within *Am*GSTF1, although as discussed earlier the Phe36 in *Am*GSTF1 lies in a different location in our structure. However Trp12 and Ile118 are non-conservatively substituted as Thr13 and Phe122 resulting in a very differently shaped H-site. Preliminary experiments (Appendix F) show that *Am*GSTF1 does not bind herbicides, most likely due to these inherent differences between the H-sites. This further enforces the hypothesis that *Am*GSTF1 does not cause MHR by a direct involvement in herbicide detoxification.[34]
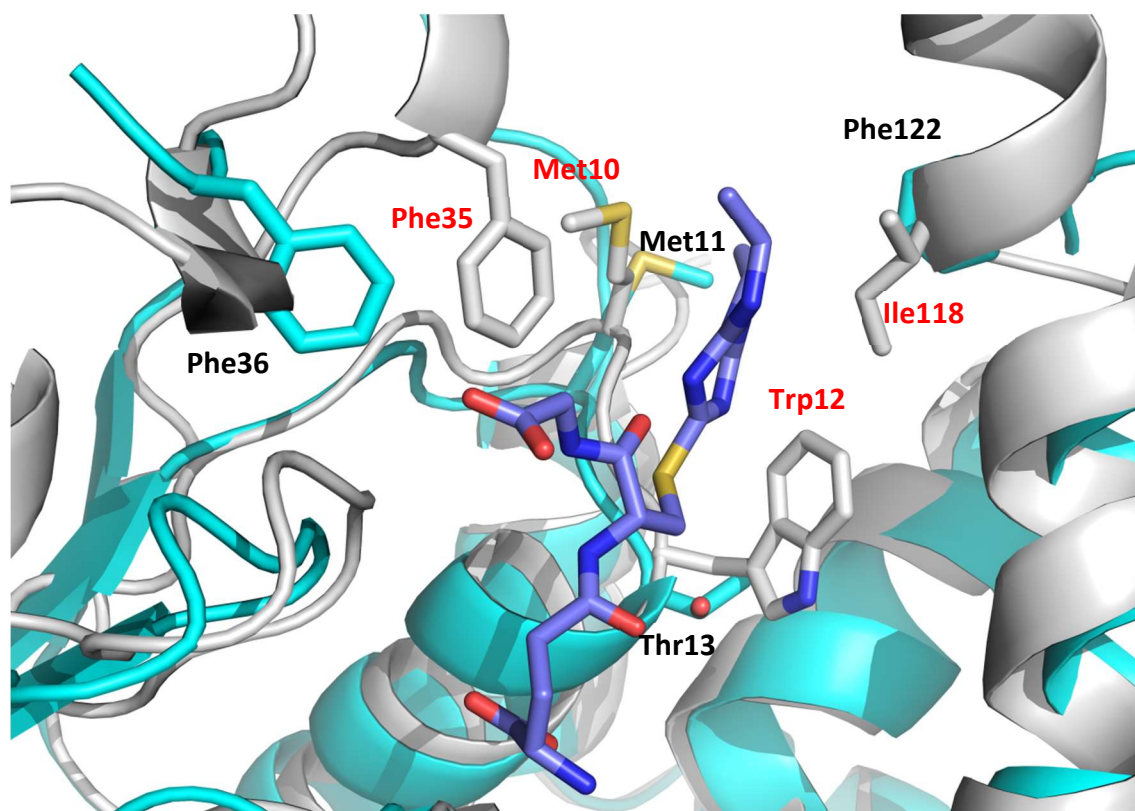
Figure 5-8: Superposition of the H-site of *Zm*GSTF1 (PDB: 1BYE; grey; red residue labels) with a bound atrazine GSH conjugate (marine) and *Am*GSTF1 (cyan; black residue labels). Produced using Pymol

5.3   Crystal Structure Determination of *Am*GSTF1 with CNBF

Figure 5-9: Morphology of CNBF treated *Am*GSTF1 crystals

Protein was alkylated with CNBF, and this was confirmed using mass spectrometry. It was observed that modification had not occurred with 100% efficiency (Figure 5-10). Trays were set up using the same conditions as for the apo protein. Crystals were obtained in the same hexagonal bipyramidal morphology as for the apo protein, however they were visibly yellow in colour (Figure 5-9). The crystals diffracted to 2.0 Å in the same hexagonal space group as for the wild type. Additional information on data processing and collection is available in Section 3.10.2 and Appendix B.
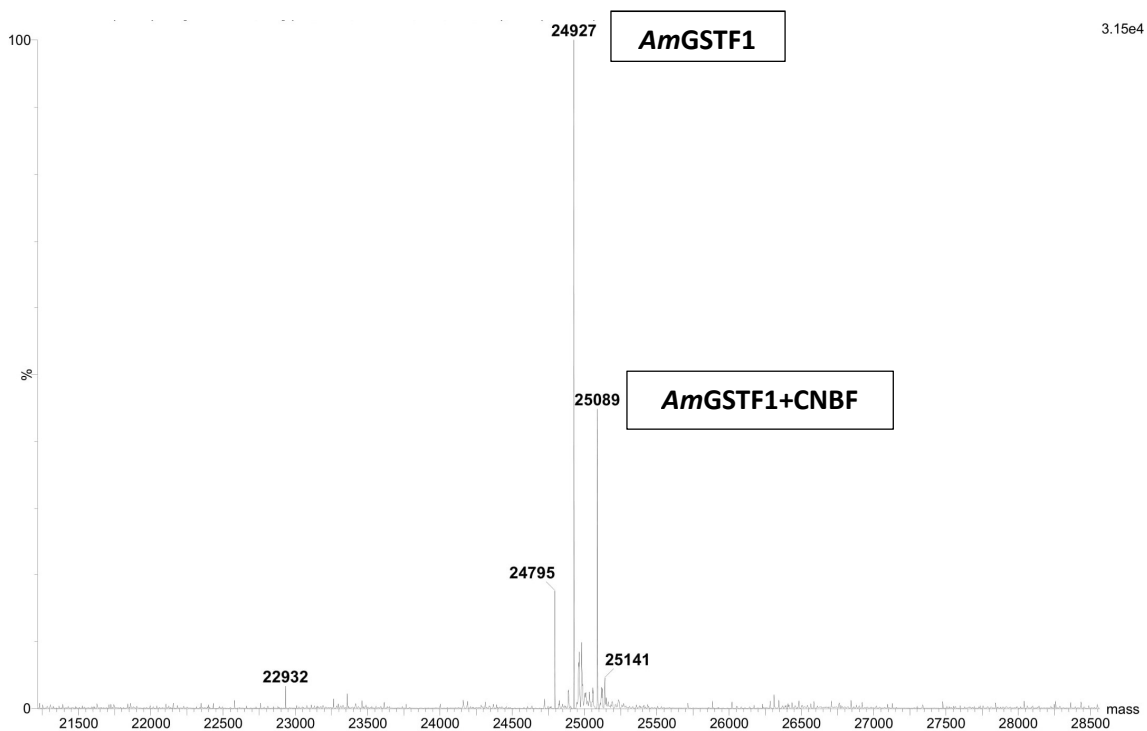
Figure 5-10: ESI Mass spectrum to confirm CNBF modification of *Am*GSTF1

The structure obtained for the CNBF bound structure is very similar to that of the apo protein, with an rmsd of only 0.19 Å for 188 equivalent Cα-atoms. The NBF-adduct was seen to couple to Cys120, as had been expected from previous mass spectrometric results (Figure 5-11a).[34] This residue is one of the last visible residues before the disordered loop that points into the active site of the neighbouring subunit (Figure 5-11b).
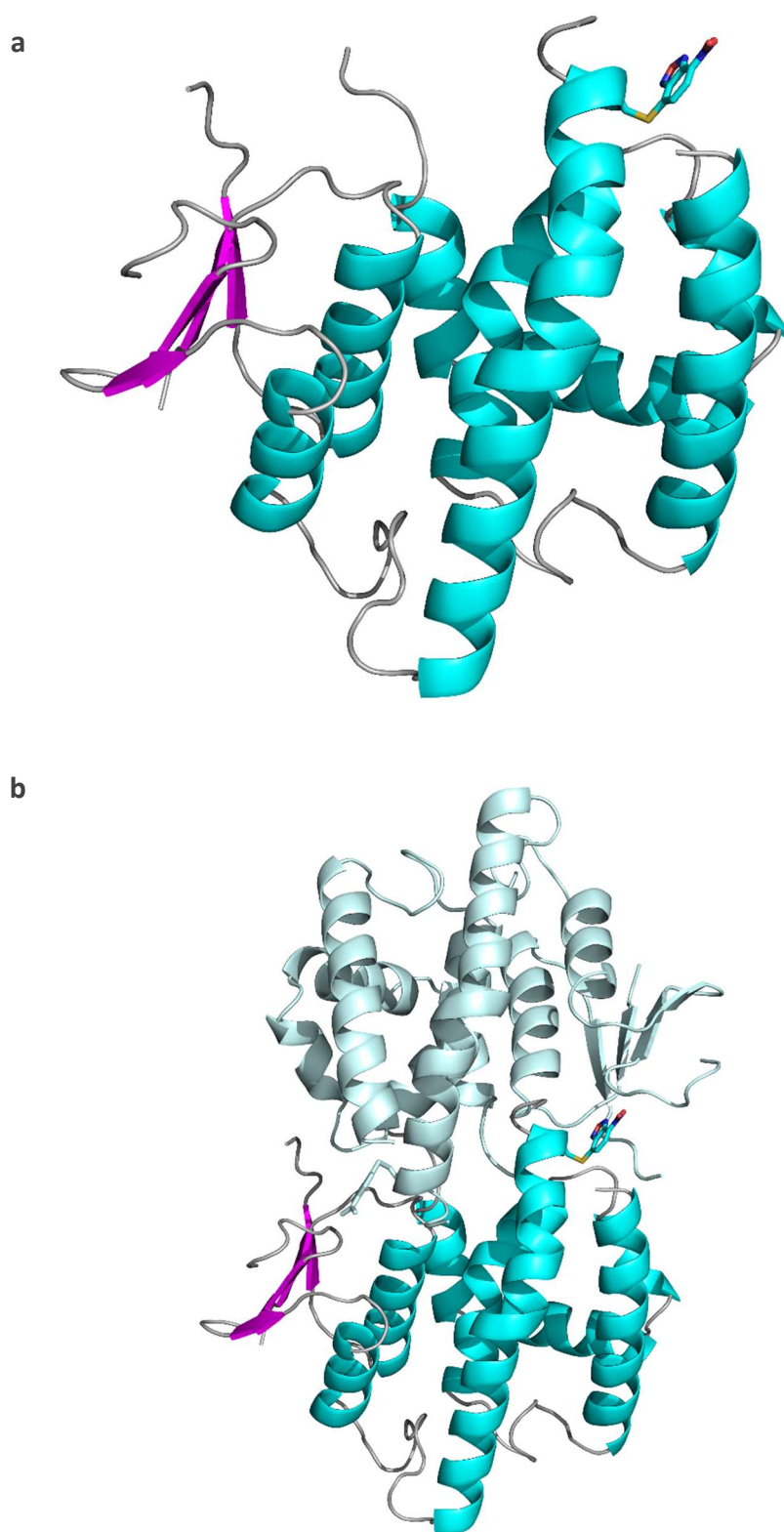
Figure 5-11: (a) Location of CNBF covalent modification on *Am*GSTF1. (b) Crystal packing of *Am*GSTF1 showing location of CNBF modification in active site of symmetry mate. Produced using Pymol.

The density for the CNBF indicated that the modification was only partially occupied (modelled at 0.7 occupancy) in the crystal (Figure 5-12). Mass spectrometry analysis of the sample used for crystallography confirms this, showing that only partial modification has occurred.
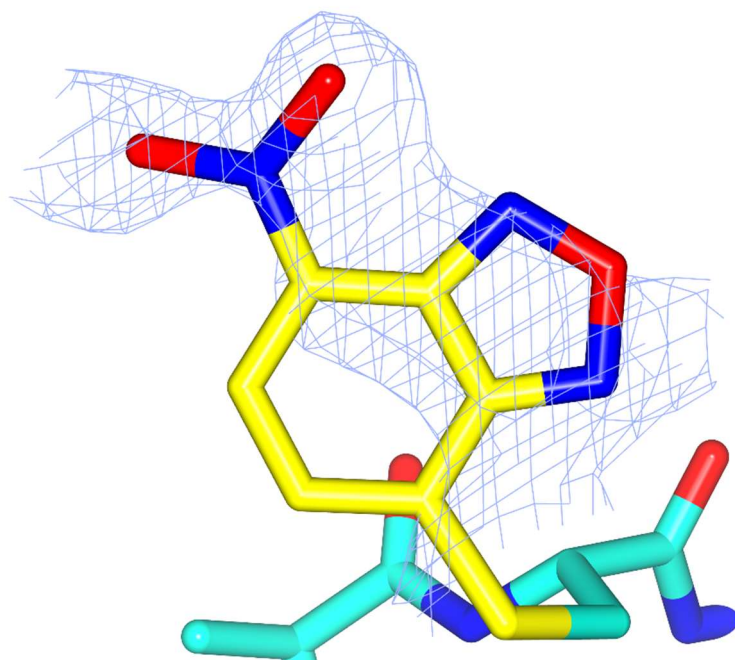
Figure 5-12: Density observed for CNBF molecule covalently linked at Cys120. $2F_o$-F_C map shown, contoured at 0.5σ. Figure created using CCP4mg.

It had been expected that CNBF modification at Cys120 might result in a structural change which would result in the observed inhibition. Surprisingly, with the exception of the covalent modification observed at Cys120, the crystal structure of the NBF-adduct does not show any major structural changes compared to the apo structure (Figure 5-13A). This could be for a number of different reasons. It is possible that CNBF is resulting in a change of structure of one of the loops that are disordered in our structure. Alternatively, it is possible that the crystal packing arrangement of the structure may be preventing us from seeing the minimum energy conformation of that loop in solution. It is possible that the CNBF binding may cause the α-4 helix to adopt a different conformation, but that the strong crystal packing interactions observed between the α-4 helices of the two crystal packing monomers prevent us from observing this effect.
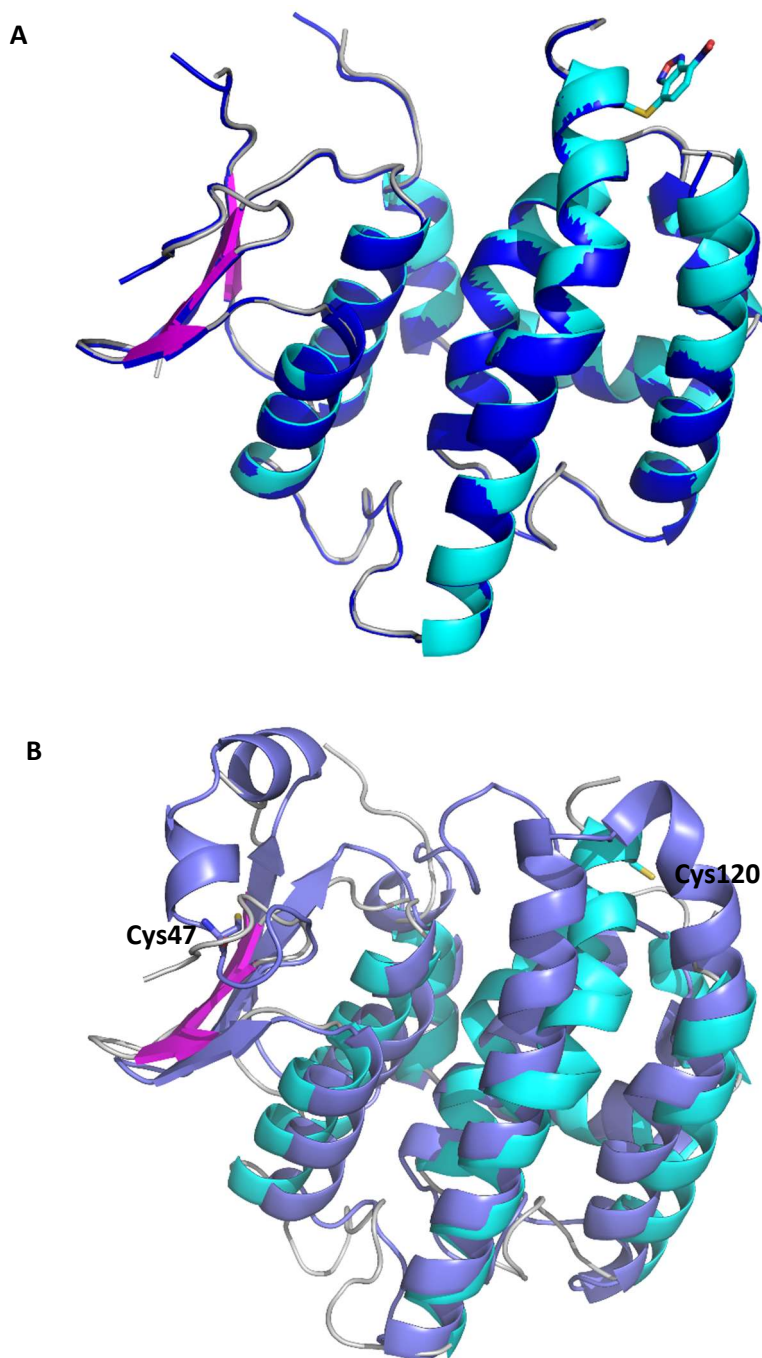
Figure 5-13A:Superposition of CNBF bound *Am*GSTF1 (cyan and pink) with Apo *Am*GSTF1 (blue). B: Superposition of *Am*GSTF1 (cyan and magenta) and *Hs*GSTP1 (slate, PDB code 3GSS) showing locations of Cys47 (*Hs*GSTP1) and Cys120 (*Am*GSTF1). Figure produced using Pymol.

Finally, it is possible that the CNBF is acting to either stabilise or destabilise the G or H sites in a way that cannot be seen from this single crystal form. This could result either in compounds no longer binding to the active site or being stabilised in the active site. It is known that for *Hs*GSTP1, the presence of GSH reduces its interaction with JNK1, reducing its ability to cause multi-drug

resistance.[83] It is thought that inhibitory drugs for *Hs*GSTP1 function by stabilising this GSH bound form. Interestingly, while CNBF also inhibits *Hs*GSTP1 via covalent modification of a cysteine (Cys47), this cysteine is located in a different area of the protein. In the *Hs*GSTP1 the Cys47 is located in the region of the loop over the G-site while in *Am*GSTF1 the cysteine is located on α4 helix in the region of the H-site (Figure 5-13). Although initial work to identify protein binding partners of *Am*GSTF1 has proved unsuccessful this has not been ruled out as a possibility. It is possible that the covalently bound CNBF is acting to stabilise the GSH form, or that it is stabilising a glutathione conjugate of CNBF (NBF-GS). We would expect NBF-GS to form spontaneously on addition of CNBF to protein *in vivo* and for it to be bound in the active site, however this cannot be seen in this structure, most likely due to the crystal packing observed.

5.3.1   Effect of CNBF modification on compound binding

Preliminary experiments were carried out to further investigate the effect that CNBF binding at Cys120 might be having on compound binding in the active site. Thermal shift assays were carried out to investigate the binding of a range of compounds to modified and unmodified protein. Compounds were selected which were predicted to bind to the active site (Figure 5-14). This comprised two flavonoid inhibitors: baicalein and MS 5-7 one of the synthesised flavonoid inhibitors with a C5 long alkyl chain and a C2 thiazole group. Ethacrynic acid, an MDR drug known to bind to the active site of *Hs*GSTP1 was also selected.[123] In addition NBF-GS was synthesised, to investigate if the alkylation had an effect on the binding of its conjugate.
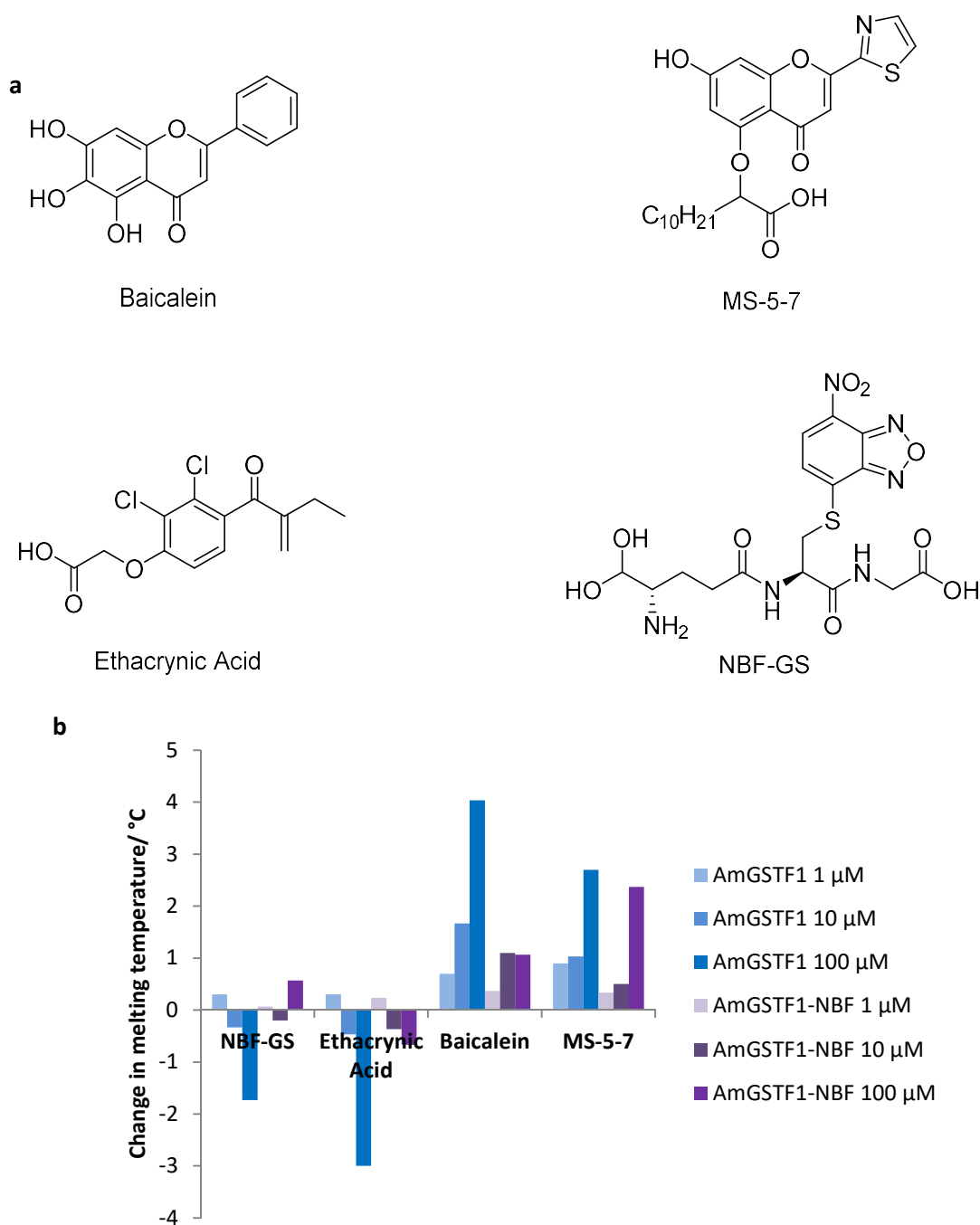
Figure 5-14: (a)Compounds used for binding assays with apo and CNBF modified protein (b)Changes of melting temperature observed with addition of various ligands to *Am*GSTF1 (blue) and CNBF modified *Am*GSTF1 (purple). Results are the average of three technical replicates.

It is clear that the CNBF modification does not prevent the binding of flavones to the protein, in both cases these are seen to stabilise the protein, increasing the T_m in a concentration dependent manner consistent with binding. CNBF modified protein was observed to have a lower T_m than for the unmodified protein, 60.7 °C as opposed to 63.6 °C, and as a result the slight differences observed in magnitudes of ΔT_m could not be used to indicate any difference in the

strength of binding. For both modified and unmodified protein, ethacrynic acid is observed to bind, but to destabilise the protein. The NBF-GS conjugate is seen to bind to and destabilise the apo protein. By contrast, it is seen to weakly stabilise the CNBF modified *Am*GSTF1, suggesting that it is possible the modification affects the binding of this conjugate. As it is not possible to determine the relative binding affinities from thermal shift assay data this result is inconclusive, and further work will need to be carried out to determine if this modification is affecting the binding affinity for the NBF-GS conjugate.

5.4   Crystal Structure Determination of *Am*GSTF1 with Flavonoid Ligands

In order to unravel the molecular basis of flavonoid inhibition, co-crystallisation experiments were carried it. Whilst initial docking studies (discussed in detail in Chapter 7) suggested that the ligands most likely bound into the active site, it was not possible to completely rule out an alternative binding position in the dimer interface of the protein, where flavonoids have been identified to bind in *At*GSTF2.[50] Co-crystal structures would then enable targeted inhibitor design.

5.4.1   Selection of ligands and initial trials

Based on the results of biophysical investigations, a selection of inhibitors was selected for crystallisation trials (Figure 5-15). As initial trials with SYG8 had resulted in precipitation of the compound,[85] solubility was also considered closely when selecting compounds. Compounds with a thiazole group in place of the B ring had proved to be significantly more soluble than any others synthesised, while maintaining strong binding, and as a result the majority of ligands selected contained this group. Compounds with a long C_{10} carbon chain had proved to be highly active both *in vivo* and *in vitro*, and as a result it was desirable to identify what role this long chain might be playing. MS4-48 was selected as it contained this long C_{10} chain, in addition to the thiazole ring which increased its solubility. As the long chain dramatically decreased solubility, MS4-33 was selected as this contained a shorter C_1 chain which would help to identify the location this chain might be binding to. MS4-31-2 was selected as it was a smaller ligand, which showed good binding and solubility. Baicalein was also selected as this was the commercial flavonoid which had performed best in the biophysical binding tests, it could be

bought in large quantities, and used in testing to reduce the concern about a limited supply of the synthesised flavonoid inhibitors.
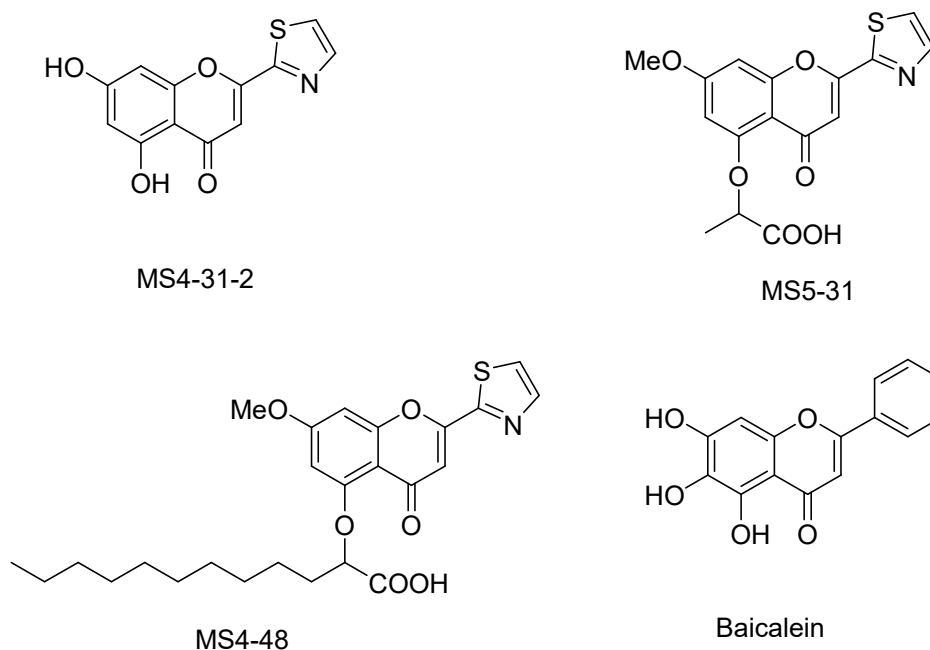
Figure 5-15: Structures of small molecule compounds selected for crystallography trials

Because of the low solubility of many of the ligands, it was necessary to dissolve the compounds in DMSO. As a result, tests were carried out to determine the maximum amount of DMSO that would be tolerated by the protein and it still crystallise. The same conditions were used as for crystallising the apo protein, and DMSO was introduced with end concentrations of 0.5%, 1.25%, 2.5%, 5% and 10%. While crystals were visibly much smaller and grew much more slowly with increasing DMSO concentration, crystals still grew at up to 2.5% DMSO concentration in the same morphology as for the wild type. However, for 5% and 10% DMSO concentration the protein was seen to precipitate and no crystals were formed. As a result an end DMSO concentration of 2.5% DMSO was used within crystal conditions to help solubilise the compounds.

Initially, compounds dissolved in DMSO and protein buffer were added directly to drops at a concentration 5× that of the protein. However, compounds were seen to precipitate within 20 mins using this method. As a result, proteins were pre-incubated with compound for either 4 h or overnight, and centrifuged to remove undissolved compound before trays were set up. This strategy proved to be much more successful, and whilst in some wells, low levels of compound precipitation was seen, this was not sufficient to prevent crystallisation, or prevent crystals being

collected for analysis. As a result of this low compound solubility soaking trials were not attempted.

Crystals were selected from a range of conditions, containing a variety of ligands and different incubation times with the protein. All crystals were seen to adopt the same morphology observed for the apo and CNBF structures. 9 datasets were collected ranging from 1.52-2.4 Å. Full details are available in Appendix G. Structures were solved using molecular replacement against the apo protein and data was refined to obtain a difference map of sufficient quality to identify if ligand density was present. In each case it was determined that no ligand was bound.

The lack of ligand binding observed was thought to be as a result of the crystal packing which prevented small molecule binding in the active site. The crystal packing observed in other areas of the protein was as would be biologically expected, and the other contacts observed are consistent with the structure of the active dimer. Therefore this should not prevent observation of density for ligands binding in other areas. As a result it seems likely that the flavonoid ligands are binding in the G or H-site, and a binding pose in the L-site or an alternative allosteric site can most likely be ruled out.

5.4.2   Crystal packing analysis

The majority of crystals were obtained in a $P6_3 2 2$ space group. A crystal mate is observed, with the α4 and α5 helices of another monomer situated in the active site of the first monomer occupying both the G and H sites (Figure 5-16a). As can be seen from the surface plot this forms a tight interaction which would prevent small molecule binding (Figure 5-16b).
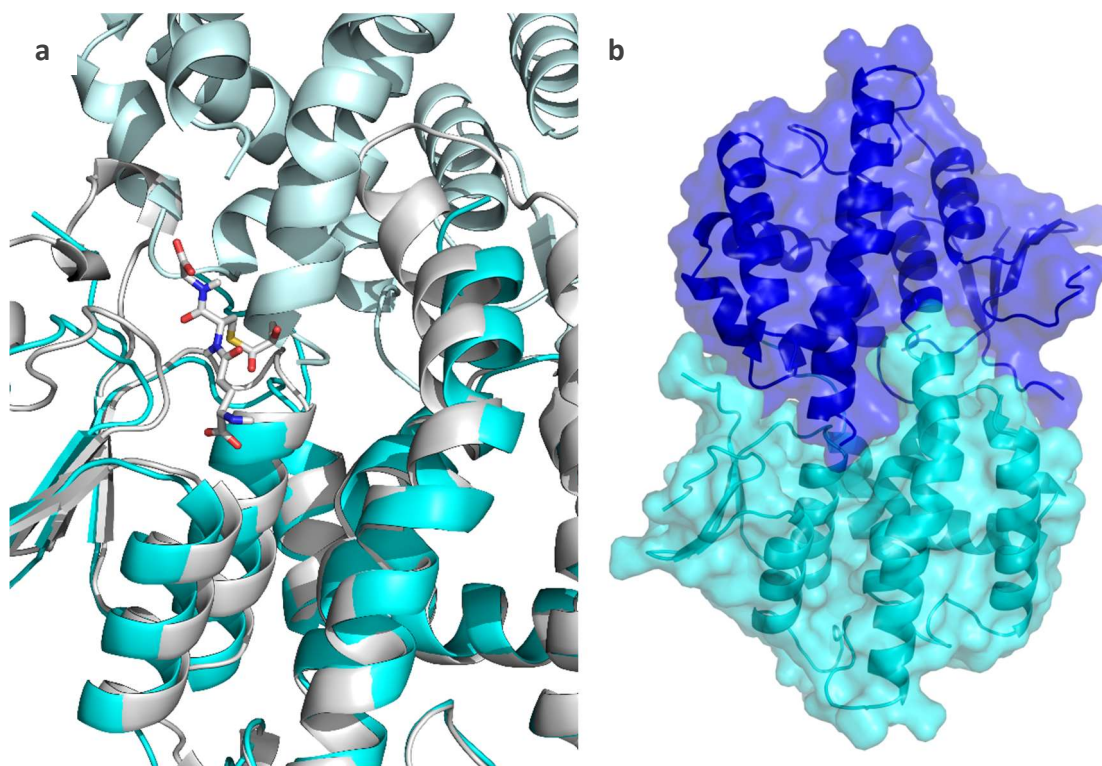
Figure 5-16: (a) Superposition of *Am*GSTF1 and symmetry mate (cyan and light cyan) with *Zm*GSTF1 (grey) and its bound ligand, highlighting location of active site. (b) Surface representation of *Am*GSTF1 and symmetry mate (cyan and blue)

While an alternative space group of R3 2 was also observed for some datasets this posed the same problem. The interactions observed in the hexagonal datasets were highly conserved, and the crystal mate was in an almost identical position, with an rmsd of only 0.31 Å, again completely blocking the active sites (Figure 5-17).
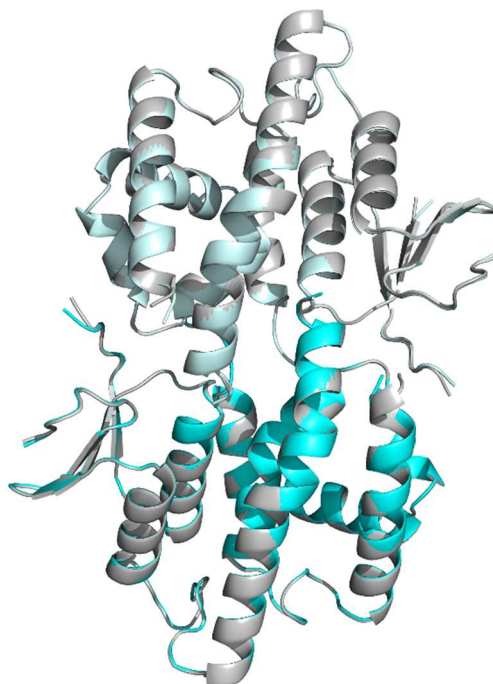
Figure 5-17: Superposition of hexagonal (cyan and light cyan) and rhomboedric (grey) crystal forms

5.4.3 Experiments to identify a new crystal packing form for *Am*GSTF1

As attempts to co-crystallise *Am*GSTF1 with compounds in the active site in the current crystal form had proved unsuccessful, it was attempted to identify crystals with a different crystal packing form. A variety of 96 well high throughput crystallography screens were set up using either protein alone or protein pre-incubated with one of the four ligands used for previous co-crystallography experiments, in an attempt to encourage the formation of crystals with ligands bound.

Whilst there were no conditions in which crystals were produced in trays with ligand but not apo trays, there were a series of promising hits where the crystals looked visibly very different to the previous hexagonal ones obtained. Of these, four different forms were of sufficient size to take for testing on i03 beamline at Diamond Light Source (Figure 5-18). Two of the four tested crystals diffracted in a pattern that was visibly different to the hexagonal form, whilst the remaining two did not diffract. Unfortunately, in both the cases where diffraction did occur, none was seen above approximately 7 Å. Attempts to optimise these crystals also proved unsuccessful, with no better crystals being observed in optimisation plates set up around each of the conditions.
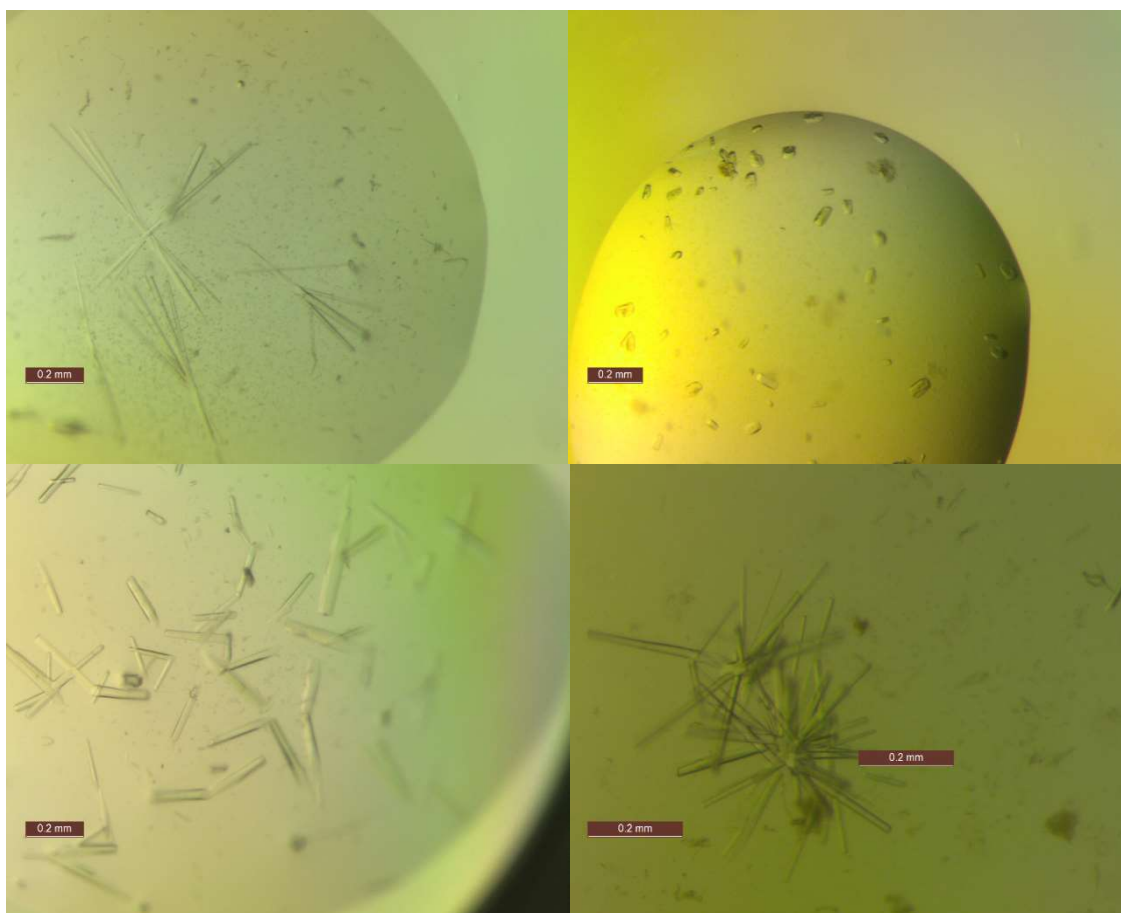
Figure 5-18: Morphologies of *Am*GSTF1 crystals observed

5.5   Conclusions

It has been possible to obtain a higher resolution structure of wild type *Am*GSTF1, with the structure being determined to 1.5 Å. This has demonstrated that it has a structure that is highly conserved with that of other phi class GSTs.[50,56,124] As expected, it has a highly conserved G-site, but significant differences exist in the H-sites. Notably, this structure still has two disordered loops, one of which is in the region of the G-site, preventing a full model of the structure from being determined.

Higher resolution data has also been obtained for *Am*GSTF1 modified with CNBF, with the structure being determined to 2.0 Å. From this structure it has been possible to demonstrate the covalent modification of Cys120 by CNBF. No other major structural changes are observed in the structure compared to the apo structure. Using thermal shift assays it has been possible to rule out that CNBF modification is preventing compound binding in the active site, however it is possible that it may be affecting the binding of an NBF-GS conjugate. As no other structural changes have been observed from the apo structure, it has not been possible to determine the method by which the covalent modification could be having an inhibitory effect. It is possible that changes could be occurring in disordered loops in the region of both the CNBF molecule and active site, which are obscured as a result of crystal packing.

Thus far, it has not been possible to solve co-crystal structures of *Am*GSTF1 in complex with flavonoid ligands. It has been determined that this is most likely the result of the crystal packing observed for wild type *Am*GSTF1. A crystal packing mate forms a tight interaction in the active site preventing the binding of small molecules. As no density was observed for flavonoid ligands in other locations this has identified the G or H-sites as the most probable locations for flavonoid inhibitor binding.

Work to disrupt the crystal packing of wild type *Am*GSTF1 has proven unsuccessful, possibly as a result of the size and predicted strength of the crystal packing interface. In any cases where crystals of different morphologies have been produced it has not been possible to optimise these sufficiently to collect data that would enable the structure to be solved. As a result it has not yet been possible to solve a complete structure of *Am*GSTF1 including loops or to study binding in the active site. The next chapter of this thesis will focus on the design of mutants to specifically disrupt the crystal packing observed.

6 Crystallography of Mutant AmGSTF1

As discussed in chapter 5, previous models of *Am*GSTF1 were incomplete, displaying two disordered loops, and with the crystal packing form thought to be preventing small molecule binding in the active site. Work attempting to identify an alternative crystal packing form of wild type (WT) *Am*GSTF1 had proved to be unsuccessful. As a result, it was decided to use a protein engineering approach to attempt to disrupt this crystal packing form.

To modify the protein as little as possible it was decided to design single site mutants to disrupt the contacts observed between crystal mates. A similar approach has previously been used successfully to crystallise a fluorescein-scFv complex.[125] The native antibody was found to crystallise in a form where a loop occluded its binding site. By mutation of an exposed tryptophan which was involved in crystal packing it was possible to crystallise the protein in a different form with the binding site exposed. A similar method was used, using single site mutations to disrupt contacts, promote an alternative crystal form and expose the binding site.

6.1 Design and production of mutants

Detailed analysis of the crystal packing from the wild type *Am*GSTF1 structure identified contacts that occurred between a monomer and a symmetry mate that was not the functional dimer (Figure 6-1a).

The main contacts were formed between aromatic residues in the α4 helix interacting with the same helix in another monomer. Interactions were seen to occur predominantly between Tyr118 and Phe122 of opposite monomers (Figure 6-1b) with the rings lying 3.5 Å apart allowing them to form π interactions.
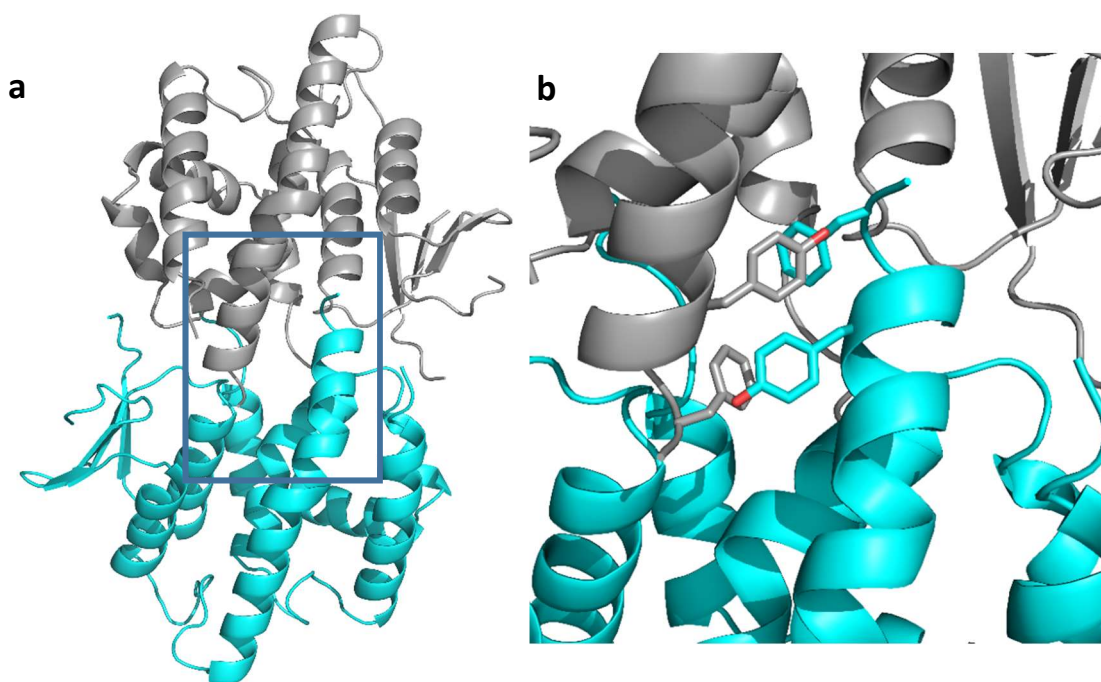
Figure 6-1: (a) Crystal packing interactions observed in *Am*GSTF1. *Am*GSTF1 is shown in cyan with its symmetry mate in grey. (b) Tyr118 and Phe122 are shown to show the π-interactions formed between monomers. (b) shows a zoomed in view of the area shown in the blue box from (a). Figure produced using Pymol.

Two single site mutants were constructed which replaced one of the residues, and one double site mutant which replaced both residues. In order to disrupt the π interactions observed the residues were mutated to non-aromatic ones. Phe122 was mutated to a threonine, as this enabled the formation of hydrogen bonds with a carbonyl on the backbone of the protein which might stabilise the α-helix and prevent loss of secondary structure. As no hydrogen bonds could be formed with the backbone at the Tyr118 position, this residue was mutated to a serine to maintain the hydrophilic nature.

Site directed mutants were generated using PCR, expressed in *E. coli*, and purified using GSH agarose affinity chromatography (see sections 3.2.4 and 3.3). In each case proteins bound to the GSH agarose column. This purification method relies on the protein's inherent ability to bind GSH. The fact that the mutants bind the column indicates they are correctly folded and have a correctly folded G-site. However in each case less protein was purified than for the wild type protein, with the mutants binding less efficiently to the column, and more protein passing through the column in the flow through. This suggests that their binding to the column may have been slightly impaired. This was particularly noticeable in the Y118S mutant where purification provided only around 3 mg of pure protein per litre of recombinant *E. coli*, compared to around

10-12 mg of wild type *Am*GSTF1. Protein purity was confirmed as sufficient for crystallisation experiments using SDS-PAGE gel (Figure 6-2).
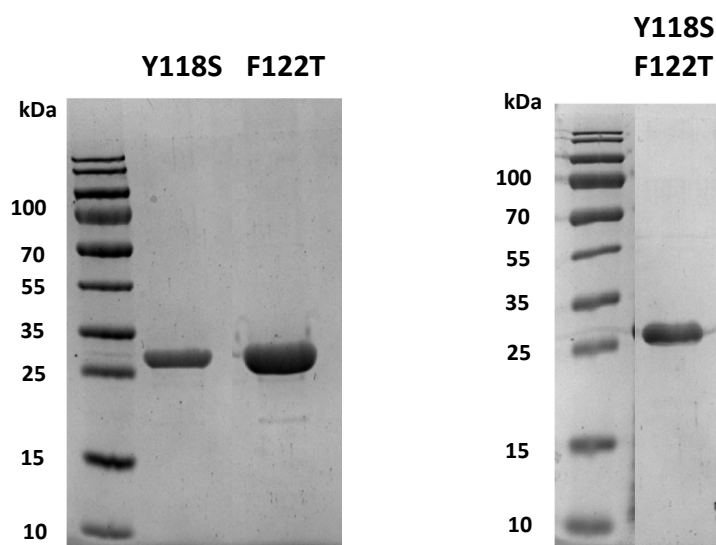
Figure 6-2: SDS-PAGE to demonstrate purity of mutants of *Am*GSTF1. Expected molecular weights: Y118S=25029 Da, F122T=25059 Da, Y118S F122T= 29483 Da.

6.1.1   Characterisation of mutants

In order for the mutant proteins to be used for further biophysical and crystallisation experiments it was important that they were correctly folded, with a particular focus on a correctly folded active site. As purification of the mutants was performed using GSH agarose affinity this required the protein to have a correctly formed G-site in order to bind to the column. Further characterisation was also carried out to confirm this.

Circular dichroism was used to confirm that the tertiary structure of the protein was similar to that of the wild type and that the protein had a fully folded conformation.[126] Each mutant gave spectra clearly indicative of a folded protein and which closely matches the spectra of the wild type protein in each case (Figure 6-3). The spectra produced are typical of proteins which are predominantly α-helical in nature.[127]
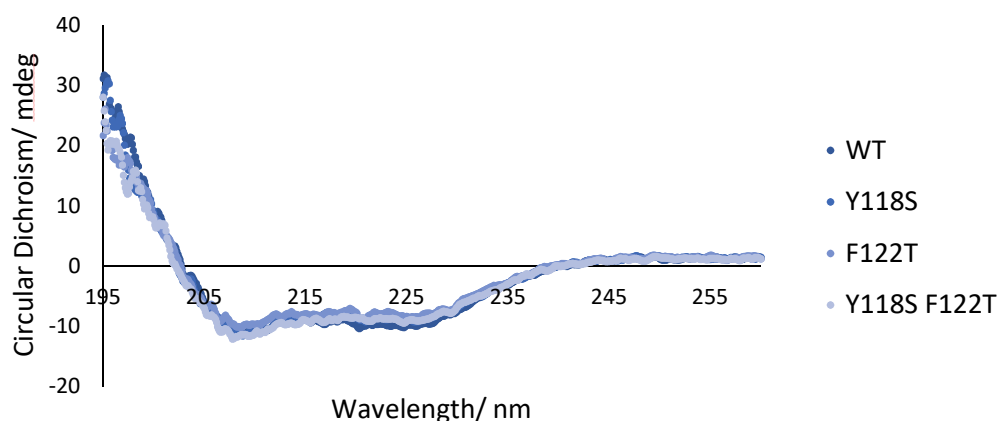
Figure 6-3: Spectra obtained from circular dichroism of wild type and mutant *Am*GSTF1 protein samples

The conjugative activity of each protein was then tested using the CDNB assay (Figure 6-4). In each case the mutants were found to show activity, albeit reduced compared to that of the wild type protein. As the Y118S mutant had been seen to less efficiently bind to the GSH agarose column it had been expected that this mutant would have the most reduced activity of the mutants, however this was not the case. While the Y118S mutant had lower activity than the wild type protein, probably due to this weaker binding to GSH, it had a markedly higher activity than the F122T or double mutant.
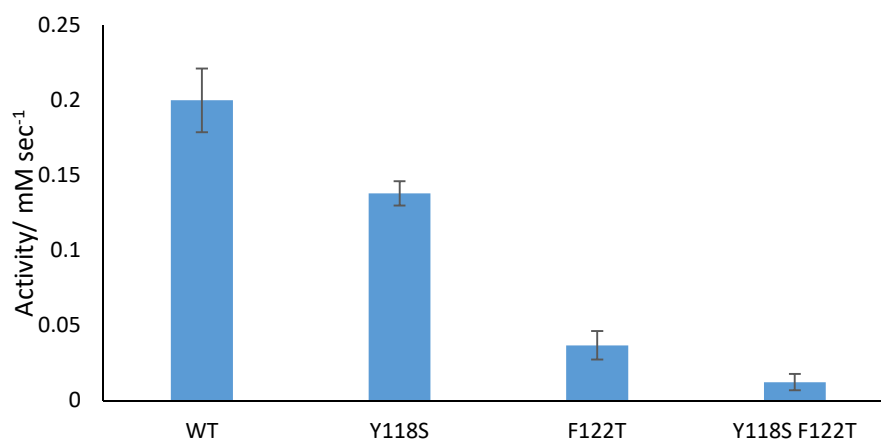
Figure 6-4: Activity of mutant *Am*GSTF1 enzymes. Results are averages of three replicates, and error bars show one standard deviation.

The results from protein purification, circular dichroism and activity assays suggested that the mutations were unlikely to have affected the overall fold of the protein, nor their ability to function as GSTs. As a result it was decided to continue with these designed mutants for crystallographic studies.

6.2 Crystal structure determination of mutants

Having determined that the mutant *Am*GSTF1 proteins were correctly folded and partially functional, attempts were then made to crystallise them.

6.2.1 Crystal structure determination of F122T *Am*GSTF1

Initially trials were carried out with the F122T mutant as it was possible to purify this in higher quantities. The F122T mutant of *Am*GSTF1 was found to crystallise well in high-throughput screens, across a range of conditions, with the Morpheus screen proving most successful.[103] Crystals were produced in two morphologies: rods, and smaller prismatic crystals (Figure 6-5). Notably, crystals were not found to grow in the condition in which the original hexagonal wild type crystals had grown reliably, suggesting that the mutation had successfully disrupted the packing interactions found in the wild type structure. The crystals with the rod morphology were found to diffract best (to 2.8 Å). Detailed information about data collection and processing are available in section 3.10.3 and Appendix B. Crystals were obtained in a tetragonal modification, in space group I 4_1 2 2, with one protein chain in the asymmetric unit. As with the original wild type structure, the functional dimer can be generated from a two-fold crystallographic axis.
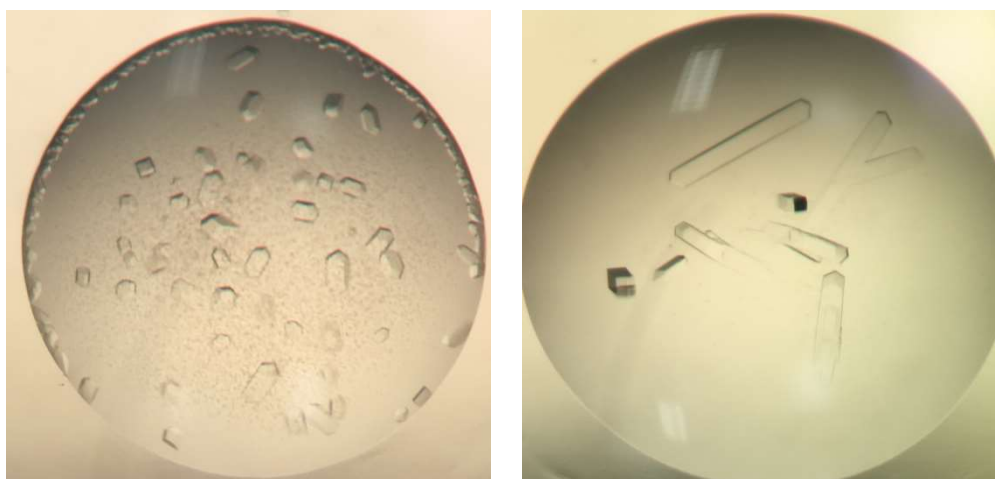
Figure 6-5: Crystal morphologies observed for F122T mutant of *Am*GSTF1- prismatic (left) and rods (right)

The structure was found to display a fold very similar to that of the wild type, however, there was clear density to build in both loops which had previously been disordered, giving a complete structure (Figure 6-6). Density was observed for 214 residues (from Ala2 to Ala215), and 6 water molecules. The model has an R factor of 0.18, an R_{free} of 0.24 and 4 Ramachadran outliers were observed. One of these was the Glu67 observed in the wild type structures, in addition to Asp60 which is observed as an outlier in all ligand bound structures and two further residues in poorly defined areas.
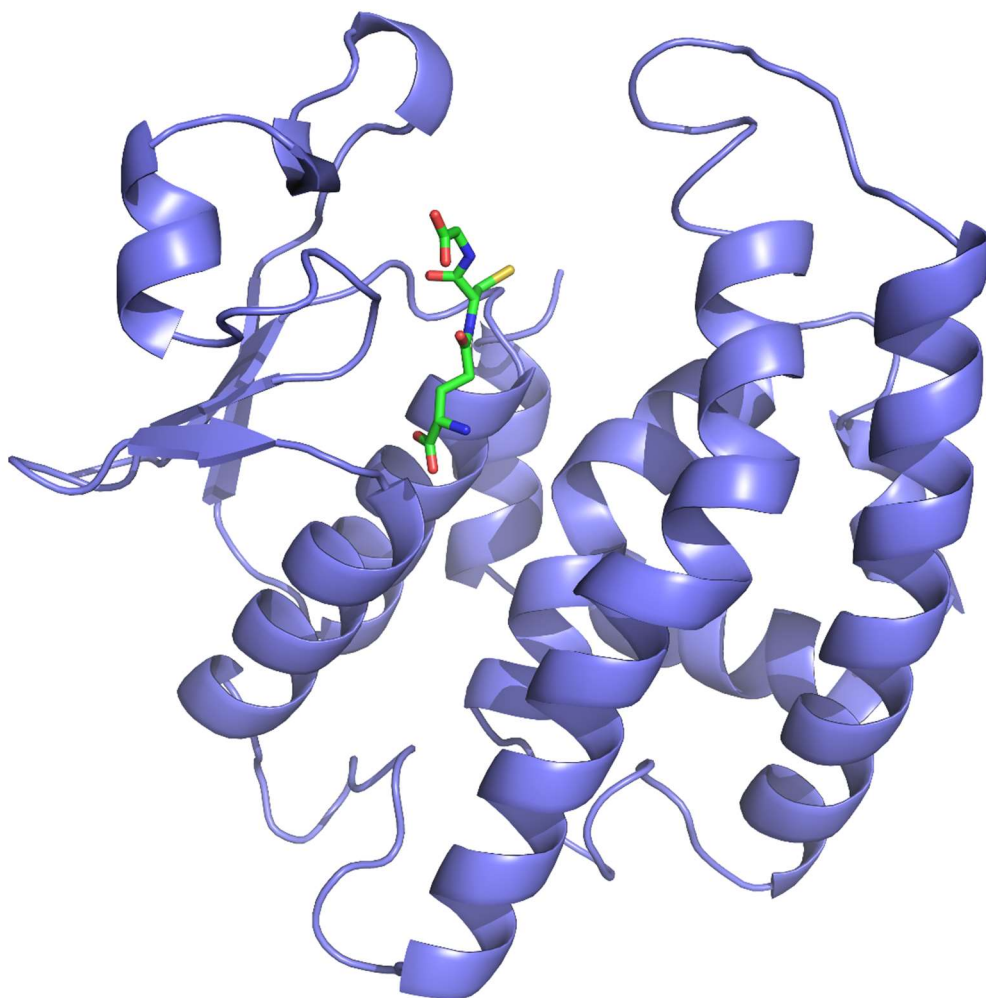
Figure 6-6: Structure of F122T mutant of *Am*GSTF1 showing structure of ordered loops and GSH molecule in the G-site, shown in green in a stick representation. Figure produced using Pymol.

Density was also seen within the G-site for a GSH molecule. The presence of this GSH molecule confirmed that the absence of small molecules in the previous wild type structure was due to the crystal packing. GSH was not added to crystal conditions and therefore must have remained from expression and purification.

6.2.1.1   Comparison with wild type and maize structures

Unsurprisingly, the F122T *Am*GSTF1 is seen to have a structure very similar to that of the wild type, with an rmsd of 1.12 Å for the 186 equivalent Cα atoms that make up the structure without the loops (Figure 6-7). The main visible differences between the structures are in the areas around the now ordered loop regions. In the area surrounding the loop in the area of the G-site the structure is seen to diverge significantly from Asn33 until Asn37 where the wild type loop became disordered. It is also seen to differ from Ala49, the first visible amino acid after the disordered loop to Ala57. Differences are also seen in the upper part of the α4 helix. In the WT structure, this is seen to bend further away from the H-site than in the F122T structure. In addition the C terminus of the F122T mutant is disordered within our structure, with the last visible residue being Ala215 compared to the full length protein observed in the wild type (to Ala219). As this C terminus is not thought to be related to ligand binding the fact that this area is disordered does not pose a problem.
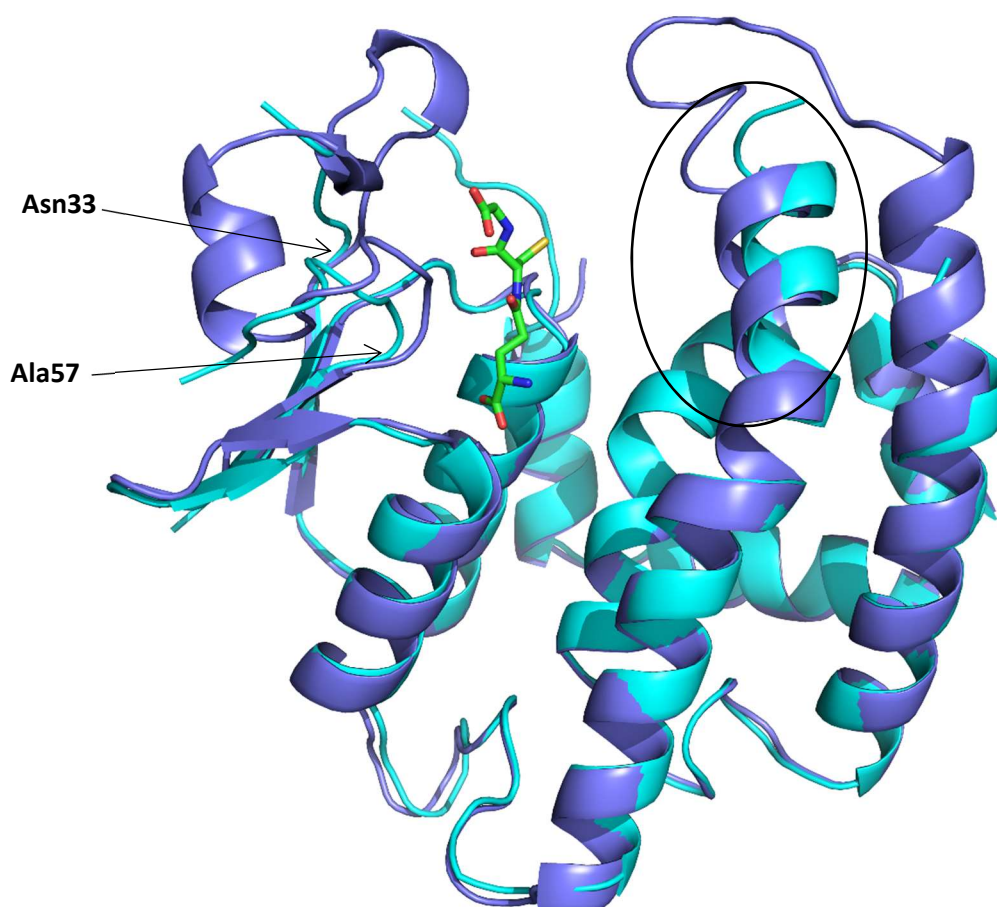
Figure 6-7: Superposition of structure of F122T mutant of *Am*GSTF1 (slate and green) with WT structure (cyan). Areas showing divergence between the structures are highlighted. Figure produced using Pymol.

The structure is found to closely resemble that of *Zm*GSTF1,[56] with the structures having an overall rmsd of 0.97 Å for 208 equivalent Cα-atoms, including both the core structure and loops (Figure 6-8).[123] Two insertions are seen in the *Am*GSTF1 structure compared to the *Zm*GSTF1 structure, explaining the structural differences in these areas.

Figure 6-8: Superposition of F122T *Am*GSTF1 (slate and green) with *Zm*GSTF1 maize (marine; PDB code: 1AXD). Areas with amino acid insertions are marked with black rings. Figure produced using Pymol.

The loop surrounding the G-site is found to closely resemble the loop seen in this structure of ligand bound *Zm*GSTF1. This is also the case for the upper part of the α4 helix. This difference between the wild type and F122T structures could be for one of two reasons. Either as a result of the crystal packing in the wild type, or as a result of the amino acid substitution in the mutant. The mutant shows the same positioning of this helix as both *Zm*GSTF1 and also Arabidopsis *At*GSTF2.[50] This suggests that this is the more likely positioning of this helix in *Am*GSTF1 and that

the difference in the wild type is caused as a result of crystal packing. The fact the change occurs lower down the helix than the mutation further supports this, making the difference in the F122T less likely to be as a result of the single site mutation.

We observe small structural differences between the *Am*GSTF1 and *Zm*GSTF1 structures as a result of two short insertions in the *Am*GSTF1 sequence. The first, a 3 residue insertion in the linker region, and the second, occurring in the loop region in domain II between the α4 and α5 helices.

The fact that the F122T mutant has a structure very similar to both the wild type and maize structures, and that a GSH is bound in the active site suggests that the reduced activity observed in the CDNB assay should not be of concern. A crystal structure of *c*GSTA1-1 in complex with a GSH-CDNB conjugate provides a possible explanation for this lack of activity (PDB code 1VF3).[128] The CDNB is seen to form a π-π interaction with Phe111 which is equivalent to Phe122 in *Am*GSTF1. Substituting this residue for a non-aromatic one would be likely to have a dramatic effect on the binding of CDNB in the active site. As a result it is most likely that the considerably lower activity of the F122T mutant and Y118S F122T double mutant was due to a reduced binding affinity of CDNB rather than an inability to function as a GST.

6.2.1.2   Analysis of the GSH binding site
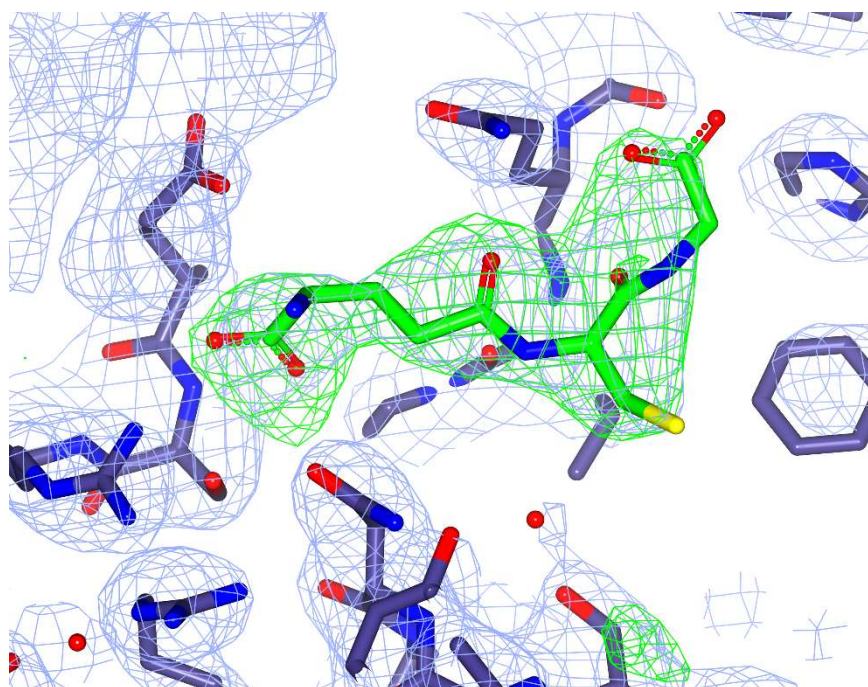
Figure 6-9: Omit map for the GSH molecule in the G-site of the F122T mutant. $2F_o$-F_c (blue) contoured at σ=1 and F_o-F_c (red and green) at σ=3. Figure produced using CCP4mg.

In the unbiased density there is clear density observed in the G-site for a GSH molecule (Figure 6-9). The GSH molecule is seen to form a number of hydrogen bond interactions with the loop over the G-site, as well as with the two α-helices which surround it. It is seen to form an extensive network of hydrogen bonds with hydrogen bonding side chains and the protein backbone, as well as hydrophobic interactions with three further residues (Figure 6-10).
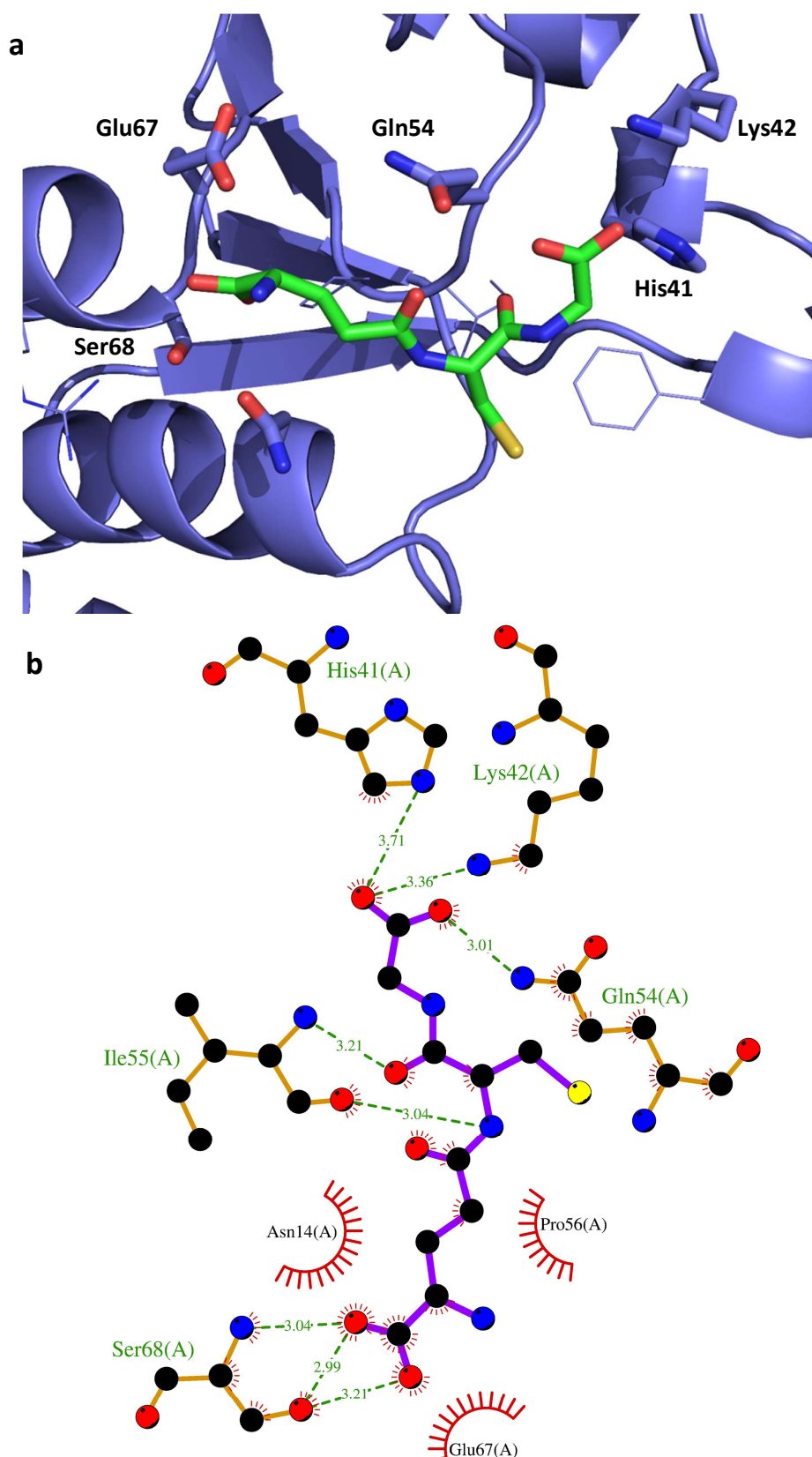
Figure 6-10: A) G-site of F122T mutant of *Am*GSTF1 (slate) with residues involved in binding of GSH (green) shown in stick representation. Figure produced using Pymol. B) LigPlot to show interactions of GSH with the active site of the F122T mutant of *Am*GSTF1. Hydrogen bonds are shown as green dashed lines and hydrophobic interactions as red combs.

As has previously been discussed, the G-site is highly conserved between AmGSTF1 and *Zm*GSTF1 and, the GSH molecule is seen to form very similar interactions. The main difference occurs in the location of the sulphur atom of the GSH or conjugated GSH molecule. In *Zm*GSTF1 this is observed to lie close to and form interactions with the catalytic Ser12 residue while in *Am*GSTF1 it does not. It is plausible that this difference is as a result of the fact that the *Zm*GSTF1 structure is of a conjugated GSH-atrazine molecule. It seems likely that during a conjugation reaction we could expect the carbon-sulphur bond of the GSH in *Am*GSTF1 to rotate to allow the sulphur to form an interaction with the Ser12 residue. The loop over the binding site is also seen to adopt a very similar structure. In *Zm*GSTF1 this is seen to have an induced fit on GSH binding, and it is likely the same would occur in *Am*GSTF1.

6.2.2   Crystal structure determination of Y118S *Am*GSTF1

Having identified the conditions where the F122T mutant would crystallise, crystallisation trials were then set up for the Y118S mutant using the conditions which had proven most effective for F122T crystallisation. As the majority of crystals of F122T *Am*GSTF1 were observed using the Morpheus screen, and due to a limited supply of Y118S protein, this mutant was exclusively screened using these conditions. Crystals of the Y118S mutant were observed in a number of wells, most of which also produced crystals of the F122T mutant, although notably under far fewer conditions than had been observed for the F122T mutant. Crystals had the same morphologies as for the F122T mutant (Figure 6-11). Crystals in the prismatic morphology diffracted best to 2.6 Å. Crystals were obtained in a tetragonal modification with space group I 4_1 2 2. Both the space group and unit cell observed are identical to that for the F122T mutant.
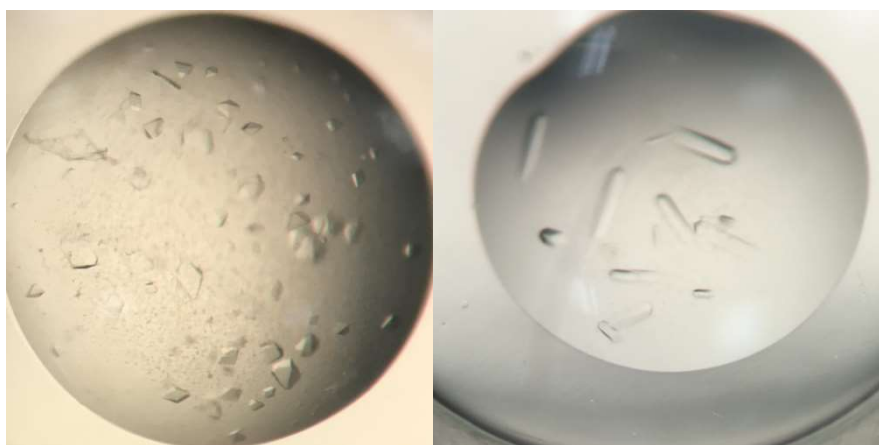
Figure 6-11: Crystal morphologies observed for Y118S mutant of *Am*GSTF1-prismatic (left) and rods (right)

The structure is of excellent quality, 2 Ramachadran outliers are observed: Glu67, Asp60 both seen as features. The protein adopts the typical GST fold (Figure 6-12), and again density was observed for the loop over the G site. Density was also observed again for a GSH molecule in the binding site, although this was refined at only 0.7 occupancy (Figure 6-13). Interestingly, the Y118S mutation appears to have resulted in the top of the α4 helix becoming disordered, with no density is observed for residues from Phe122 to Thr131. This suggests that interactions formed by this Tyr118 residue are essential to the structure of this helix.
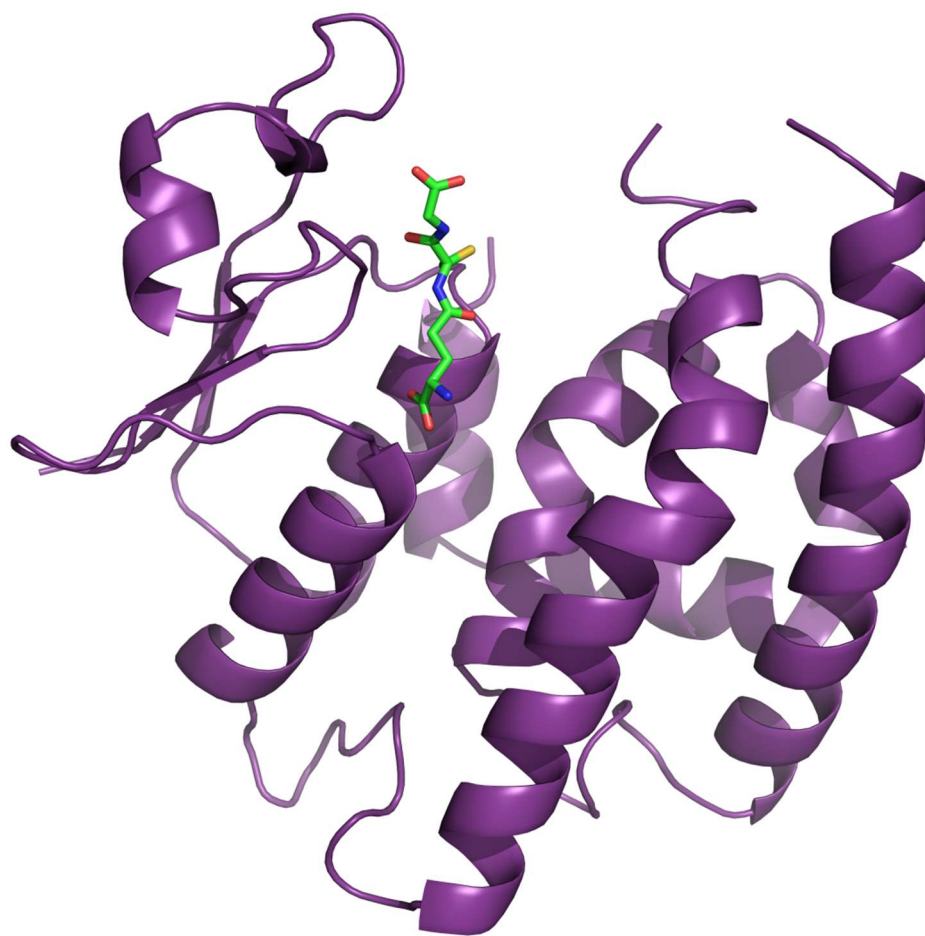
Figure 6-12: Structure of Y118S mutant of *Am*GSTF1 (purple) with a GSH molecule bound in the G-site shown in green in stick form. Figure produced using Pymol.
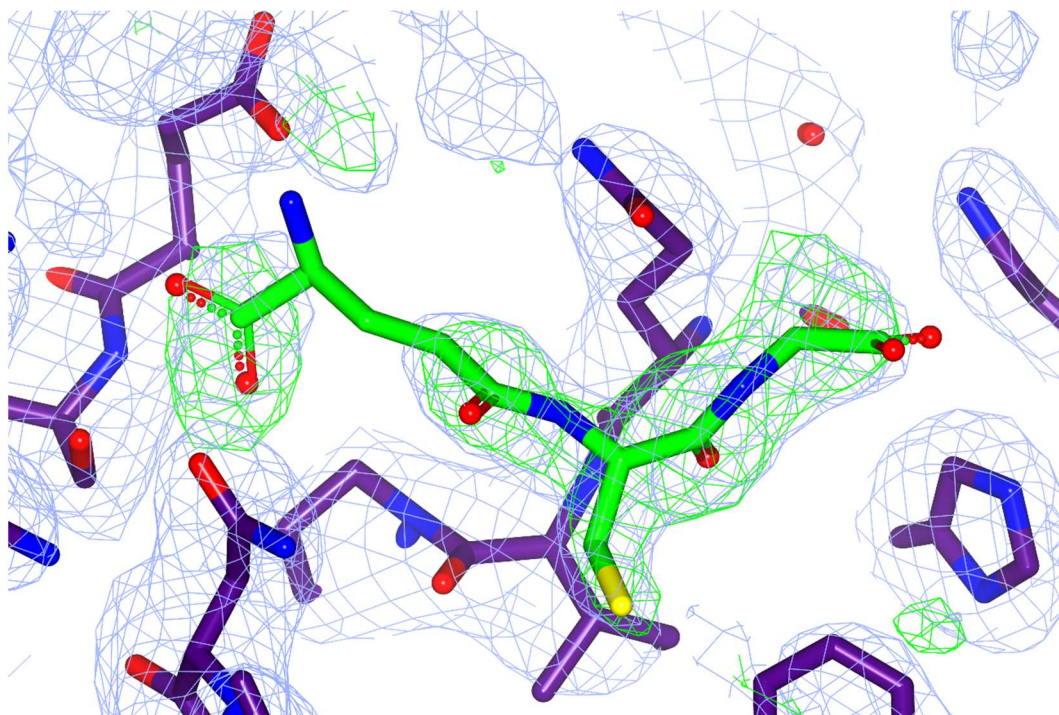
Figure 6-13: Omit map for GSH molecule bound in the G-site of the Y118S mutant of *Am*GSTF1. 2F$_o$-F$_c$ (blue) contoured at σ=1 and F$_o$-F$_c$ (green) at σ=3. Figure produced using CCP4mg.

With the exception of this area of disorder the structure is virtually identical to that of the F122T mutant, and they superimpose with an rmsd of 0.31 Å across 207 Cα atoms (Figure 6-14). There are no apparent differences in the location of any of the residues involved in GSH binding, except Phe36 which is rotated slightly between the structures (Figure 6-15). This is of particular interest as evidence suggests that the Y118S mutant binds GSH less efficiently than the wild type protein or other mutants. The Y118S mutant bound to the GSH agarose column less efficiently and the GSH molecule observed in the active site is only partially occupied, with weaker unbiased density than was observed for the F122T mutant. This suggests that it is possible that the structure of the α4 helix may be able to play a part in binding in the G-site, in a way that cannot be seen via crystallographic methods which capture only one conformation of the protein. This is particularly relevant in relation to a possible mechanism for CNBF inhibition, which binds to this helix (see section 5.3).
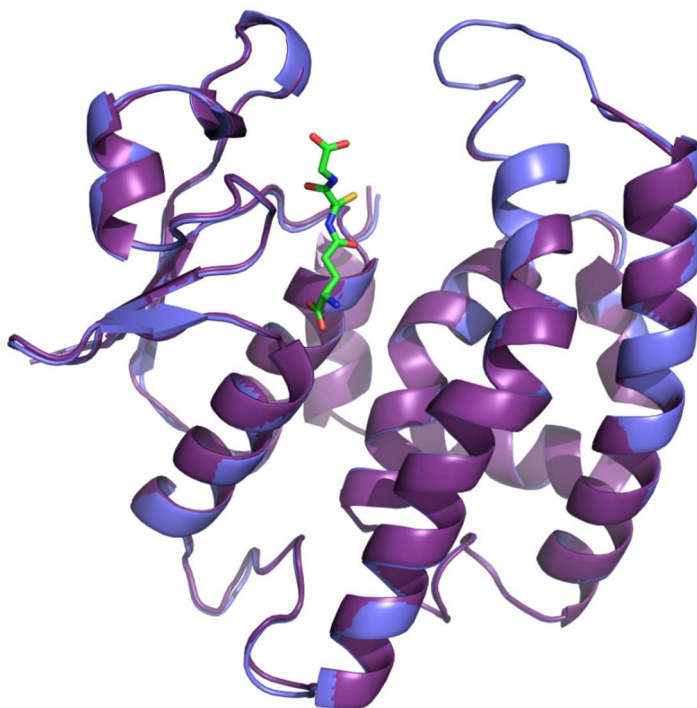
Figure 6-14: Superposition of structures of Y118S mutant (purple) and F122T mutant (slate) of
***Am*GSTF1 including the GSH molecule from the Y118S structure (green)**
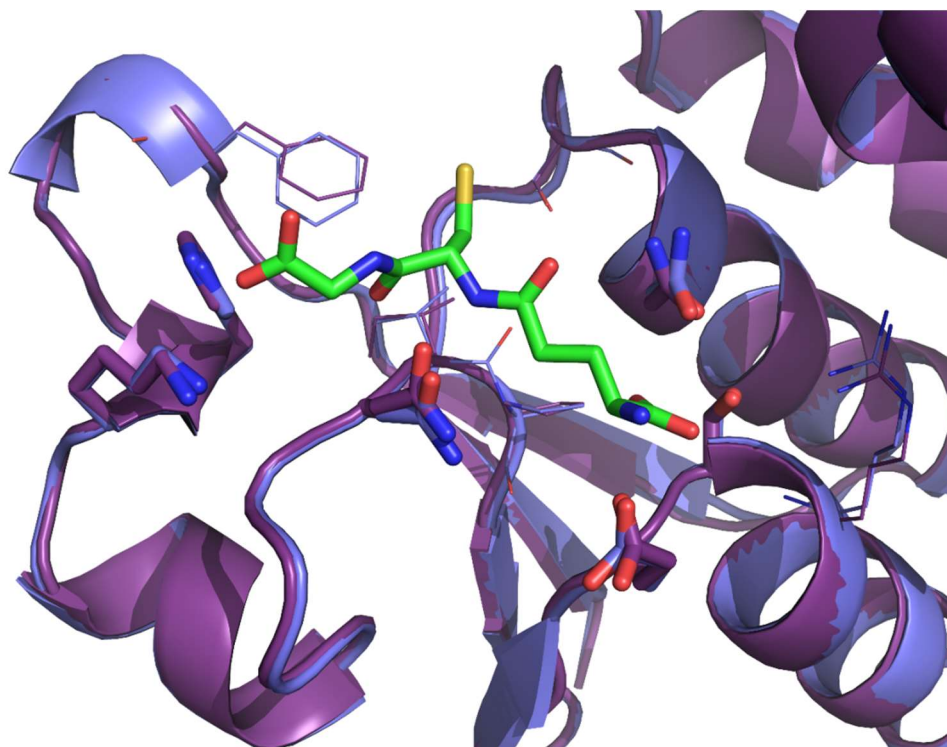
Figure 6-15: Superposition of the G-sites from the AmGSTF1 Y118S mutant (purple) and F122T mutant
(slate) with the GSH molecule from the Y118S structure (green). Residues involved in binding are
shown in stick form. Figure produced using Pymol.

6.3   Ligand binding assays with mutants

In order to use co-crystal structure determination to investigate ligand binding at the molecular level it was necessary to identify if the mutants bound the ligands. The ligand binding properties of the mutants were investigated using thermal shift assays. Due to the disorder observed at the top of the α4 helix, the F122T mutant was selected for these experiments over the Y118S mutant in order to ensure the protein structure was as close as possible to what would be expected for the wild type.

A selection of inhibitors was chosen for testing, incorporating predominantly flavonoid based inhibitors but also a NBF-GS conjugate, in order to evaluate compounds with potentially different binding poses within the H-site. The flavonoid ligands selected all contain the thiazole group, as these were most soluble (Figure 6-16). Ligands were selected which had shown varying levels of binding to the wild type protein to identify if the mutant responded in a similar manner to the wild type to the variously positioned functional groups.

MS 4-31-2

MS 5-35

MS 4-48

MS 5-6
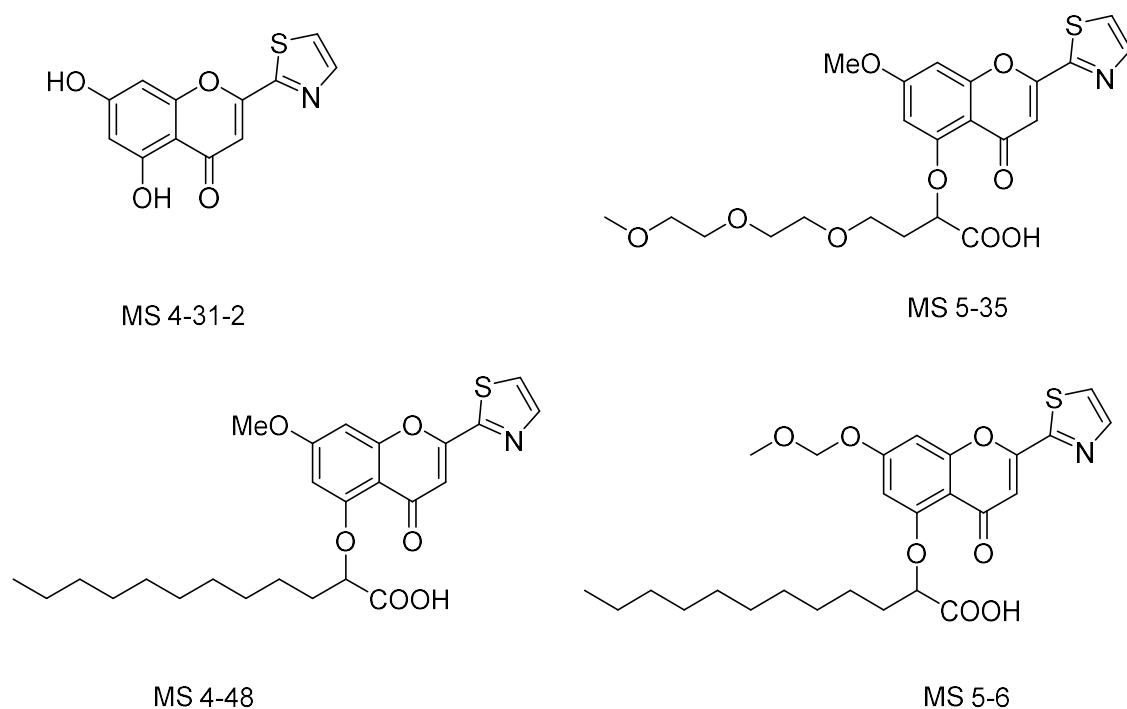
Figure 6-16: Inhibitors selected for testing in thermal shift assay with mutant *AmGSTF1*

For the F122T mutant, no thermal stabilisation is observed for any of the flavonoid inhibitors (Figure 6-17a), indicating that no binding occurs, suggesting that Phe122 was essential for

binding. This will be discussed further in Chapter 7. However, for the GSH-CNBF conjugate, a change in thermal stability was still observed (Figure 6-17b), with the F122T mutant binding in a destabilising manner.
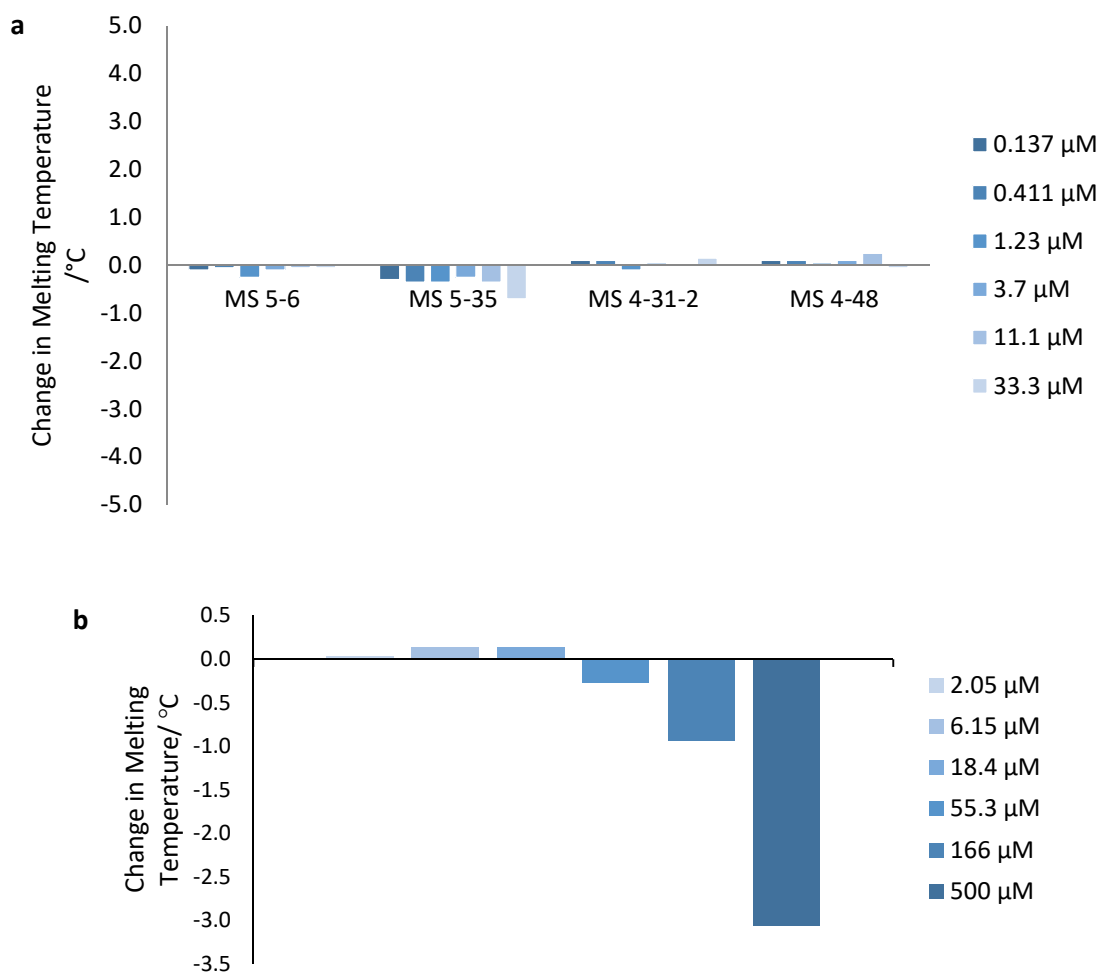
Figure 6-17: Change in melting temperature observed for F122T mutants using thermal shift assays with A) Flavonoid compounds and B) NBF-GS at a range of concentrations

As the flavonoid inhibitors had not been seen to bind to the F122T mutant they were also tested with the Y118S mutant, again no thermal stabilisation was seen suggesting the molecules did not bind (Figure 6-18). This was most likely as a result of the disorder in the upper portion of the α4 helix resulting in a change in binding.
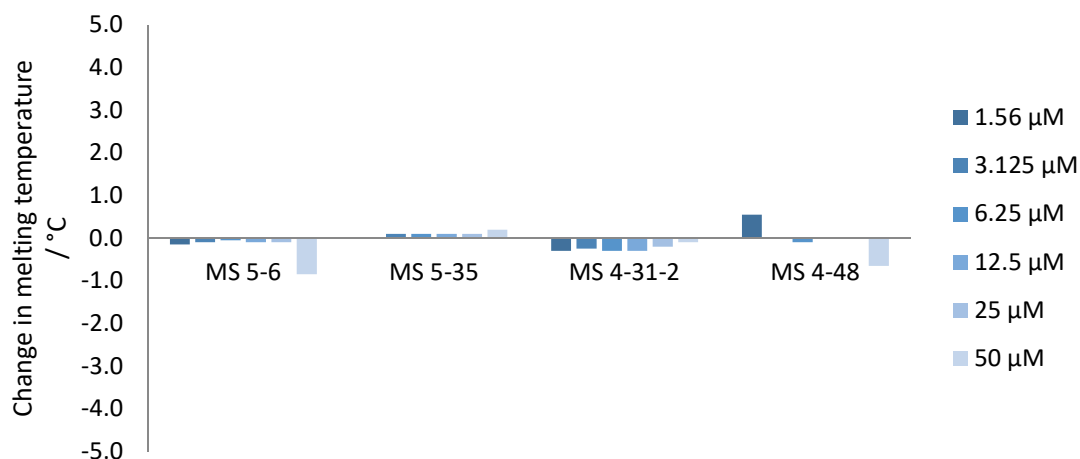
Figure 6-18: Change in melting temperature observed for Y118S mutants using thermal shift assays with flavonoid compounds

6.4 Cross-seeding with mutants

As the mutants were seen not to bind to the flavonoid inhibitors, it was necessary to resume working with the wild type protein, in order to identify whether this was due to an overall structural change between the mutants and wild type or as a result of the substituted amino acid. In some cases, groups have had success using homologous proteins for cross-seeding to encourage the crystallisation of proteins. This is particularly the case where sequence identity is high, however studies investigating metallo-β-lactamases have had successes with sequence similarity as low as 24%.[129] Seeding removes the need for spontaneous nucleation to occur, and as a result can lead to crystals forming in conditions in which they otherwise would not. By seeding with the mutant crystals, wild type protein would be more likely to form crystals in the modification of the mutant protein. As the F122T mutant had a fully ordered core structure it was selected for use for micro-cross-seeding matrix screening. This method uses micro-seeds to screen for crystallisation forms, across varied crystallisation conditions, using high throughput screens.[130–132]

6.4.1   Structure of wild type *Am*GSTF1

From high throughput screen conditions containing seeds of F122T mutant, crystals of wild type *Am*GSTF1 were produced. Crystals were produced in the same rod morphology as the F122T protein used to seed, as well as the prismatic morphology also seen for the F122T and Y118S crystals (Figure 6-19). As with the F122T mutant crystals of the rod morphology diffracted best, to 2.25 Å. Crystals were obtained in a tetragonal modification with a space group of I 4_1 2 2, the same as observed for the mutant structures. Prismatic crystals were also seen to be in a tetragonal modification with space group I 4_1 2 2, but these diffracted less well (to 2.8 Å). This may be as a result of a marginally larger unit cell C axis (104 Å as opposed to 99 Å).
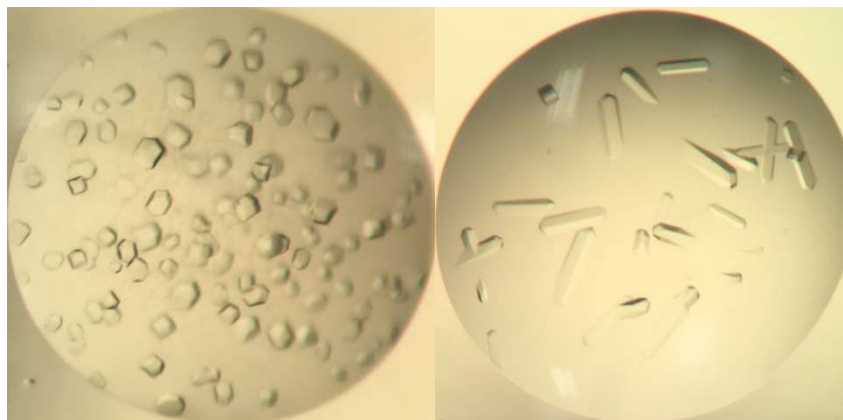
Figure 6-19: Crystal morphologies observed for WT *Am*GSTF1 after seeding with F122T protein-prismatic (left) and rods (right)

Density is seen for 214 residues (from Ala2 to Ala215), 24 water molecules and a GSH molecule. The model has an R factor of 0.23, an R_{free} of 0.28, and 4 Ramachadran outliers: Glu67 and Asp60 and two residues in the loop between α4 and α5. Unsurprisingly there is virtually no structural difference observed between the wild type structure and the F122T mutant, with an rmsd of 0.32 Å across 214 Cα atoms (Figure 6-20).

It is notable that the α4 helix adopts the same conformation as observed in the F122T mutant as opposed to the hexagonal form wild type crystal. This confirms that the conformation of this α-helix in the hexagonal wild type crystal, where this bends away from the H-site, is most likely as a result of the crystal packing observed. The other monomer packed in the active site probably results in the hexagonal form wild type structure adopting this conformation.

The location and orientation of the GSH molecule is seen to mirror that observed in the F122T mutant. It is seen to form the same interactions as in the mutant structure, and additional hydrogen bonding interactions are observed, mediated by two water molecules (Figure 6-21). The G-site is also found to be mostly unchanged from the hexagonal form wild type structure. The only major variations are that the Phe36 residue immediately before the loop over the G-site adopts an alternative conformation, and that the loop over the region is now visible.
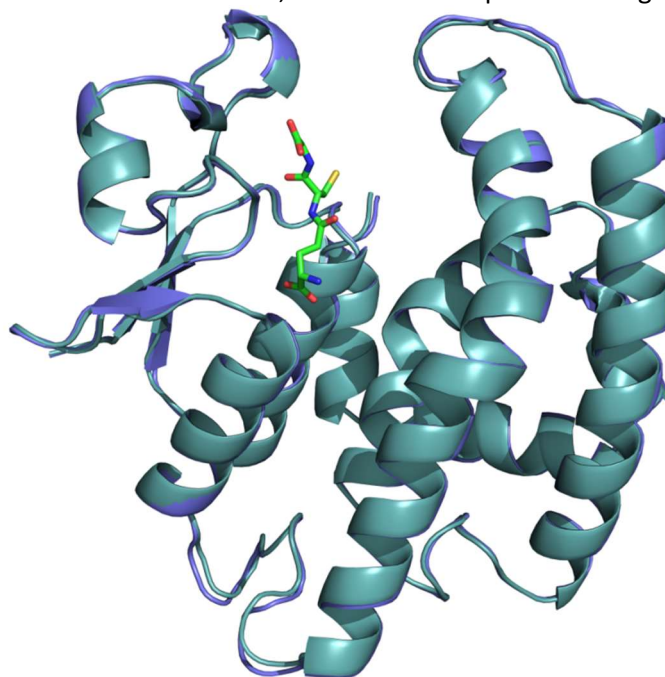
Figure 6-20: Superposition of structures of WT (light teal) and F122T mutant of *Am*GSTF1 with the GSH from WT (green). Figure produced using Pymol.
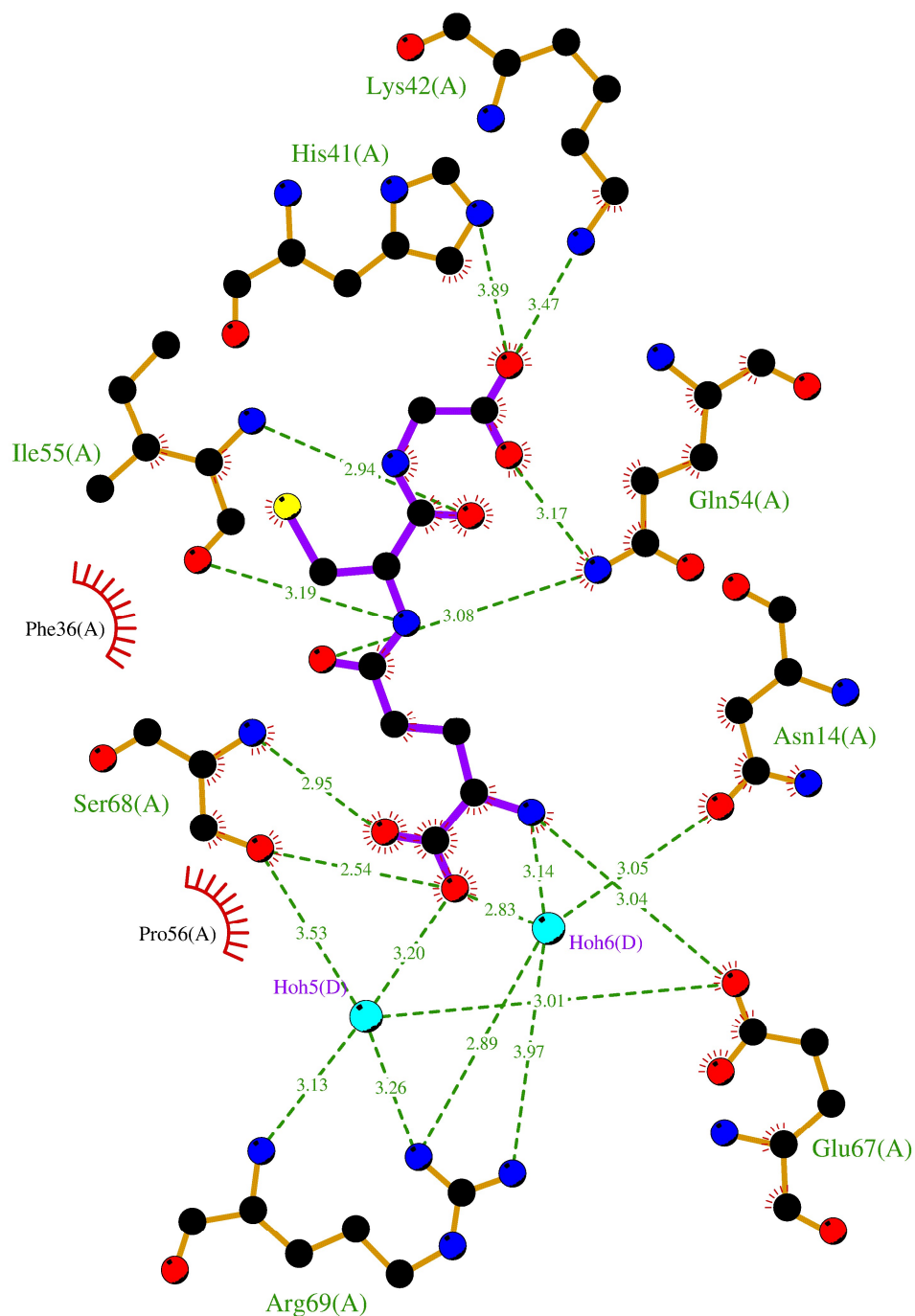
Figure 6-21: LigPlot to show the interactions observed between GSH and the active site of wild type *Am*GSTF1. Hydrogen bonds are represented as green dashed lines and hydrophobic interactions as red combs

The fact that the structure of this wild type *Am*GSTF1 is virtually identical to that of the F122T mutant also points towards a possible binding site for the flavonoid inhibitors. The only major difference is the presence of a Phe122 in the wild type as opposed to a Thr122 in the mutant. The fact that the F122T mutant is seen not to bind the flavonoid inhibitors suggests that this

aromatic Phe122 residue may be essential for forming π-π interactions with the aromatic rings in the flavonoid inhibitor.

6.4.2   Crystal structure determination of wild type *Am*GSTF1 with inhibitors

Having determined a complete structure of wild type *Am*GSTF1, it was desirable to use this to determine the structure of *Am*GSTF1 in complex with inhibitors. Soaking experiments were carried out using the tetragonal form crystals produced in seeding experiments. Experiments were carried out using both the flavonoid based ligands, in addition to a CNBF-GSH conjugate, to try to determine further the mode of action of both of these inhibitor types.

Flavonoid ligands were available as DMSO stocks and these were added to crystal conditions to give a compound concentration 5× that of the protein, and a final DMSO concentration of 5%. Unfortunately on addition of these stocks crystals were observed to crack and then dissolve. Controls were carried out using just DMSO to determine if this was as a result of the DMSO or whether the compounds binding to the protein was resulting in the crystal form becoming unstable. DMSO was added to drops at a variety of concentrations decreasing to 0.5% final concentration, and in each case crystals were seen to crack and dissolve. This suggests that these crystals are particularly sensitive to DMSO. As flavonoid ligands were only available as DMSO stocks, it was not possible to use them for further soaking experiments. In addition, their low solubility would make it difficult to use them for soaking in the absence of any DMSO, were solid forms available. This suggests that co-crystallisation may prove to be a better option for structure determination of *Am*GSTF1 in combination with a flavonoid ligand, as additional screening and optimisation would need to be carried out to identify crystals which are less DMSO sensitive.

The NBF-GS conjugate was available as a solid and as the conjugated GSH molecule means it is highly soluble this was suitable for use in soaking experiments. A small amount of solid was added directly to drops using a needle, drops were seen to rapidly turn yellow, and crystals were also seen to adopt a bright yellow colour.

6.4.2.1   Crystal structure of wild type *Am*GSTF1 in complex with NBF-GS

Following soaking in NBF-GS, a dataset was collected, and crystals were found to diffract to 2.30 Å. As with the GSH bound structure, crystals were in a tetragonal modification, with a space group of I 4_1 2 2. Further details on data processing are available in section 3.10.5 and Appendix B.
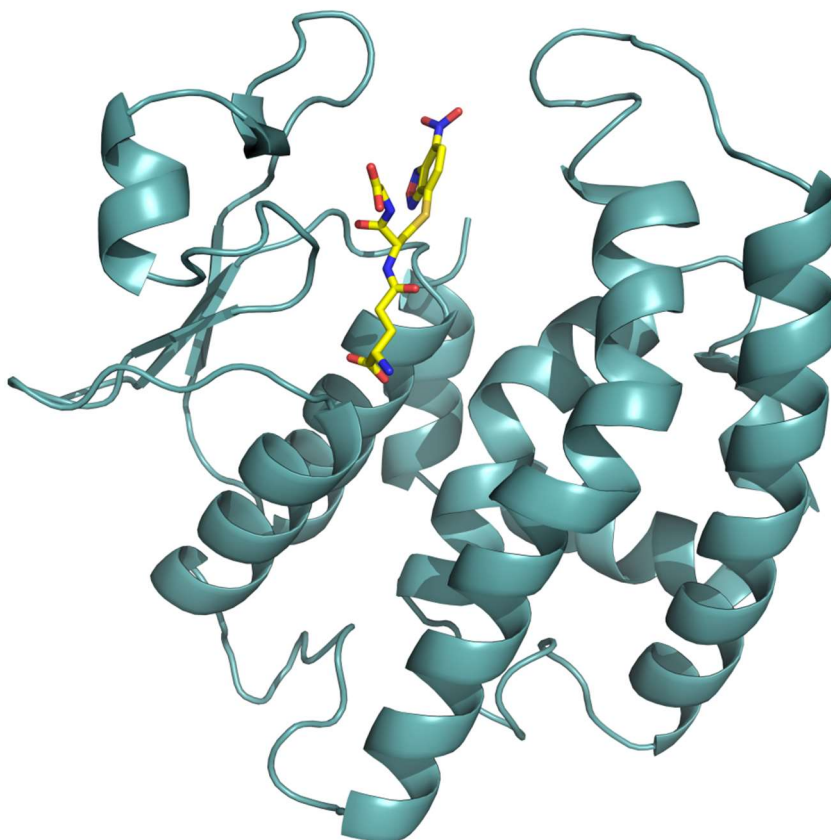
Figure 6-22: Structure of WT *Am*GSTF1 in complex with NBF-GS. Figure produced using Pymol.

Density was seen for 214 residues (from Ala2 to Ala215), and 8 water molecules (Figure 6-22). The model had an R factor of 0.21, an R_{free} of 0.25, and 4 Ramachadran outliers. The structure was found to be virtually identical to the GSH bound structure, with an rmsd of 0.23 across 214 equivalent Cα atoms. Following initial molecular replacement unbiased density for the NBF-GS molecule could clearly be seen in the active site (Figure 6-23).
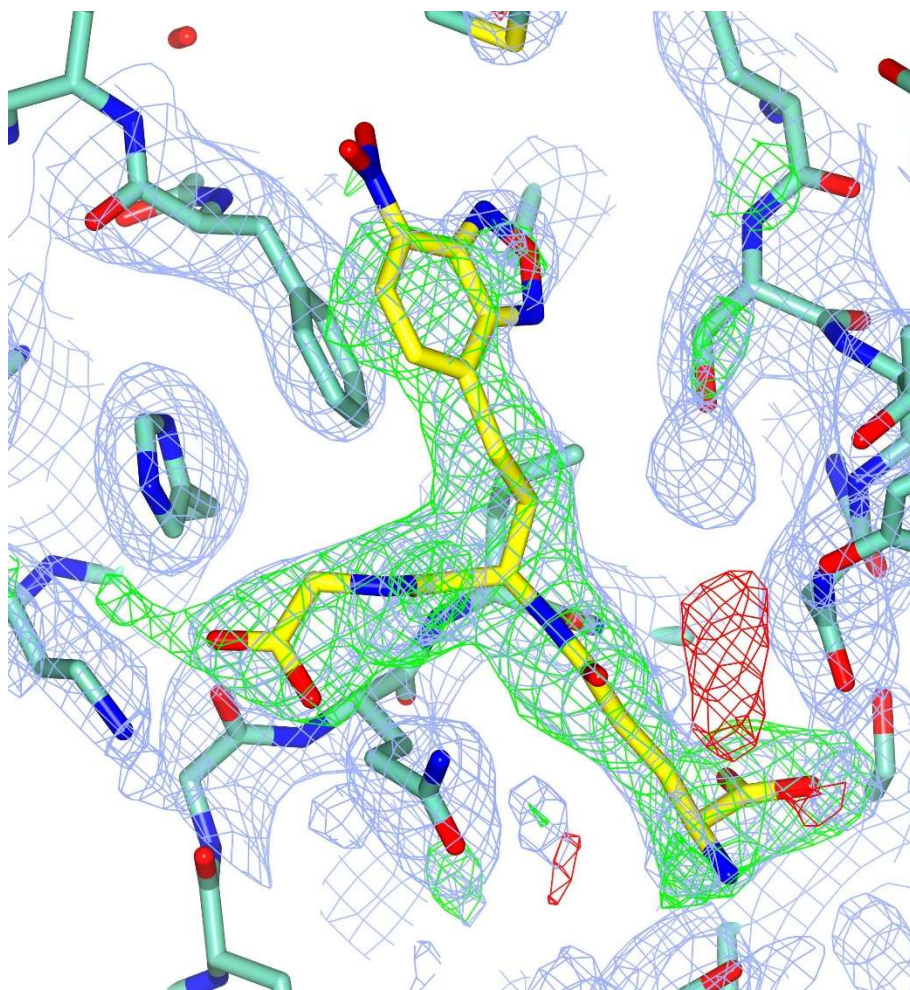
Figure 6-23: Omit map for NBF-GS ligand in _Am_GSTF1 structure. $2F_o$-F_c (blue) contoured at $\sigma=1$ and F_o-F_c (red and green) at $\sigma=3$. Figure produced using CCP4mg.

The NBF-GS molecule was seen to bind across both the G and H sites. The GSH portion of the molecule adopts the same conformation as the unconjugated GSH molecule seen in previous structures. The CNBF portion of the molecule is then seen to extend into the upper portion of the H-site (Figure 6-24). The aromatic portion of the molecule is seen to form a strong π-stacking interaction with Phe36, and the Phe36 is seen to rotate from its location in the structure with only GSH to allow this to occur (Figure 6-25). Whilst density for the Phe122 residue cannot be seen, it is probable that this would form an edge on face interaction, with a likely orientation for Phe122 lying within range for this interaction to occur. Additionally, hydrophobic interactions are seen to occur with Met126, Pro9 and the CNBF portion of the molecule (Figure 6-26).
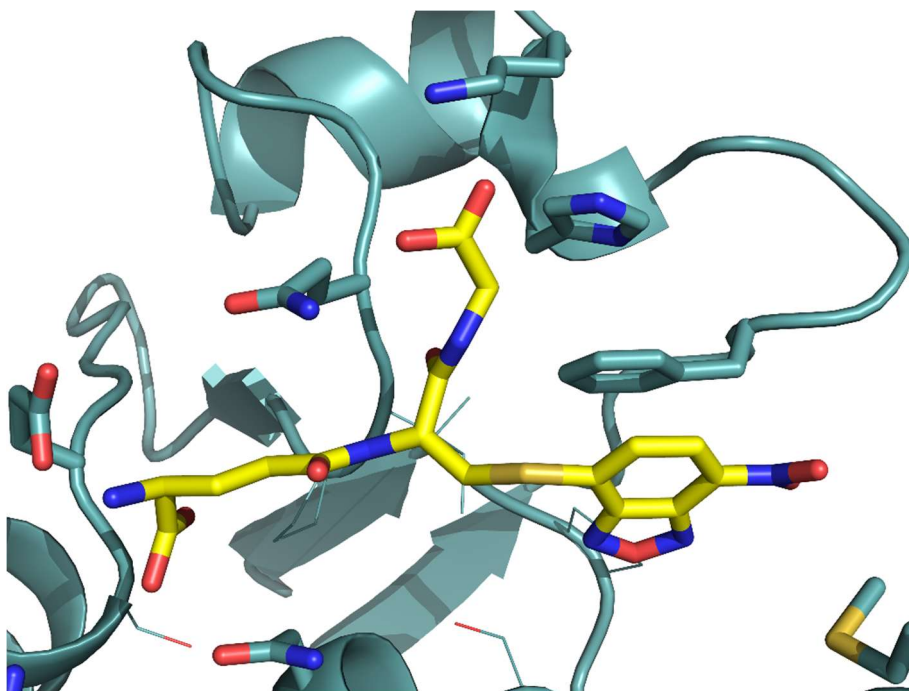
Figure 6-24: Interactions observed for NBF-GS (shown in yellow in stick form) binding in the active site of WT *Am*GSTF1 (shown in teal). Figure produced using Pymol.
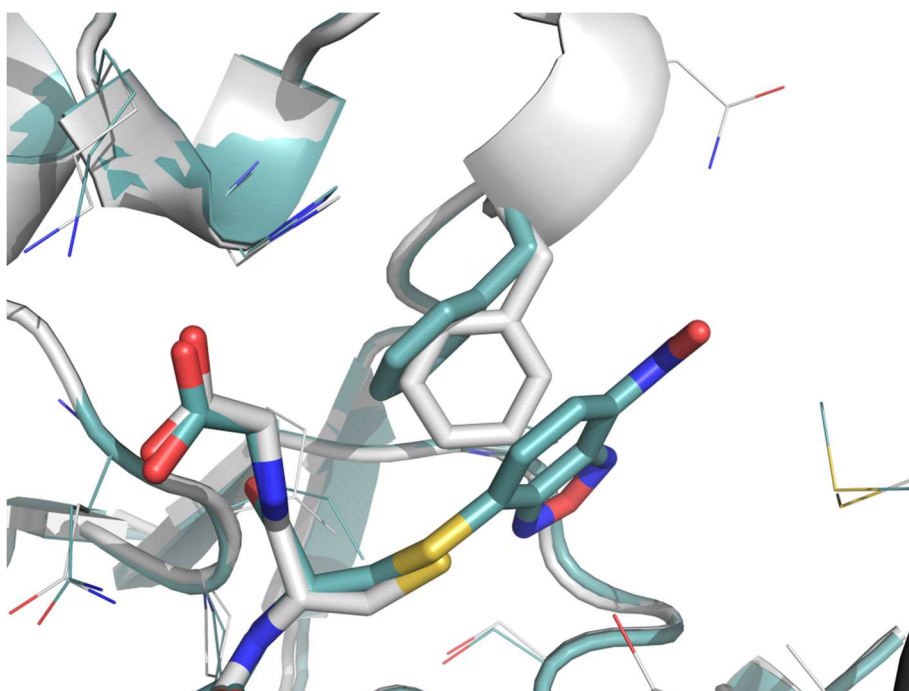
Figure 6-25: Movement of Phe36 observed in WT *Am*GSTF1 with (light teal) and without (grey) NBF-GS bound. Figure produced using Pymol.
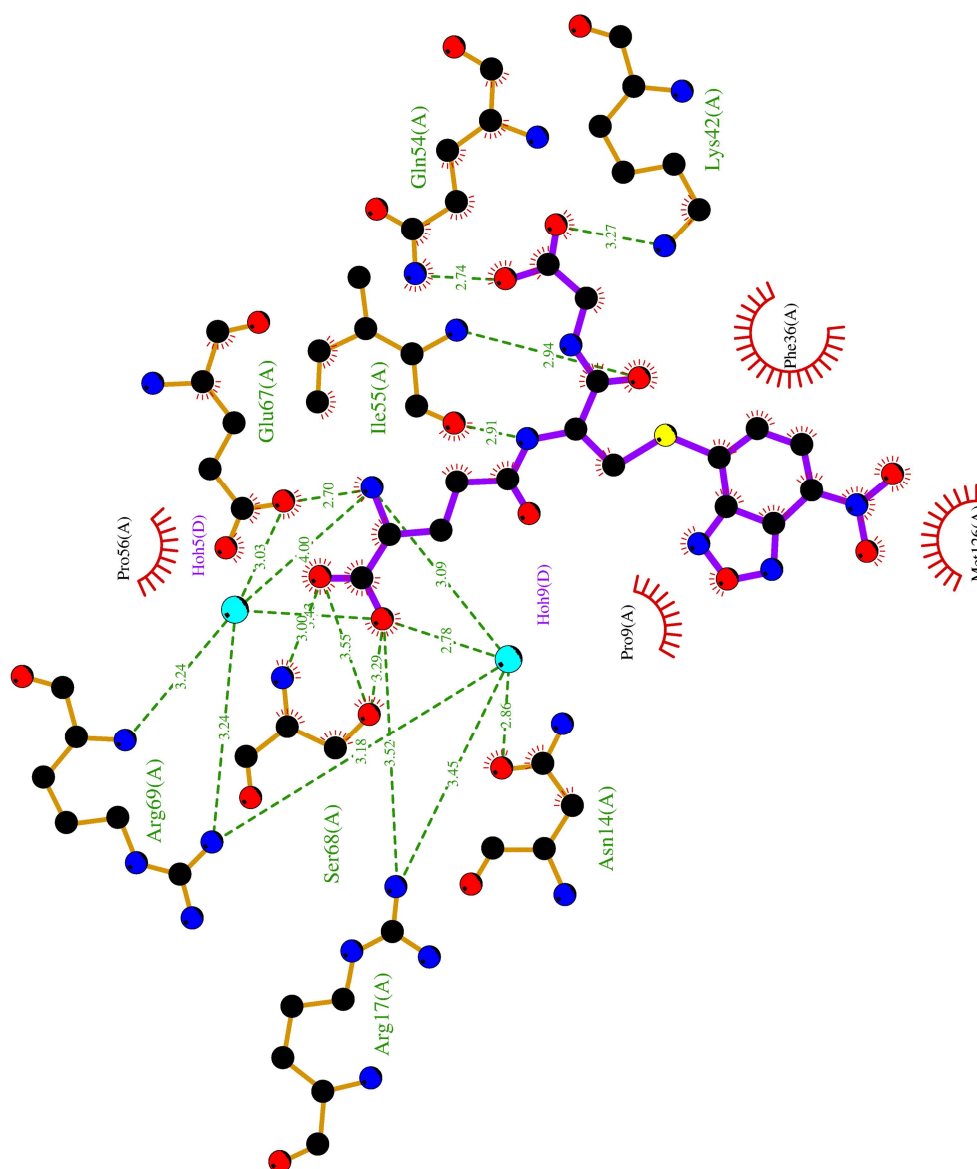
Figure 6-26: LigPlot to show interactions formed between the NBF-GS conjugate and the active site of _Am_GSTF1. Hydrogen bonds are shown as green dotted lines, and hydrophobic interactions as red combs

6.4.2.1.1 Comparison with *Hs*GSTP1 structure

The structure was then compared with structures of *Hs*GSTP1. As CNBF was originally identified as an inhibitor for this protein and MDR, it was of interest to determine whether it might be functioning in *Am*GSTF1 in the same manner. Interestingly this NBF-GS conjugate is seen to occupy a similar position to that observed of inhibitors in the active site of *Hs*GSTP1. Notably NBDHEX,[133] a derivative of CNBF, and the GSH conjugate of ethacrynic acid[123] occupy a similar location within the active site (Figure 6-27). In both these cases and the *Am*GSTF1 structure the ligands are seen to form interactions both with residues in the α4 helix and the loop over the G-site.
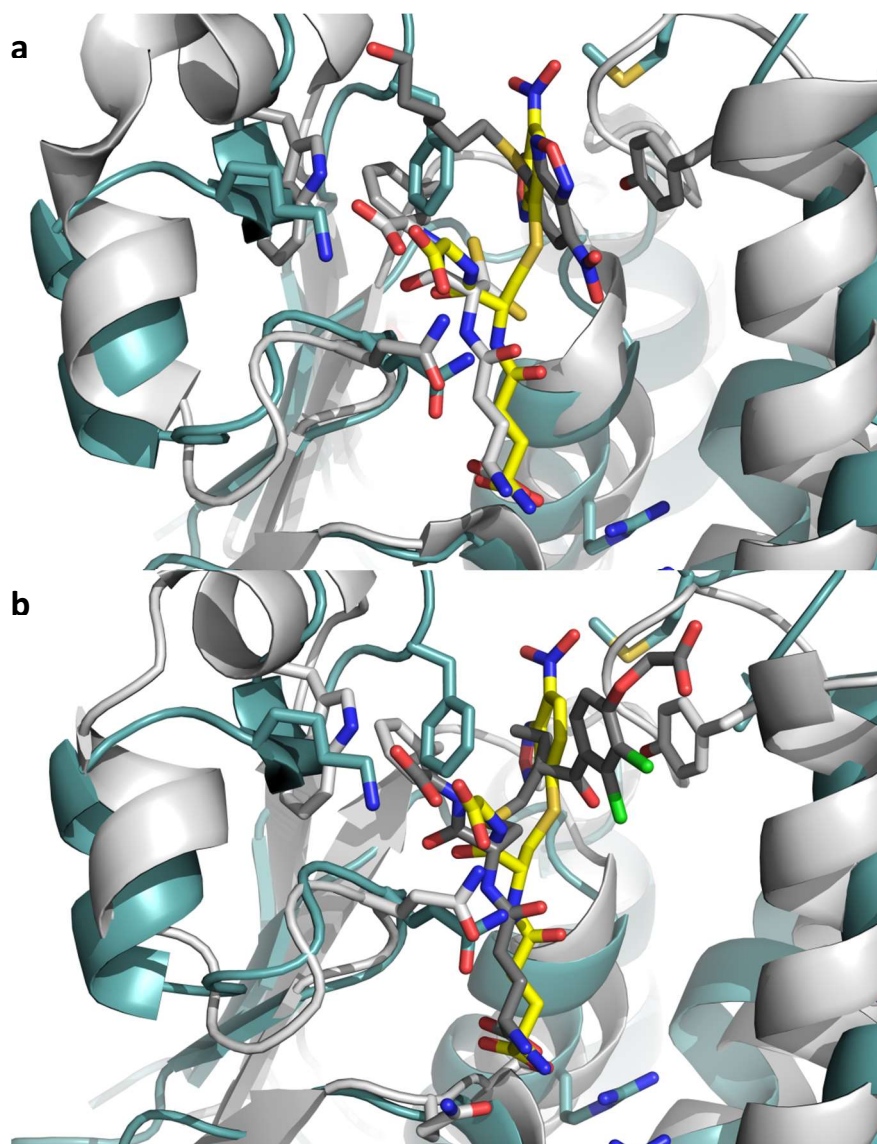
Figure 6-27: Comparison between binding site of NBF-GS (shown as yellow sticks) in AmGSTF1 (light teal) and HsGSTP1 (grey) binding site of (a)- NBDHex (PDB code:3GUS; dark grey sticks)[13] and (b) Ethacrynic acid-GSH conjugate (PDB code: 3GSS; dark grey sticks).[7] Figure produced using Pymol.

In *Hs*GSTP1 these compounds are thought to stabilise GSH binding in the active site by preventing its conjugation with natural substrate molecules and subsequent dissociation. This GSH bound form is thought to show a lower affinity to JNK1. Consequently stabilising the GSH bound form prevents *Hs*GSTP1 binding to JNK1 and in turn apoptosis.[83] That this CNBF conjugate binds in a similar location supports the hypothesis that this may also be functioning to block the active site and stabilise GSH binding. It is probable that this CNBF-GSH conjugate would work in conjunction with the covalent CNBF modification observed in the wild type protein following alkylation to achieve this.

6.5   Conclusions

In order to disrupt a crystal packing arrangement where the active site was blocked, preventing small molecule binding, a protein engineering approach was used. Two single site mutants, and one double mutant were designed to disrupt the interactions that held together this crystal packing arrangement. The two single site mutants were crystallised, and structures obtained at 2.8 Å for the F122T mutant, and 2.6 Å for the Y118S mutant. In both these cases, the designed mutations were sufficient to successfully disrupt crystal packing and the crystals were observed to be in a different modification. Notably, the active site was exposed with density for a GSH molecule observed in the G site. In addition, the previously disordered loop over the G site was seen to be ordered. These structures represented a major advance in the project. They allowed the structure of a complete active site to be determined for the first time, giving a complete model for *in silico* docking experiments to be carried out. An exposed active site made it possible to use crystals in this packing arrangement for soaking experiments to determine ligand binding sites. As the single site mutations were sufficient to disrupt crystal packing no attempt has yet been made to crystallise the double mutant, as it was preferable to use protein as consistent as possible with the wild type.

The presence of a GSH molecule confirms that the absence of small molecules in the previous wild type crystal form was as a result of crystal packing rather than lack of affinity for compounds used in the co-crystallography experiments. The GSH molecule had remained in the samples from protein purification and expression rather than being separately introduced. Even if the flavonoid molecules added to wild type protein had not bound, we would have expected to see a GSH molecule bound to the active site had it been accessible.

Thermal shift binding assays were carried out using the F122T and Y118S mutants to determine whether the flavonoid ligands were still seen to bind. In each case, no binding was observed. While this meant it would not be possible to use the mutants directly for soaking experiments, it did suggest that these residues were likely to be involved in flavonoid binding, identifying a probable binding location for these compounds in the H site.

To allow for future soaking experiments, in addition to obtaining a more biologically relevant structure, attempts were made to crystallise wild type *Am*GSTF1 in the same space group. By using seeds created from F122T crystals in the tetragonal modification, in random Microseed Matrix Screening in combination with wild type protein, it was possible to obtain crystals of wild type protein in the packing arrangement of the mutant. Using this method it was possible to grow crystals of wild type *Am*GSTF1 which diffracted to 2.25 Å, in the same space group and crystal packing arrangement as had been observed for the F122T crystals. This structure was seen to be virtually identical to that of the of the F122T mutant, confirming the biological relevance of this structure as a model for the wild type protein.

7 *In Silico* Studies of *Am*GSTF1

As no co-crystal structure of *Am*GSTF1 in complex with a flavonoid inhibitor was obtained, *in silico* docking studies were carried out to further investigate the mode of action of these compounds. For the majority of investigations the program GOLD was used.[134]

7.1 Introduction to GOLD

GOLD uses a genetic algorithm to create and score multiple ligand poses which converge to give one pose per "run". The user defines a number of runs, for example the default is 10, which would give 10 individual poses. Termination conditions are also set, which determine when the runs will stop. "Early termination" will result in runs stopping when a certain number of poses are within a certain RMSD, for example for this work, 3 solutions within 1.5 Å was used as a termination criteria. "Early termination" results in runs being stopped as soon as these conditions are met, regardless of the number of runs that were initially instructed.

GOLD has a number of built in fitness functions which are used to score poses; GoldScore, ChemScore, ASP and ChemPLP.[135] ChemPLP is the default function in GOLD and the fitness function used in this work, it will be examined in more detail here.

ChemPLP is an empirical fitness function optimised for pose prediction (Equation 7-1).

$$\text{fitnessPLP} = -\big(w_{PLP} \cdot f_{PLP} + W_{lig-clash} \cdot f_{lig-clash} + w_{lig-tors} \cdot f_{lig-tors} + f_{chem-cov} + w_{prot} \cdot f_{chem-prot} + w_{cons} \cdot f_{cons}\big)$$

$$\text{fitnessCHEMPLP} = \text{fitnessPLP} - (f_{chem-h} + f_{chem-cho} + f_{chem-met})$$

Equation 7-1: ChemPLP fitness function. This uses PLP adjusted with select bonding terms. PLP models steric complementarity, lig-clash is a heavy atom clash term, lig-tors is a ligand torsion potential. Chem-cov, a function for covalent docking, chem-prot which considers flexible side chains and waters, and cons which handles constraints are included as necessary.

PLP models the attraction and repulsion of protein and ligand heavy-atoms by designating atoms as: donor, acceptor, donor/acceptor, non-polar and metal. The interactions between these are then characterised (Table 7-1).[136]

Table 7-1: PLP interaction types

Ligand Atom Type	Protein Atom Type				
	donor	*acceptor*	*don./acc.*	*nonpolar*	*metal*
Donor	repulsive	H-bond	H-bond	buried	repulsive
Acceptor	H-bond	repulsive	H-bond	buried	metal
don/acc	H-bond	H-bond	H-bond	buried	metal
nonpolar	buried	buried	buried	nonpolar	buried

GOLD is a flexible docking program, and during the course of a run both the conformation of the protein and ligand will be altered. While ligands will rotate around any internal rotatable bonds, stereochemistry will not be altered, and as a result it is necessary to input any stereoisomers separately. At a basic level, residues in the protein capable of forming hydrogen bonds, Ser, Thr, Tyr and Lys, undergo hydrogen bond optimisation. It is further possible to increase flexibility by defining the side chains of up to 10 residues in the active site as "flexible". These can either be allowed to take on the structure of a rotamer from a library, either built in[137] or input, or can be given complete freedom to rotate around acyclic bonds.

In addition to increasing the flexibility of the protein, it is also possible to introduce a number of constraints into the docking. These can range from determining the distance between a protein and ligand, or forcing the presence of hydrogen bonds between protein and ligand in poses. These constraints are best used when a large amount of information is available about the mode of ligand docking, and as a result were not used for this work.

7.2 Docking using Chimeric GSTs

As a complete crystal structure had not originally been solved, with the initial wild type structure having two disordered loops, two options were available. Either to use an incomplete structure, or to create a chimeric structure using the incomplete structure for *Am*GSTF1 in conjunction with another structurally similar protein. As one of the hypothesised locations for the binding sites was located in the active site, defined by one of the disordered loops, it was decided to create a chimeric model to allow for most accurate docking.

7.2.1   Designing a chimeric *Am*GSTF1

While only one structure of *Am*GSTF1 existed prior to this work, there are a large number of crystal structures for other GSTs which can be found on the PDB. The sequence similarities of these GSTs were compared to *Am*GSTF1 to identify which would provide the most likely positioning of the loops. Only phi class GSTs were considered, as whilst GSTs show high structural similarity between classes the differences that do exist occur predominantly in these loop regions. Structures of phi class plant GSTs exist for *Arabidopsis thaliana*[50,138], *Populus trichocarpa* (poplar)[139] and *Zea mays* (maize).[56,124] A sequence alignment can be seen in Appendix D.

Of these the structures from maize were found to be the most similar, with a sequence similarity of 63.4%. A number of regions are highly conserved, and the structure shows an rmsd of only 1.27 Å for 182 equivalent Cα atoms from the partial *Am*GSTF1 structure. As a result they were used to create a chimeric structure.

Ideally it would have been possible to mutate the amino acids from the maize GST, to those present in the *Am*GSTF1 sequence, *in silico*. However, due to an amino acid insertion in one of the loops this was not possible. As a result ModWeb,[106,140] an online server that calculates likely structures of proteins based on the best available PDB model, was used to determine the most likely structure. Structures were created based on two maize structures: 1AXD and 1AW9. 1AXD is a structure for a Glutathione bound GST, while 1AW9 is for an apo-GST (Figure 7-1). The main difference between the structures exists in the loop over the G-site which is observed to undergo a large conformational change on GSH binding. As it was unknown whether the ligands would bind to the structure in the presence or absence of GSH, separate structures were produced based on each of the two conformations, for use in docking.
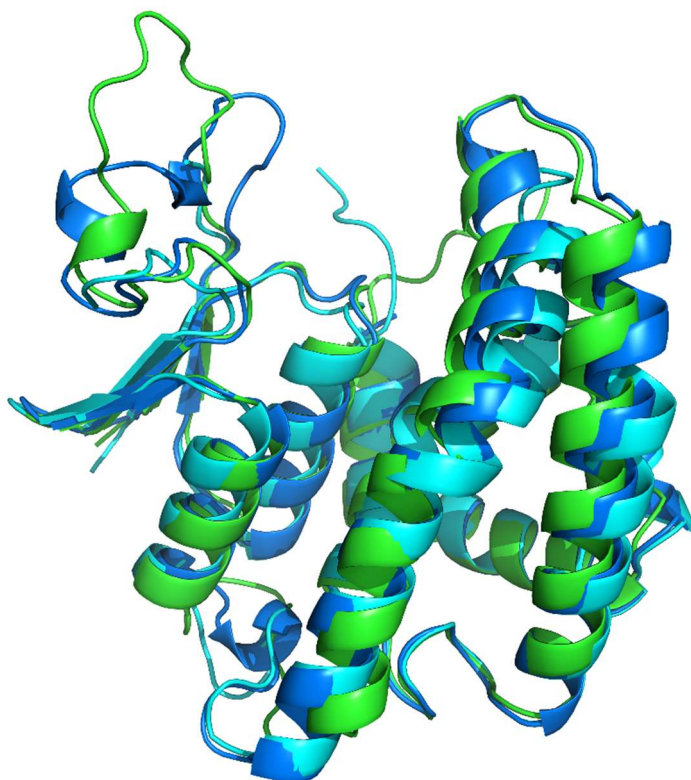
Figure 7-1: Superposition of *Am*GSTF1 (cyan) and *Zm*GSTF1 PDB code 1AXD (marine) and 1AW9 (green). Figure produced in Pymol.

In order to identify the most accurate possible structure for *Am*GSTF1, the loop regions created using ModWeb[106] were removed and inserted into the original *Am*GSTF1 structure. A round of geometry minimisation was then carried out in Phenix.[141] This idealises the geometry of a model according to standard geometry restraints, ensuring the bond geometry was standard in the areas where the loop regions of chimeric GSTs were joined. The chimeric GSTs produced were then used for docking in GOLD (**Error! Reference source not found.**).
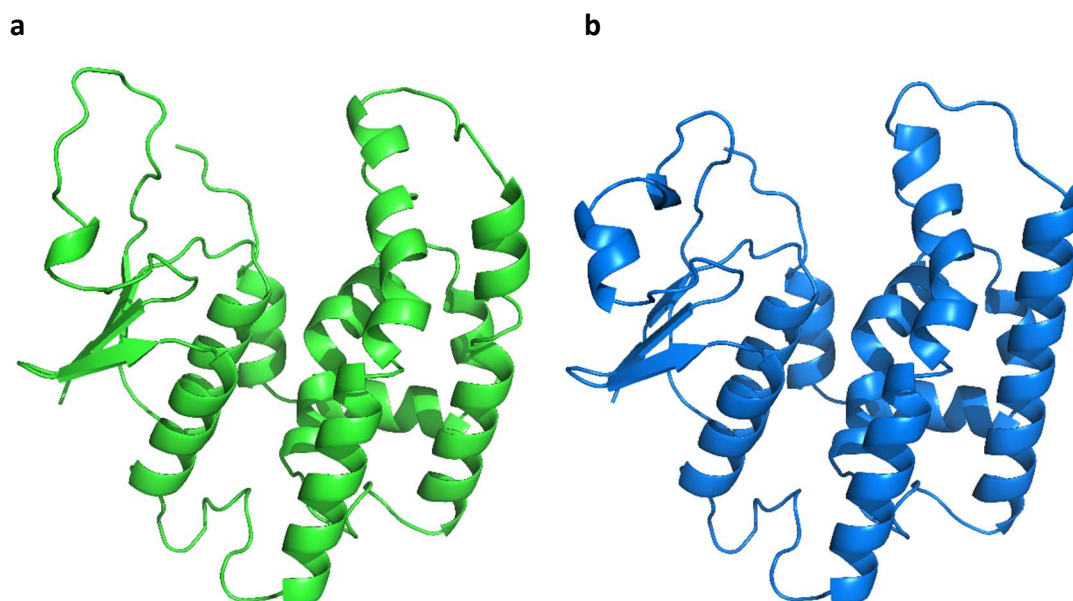
a   **b**

Figure 7-2: Chimeric structures of *Am*GSTF1 based on (a): PDB 1AW9 (green) and (b): PDB 1AXD (marine). Figure produced in Pymol.

7.2.2   Docking in the active site

Docking was initially carried out using both the 1AXD and 1AW9 chimeric models, to identify if either gave a more likely docking pose for an inhibitor in the active site. As insufficient density had been seen to model Tyr118, a residue in the H site in the original structure, this was inserted manually, and given flexibility within the docking to adopt any rotamer from the built in library. This library contains the most commonly observed side-chain conformations for naturally occurring amino acids.[137] This gave a complete structure of the G and H-site with all necessary relevant residues to be used in docking experiments (**Error! Reference source not found.**).
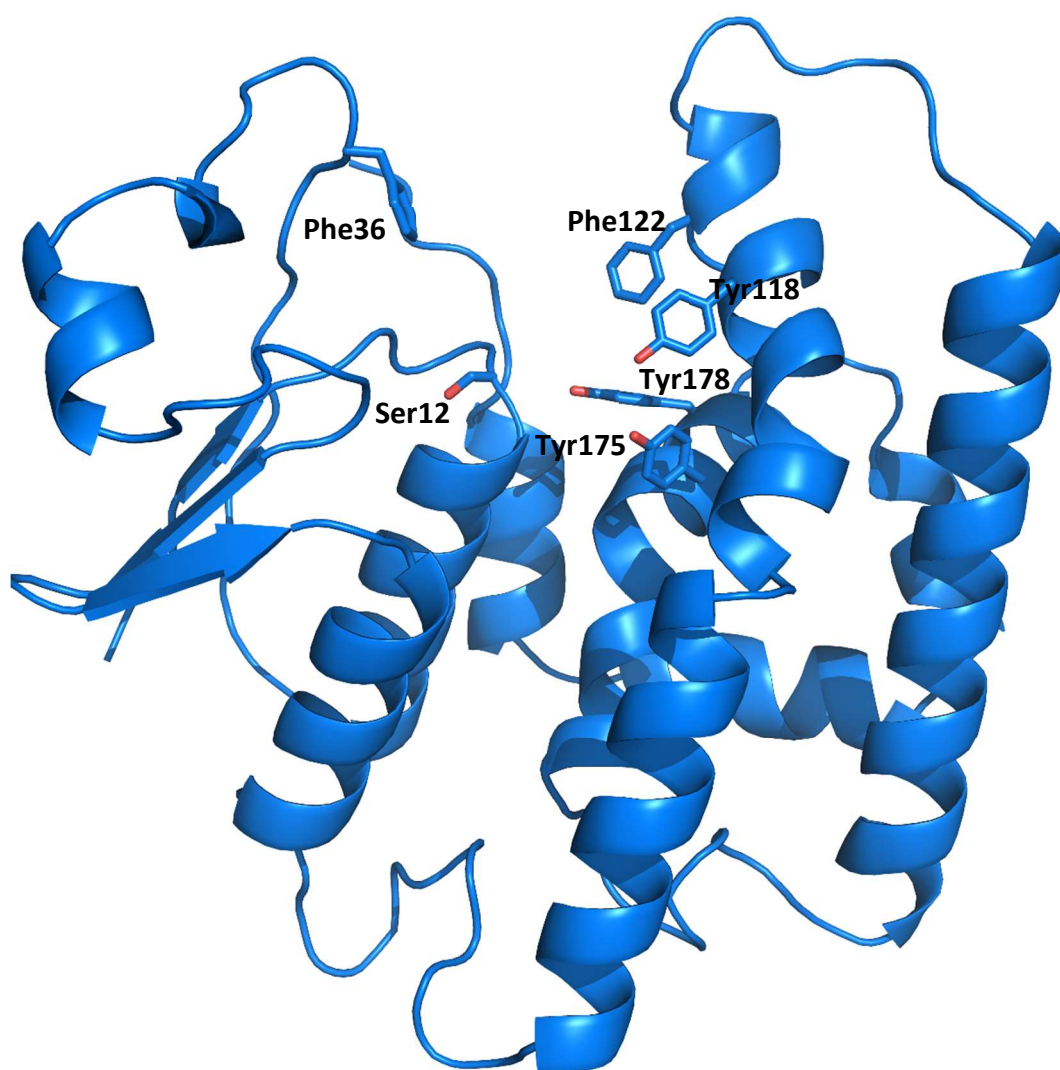
Figure 7-3: Structure of *Am*GSTF1 based on 1AXD, showing active site and location of key residues. Figure produced in Pymol.

Initial screening was carried out using two different flavonoid molecules (**Error! Reference source not found.**). Whilst it had been shown to bind less strongly, an inhibitor with a short alkyl chain at the C5 position was used (Error! Reference source not found.), as docking programmes are known to deal better with compounds with limited flexibility. In addition the inhibitor containing the full C_{10} chain was used (Error! Reference source not found.), to see if a satisfactory pose could be determined for this hydrophobic chain.
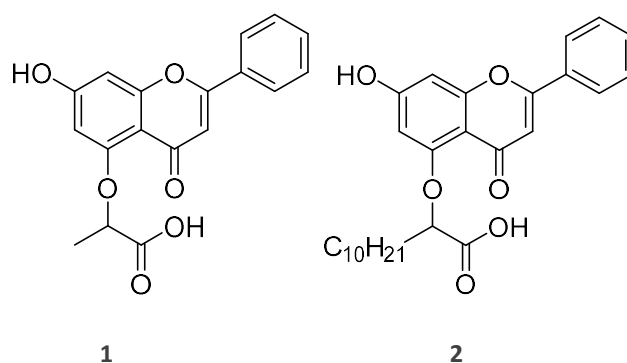
Figure 7-4: Inhibitors used for docking studies

For both proteins, the binding site was defined as a 10 Å radius around the catalytic Ser12 residue. As no information was known about binding pose, all waters were removed from the docking, and with the exception of the flexible Tyr118 residue all other settings were kept at their defaults, to avoid encouraging false interactions. ChemPLP was used as the scoring function, this is the newest GOLD scoring function, it is the fastest and has been seen to give the most accurate poses in most cases.[136]

Interestingly, the two different chimeric protein models produced very different binding poses, and these will be discussed here.

1AW9 (apo) based chimeric model

Using both compounds 1 and 2, the majority of poses observed had them located outside the G or H sites. For the short chain inhibitor, one pose was obtained with the inhibitor situated in the G-site, binding to the loop in the extended form which is observed in the absence of GSH (Figure 7-5). It is possible that the contacts formed by the inhibitor in this pose would act to stabilise the protein in this conformation, preventing GSH binding and therefore activity. However, as there is no known case of anything binding to the G-site with the exception of GSH or a GSH mimic, in which case a conformational change of this loop is seen, this seems unlikely. The lack of consistency observed between poses also suggests that this pose is unlikely to be adopted by the inhibitor *in vivo*.
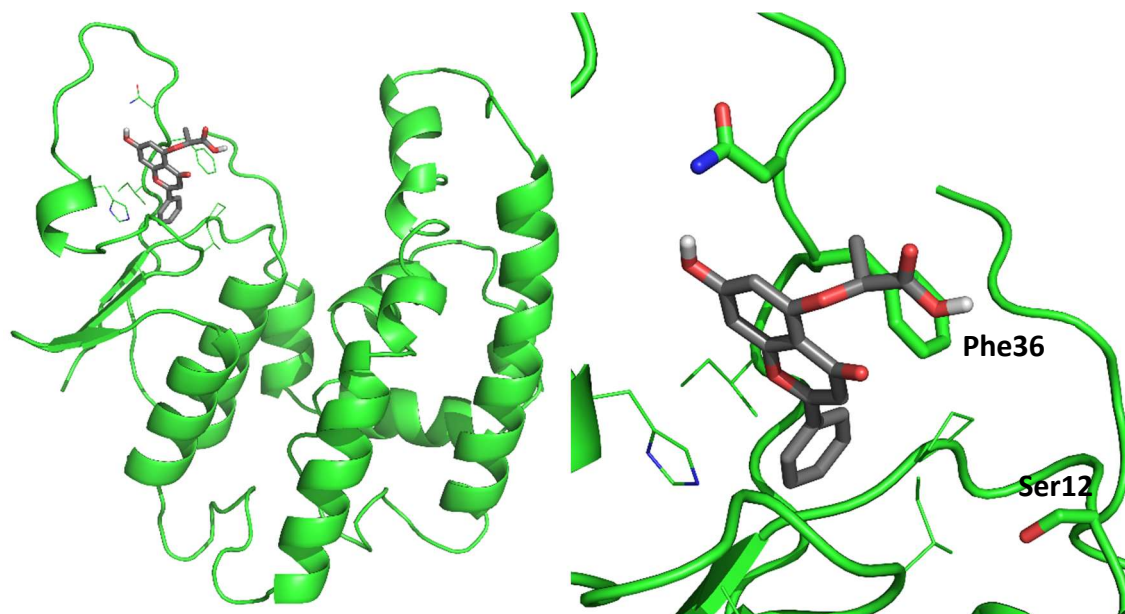
Figure 7-5: Docking pose for flavonoid inhibitor in G-site of 1AW9 chimeric structure. The structure is shown in green and the inihibitor in grey in stick representation. Residues most involved in binding are highlighted as sticks, in addition to Ser12. Figure produced in Pymol.

1AXD (ligand bound) based chimeric model

For compound **1**, a relatively consistent binding pose was identified in the H-site of the protein, with the inhibitor forming a range of interactions with the α4 helix of the protein (**Error! Reference source not found.**). The majority of poses for both stereoisomers are consistent, with the inhibitor forming π-π interactions with Phe122 and Tyr118. While these poses are consistent with the observation that the A ring must be aromatic,[85] it is seen to form a π-stacking interaction with Phe122, this pose is inconsistent with several of the other observations from the SAR data. The B-ring is not observed to form any π interactions, which would be expected as it needs to be aromatic, and the hydroxyl at the C7 position does not form a hydrogen bond. As a result it is relatively unlikely that the binding pose predicted represents an accurate pose for the flavonoid inhibitor within the active site.
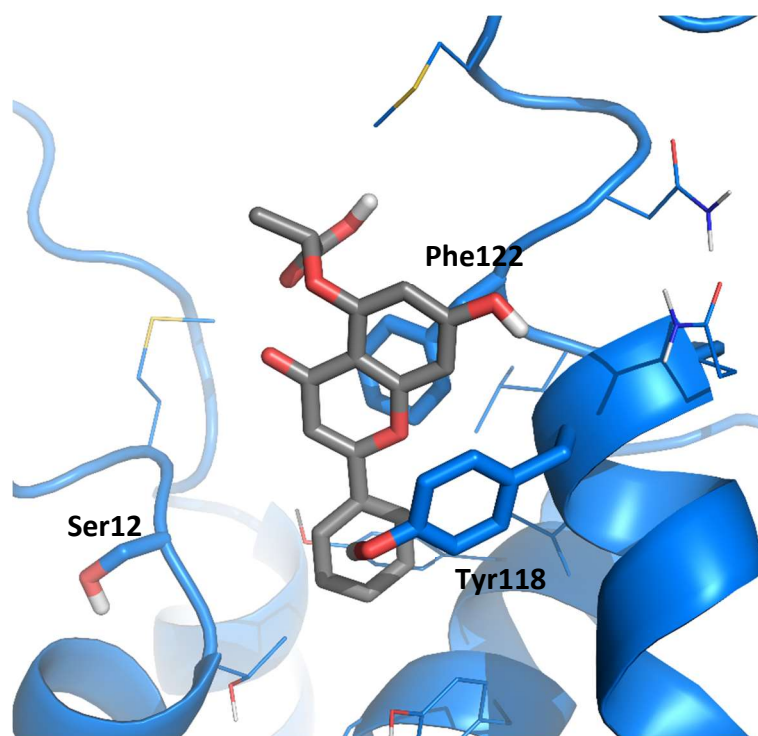
Figure 7-6: Pose for short chain ligand in H-site of chimeric 1AXD structure obtained for docking with rigid active site with flexible Tyr118 and Phe122 residues. Figure produced in Pymol.

For compound **2**, the poses are seen to be more varied, with an rmsd of up to 13.1 Å observed for the poses, as opposed to only 5.2 Å for the short chain inhibitor. While some of the poses observed for the core of the structure are similar to those observed for the short chain structure, with this part of the structure forming interactions with the α4 helix, the majority adopt different orientations where the majority of interactions are caused exclusively by the long chain. This would be inconsistent with the core of the molecule having a strong inhibitory effect on its own. In addition, these poses are clearly unsatisfactory, with a large number of short contacts observed between the protein and ligand in each case. As a result, further investigations using this model were carried out using exclusively compound **1**, as this had provided better poses. Once a suitable prospective pose had been identified for the core of the molecule, it was hoped it would then be possible to identify possible areas of hydrophobicity where the long alkyl chain might bind.

Having demonstrated the importance of the aromaticity of both the C-ring and B-ring in SAR studies, a second round of docking was carried out allowing all aromatic residues in the defined binding site freedom to adopt any rotamer within the built in library. It was likely that on ligand binding these residues might rotate to form optimal contacts, and allowing for this movement might give a more accurate portrayal of a possible binding pose.

For the two different isomers two different binding poses were observed. For the R isomer in the predominant pose the compound binds outside of the G or H site, adopting a position where it stacks with the Phe36 lying between it and the C-terminal tail of the protein (Figure 7-7). This pose seems unlikely; while it has a similar fitness score to those observed with fixed residue docking, it is hard to imagine how the molecule would have an effect from this location.
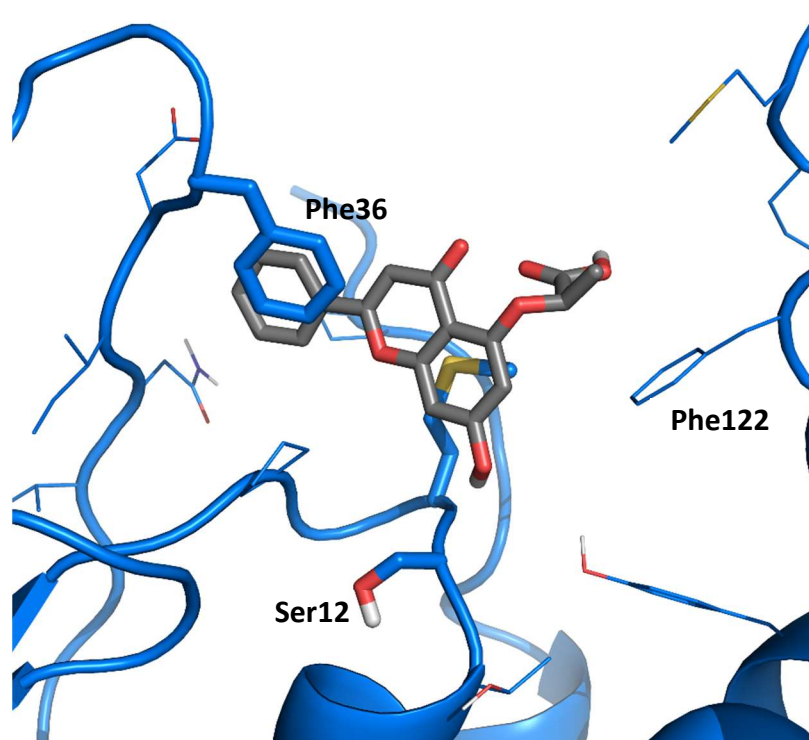

Figure 7-7: Docking pose obtained for R-isomer using Chimeric 1AXD structure with flexible aromatic active site residues. Figure produced in Pymol.

For the S-isomer a more probable binding location was observed, with the compound again placed in the H-site, but in a different orientation than had been observed for the rigid docking (Figure 7-8). The C-ring forms a π-stacking interaction with Phe122 and the acid group forms a hydrogen bond with Tyr175. The B-ring is close to forming π-interactions with either Phe177 or Tyr178, which would explain the increased activity seen with aromaticity in this ring. It is possible one of these residues would, in reality, adopt a slightly different conformation on ligand binding, allowing a strong π-π interaction to form.
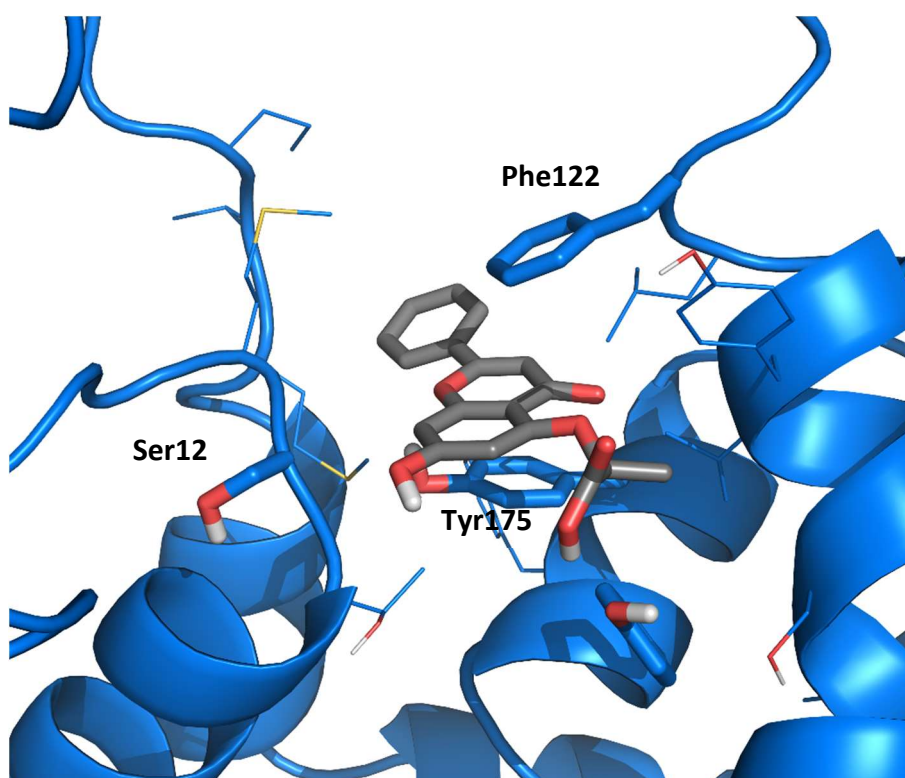
Figure 7-8: Docking pose obtained for S-isomer using Chimeric 1AXD structure with flexible aromatic active site residues. Figure produced in Pymol.

Overall, this pose for the S-isomer represents a likely pose observed for docking in the active site, indicating that the flavonoid ligands may bind to this area in the H-site.

7.2.3   Docking in the dimer interface

In addition to the active sites which are involved in conjugative activity, GSTs possess an additional binding site: the L-site. In *At*GSTF2, Quercetin, a flavonoid, has been identified as binding in this site,[50] and as a result this was considered as a possible binding site for our flavonoid inhibitors (**Error! Reference source not found.**). The L-site is most often involved in transport of molecules and as it was suspected *Am*GSTF1 might be causing MHR via a signalling effect[34] it was possible these inhibitors could be preventing binding of other small molecules to this site and thus any signalling effect.
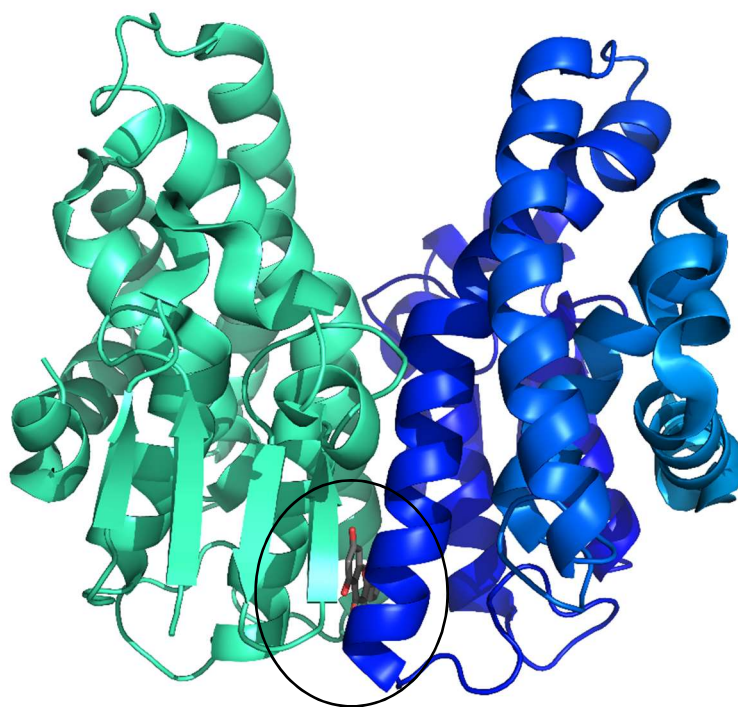
Figure 7-9: Structure of *At*GSTF2 (PDB: 5ADV) showing location of L-site with quercetin bound. Each monomer is shown as a different colour, and the quercetin molecule is in grey in a stick representation. Figure produced in Pymol.

A docking site was defined based on the binding site of quercetin in *At*GSTF2. An area of 10 Å around Glu89, which was located adjacent to the quercetin when *Am*GSTF1 was superimposed with *At*GSTF2, was used.

The fitness scores observed for poses in the L-site were significantly lower than those observed for the H-site (around 35-40 rather than 45-50). For each of the 10 poses of the inhibitor, the program placed the inhibitor on the surface of the protein as opposed to in a similar location to the quercetin in *At*GSTF2 (Figure 7-10). The fact that no satisfactory pose was identified for the inhibitor in the dimer interface suggests that this is unlikely to be its binding location.
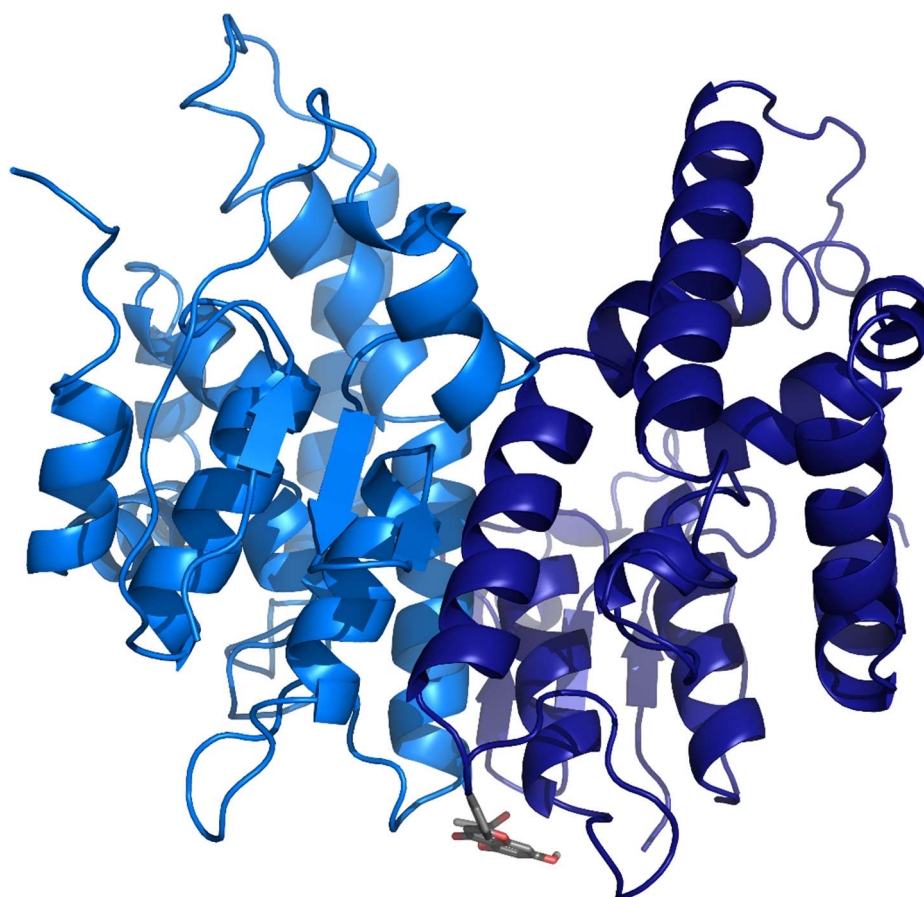
Figure 7-10: Pose obtained for docking in the L-site. The ligand is seen on the exterior of the protein as opposed to finding a stable pose within the L-site. The two monomers of *At*GSTF2 are shown in different shades of blue, and the quercetin molecule is shown in grey in stick representation. Figure produced in Pymol.

7.3   Docking using a mutant *Am*GSTF1 structure

Having obtained a structure of the F122T mutant, this was used for *in silico* modelling in place of the chimeric models. As the mutation was located in the active site, the Thr122 in the structure was mutated back to a Phe122, and this residue given flexibility in docking runs.

7.3.1   Comparison between chimeric and F122T structures

A number of significant differences were identified between the chimeric model and the F122T structure. As the F122T structure had GSH bound, it adopted a structure most similar to the chimeric structure based on PDB 1AXD, the ligand bound *Zm*GSTF1 structure. While the structures were similar, a number of differences resulted in an rmsd of 1.97 Å across 213

equivalent Cα atoms between the structures (Figure 7-11). The overall structure of the loop over the G-site was well predicted, with the fold very accurately matching that observed in the F122T structure, however, the placement varied significantly. The loop in the chimeric structure is seen to lie 3.5 Å from the position of the loop in the F122T structure, opening up the binding site. A difference is also observed in the positioning of the α4 helix and associated loop. The difference in the α4 helix observed between the different crystal forms has again resulted in the chimeric structure having a more open active site, with a small difference of 1.6 Å observed at the Phe122 position, but with this extending to a maximum of 4.8 Å at the Met126 position. These two changes result in the active site, where docking is taking place, adopting a significantly different shape in the chimeric GST as opposed to the F122T complete structure.
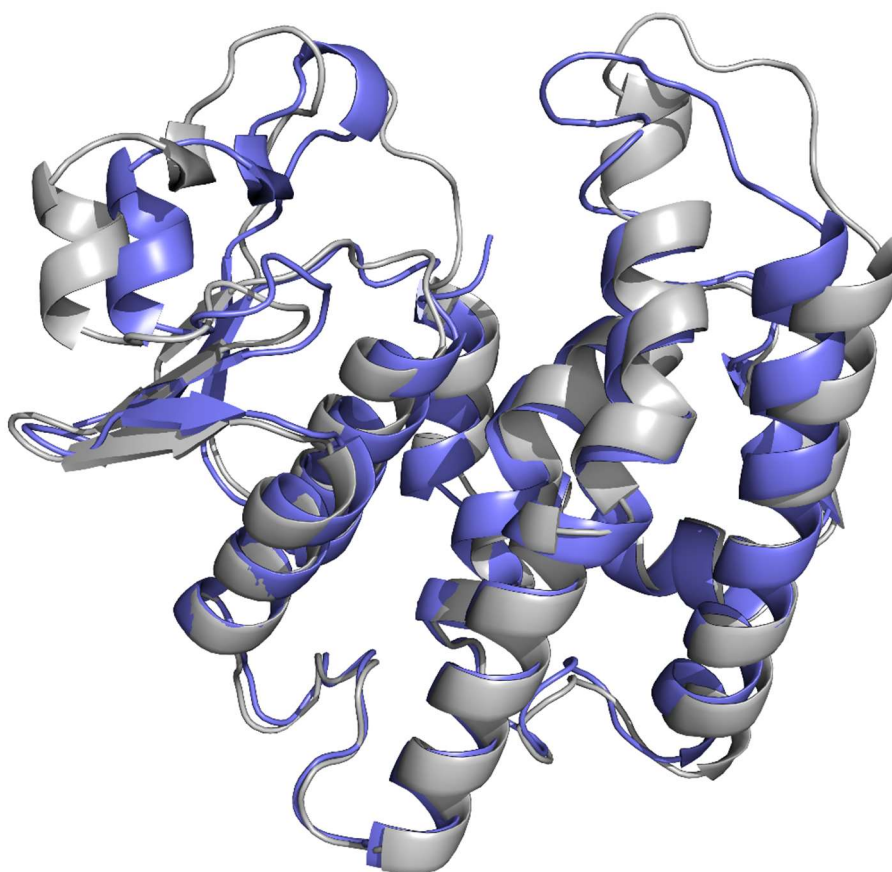

Figure 7-11: Superposition of Chimeric structure based on 1AXD (grey) and F122T mutant (slate). Figure produced in Pymol.

7.3.2   Docking results using F122T structure

Docking was initially carried out using a predominantly rigid setup, where only Tyr118 and the reinserted Phe122 were given flexibility. Tyr118 was included as poor density had been seen for it in the structure. As before, the flavonoid with the short alkyl chain was used for investigations. For the S-isomer, a very consistent pose was observed (Figure 7-12), with the docking terminating early as the top 3 poses were within 1.5 Å of each other.
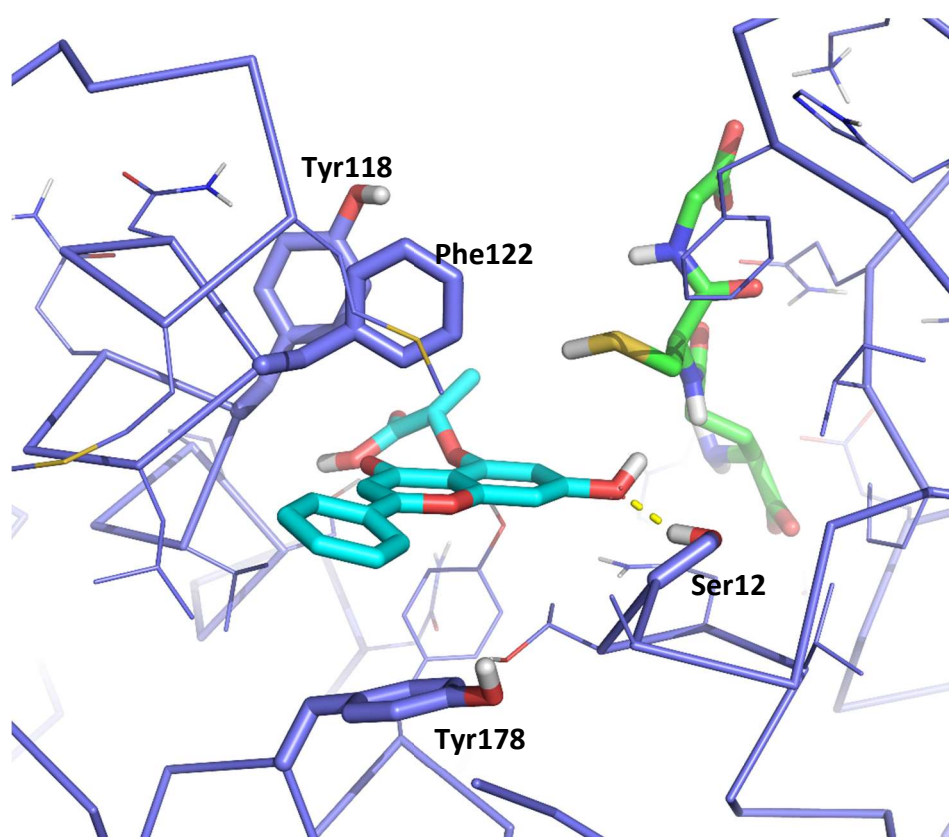
Figure 7-12: Docking pose observed for S-isomer using F122T structure with flexible Tyr118 and Phe122 residues. Key residues involved in binding are shown in stick representation, and a hydrogen bond formed between the C7-OH and Ser12 is shown as a yellow dashed line. GSH is shown in green and the flavonoid inhibitor in cyan, both in stick representation. Figure produced in Pymol.

For the R-isomer, while the run did not terminate early, the molecule was also found to adopt a relatively consistent position. The majority of solutions were found to lie within an rmsd of 4.4 Å, with the exception of two poses where the ring system is seen to flip (Figure 7-13). The level of consistency observed for both isomers is indicative of a better pose than was observed using the chimeric models, and the fitness scores observed are also significantly higher (around 75 as opposed to 50 for the chimeric models).
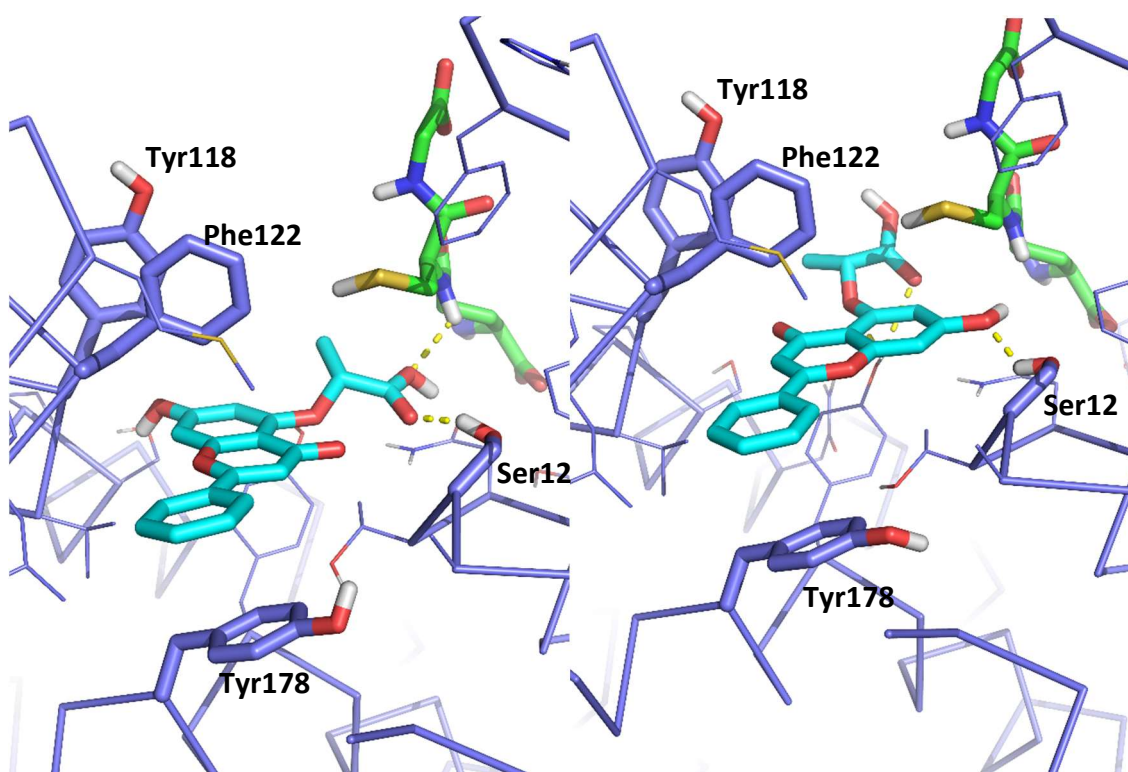
Figure 7-13: The two poses obtained for the central ring system for the R-isomer using the F122T structure with flexible Tyr118 and Phe122 residues. Residues involved in binding are shown in stick representation, and hydrogen bonds are shown as yellow dashed lines. GSH is shown in green and the flavonoid inhibitor in cyan, both in stick representation. Figure produced in Pymol.

The poses observed were also consistent with the SAR data. The C ring is observed to be involved in edge-face π interactions with Phe122, which is consistent with the need for this ring to be aromatic. In addition, using this new structure, the B-ring is seen to form π-interactions with Tyr178, consistent with the reduction of activity observed if this ring was not aromatic.

7.3.2.1   Flexible active site

Having achieved a consistent and probable pose, a second docking was carried out where the active site was given more flexibility. A selection of residues observed to form interactions with the inhibitor were given total flexibility in the docking to determine if changes might occur in the active site on compound binding.

With flexible residues, the ligand is seen to adopt an almost identical pose in the active site. As was expected, due to the consistency of previous poses no major changes in residue location were observed. However several small changes in conformation were observed. Most noticeably

the Phe122 residue is seen to adopt a different conformation where it forms stronger π-stacking interactions with the C-ring. In addition the Tyr178 residue is seen to move nearer to the B-ring, allowing for more efficient interactions to occur. The effect of these increased interactions can also be observed in the fitness score, which is significantly higher than that observed for the rigid docking, increasing to a maximum of 95 for the top pose.

Three predominant interactions are observed which are consistent between the R and S isomer. Firstly, and most significantly, a π-π interaction is observed between the C ring and the Phe122 residue. Aromaticity of this ring has been seen to be essential to compound binding, supporting the observation from docking that this would form π-interactions. Significantly, the F122T mutant is seen not to bind to the flavonoid compounds, further supporting this positioning for the flavonoid ring. Secondly, the oxygen of the C7 hydroxyl group is seen to form a hydrogen bond with the catalytic Ser12 residue, acting as a hydrogen bond acceptor. This is again consistent with the SAR, which demonstrates no significant loss in activity with alkylation of this group. The fact that addition of electron donating groups to this hydroxyl group result in a loss of activity can be explained due to sterics. While there is space in the region of this C7-OH group for additional groups, these would likely result in steric effects, explaining this activity decrease. Finally, the B-ring is seen to form a π stacking interaction with Tyr178. The fact that these interactions are consistent across all poses, as well as being consistent with the SAR studies give confidence in the docking (Figure 7-14).

Only one significant difference is observed between the poses for the two isomers. For the S-isomer, in the top ranked pose, a hydrogen bond network is seen to form with the α-alkoxy carboxylate group. Hydrogen bonds form between the hydroxyl group of the carboxylic acid and both Tyr175 and the amine group of Asn14, with this group acting as both a hydrogen bond acceptor and donor. A less extensive network is observed for the R-isomer, with the group forming a hydrogen bond with only the Tyr175 residue.
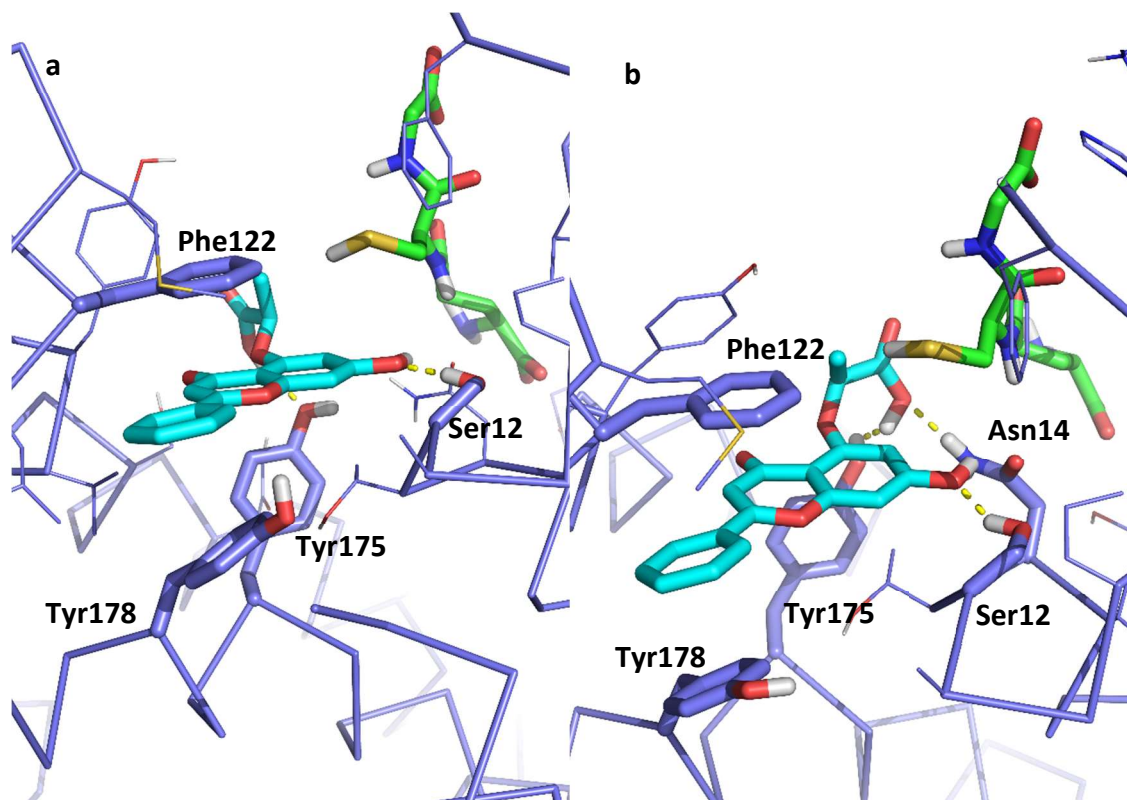
Figure 7-14: Poses achieved for docking into F122T structure with flexible active site residues (a): R-isomer (b): S-isomer. Protein is shown in slate with residues involved in binding shown in stick representation. The flavonoid inhibitor is shown in cyan, and the GSH molecule in green. Hydrogen bonds are shown as yellow dashed lines. Figure produced in Pymol.

Using inhibitor **2** with the long alkyl chain was once again unsuccessful. The large amount of flexibility in the inhibitor resulted in no consistent pose being observed. Of the poses that were observed none was identified as being significantly more plausible than the others. Using inhibitor **1** the pose observed has a hydrophobic region near its binding location where it is expected this hydrophobic alkyl chain would bind.

By way of confirmation this docking was also carried out using Glide,[142] another of the leading programs for *in silico* modelling. This resulted in a very similar pose being observed, giving further confidence in the top pose generated by GOLD. In addition, a 5 ns molecular dynamics simulation was carried out using the Desmond software package,[143] which demonstrated that the pose was energetically favourable, with the ligand remaining in the active site for the duration of the simulation.

This pose, with the flavonoid inhibitors forming a complex in the active site points towards a possible mode of action. In this pose this inhibitor would block the active site, preventing

substrate binding. It would also stabilise the GSH molecule bound in the active site of the protein. As a large structural change is seen to occur on GSH binding,[56] it is possible that in this GSH bound form *Am*GSTF1 would be unable to play its causative role in MHR. Notably this is seen to be the case in *Hs*GSTP1, with the GSH bound form unable to play its role in MDR.[83]

7.4 Conclusions

Initial *in silico* screening was carried out using chimeric *Am*GSTF1 models created from the partial *Am*GSTF1 structure, in addition to a model based on *Zm*GSTF1. Using this model it was possible to identify that the active site, and the H-site in particular, was most likely to be the site of flavonoid inhibitor binding. It was also possible to rule out the L-site as a probable location for inhibitor binding. However, using this model it was not possible to identify a pose consistent with all the SAR observed.

Having obtained a structure for the F122T mutant this structure, with the Phe122 residue reinserted, was used for all subsequent docking experiments. From these experiments it was possible to determine a probable site of flavonoid inhibitor binding. A pose was observed in the H-site which was consistent with the SAR observed from thermal shift assays and activity assays. In addition, as the pose prominently features π-stacking interactions with the Phe122 residue, it is further supported by the observation that the flavonoid compounds do not bind to the F122T mutant.

Docking experiments using compounds containing the long alkyl chain were found to be inconclusive. The docking programs do not respond well to the high degree of flexibility in this area, resulting in high levels of variation between poses. However, the poses observed for the core of the compound are situated in the vicinity of a highly hydrophobic region, and this may provide a likely binding site for this long alkyl chain.

Based on the pose identified as most probable from the docking experiments it is possible to propose a possible mode of action for the inhibitors. Evidence suggests that flavonoid inhibitors bind to the H-site, which would both block other compounds from binding, thus preventing the catalytic activity of the enzyme. In addition they would also act to stabilise the GSH bound form of the protein, which is also thought to be an important feature of MDR inhibitors for *Hs*GSTP1.[117]

8 Conclusions and Further Work

8.1 Apo structure of *Am*GSTF1

Building on previous work which had determined an apo crystal structure of AmGSTF1, this work extended the resolution to near atomic detail to 1.5 Å. The structure was seen to be typical of a phi class GST showing a high level of similarity to other phi class plant GST structures. While, as had been observed in the previous lower resolution structure, the crystal packing form resulted in two disordered loops, the higher resolution enabled density to be observed for an additional 3 residues in these loop regions. Additionally, the crystal packing in this modification was found to prevent small molecule binding in the active site. Extensive crystallisation experiments were undertaken to identify a new crystal form using new crystallographic conditions and co-crystallography, however no crystals were obtained of sufficient quality in any other packing form. Analysis of the interactions observed in the crystal using PISA[113] predicts that this non-biological interface is almost as strong as that seen in the biological dimer. The strength of this interaction provides a possible explanation for why it is observed so consistently.

As it was not possible to disrupt crystal packing using different crystallisation conditions a series of mutants were rationally designed. Crystals for two of these mutants were obtained, in both cases these were in a different crystal modification and space group to the wild type protein. The structures of the F122T and Y118S mutants were solved to a resolution of 2.8 Å and 2.6 Å respectively. For both mutants the loop over the active site could be seen and the active site was occupied by a GSH molecule. For the F122T structure the second loop region, between the α4 and α5 helices, was also seen to be ordered, although for the Y118S structure the mutation resulted in an associated region of disorder.

In order to obtain a wild type structure in the crystal packing form of the mutants, random matrix microseeding with seeds of the F122T mutant and wild type protein was carried out. This protocol produced crystals of wild type protein in the tetragonal modification, allowing for the first time the complete structure of wild type *Am*GSTF1 with ordered loops and a GSH molecule in the G-site, to be determined to 2.3 Å. This structure confirms the validity of the F122T mutant as a model for the wild type protein. Importantly, this new crystal form allowed for soaking experiments to be carried out to determine small molecule binding sites.

This method of cross-seeding wild type proteins with mutants is highly applicable for use in other protein systems. It allows for the possibility of mutating amino acids within the active site, which may be essential for binding or protein function, to gain a desirable crystal form. This desirable crystal form can then be used to seed wild type protein, giving a structure that is more biologically relevant and can be used for co-crystallography experiments.

As the structure has some residues in the H-site with insufficient density to model side chains, future work should focus on optimising crystal conditions to obtain a structure at higher resolution than the current 2.3 Å. This structure could also be a starting point for *in silico* ligand discovery, enabling novel inhibitors to be identified.

8.2   CNBF Inhibitors

CNBF had previously been identified to form an adduct with the Cys120 residue of AmGSTF1 using mass spectrometry. In this work, a crystal structure has been obtained at 2.0 Å. This is in the hexagonal crystal modification and has two disordered loops, including one in the active site area. This structure shows density at Cys120 for a covalently bound CNBF molecule albeit at lower occupancy. However, no other major differences were observed from the apo structure and it was not possible to determine the mechanism by which this covalent adduct was causing inhibition.

Binding assays using CNBF-modified and unmodified protein determined that CNBF did not prevent small molecule binding in the protein active site. However, using thermal shift assays it was not possible to ascertain if the covalent modification affected the binding constants for these molecules. Future work should use isothermal titration calorimetry to determine if CNBF covalent modification affects the binding constant of small molecules in the active site, in particular a NBF-GS conjugate.

By soaking wild type crystals obtained from cross-seeding with F122T protein, it was possible to obtain a co-crystal structure for *Am*GSTF1 in complex with a NBF-GS conjugate with the structure solved to 2.3 Å. The conjugate was seen to bind in the active site, occupying a very similar position to inhibitors for *Hs*GSTP1. Identification of this binding site in the active site suggests that CNBF may be working to inhibit the protein by two means possibly acting in combination; through covalent modification of the protein and by blocking the active site. This is supported

by work from biophysical studies demonstrating that CNBF modification at Cys120 does not prevent NBF-GS binding suggesting that *in vivo* the protein is likely to both be alkylated at Cys120 and have NBF-GS bound in the active site. Further work should focus on trying to obtain a crystal structure of *Am*GSTF1 with both a Cys120-NBF adduct and NBF-GS bound in the active site. Seeds of the F122T mutant should be used alongside CNBF-modified protein to promote the crystal packing form that allows for small molecule binding in the active site.

8.3 Flavonoid Inhibitors

Thermal shift assays were used to characterise the binding of ligands to *Am*GSTF1. This data acted in a complementary manner to the CDNB inhibition data that had previously been collected. It also validates the data from these inhibition assays, as the two orthogonal assays are in agreement. Using this data it was possible to carry out detailed SAR analysis, identifying which aspects of the flavonoid structure were essential for binding and activity.

In order to determine the likely binding site for flavonoid inhibitors the structure of the F122T mutant was used for *in silico* docking experiments. It was possible to propose a likely binding pose for flavonoid ligands in the H site of *Am*GSTF1. The pose agrees with the observations from SAR. In addition, the pose suggests that the flavonoid inhibitors form a π-interaction with Phe122. The F122T mutants, with this residue substituted for an aliphatic amino acid which could not form π interactions, did not bind flavonoid inhibitors. By contrast in thermal shift assays the F122T mutants were observed to bind NBF-GS. The structure of wild type protein in complex with NBF-GS demonstrates that Phe122 is not essential for NBF-GS binding, further supporting the argument that this residue plays a vital role in the binding site for flavonoid inhibitors.

Initial attempts at obtaining a co-crystal structure of *Am*GSTF1 in combination with a flavonoid ligand by soaking were unsuccessful as the crystals were highly intolerant of DMSO, and all flavonoids were available only as DMSO stocks. Future work should focus on obtaining a co-crystal structure using other methods. Soaking using solid compounds, the method used for the NBF-GS co-crystal structure, is a possibility. However due to limited solubility of the compounds this is unlikely to be successful. Due to this very limited solubility co-crystallisation is more likely to be successful than soaking. Crystallisation trials of protein with ligands in DMSO would identify conditions where crystals could grow in the presence of DMSO, allowing this solvent to be used to aid the solubility of the compounds. As *in silico* modelling was not able to provide a

conclusive binding position for the long alkyl chain contained within the compound, co-crystallography experiments would be particularly important with regards to determining where this may be binding. In turn this would help explain the role this functional group plays in increasing activity of the inhibitors.

In addition to co-crystallisation efforts, future work could focus on further validating the pose obtained from *in silico* modelling using orthogonal methods. Single site mutants could be created for those residues identified in docking experiments as playing a major role in binding. Binding assays using flavonoid inhibitors could then be carried out to determine the importance of the various residues in compound binding. While these tests could be done using thermal shift assays, any changes in T_m between mutant proteins would make comparison of binding difficult. Preliminary work using microscale thermophoresis with *Am*GSTF1 and flavonoid inhibitors showed promising results and this may prove a useful method to obtain comparable binding constants for mutant and wild type proteins.

8.4   Role of *Am*GSTF1

While the work presented in this thesis has allowed for a better understanding of the mode of action of inhibitors for *Am*GSTF1, in particular their binding site within the protein, it is still only possible to speculate on how these molecules function to inhibit Multiple Herbicide Resistance. This work suggests that inhibition can occur either via blocking the H-site (in the case of flavonoid inhibitors) or via covalent modification in combination with blocking the H-site (in the case of CNBF). It has been shown that both classes of inhibitors are not inherently phytotoxic to black grass but are able to act as synergists alongside herbicides to restore herbicidal activity. That the inhibitors give a response both *in vitro* against *Am*GSTF1 and *in vivo* as synergists supports the hypothesis that it is their action against *Am*GSTF1 which causes this response.

However, the role of *Am*GSTF1 in causing Multiple Herbicide Resistance is still poorly understood. Future work should focus on unravelling the overall function *Am*GSTF1 is playing in MHR. Preliminary work carried out in this thesis suggests that herbicides do not bind to *Am*GSTF1 supporting previous work that suggested that *Am*GSTF1 does not cause MHR by direct detoxification of herbicides.[34] As inhibitors have been identified as binding within the H-site this suggests possible roles for *Am*GSTF1 within MHR. It is possible that they are acting to inhibit activity of the protein towards an as yet unidentified molecule by blocking the H-site, supporting

the hypothesis that the protein might be acting in a signalling role. Alternatively *Am*GSTF1 may be involved in protein-protein interactions as identified for *Hs*GSTP1 in MDR. Future work should focus on identifying whether any other proteins may be involved in the MHR pathway. In this case the inhibitors may be acting in a manner similar to those for MDR by preventing these protein-protein interactions from occurring.

By further understanding the role *Am*GSTF1 is playing will in turn allow for better design of inhibitors targeting this mode of action. It will also allow for a greater understanding of the role played by both CNBF and the flavonoid inhibitors. Only when there is a deeper understanding of the role played by *Am*GSTF1 in Multiple Herbicide Resistance will it be possible to fully determine the mechanism by which these inhibitors function.

References

1. Quastel, J. H. 2,4-Dichlorophenoxyacetic Acid (2,4-D) as a Selective Herbicide. in *Agricultural control chemicals* **1**, 244–249 (American Chemical Society, 1950).

2. Davies, J. & Caseley, J. C. Herbicide safeners: a review. *Pestic Sci* **55**, 1043–1058 (1999).

3. Sherwani, S. I., Arif, I. A. & Khan, H. A. Modes of Action of Different Classes of Herbicides. *Herbicides, Physiology of Action, and Safety* (2015). doi:10.5772/61779

4. Délye, C., Jasieniuk, M. & Le Corre, V. Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* **29**, 649–658 (2013).

5. Weed Science Society of America-Herbicide Site of Action (SOA) Classification List. (2018).

6. Sandermann, H. Plant metabolism of xenobiotics. *Trends in Biochem Sci* **17**, 82–84 (1992).

7. Yuan, J. S., Tranel, P. J. & Stewart Jr, C. N. Non-target-site herbicide resistance: a family business. *Trends Plant Sci.* **12**, 6–13 (2007).

8. Schuler, M. A. & Werck-Reichhart, D. Functional Genomics of P450S. *Annu Rev Plant Biol* **54**, 629–667 (2003).

9. Schuler, M. A. Plant Cytochrome P450 Monooxygenases. *Crit Rev Plant Sci* **15**, 235–284 (1996).

10. L. Lamoureux, G., H. Shimabukuro, R. & Stuart Frear, D. Glutathione and glucoside conjugation in herbicide selectivity. in *Herbicide Resistance in Weeds and Crops* 227–261 (1991). doi:10.1016/B978-0-7506-1101-5.50020-6

11. Bowles, D., Isayenkova, J., Lim, E.-K. & Poppenberger, B. Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant. Biol* **8**, 254–263 (2005).

12. Lamoureux, G. L. & Rusness, D. G. Tridiphane [2-(3,5-dichlorophenyl)-2-(2,2,2-trichloroethyl)oxirane] an atrazine synergist: Enzymatic conversion to a potent glutathione S-transferase inhibitor. *Pest. Biochem. Physiol.* **26**, 323–342 (1986).

13. Martinoia, E., Grill, E., Tommasini, R., Kreuz, K. & Amrhein, N. ATP-dependent glutathione S-conjugate 'export' pump in the vacuolar membrane of plants. *Nature* **364**, 247–249 (1993).

14. Heap, I. *The International survey of herbicide resistant weeds*. (2018).

15. Powles, S. B. & Yu, Q. Evolution in Action: Plants Resistant to Herbicides. *Annu. Rev. Plant. Biol.* **61**, 317–347 (2010).

16. Vila-Aiub, M. M., Neve, P. & Powles, S. B. Fitness costs associated with evolved herbicide resistance alleles in plants. *New Phytol.* **184**, 751–767 (2009).

17. Pang, S. S., Duggleby, R. G., Schowen, R. L. & Guddat, L. W. The crystal structures of Klebsiella pneumoniae acetolactate synthase with enzyme-bound cofactor and with an unusual intermediate. *J. Biol. Chem.* **279**, 2242–2253 (2004).

18. McCourt, J. A., Pang, S. S., King-Scott, J., Guddat, L. W. & Duggleby, R. G. Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. *Proc Natl Acad Sci U S A* **103**, 569–573 (2006).

19. Gronwald, J. W. Lipid Biosynthesis Inhibitors. *Weed Sci.* **39**, 435–449 (1991).

20. Zhang, H. L., Tweel, B. & Tong, L. Molecular basis for the inhibition of the carboxyltransferase domain of acetyl-coenzyme-A carboxylase by haloxyfop and diclofop. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 5910–5915 (2004).

21. Délye, C., Zhang, X.-Q., Michel, S., Matéjicek, A. & Powles, S. B. Molecular bases for sensitivity to acetyl-coenzyme A carboxylase inhibitors in black-grass. *Plant Physiol.* **137**, 794–806 (2005).

22. Dill, G. M. Glyphosate-resistant crops: history, status and future. *Pest Manag Sci* **61**, 219–224 (2005).

23. Sammons, R. D. & Gaines, T. A. Glyphosate resistance: state of knowledge. *Pest Manag Sci* **70**, 1367–1377 (2014).

24. Healy-Fried, M. L., Funke, T., Priestman, M. A., Han, H. & Schönbrunn, E. Structural basis of glyphosate tolerance resulting from mutations of Pro101 in Escherichia coli 5-enolpyruvylshikimate-3-phosphate synthase. *J. Biol. Chem.* **282**, 32949–32955 (2007).

25. Gaines, T. A. *et al.* Gene amplification confers glyphosate resistance in Amaranthus palmeri. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1029–1034 (2010).

26. Feng, P. C. C. *et al.* Investigations into glyphosate-resistant horseweed (Conyza canadensis): retention, uptake, translocation, and metabolism. *Weed Sci.* **52**, 498–505 (2004).

27. Yu, Q., Han, H., Nguyen, L., Forster, J. W. & Powles, S. B. Paraquat resistance in a Lolium rigidum population is governed by one major nuclear gene. *Theor. Appl. Genet.* **118**, 1601–1608 (2009).

28. Plaisance, K. L. & Gronwald, J. W. Enhanced catalytic constant for glutathione S-transferase (atrazine) activity in an atrazine-resistant Abutilon theophrasti biotype. *Pest. Biochem. Physiol.* **63**, 34–49 (1999).

29. Powles, S. B., Wilson, H. P. & Westwood, J. H. Herbicide cross resistance and multiple resistance in plants. Available at: https://hracglobal.com/files/Herbicide-Cross-Resistance-and-Multiple-Resistance-in-Plants.pdf. (Accessed: 4th January 2019)

30. Whaley, C. M., Wilson, H. P. & Westwood, J. H. A New Mutation in Plant ALS Confers Resistance to Five Classes of ALS-Inhibiting Herbicides. *Weed Science* **55**, 83–90 (2007).

31. Kemp, M. S. (University of B., Moss, S. R. & Thomas, T. H. Herbicide resistance in Alopecurus myosuroides. *ACS. Symp. Ser. Am. Chem. Soc.* (1990).

32. Letouzé, A. & GASQUEZ, J. Enhanced activity of several herbicide-degrading enzymes: a suggested mechanism responsible for multiple resistance in blackgrass (Alopecurus myosuroides Huds.). *Agronomie* **23**, 601–608 (2003).

33. Cummins, I., Cole, D. J. & Edwards, R. A role for glutathione transferases functioning as glutathione peroxidases in resistance to multiple herbicides in black-grass. *Plant J.* **18**, 285–292 (1999).

34. Cummins, I. *et al.* Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5812–5817 (2013).

35. Öztetik, E. A Tale of Plant Glutathione S-Transferases: Since 1970. *Bot. Rev* **74**, 419–437 (2008).

36. Sappl, P. G. *et al.* The Arabidopsis glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress: Genomic and reverse genetic analysis of plant GSTs. *Plant J.* **58**, 53–68 (2009).

37. Basantani, M. & Srivastava, A. Plant glutathione transferases — a decade falls short. *Can. J. Bot.* **85**, 443–456 (2007).

38. Frova, C. The plant glutathione transferase gene family: Genomic structure, functions, expression and evolution. *Physiol. Plant.* **119**, 469–479 (2003).

39. Salinas, A. E. & Wong, M. G. Glutathione S-transferases--a review. *Curr. Med. Chem.* **6**, 279–309 (1999).

40. Bartling, D., Radzio, R., Steiner, U. & Weiler, E. W. A glutathione S-transferase with glutathione-peroxidase activity from Arabidopsis thaliana. *Eur. J. Biochem.* **216**, 579–586 (1993).

41. Eshdat, Y., Holland, D., Faltin, Z. & Ben-Hayyim, G. Plant glutathione peroxidases. *Physiol. Plant.* **100**, 234–240 (1997).

42. Cummins, I., Moss, S., Cole, D. J. & Edwards, R. Glutathione transferases in herbicide-resistant and herbicide-susceptible black-grass (Alopecurus myosuroides). *Pestic. Sci.* **51**, 244–250 (1997).

43. Thom, R., Dixon, D. P., Edwards, R., Cole, D. J. & Lapthorn, A. J. 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 catabolism1. *J Mol Biol* **308**, 949–962 (2001).

44. Marrs, K. A., Alfenito, M. R., Lloyd, A. M. & Walbot, V. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature* **375**, 397–400 (1995).

45. Chen, Z., Silva, H. & Klessig, D. F. Active oxygen species in the induction of plant systemic acquired resistance by salicylic acid. *Science* **262**, 1883–1886 (1993).

46. Kauss, H. & Jeblick, W. Pretreatment of Parsley Suspension Cultures with Salicylic Acid Enhances Spontaneous and Elicited Production of H2O2. *Plant Physiol.* **108**, 1171–1178 (1995).

47. Dudler, R., Hertig, C., Rebmann, G., Bull, J. & Mauch, F. A pathogen-induced wheat gene encodes a protein homologous to glutathione-S-transferases. *Mol. Plant Microbe Interact.* **4**, 14–18 (1991).

48. Litwack, G., Ketterer, B. & Arias, I. M. Ligandin: a Hepatic Protein which Binds Steroids, Bilirubin, Carcinogens and a Number of Exogenous Organic Anions. *Nature* **234**, 466–467 (1971).

49. Jones, A. M. Auxin-Binding Proteins. *Annu. Rev. Plant. Physiol. Plant Mol. Biol* **45**, 393–420 (1994).

50. Ahmad, L., Rylott, E. L., Bruce, N. C., Edwards, R. & Grogan, G. Structural evidence for Arabidopsis glutathione transferase AtGSTF2 functioning as a transporter of small organic ligands. *FEBS Open Bio* **7**, 122–132 (2017).

51. Fernández-Cañón, J. M. & Peñalva, M. A. Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. *J. Biol. Chem.* **273**, 329–337 (1998).

52. Johansson, A. S. & Mannervik, B. Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. *J. Biol. Chem.* **276**, 33061–33065 (2001).

53. Kanaoka, Y. *et al.* Cloning and crystal structure of hematopoietic prostaglandin D synthase. *Cell* **90**, 1085–1095 (1997).

54. Paumi, C. M., Smitherman, P. K., Townsend, A. J. & Morrow, C. S. Glutathione S-transferases (GSTs) inhibit transcriptional activation by the peroxisomal proliferator-activated receptor gamma (PPAR gamma) ligand, 15-deoxy-delta 12,14prostaglandin J2 (15-d-PGJ2). *Biochemistry* **43**, 2345–2352 (2004).

55. Jowsey, I. R., Smith, S. A. & Hayes, J. D. Expression of the murine glutathione S-transferase alpha3 (GSTA3) subunit is markedly induced during adipocyte differentiation: activation of

the GSTA3 gene promoter by the pro-adipogenic eicosanoid 15-deoxy-Delta12,14-prostaglandin J2. *Biochem. Biophys. Res. Commun.* **312**, 1226–1235 (2003).

56. Neuefeind, T., Huber, R., Dasenbrock, H., Prade, L. & Bieseler, B. Crystal structrure of herbicide-detoxifying maize glutathione S-transferase-I in complex with lactoylglutathione: evidence for an induced-fit mechanism. *J. Mol. Biol.* **274**, 446–453 (1997).

57. Valenzuela-Chavira, I. *et al.* Insights into ligand binding to a glutathione S-transferase from mango: Structure, thermodynamics and kinetics. *Biochimie* **135**, 35–45 (2017).

58. Tossounian, M. A. *et al.* Disulfide bond formation protects Arabidopsis thaliana glutathione transferase tau 23 from oxidative damage. *Biochim Biophys Acta Gen Subj* **1862**, 775–789 (2018).

59. Tars, K. *et al.* Structural basis of the suppressed catalytic activity of wild-type human glutathione transferase T1-1 compared to its W234R mutant. *J Mol Biol* **355**, 96–105 (2006).

60. Do, H. *et al.* Structural understanding of the recycling of oxidized ascorbate by dehydroascorbate reductase (OsDHAR) from Oryza sativa L. japonica. *Sci Rep* **6**, 19498–19498 (2016).

61. Bodra, N. *et al.* Arabidopsis thaliana dehydroascorbate reductase 2: Conformational flexibility during catalysis. *Sci Rep* **7**, 42494–42494 (2017).

62. Lallement, P. A. *et al.* Structural and enzymatic insights into Lambda glutathione transferases from Populus trichocarpa, monomeric enzymes constituting an early divergent class specific to terrestrial plants. *Biochem J* **462**, 39–52 (2014).

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

64. Holmgren, A., Söderberg, B. O., Eklund, H. & Brändén, C. I. Three-dimensional structure of Escherichia coli thioredoxin-S2 to 2.8 A resolution. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2305–2309 (1975).

65. Eklund, H. *et al.* Structure of oxidized bacteriophage T4 glutaredoxin (thioredoxin). Refinement of native and mutant proteins. *J. Mol. Biol.* **228**, 596–618 (1992).

66. Epp, O., Ladenstein, R. & Wendel, A. The Refined Structure of the Selenoenzyme Glutathione Peroxidase at 0.2-nm Resolution. *Eur. J. Biochem.* **133**, 51–69 (1983).

67. Erhardt, J. & Dirr, H. Native Dimer Stabilizes the Subunit Tertiary Structure of Porcine Class pi Glutathione S-transferase. *Eur. J. Biochem.* **230**, 614–620 (1995).

68. Caccuri, A. M. *et al.* Catalytic Mechanism and Role of Hydroxyl Residues in the Active Site of Theta Class Glutathione S-Transferases Investigation of Ser-9 and Tyr-113 in a Glutathione-S-Transferase from the Australian Sheep Blowfly, Lucila cuprina. *J. Biol. Chem.* **272**, 29681–29686 (1997).

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

70. Armstrong, R. N. Glutathione S-Transferases: Structure and Mechanism of an Archetypical Detoxication Enzyme. in *Advances in Enzymology and Related Areas of Molecular Biology* (ed. Meister, A.) 1–44 (John Wiley & Sons, Inc., 1994).

71. Dixon, D. P., Lapthorn, A. & Edwards, R. Plant glutathione transferases. *Genome Biol.* **3**, reviews3004.1 (2002).

72. McTigue, M. A., Williams, D. R. & Tainer, J. A. Crystal Structures of a Schistosomal Drug and Vaccine Target: Glutathione S-Transferase from Schistosoma japonica and its Complex with the Leading Antischistomal Drug Praziquantel. *J. Mol. Biol.* **246**, 21–27 (1995).

73. Graminski, G. F., Zhang, P., Sesay, M. A., Ammon, H. L. & Armstrong, R. N. Formation of the 1-(S-glutathionyl)-2,4,6-trinitrocyclohexadienate anion at the active site of glutathione S-transferase: evidence for enzymic stabilization of sigma-complex intermediates in nucleophilic aromatic substitution reactions. *Biochemistry* **28**, 6252–6258 (1989).

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

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

76. Moss, S. *Rothamsted Research Report*. (2013).

77. Hicks, H. L. *et al.* The factors driving evolved herbicide resistance at a national scale. *Nat. Ecol. Evol.* **2**, 529 (2018).

78. Geissel, H. Problem Weeds in Cereals. (2004). Available at: http://www.crsbooks.net/agscience/weeds.html. (Accessed: 4th January 2019)

79. Anderson, M. P. & Gronwald, J. W. Atrazine Resistance in a Velvetleaf (Abutilon theophrasti) Biotype Due to Enhanced Glutathione S-Transferase Activity. *Plant Physiol.* **96**, 104–109 (1991).

80. Mahajan, S. & Atkins, W. M. The chemistry and biology of inhibitors and pro-drugs targeted to glutathione S-transferases. *Cell. Mol. Life Sci.* **62**, 1221–1233 (2005).

81. Tew, K. D. *et al.* The Role of Glutathione S-transferase P in signaling pathways and S-glutathionylation in Cancer. *Free Radic Biol Med* **51**, 299–313 (2011).

82. Ricci, G. *et al.* Structural Flexibility Modulates the Activity of Human Glutathione Transferase P1-1 Role of Helix 2 Flexibility in the Catalytic Mechanism. *J. Biol. Chem.* **271**, 16187–16192 (1996).

83. De Luca, A., Federici, L., De Canio, M., Stella, L. & Caccuri, A. M. New Insights into the Mechanism of JNK1 Inhibition by Glutathione Transferase P1-1. *Biochemistry* **51**, 7304–7312 (2012).

84. Townsend, D. M. & Tew, K. D. The role of glutathione-S-transferase in anti-cancer drug resistance. *Oncogene* **22**, 7369–7375 (2003).

85. Straker, H. E. Chemical methods towards the investigation of multiple herbicide resistance in black grass (Alopecurus myosuroides). (Durham University, 2014).

86. Schwarz, M. Exploring the molecular basis of multiple herbicide resistance in black grass (Alopecurus myosuroides). (Durham University, 2017).

87. Freitag-Pohl, S. *Unpublished results*.

88. Protein identification and analysis tools on the expasy server. in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005).

89. Dixon, D. P., Hawkins, T., Hussey, P. J. & Edwards, R. Enzyme activities and subcellular localization of members of the Arabidopsis glutathione transferase superfamily. *J Exp Bot* **60**, 1207–1218 (2009).

90. Grøftehauge, M. K., Hajizadeh, N. R., Swann, M. J. & Pohl, E. Protein–ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta. Crystallogr. D Struct. Biol.* **71**, 36–44 (2015).

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

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

93. Mosca, R. & Schneider, T. R. RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes. *Nucleic Acids Res* **36**, W42–W46 (2008).

94. Mosca, R., Brannetti, B. & Schneider, T. R. Alignment of protein structures in the presence of domain motions. *BMC Bioinformatics* **9**, 352 (2008).

95. Kabsch, W. XDS. *Acta. Crystallogr. D Struct. Biol.* **66**, 125–132 (2010).

96. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta. Crystallogr. D Struct. Biol.* **67**, 282–292 (2011).

97. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta. Crystallogr. D Struct. Biol.* **69**, 1204–1214 (2013).

98. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Cryst* **40**, 658–674 (2007).

99. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta. Crystallogr. D Struct. Biol.* **67**, 235–242 (2011).

100. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta. Crystallogr. D Struct. Biol.* **66**, 486–501 (2010).

101. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta. Crystallogr. D Struct. Biol.jligan* **60**, 2184–2195 (2004).

102. Lebedev, A. A. *et al.* JLigand: a graphical tool for the CCP4 template-restraint library. *Acta. Crystallogr. D Struct. Biol.* **68**, 431–440 (2012).

103. Gorrec, F. The MORPHEUS protein crystallization screen. *J Appl Cryst* **42**, 1035–1042 (2009).

104. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta. Crystallogr. D Struct. Biol.* **62**, 1002–1011 (2006).

105. Long, F. *et al.* AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol* **73**, 112–122 (2017).

106. Fiser, A. & Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. in *Methods in Enzymology* **374**, 461–491 (Academic Press, 2003).

107. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).

108. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212–2221 (2007).

109. Matulis, D., Kranz, J. K., Salemme, F. R. & Todd, M. J. Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor. *Biochemistry* **44**, 5258–5266 (2005).

110. Bullock, A. N. *et al.* Structural Basis of Inhibitor Specificity of the Human Protooncogene Proviral Insertion Site in Moloney Murine Leukemia Virus (PIM-1) Kinase. *J. Med. Chem.* **48**, 7604–7614 (2005).

111. Vedadi, M. *et al.* Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **103**, 15835–15840 (2006).

112. Palusiak, M. & Grabowski, S. J. Methoxy group as an acceptor of proton in hydrogen bonds. *J. Mol. Struct.* **642**, 97–104 (2002).
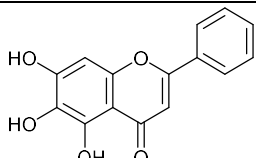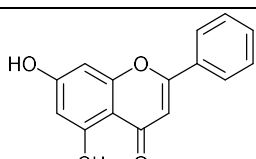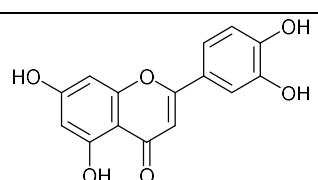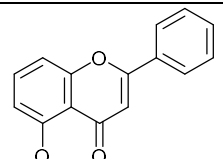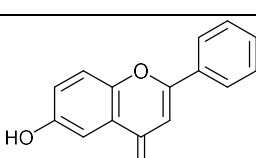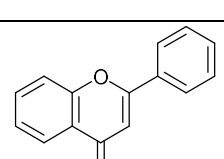
113. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

114. Fabrini, R. *et al.* Monomer-dimer equilibrium in glutathione transferases: a critical re-examination. *Biochemistry* **48**, 10473–10482 (2009).

115. Nair, D. P. *et al.* The Thiol-Michael Addition Click Reaction: A Powerful and Widely Used Tool in Materials Chemistry. *Chem. Mater.* **26**, 724–744 (2014).

116. Fluorescence Anisotropy. in *Principles of Fluorescence Spectroscopy* (ed. Lakowicz, J. R.) 353–382 (Springer US, 2006). doi:10.1007/978-0-387-46312-4_10

117. Okamura, T. *et al.* Phosphorylation of Glutathione S-Transferase P1 (GSTP1) by Epidermal Growth Factor Receptor (EGFR) Promotes Formation of the GSTP1-c-Jun N-terminal kinase (JNK) Complex and Suppresses JNK Downstream Signaling and Apoptosis in Brain Tumor Cells. *J. Biol. Chem.* **290**, 30866–30878 (2015).

118. Thom, R. *et al.* Structure of a Tau Class Glutathione S-Transferase from Wheat Active in Herbicide Detoxification,. *Biochemistry* **41**, 7008–7020 (2002).

119. Cummins, I. *et al.* Cloning, characterization and regulation of a family of phi class glutathione transferases from wheat. *Plant Mol Biol* **52**, 591–603 (2003).

120. Prade, L., Huber, R. & Bieseler, B. Structures of herbicides in complex with their detoxifying enzyme glutathione S-transferase - explanations for the selectivity of the enzyme in plants. *Structure* **6**, 1445–1452 (1998).

121. Oakley, A. J. *et al.* The crystal structures of glutathione S-transferases isozymes 1–3 and 1–4 from Anopheles dirus species B. *Protein Sci* **10**, 2176–2185 (2001).

122. Cole, D. J. Detoxification and activation of agrochemicals in plants. *Pestic. Sci.* **42**, 209–222 (1994).

123. Oakley, A. J. *et al.* The three-dimensional structure of the human Pi class glutathione transferase P1-1 in complex with the inhibitor ethacrynic acid and its glutathione conjugate. *Biochemistry* **36**, 576–585 (1997).

124. Neuefeind, T. *et al.* Cloning, sequencing, crystallization and X-ray structure of glutathione S-transferase-III from Zea mays var. mutin: a leading enzyme in detoxification of maize herbicides. *J. Mol. Biol.* **274**, 577–587 (1997).

125. Honegger, A., Spinelli, S., Cambillau, C. & Plückthun, A. A mutation designed to alter crystal packing permits structural analysis of a tight-binding fluorescein-scFv complex. *Protein Sci.* **14**, 2537–2549 (2005).

126. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–139 (2005).
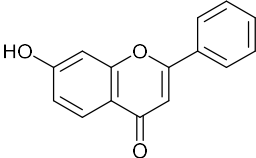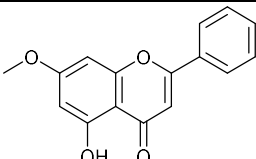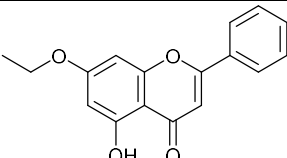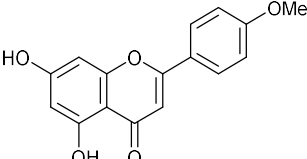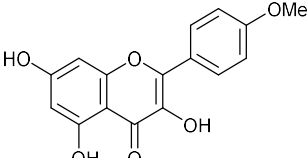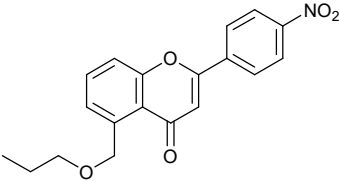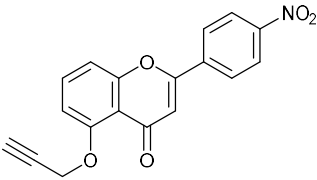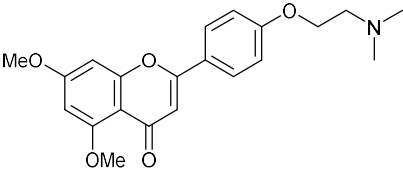
127. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876–2890 (2006).

128. Lin, S., Lo, Y., Tam, M. & Liaw, Y. PDB 1VF3, To be published.

129. Abuhammad, A. *et al.* "To Cross-Seed or Not To Cross-Seed": A Pilot Study Using Metallo-β-lactamases. *Cryst. Growth Des.* **17**, 913–924 (2017).

130. Ireton, G. C. & Stoddard, B. L. Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta. Crystallogr. D Struct. Biol.* **60**, 601–605 (2004).

131. D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Cryst F* **70**, 1117–1126 (2014).

132. D'Arcy, A., Villard, F. & Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta. Crystallogr. D Struct. Biol.* **63**, 550–554 (2007).

133. Federici, L. *et al.* Structural basis for the binding of the anticancer compound 6-(7-nitro-2,1,3-benzoxadiazol-4-ylthio)hexanol to human glutathione s-transferases. *Cancer Res* **69**, 8025–8034 (2009).

134. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).

135. The GOLD User Guide (5.6.3 GOLD Release) - The Cambridge Crystallographic Data Centre (CCDC). Available at: https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/44a5b30d4b6c41cc98b7f01962542ff4.pdf. (Accessed: 28th October 2018)

136. Korb, O., Stützle, T. & Exner, T. E. Empirical Scoring Functions for Advanced Protein–Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **49**, 84–96 (2009).

137. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate rotamer library. *Proteins* **40**, 389–408 (2000).

138. Reinemer, P. *et al.* 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. *J Mol Biol* **255**, 289–309 (1996).

139. Pégeot, H. *et al.* Structural plasticity among glutathione transferase Phi members: natural combination of catalytic residues confers dual biochemical activities. *FEBS J* **284**, 2442–2463 (2017).

140. Pieper, U. *et al.* ModBase, a database of annotated comparative protein structure models and associated resources. *Nucl. Acids Res.* **42**, D336–D346 (2014).
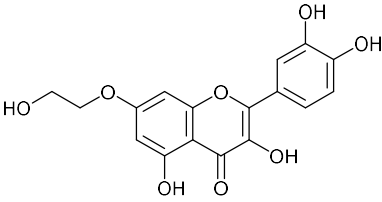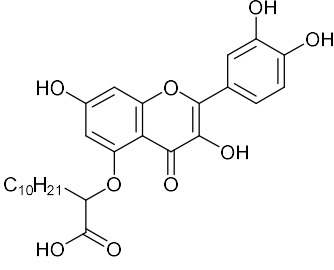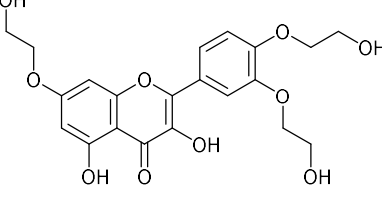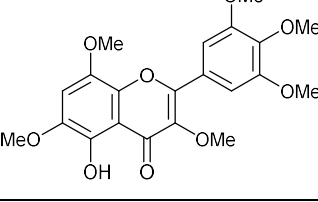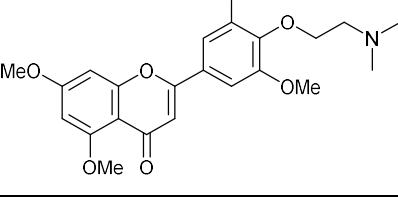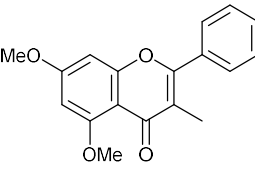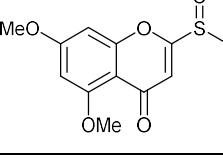
141. DiMaio, F. *et al.* Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat Meth* **10**, 1102–1104 (2013).

142. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).

143. Bowers, K. *et al.* Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. in *ACM/IEEE SC 2006 Conference (SC'06)* 43–43 (IEEE, 2006). doi:10.1109/SC.2006.54

144. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* **42**, W320–W324 (2014).

Appendix A: Biophysical Data
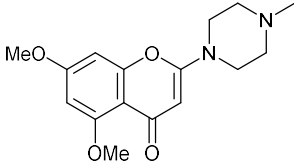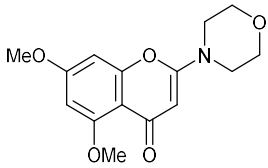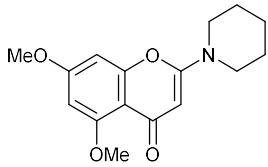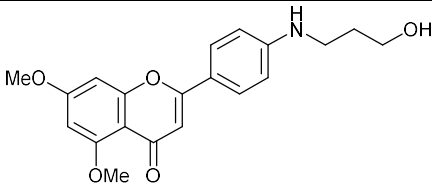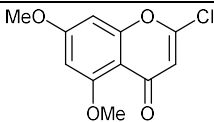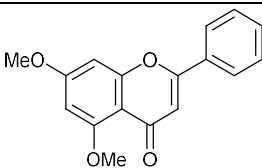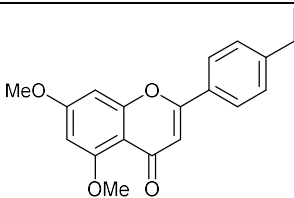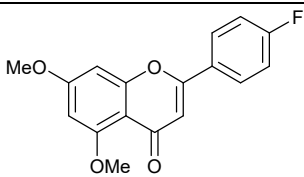
B1: Complete data from thermal shift assays and CDNB inihibtion assays. Data from CDNB assays performed by Maria Schwarz (MS compounds) or Hannah Straker (COM and SYG compounds). N.D. shows where data was not determined.

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
COMMERICALLY AVAILABLE FLAVONOIDS							
COM1		0.4	1.7	5.3	N.D.	42	N.D.
COM2		0.1	0.1	0.2	N.D.	70	96
COM3		0.4	1.4	4.8	N.D.	N.D.	N.D.
COM4		0.5	-0.1	-0.7	N.D.	N.D.	N.D.
COM5		0.4	0.0	-0.2	N.D.	N.D.	61
COM6		0.3	0	-0.2	N.D.	N.D.	14
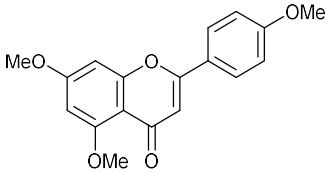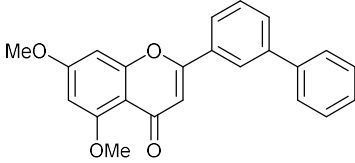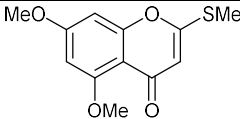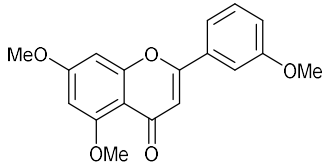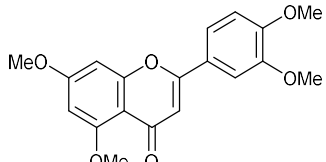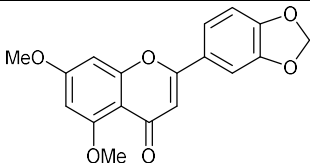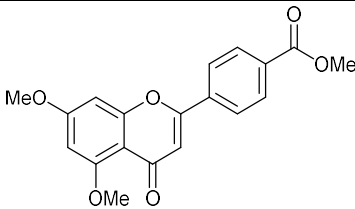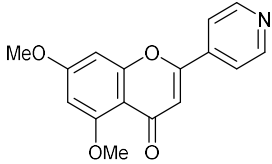
Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
COM7		0.2	0.5	0.5	N.D.	N.D.	69
COM8		0.4	0.2	0.1	N.D.	25	98
SYNGENTA LIBRARY COMPOUNDS							
SYG1		0.1	0	0.1	N.D.	6	N.D.
SYG2		**0.3**	0.4	0.6	N.D.	52	N.D.
SYG3		0.5	0.4	-0.4	N.D.	71	N.D.
SYG4		0.6	0.5	0.9	N.D.	8	N.D.
SYG5		0.2	0.2	0.7	N.D.	6	N.D.
SYG6		0.1	0.3	0.5	N.D.	7	N.D.
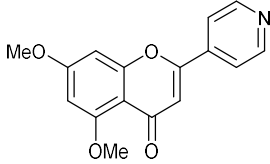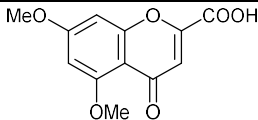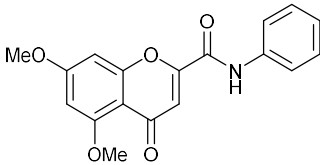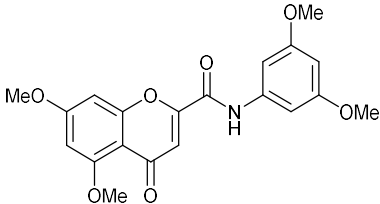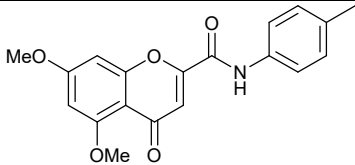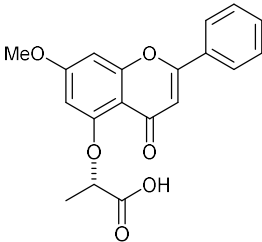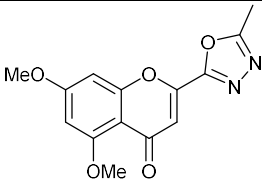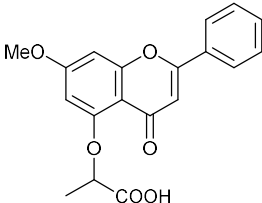
Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 µM	10 µM	100 µM	1 µM	10 µM	100 µM
SYG7		0.2	0	0.7	N.D.	2	N.D.
SYG8		0.2	2.2	1.1	N.D.	100	N.D.
SYG9		-0.3	-0.2	0.2	N.D.	40	N.D.
SYG10		0	0.2	0.1	N.D.	12	N.D.
SYG11		0.1	0.1	0.4	N.D.	7	N.D.
MARIA SCHWARZ COMPOUNDS							
MS-1-109		0.1	0.2	0.2	N.D.	17	N.D.
MS-1-12-1		0.2	0.1	0.4	N.D.	0	55
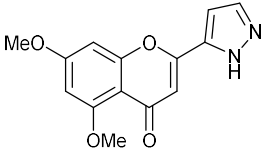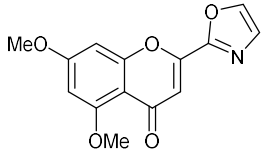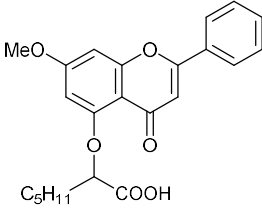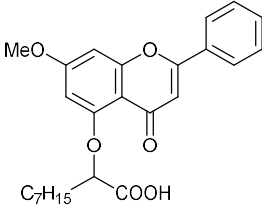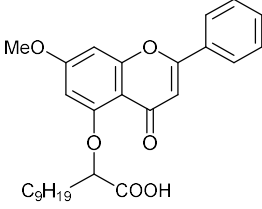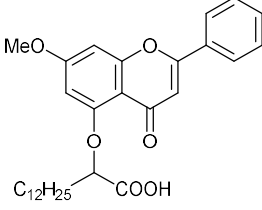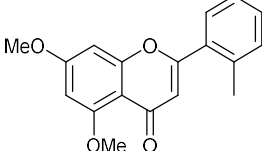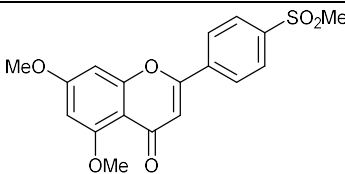
Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-1-12-2		0.2	0.2	0	N.D.	8	59
MS-1-123		0.2	0	0.2	N.D.	25	98
MS-1-129		0.2	0.4	0.6	N.D.	25	N.D.
MS-1-134		0.2	0.5	1.2	71	99	N.D.
MS-1-134 E1		0.2	0.2	0.5	46	N.D.	N.D.
MS-1-134 E2		0.2	0.5	1.2	68	N.D.	N.D.
MS-1-139		0.1	0.3	1.1	6	39	N.D.
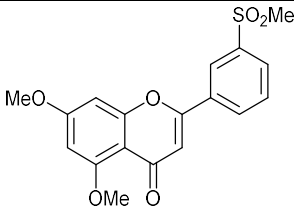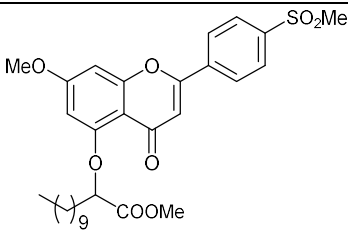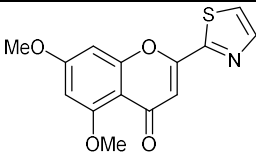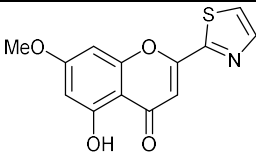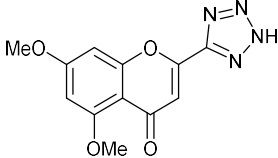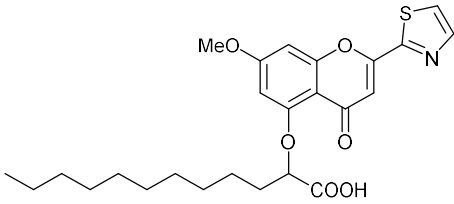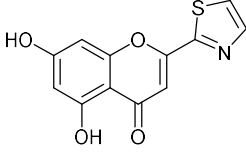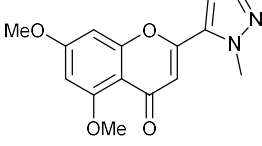
Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 µM	10 µM	100 µM	1 µM	10 µM	100 µM
MS-1-16		0	0.2	0.1	N.D.	N.D.	16
MS-1-18		0.1	0.2	0.2	N.D.	0	37
MS-1-19		0.2	0.1	0.3	N.D.	12	67
MS-1-20		0.2	0.2	0.1	N.D.	N.D.	12
MS-1-46		0.3	0.3	0.4	N.D.	32	5
MS-1-51		0.2	0.3	0.3	N.D.	20	70
MS-1-53-10		0.3	0.1	0	N.D.	12	80
MS-1-53-2		0.2	0.3	0.3	N.D.	23	78
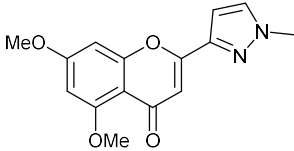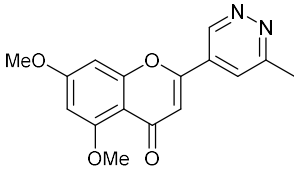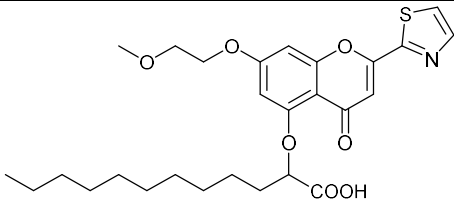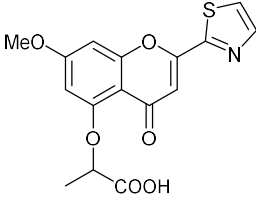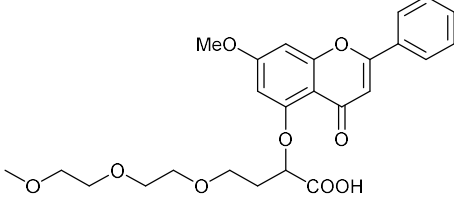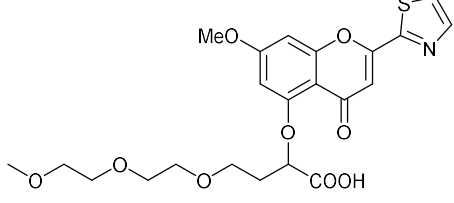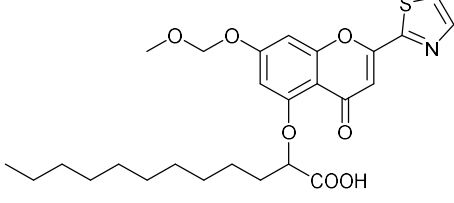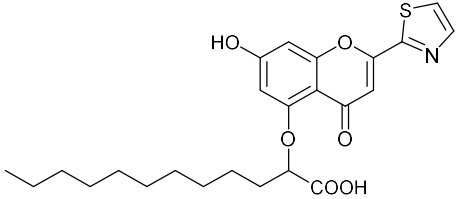
Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-1-53-4		0.2	0.2	0.7	N.D.	15	100
MS-1-53-6		0.1	0.1	-0.1	N.D.	12	87
MS-1-6		0.1	0.2	0	N.D.	N.D.	13
MS-1-66A		0.3	0.2	-0.2	N.D.	24	79
MS-1-66-B		0.3	0	0.2	N.D.	11	100
MS-1-66C		0.3	0.2	0.3	N.D.	12	33
MS-1-66-D		0.3	0.3	0.3	N.D.	20	58
MS-1-71		0.2	0.3	0.2	N.D.	10	50

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-1-71		0.2	0.3	0.1	N.D.	10	50
MS-1-79		-0.1	0	0.1	N.D.	N.D.	N.D.
MS-1-79		0.2	0.0	0.0	N.D.	22	49
MS-1-80		0.2	0.5	0.1	N.D.	19	56
MS-1-82		0.2	0.1	0.3	N.D.	34	64
MS-3-12		0.3	0.2	0.5	N.D.	5	N.D.
MS-3-128		-0.1	-0.2	-0.3	N.D.	8	53
MS-3-13		0.2	0.2	1.1	N.D.	14	N.D.

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-3-132A		-0.1	-0.1	0	N.D.	6	59
MS-3-140		-0.1	-0.2	-0.2	N.D.	11	64
MS-3-21		0.2	0.4	2.3	4	69	N.D.
MS-3-22		0.3	0.7	2.9	40	86	N.D.
MS-3-24		0.4	0.7	0.9	58	97	N.D.
MS-3-24		0.2	0.7	1.43	75	90	N.D.
MS-3-41		-0.1	0	0	N.D.	9	36
MS-3-42		0.1	0.2	0.3	N.D.	14	35

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-3-43		0.2	0.1	0.3	N.D.	11	60
MS-3-50		0.2	0.2	0.9	40	94	N.D.
MS-4-30-1		-0.4	-0.5	-0.5	N.D.	23	83
MS-4-30-2		0.4	0.4	0.1	N.D.	23	61
MS-4-4		-0.2	-0.3	-0.4	N.D.	13	53
MS-4-48		0.8	1.8	2.8	56	95	N.D.
MS-4-50		0.3	0.4	0.9	N.D.	26	89
MS-4-8-1		0	0	-0.1	N.D.	19	64

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 µM	10 µM	100 µM	1 µM	10 µM	100 µM
MS-4-8-2		0	0	-0.2	N.D.	9	58
MS-4-9		-0.1	-0.1	-0.3	N.D.	8	64
MS-5-30		0.7	1.6	4.3	52	95	N.D.
MS-5-31		0.6	0.5	0.7	N.D.	30	92
MS-5-34		1.2	1.3	2.1	N.D.	13	93
MS-5-35		0.7	0.8	1.0	N.D.	14	98
MS-5-6		0.7	1.6	3.8	67	98	N.D.

Compound Code	Structure	Thermal Shift Assay ΔT/ °C			CDNB Inhibition / %		
		1 μM	10 μM	100 μM	1 μM	10 μM	100 μM
MS-5-7		1.3	1.7	4.2	72	100	N.D.

Appendix B: Crystallographic Tables

B1: Crystallographic data tables for crystal structures of wild type *Am*GSTF1 (structures discussed in chapter 4)

	*Am*GSTF1 WT apo	*Am*GSTF1 WT covalent CNBF
Data Collection		
Space Group	P 6_3 2 2	P 6_3 2 2
Unit cell dimensions [Å], [˚]	a=b=103.69, c=78.77 $\alpha=\beta=90, \gamma=120$	a=b=104.088, c=78.82 $\alpha=\beta=90, \gamma=120$
Resolution (Å)[a]	51-1.53 (1.56-1.53)	59.34 -2.00 (2.05-2.00)
Beamline	DLS i04-1	DLS i04-1
R_{merge}[a]	0.03434 (0.7881)	0.091 (0.734)
rMeas[a]	0.036887 (0.88657)	0.098 (0.802)
rPim (pooled Friedels)[a]	0.010 (0.27)	0.041 (0.338)
I / σI[a]	27.326 (1.992)	13.6 (2.5)
Completeness (%)[a]	100.0 (99.46)	95.7 (97.5)
Redundancy	12.568 (10.348)	
Refinement		
No. reflections (all/free)	38090/1777	16633/868
R_{work} / R_{free}	0.193/0.217	0.19/0.237
No. of non H atoms Protein Water Ligand	1540 112 -	1495 59 12
B-factors Protein Water Ligand	32.1 39.2 -	35.7 39.2 63.1
R.m.s. deviations Bond lengths (Å) Bond angles (°)	0.0187 1.936	0.0161 1.698

[a]The values in parentheses correspond to the outermost resolution shell

B2: Crystallographic data tables for crystal structures of mutant *Am*GSTF1, and structures obtained by cross-seeding with mutant *Am*GSTF1 (structures discussed in chapter 5)

|	F122T (GSH bound)	Y118S (GSH bound)	*Am*GSTF1 WT (GSH bound)	*Am*GSTF1 WT (NBF-GS bound)
Data Collection				
Space Group	I 4_1 2 2	I 4_1 2 2	I 4_1 2 2	I 4_1 2 2
Unit cell dimensions [Å], [˚]	a=b=112.590, c=104.300 $\alpha=\beta=\gamma=90$	a=b=112.611, c=103.646 $\alpha=\beta=\gamma=90$	a=b=112.877, c=101.246 $\alpha=\beta=\gamma=90$	a=b=112.896, c=101.669 $\alpha=\beta=\gamma=90$
Resolution (Å)[a]	43.62-2.78 (2.93-2.78)	79.63-2.60 (2.72-2.60)	79.82-2.25 (2.32-2.25)	35.70-2.30 (2.38-2.30)
Beamline	DLS i24	DLS i04-1	DLS i04-1	DLS i04-1
R_{merge}[a]	0.040 (0.456)	0.093 (1.017)	0.065 (0.821)	0.069 (0.971)
rMeas[a]	0.057 (0.645)	0.096 (1.059)	0.066 (0.032)	0.070 (0.989)
rPim[a]	0.040 (0.456)	0.019 (0.209)	0.013 (0.007)	0.019 (0.186)
I / σI[a]	10.7 (1.3)	22.7 (3.7)	29.0 (3.9)	26.5 (3.7)
Completeness (%)[a]	99.8 (100)	100 (99.5)	100.0 (99.9)	100 (98.6)
Multiplicity	1.8 (1.6)	26.1 (22.1)	25.8 (21.4)	26.4 (22.4)
Refinement				
No. reflections (all/free)	8718/478	10565/ 549	15823/ 765	14901/ 715
R_{work} / R_{free}	0.18/0.24	0.20/ 0.28[b]	0.23/ 0.28	0.21/ 0.25
No. of non H atoms Protein Water Ligand	1683 6 20	1637 18 20	1678 24 20	1696 8 32
B-factors Protein Water Ligand	86.1 76.2 91.0	41.9 66.7 69.5	57.5 52.5 67.2	69.6 58.7 72.8
R.m.s. deviations Bond lengths (Å) Bond angles (°)	0.0104 1.587	0.0174 1.973	0.0191 2.108	0.0179 1.961

[a]The values in parentheses correspond to the outermost resolution shell

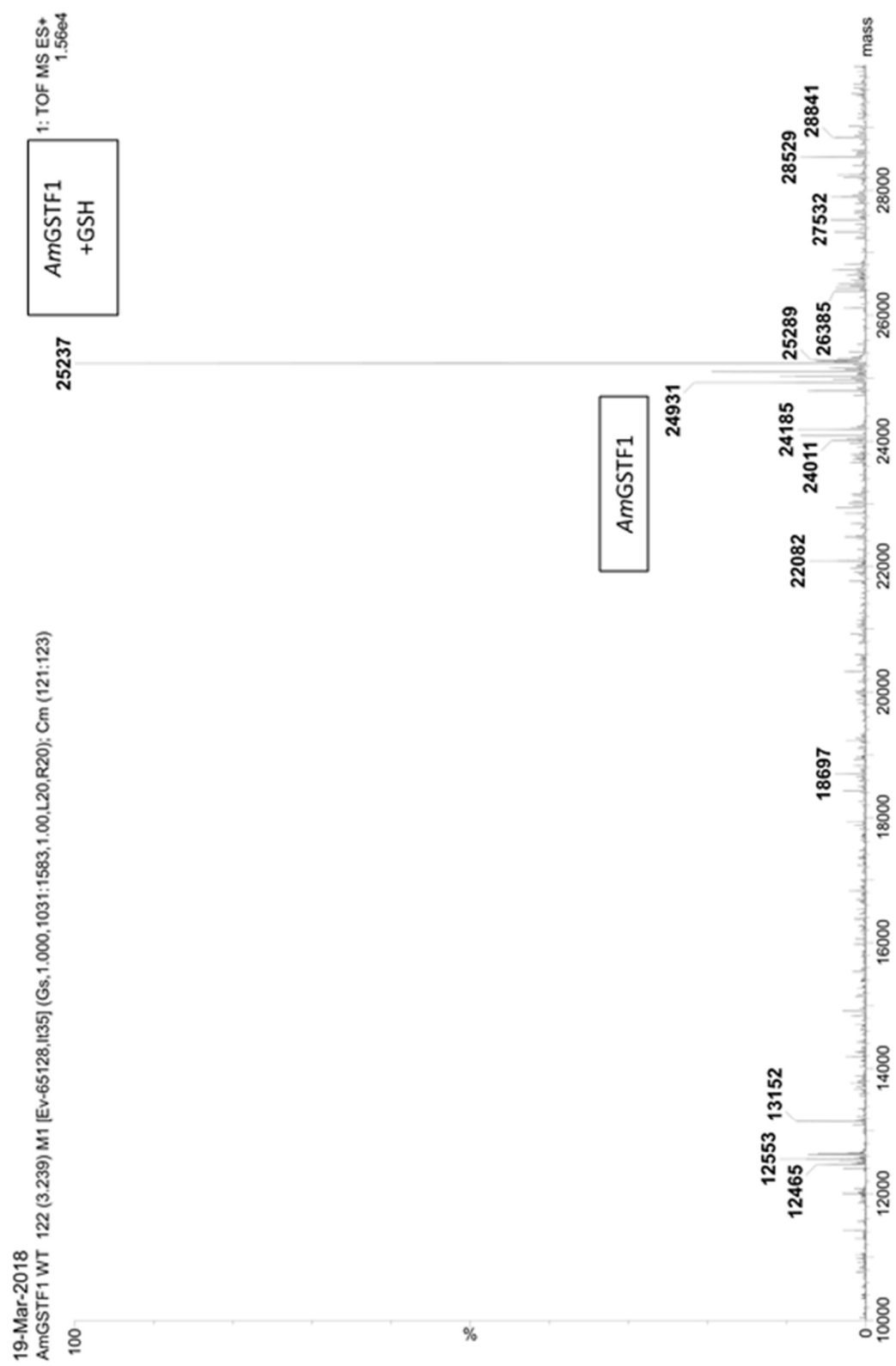[b]Refined using 5 cycles of automatic TLS paramaters
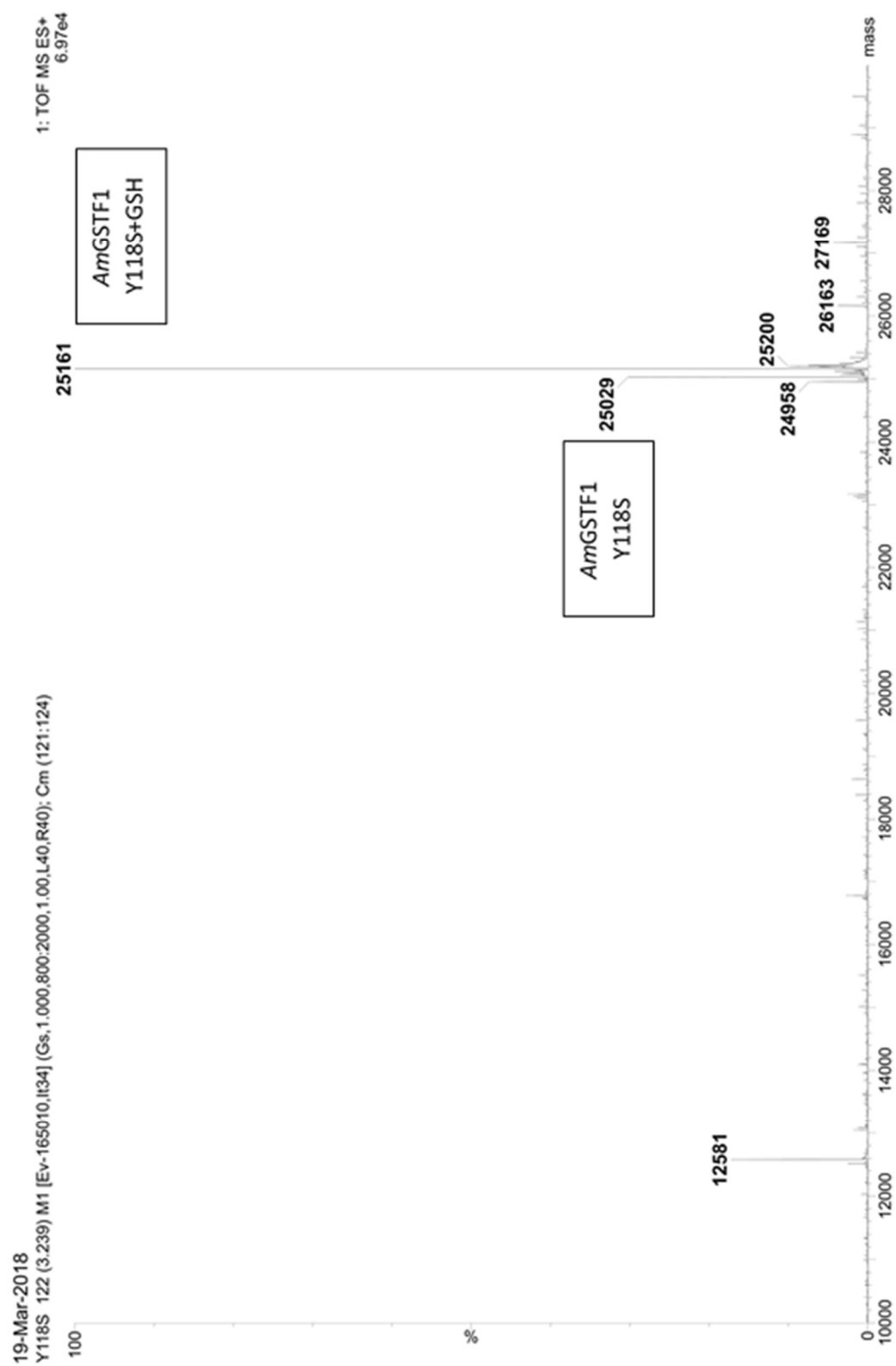
Appendix C: Mass Spectrometry Data
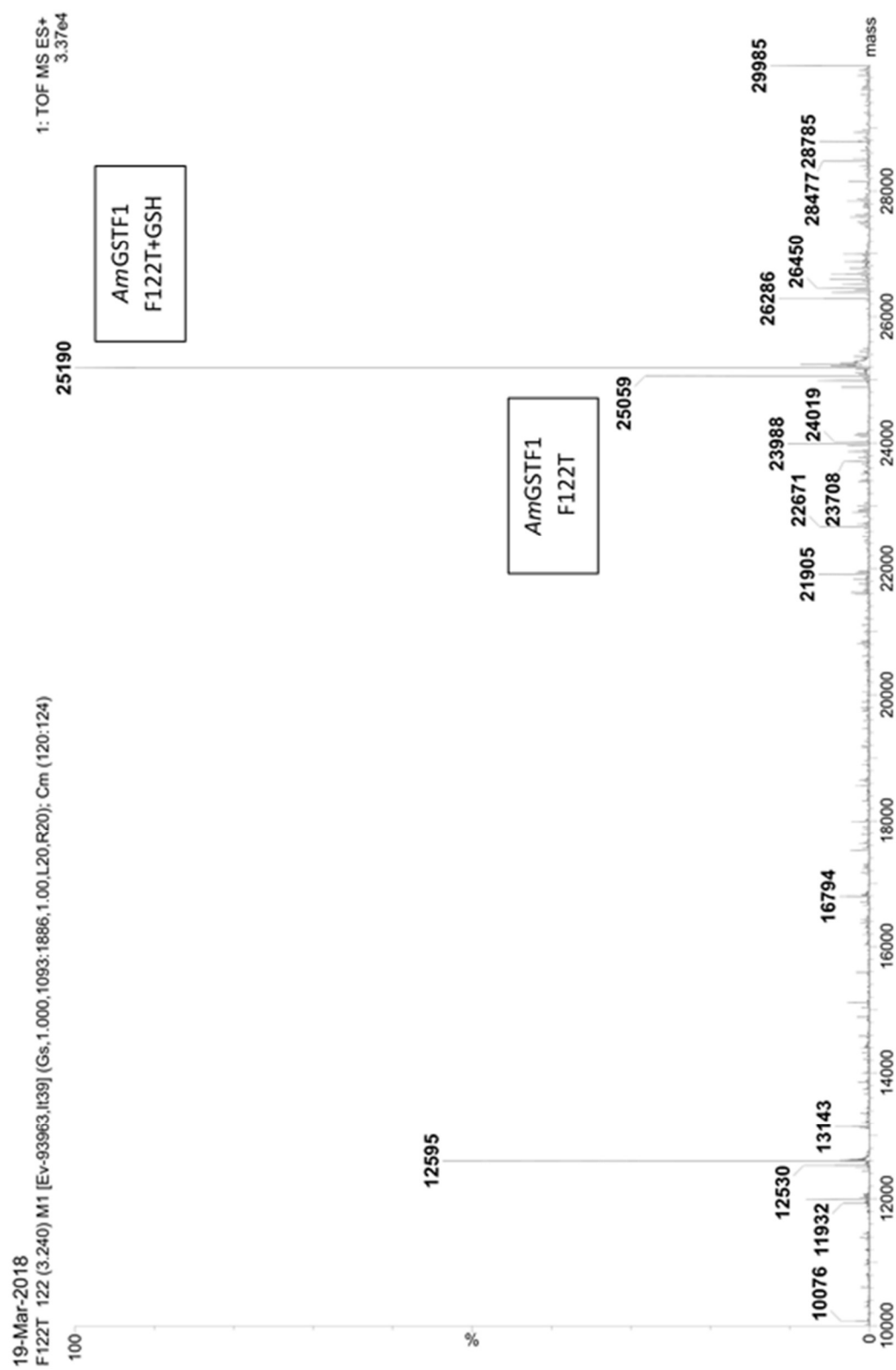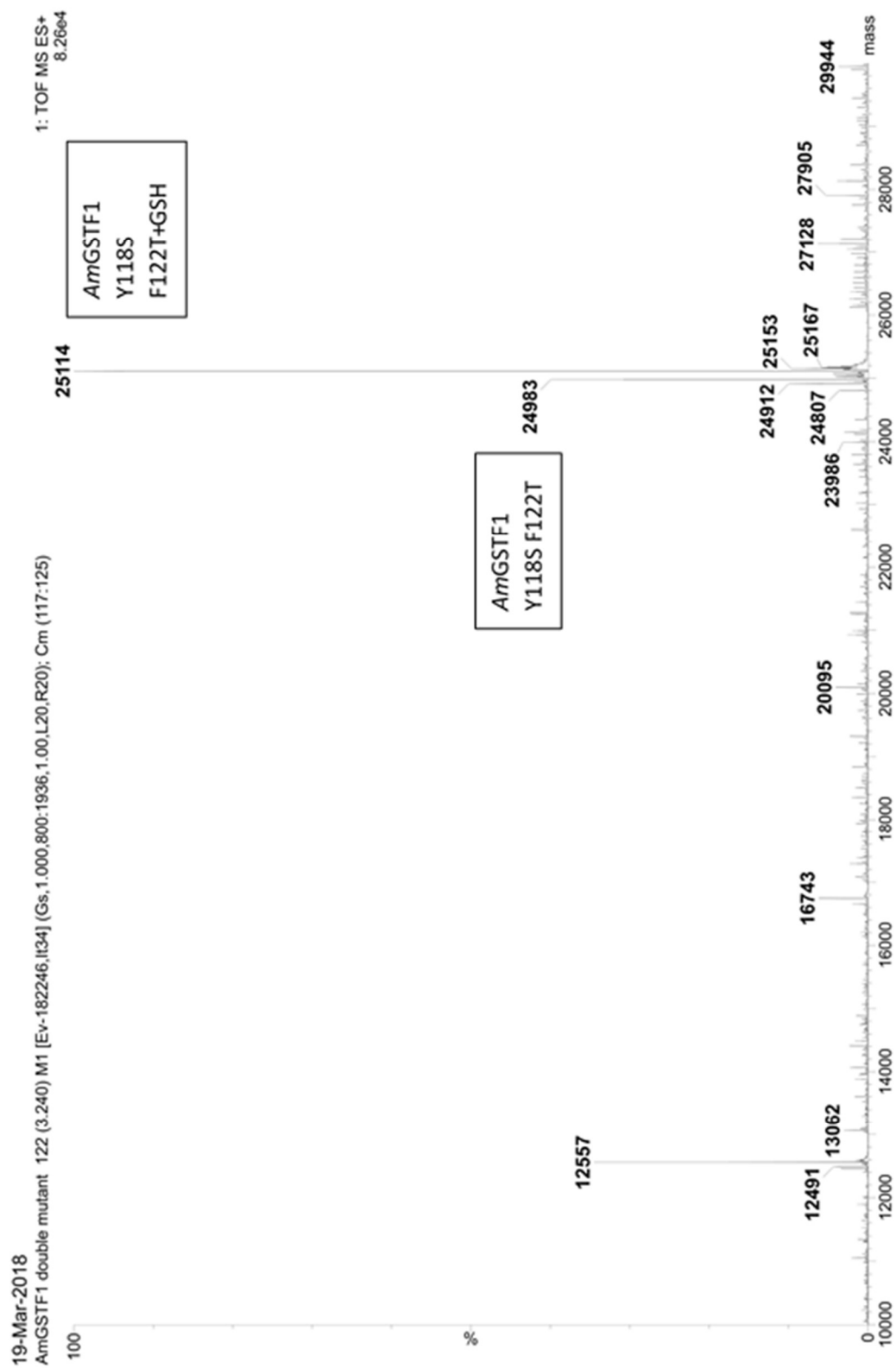
C1: Strep Tagged *Am*GSTF1
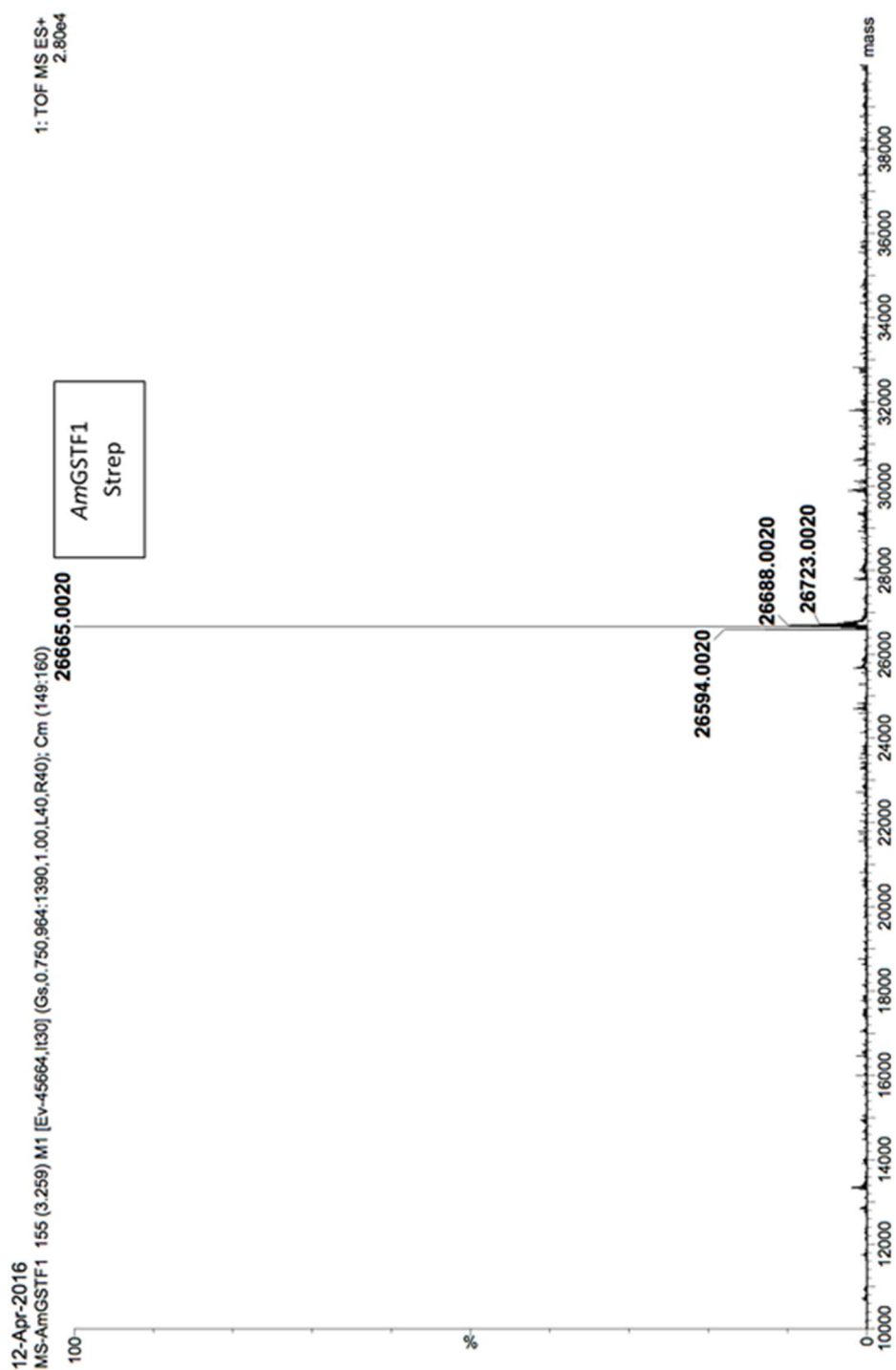
C2: Untagged *Am*GSTF1

C3: *Am*GSTF1 Y118S Mutant

C4: *Am*GSTF1 F122T Mutant

C5: *Am*GSTF1 Y118S F122T Mutant

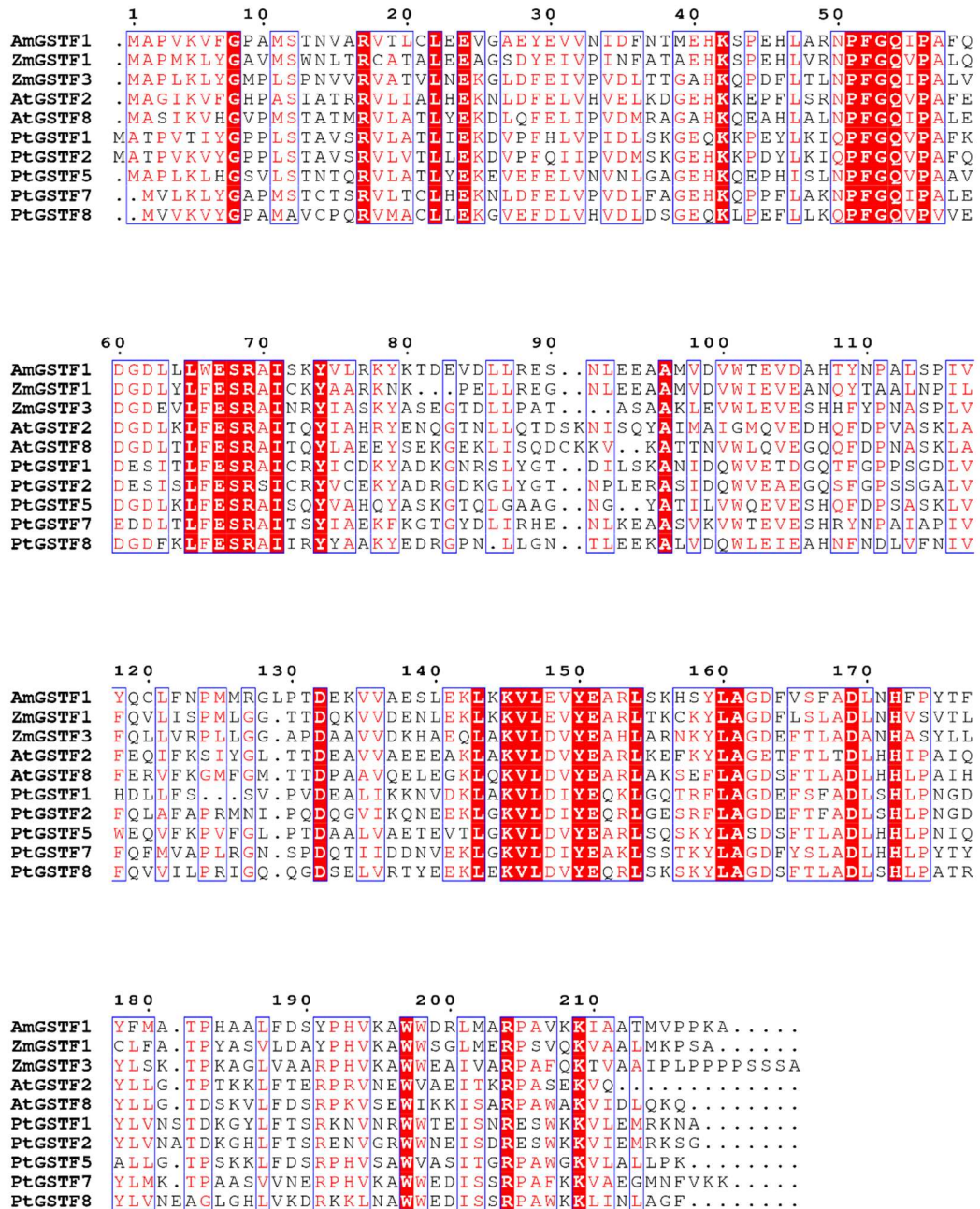# Appendix D: Sequence Alignment

Figure D1: Alignment of sequences of relevant GSTs. Sequences aligned using Clustal Omega. Figure created using ESPript.[144]

Table D1: Sequence similarity to *Am*GSTF1. Calculated using NCBI BLAST.

Protein	Sequence Similarity to AmGSTF1/ %
*Zm*GSTF1	63
*Zm*GSTF3	47
*At*GSTF2	41
*At*GSTF8	42
*Pt*GSTF1	38
*Pt*GSTF2	38
*Pt*GSTF5	46
*Pt*GSTF7	52
*Pt*GSTF8	44

Appendix E: Protein and DNA Sequences

*Am*GSTF1 Wild Type (strep tagged) Protein Sequence

MASWSHPQFEKGLINHMAPVKVFGPAMSTNVARVTLCLEEVGAEYEVVNIDFNTMEHKSPEHLARNPFG
QIPAFQDGDLLLWESRAISKYVLRKYKTDEVDLLRESNLEEAAMVDVWTEVDAHTYNPALSPIVYQCLFNP
MMRGLPTDEKVVAESLEKLKKVLEVYEARLSKHSYLAGDFVSFADLNHFPYTFYFMATPHAALFDSYPHVKA
WWDRLMARPAVKKIAATMVPPKA

(Strep tag is shown in red)

*Am*GSTF1 Wild Type (strep tagged) Nucleotide Sequence

ATGGCTAGCTGGAGCCACCCGCAGTTCGAGAAAGGCTTAATTAACCATATGGCGCCGGTGAAAGTTTT
TGGCCCGGCCATGAGCACCAACGTTGCACGTGTGACCCTGTGCCTGGAAGAAGTTGGTGCAGAATATG
AAGTTGTGAATATTGATTTCAACACCATGGAACATAAAAGCCCGGAACACCTGGCGCGCAATCCGTTTG
GCCAGATCCCGGCCTTCCAGGATGGTGATCTGCTGCTGTGGGAAAGTCGTGCAATTAGCAAATATGTTC
TGCGCAAATATAAAACCGATGAAGTTGATCTGCTGCGTGAAAGCAACCTGGAAGAAGCAGCGATGGT
GGATGTTTGGACCGAAGTTGATGCCCATACCTATAATCCGGCACTGAGCCCGATCGTGTATCAGTGTCT
GTTTAACCCGATGATGCGCGGCCTGCCGACCGATGAAAAAGTTGTTGCAGAAAGCCTGGAAAAACTGA
AAAAAGTGCTGGAAGTTTATGAAGCGCGTCTGAGCAAACACAGCTATCTGGCCGGTGATTTCGTTAGC
TTTGCAGATCTGAATCATTTCCCGTATACCTTTTATTTCATGGCAACCCCGCACGCGGCCCTGTTTGATA
GCTATCCGCATGTGAAAGCATGGTGGGATCGCCTGATGGCACGTCCGGCAGTGAAGAAAATCGCCGC
GACCATGGTGCCGCCGAAAGCC

(Strep tag is shown in red)

*Am*GSTF1 Wild Type (untagged) Protein Sequence

MAPVKVFGPAMSTNVARVTLCLEEVGAEYEVVNIDFNTMEHKSPEHLARNPFGQIPAFQDGDLLLWESRA
ISKYVLRKYKTDEVDLLRESNLEEAAMVDVWTEVDAHTYNPALSPIVYQCLFNPMMRGLPTDEKVVAESLEK
LKKVLEVYEARLSKHSYLAGDFVSFADLNHFPYTFYFMATPHAALFDSYPHVKAWWDRLMARPAVKKIAAT
MVPPKA

*Am*GSTF1 Wild Type (untagged) Nucleotide Sequence

ATGGCGCCGGTGAAAGTTTTTGGCCCGGCCATGAGCACCAACGTTGCACGTGTGACCCTGTGCCTGGA
AGAAGTTGGTGCAGAATATGAAGTTGTGAATATTGATTTCAACACCATGGAACATAAAAGCCCGGAAC
ACCTGGCGCGCAATCCGTTTGGCCAGATCCCGGCCTTCCAGGATGGTGATCTGCTGCTGTGGGAAAGT
CGTGCAATTAGCAAATATGTTCTGCGCAAATATAAAACCGATGAAGTTGATCTGCTGCGTGAAAGCAAC
CTGGAAGAAGCAGCGATGGTGGATGTTTGGACCGAAGTTGATGCCCATACCTATAATCCGGCACTGAG
CCCGATCGTGTATCAGTGTCTGTTTAACCCGATGATGCGCGGCCTGCCGACCGATGAAAAAGTTGTTGC
AGAAAGCCTGGAAAAACTGAAAAAAGTGCTGGAAGTTTATGAAGCGCGTCTGAGCAAACACAGCTATC

TGGCCGGTGATTTCGTTAGCTTTGCAGATCTGAATCATTTCCCGTATACCTTTTATTTCATGGCAACCCCG
CACGCGGCCCTGTTTGATAGCTATCCGCATGTGAAAGCATGGTGGGATCGCCTGATGGCACGTCCGGC
AGTGAAGAAAATCGCCGCGACCATGGTGCCGCCGAAAGCC

*Am*GSTF1 Y118S Mutant Protein Sequence

MAPVKVFGPAMSTNVARVTLCLEEVGAEYEVVNIDFNTMEHKSPEHLARNPFGQIPAFQDGDLLLWESRA
ISKYVLRKYKTDEVDLLRESNLEEAAMVDVWTEVDAHTYNPALSPIVSQCLFNPMMRGLPTDEKVVAESLEK
LKKVLEVYEARLSKHSYLAGDFVSFADLNHFPYTFYFMATPHAALFDSYPHVKAWWDRLMARPAVKKIAAT
MVPPKA

*Am*GSTF1 Y118S Mutant Nucleotide Sequence

ATGGCGCCGGTGAAAGTTTTTGGCCCGGCCATGAGCACCAACGTTGCACGTGTGACCCTGTGCCTGGA
AGAAGTTGGTGCAGAATATGAAGTTGTGAATATTGATTTCAACACCATGGAACATAAAAGCCCGGAAC
ACCTGGCGCGCAATCCGTTTGGCCAGATCCCGGCCTTCCAGGATGGTGATCTGCTGCTGTGGGAAAGT
CGTGCAATTAGCAAATATGTTCTGCGCAAATATAAAACCGATGAAGTTGATCTGCTGCGTGAAAGCAAC
CTGGAAGAAGCAGCGATGGTGGATGTTTGGACCGAAGTTGATGCCCATACCTATAATCCGGCACTGAG
CCCGATCGTGTATCAGTGTCTGTTTAACCCGATGATGCGCGGCCTGCCGACCGATGAAAAGTTGTTGC
AGAAAGCCTGGAAAAACTGAAAAAAGTGCTGGAAGTTTATGAAGCGCGTCTGAGCAAACACAGCTATC
TGGCCGGTGATTTCGTTAGCTTTGCAGATCTGAATCATTTCCCGTATACCTTTTATTTCATGGCAACCCCG
CACGCGGCCCTGTTTGATAGCTATCCGCATGTGAAAGCATGGTGGGATCGCCTGATGGCACGTCCGGC
AGTGAAGAAAATCGCCGCGACCATGGTGCCGCCGAAAGCCTAA

*Am*GSTF1 F122T Mutant Protein Sequence

MAPVKVFGPAMSTNVARVTLCLEEVGAEYEVVNIDFNTMEHKSPEHLARNPFGQIPAFQDGDLLLWESRA
ISKYVLRKYKTDEVDLLRESNLEEAAMVDVWTEVDAHTYNPALSPIVYQCLTNPMMRGLPTDEKVVAESLEK
LKKVLEVYEARLSKHSYLAGDFVSFADLNHFPYTFYFMATPHAALFDSYPHVKAWWDRLMARPAVKKIAAT
MVPPKA

*Am*GSTF1 F122T Mutant Nucleotide Sequence

ATGGCGCCGGTGAAAGTTTTTGGCCCGGCCATGAGCACCAACGTTGCACGTGTGACCCTGTGCCTGGA
AGAAGTTGGTGCAGAATATGAAGTTGTGAATATTGATTTCAACACCATGGAACATAAAAGCCCGGAAC
ACCTGGCGCGCAATCCGTTTGGCCAGATCCCGGCCTTCCAGGATGGTGATCTGCTGCTGTGGGAAAGT
CGTGCAATTAGCAAATATGTTCTGCGCAAATATAAAACCGATGAAGTTGATCTGCTGCGTGAAAGCAAC
CTGGAAGAAGCAGCGATGGTGGATGTTTGGACCGAAGTTGATGCCCATACCTATAATCCGGCACTGAG
CCCGATCGTGTATCAGTGTCTGTTTAACCCGATGATGCGCGGCCTGCCGACCGATGAAAAGTTGTTGC
AGAAAGCCTGGAAAAACTGAAAAAAGTGCTGGAAGTTTATGAAGCGCGTCTGAGCAAACACAGCTATC
TGGCCGGTGATTTCGTTAGCTTTGCAGATCTGAATCATTTCCCGTATACCTTTTATTTCATGGCAACCCCG
CACGCGGCCCTGTTTGATAGCTATCCGCATGTGAAAGCATGGTGGGATCGCCTGATGGCACGTCCGGC
AGTGAAGAAAATCGCCGCGACCATGGTGCCGCCGAAAGCCTAA

*Am*GSTF1 Y118S F122T Mutant Protein Sequence

MAPVKVFGPAMSTNVARVTLCLEEVGAEYEVVNIDFNTMEHKSPEHLARNPFGQIPAFQDGDLLLWESRA
ISKYVLRKYKTDEVDLLRESNLEEAAMVDVWTEVDAHTYNPALSPIVSQCLTNPMMRGLPTDEKVVAESLEK
LKKVLEVYEARLSKHSYLAGDFVSFADLNHFPYTFYFMATPHAALFDSYPHVKAWWDRLMARPAVKKIAAT
MVPPKA

*Am*GSTF1 Y118S F122T Mutant Nucleotide Sequence

ATGGCGCCGGTGAAAGTTTTTGGCCCGGCCATGAGCACCAACGTTGCACGTGTGACCCTGTGCCTGGA
AGAAGTTGGTGCAGAATATGAAGTTGTGAATATTGATTTCAACACCATGGAACATAAAAGCCCGGAAC
ACCTGGCGCGCAATCCGTTTGGCCAGATCCCGGCCTTCCAGGATGGTGATCTGCTGCTGTGGGAAAGT
CGTGCAATTAGCAAATATGTTCTGCGCAAATATAAAACCGATGAAGTTGATCTGCTGCGTGAAAGCAAC
CTGGAAGAAGCAGCGATGGTGGATGTTTGGACCGAAGTTGATGCCCATACCTATAATCCGGCACTGAG
CCCGATCGTGTATCAGTGTCTGTTTAACCCGATGATGCGCGGCCTGCCGACCGATGAAAAAGTTGTTGC
AGAAAGCCTGGAAAAACTGAAAAAAGTGCTGGAAGTTTATGAAGCGCGTCTGAGCAAACACAGCTATC
TGGCCGGTGATTTCGTTAGCTTTGCAGATCTGAATCATTTCCCGTATACCTTTTATTTCATGGCAACCCCG
CACGCGGCCCTGTTTGATAGCTATCCGCATGTGAAAGCATGGTGGGATCGCCTGATGGCACGTCCGGC
AGTGAAGAAAATCGCCGCGACCATGGTGCCGCCGAAAGCCTAA

*At*GSTF8 (Strep tagged) Protein Sequence

MASWSHPQFEKGLINHMASIKVHGVPMSTATMRVLATLYEKDLQFELIPVDMRAGAHKQEAHLALNPFG
QIPALEDGDLTLFESRAITQYLAEEYSEKGEKLISQDCKKVKATTNVWLQVEGQQFDPNASKLAFERVFKGM
FGMTTDPAAVQELEGKLQKVLDVYEARLAKSEFLAGDSFTLADLHHLPAIHYLLGTDSKVLFDSRPKVSEWIK
KISARPAWAKVIDLQKQ

(Strep tag is shown in red)

*At*GSTF8 (Strep tagged) Nucleotide Sequence

ATGGCTAGCTGGAGCCACCCGCAGTTCGAGAAAGGCTTAATTAACCATATGGCCAGTATCAAGGTTCA
CGGAGTCCCCATGTCCACCGCCACAATGCGCGTCCTCGCTACTCTTTACGAGAAAGATCTCCAGTTCGA
GCTCATCCCCGTCGATATGAGAGCCGGTGCCCACAAGCAAGAGGCCCACCTTGCCCTCAACCCCTTCGG
TCAAATTCCTGCTCTCGAGGACGGTGATTTGACGCTTTTCGAGTCAAGAGCCATCACACAGTACCTAGC
AGAGGAATACAGTGAGAAAGGTGAAAAGCTTATCTCCCAAGACTGCAAGAAAGTCAAGGCAACCACT
AACGTATGGCTTCAAGTTGAAGGTCAACAGTTTGACCCTAACGCCTCTAAGCTTGCCTTCGAGCGTGTC
TTTAAAGGCATGTTCGGCATGACCACTGACCCTGCCGCTGTCCAAGAGCTCGAAGGTAAGCTCCAGAA
AGTCTTGGATGTCTACGAGGCTAGGCTCGCCAAATCTGAGTTCTTGGCTGGTGATTCCTTCACTCTTGCT
GATCTTCACCACCTCCCAGCCATCCATTACTTGTTGGGTACTGACTCCAAGGTGCTCTTTGACTCTCGCCC
TAAGGTTAGCGAGTGGATTAAGAAGATCTCTGCCAGGCCTGCTTGGGCTAAGGTTATTGACCTCCAGA
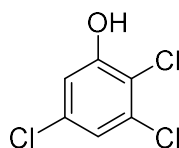AGCAG

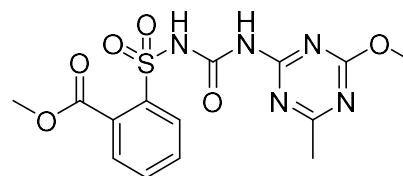(Strep tag is shown in red)

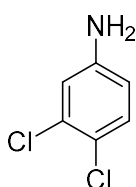# Appendix F: Herbicide Binding

H1: Herbicide Compounds Tested
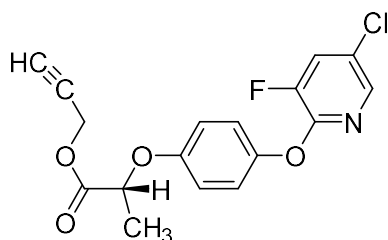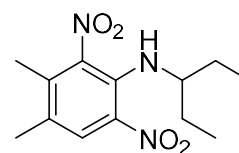

Clodinafop

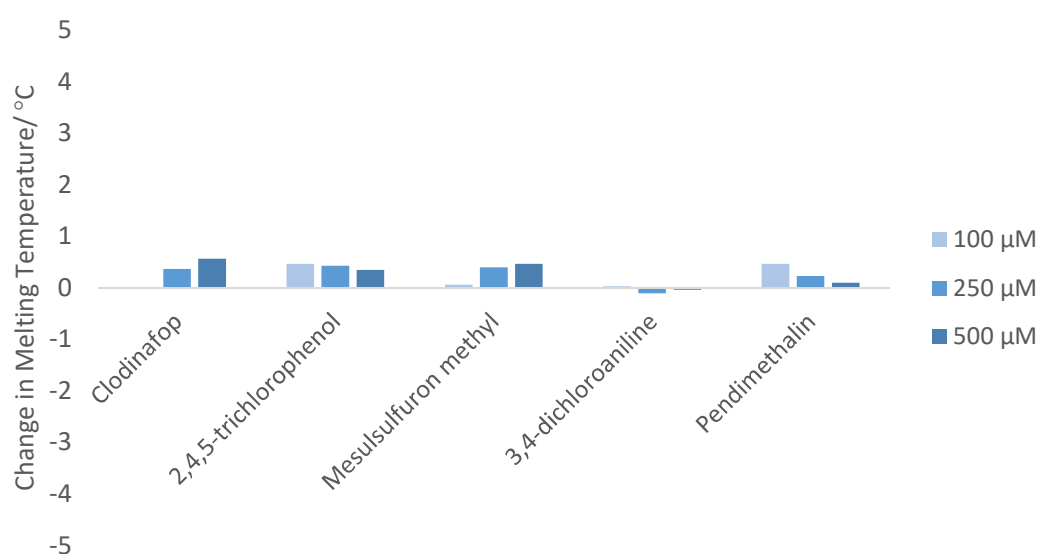2,4,5-trichlorophenol

Mesulfuron Methyl

3,4-dichloroaniline

Pendimethalin

H2: Thermal Shift Data for Herbicide binding to *Am*GSTF1

Appendix G: Apo Structures of *Am*GSTF1

G1: Details of datasets collected of apo *Am*GSTF1 from co-crystallography experiments using Flavonoid Inhibitors.

Dataset Number	Resolution/ Å	Ligand used for co-crystallography	Completeness/ %	R	R_{free}
1	1.52	COM1	99.9	0.23	0.25
2	1.55	MS-4-50	100	0.20	0.22
3	1.58	MS-4-48	100.0	0.25	0.27
4	1.59	MS-4-30-1	100.0	0.27	0.30
5	1.61	MS-4-50	100.0	0.27	0.29
6	1.7	COM1	100	0.24	0.28
7	1.75	MS-4-48	99.9	0.27	0.31
8	1.75	MS-4-30-1	93.7	0.26	0.28
9	2.4	MS-5-31	99.6	0.21	0.27