Structural Studies of Glutathione S-transferases from Plants

by

Russell Thom

UNIVERSITY
of
GLASGOW

A thesis submitted for the degree of Doctor of Philosophy in

The Faculty of Science, University of Glasgow

Declaration

This thesis has been written in accordance with the University of Glasgow regulations, has not been presented for a degree at any other university and is original except where indicated otherwise by reference in the text. The work contained within is the author's own except where work was done in collaboration as indicated.

Signed  Date 

Russell Thom **15.07.2001**

Summary

Glutathione S-transferases (GSTs) [EC 2.5.1.18] are a ubiquitous family of enzymes, widely distributed in most forms of life: bacteria, fungi, parasites, yeast, insects, mammals and higher plants, that catalyse the nucleophilic addition of the tripeptide glutathione (γ-Glu-Cys-Gly) to a wide variety of endobiotic and xenobiotic electrophiles.

In October 1997 when this work began, GSTs were classified into five families; Alpha, Mu, Pi, Theta and Sigma and a representative structure had been solved for each class. However, cDNA searching was beginning to indicate that more classes of GSTs existed.

Enzyme kinetic data for plant GSTs had also indicated that particular GSTs which shared closer sequence identity to each had distinct substrate specificities. However, it was unclear from the only plant GST structurally determined (AtGSTF2 determined in 1996) the structural features that might be promoting these substrate specificities.

The aim of this thesis was to structurally characterise a range of plant GSTs to increase the understanding as to which structural features determine substrate specificity.

During the course of this work, a total of ten plant GSTs from species including Maize, Wheat, Carnation, Petunia and *Arabidopsis thaliana* were overexpressed in *E. coli.* Nine of these overexpressed GSTs were purified to produce protein for crystallisation using a range of matrices such as Orange A agarose, S-hexyl glutathione and metal chelate followed by ion-exchange and/or gel filtration chromatography depending on the particular GST.

Initial crystallisation conditions for seven of these purified GSTs were found by sparse matrix screening. Crystallisation conditions providing crystals suitable for X-ray diffraction experiments were determined for four of the GSTs under study. The data gained from these diffraction experiments enabled the solution of four different plant GST three-dimensional structures.

The first structure to be solved was ZmGSTF1, a GST isoenzyme constitutively expressed in Maize. This isoenzyme, which is able to metabolise the commercially important herbicide atrazine, was solved in two crystal forms belonging to the spacegroups P432 and $P2_1$ to a resolution of 2.8Å and 2.9Å respectively. Initial phases allowing calculation of electron density maps were calculated by molecular replacement with the program AmoRe.

The second structure to be solved was AtGSTZ1 from *Arabidopsis thaliana*. Data for this GST isoenzyme, believed to take part in endogenous metabolism, was collected to 1.65Å resolution and phase information was determined using single isomorphous replacement with anomalous scattering (SIRAS).

The third GST structure determined (AtGSTT1), again from *Arabidopsis thaliana*, was found to share significant homology with the mammalian Theta class GSTs. The 2.3Å structure was determined following calculation of phase information by the molecular replacement program EPMR, which uses evolutionary programming to perform a limited search of six dimensional space.

The fourth structure determined was a GST isoenzyme from wheat, able to metabolise the commercially important herbicide fenoxaprop. This structure determined to 2.3Å resolution contained a molecule of the substrate analog S-hexyl glutathione bound in the active site. Phase information for this isoenzyme was again calculated using the molecular replacement program EPMR.

These structural models provide a detailed understanding of the structure determinants of a variety of GSTs which dictate the different substrate

specificity's of GSTs and provide suggestions for the rational design of GSTs to improve herbicide selectivity in crops.

In addition, this study examined the possible roles of GSTs within endogenous metabolic pathways and presents material relating to the in-vivo role of plant GSTs and their binding to endogenous substrates.

Acknowledgements

My thanks go to a great number of people who have provided assistance and advice throughout this project.

In particular I thank my supervisor Dr Adrian Lapthorn for his enthusiasm and for generally being 'a good supervisor whatever that may be', throughout all parts of the project. My collaboration with Durham University was extremely productive and I am indebted to Dr David Dixon and Dr Robert Edwards for the constructs they supplied, helpful discussions and guidance throughout the course of the project. My thanks also to Professor Virginia Walbot and Dr Lukas Müller for the kind gifts of their constructs AN9 and ZmGSTF4. In addition, I thank AVENTIS Crop Science Ltd. for funding the project and in particular Dr David Cole for his enthusiasm, support and discussions.

I owe much to all the members of the Protein Crystallography Group including; Neil Isaac, Andy Freer, Alex Rosack, Paul Emsley, Dina Fotinou, Jeremy Beauchamp, Neil Hanlon, Isobel Black, John Maclean, Derek Thomson, David Robinson and Claire Martin for being both patient and sharing their crystallographic experience with me. I also thank them for making the last three years a lot of fun.

Finally I thank my mum and dad for nurturing my curiosity and always believing that one day my thesis would be finished. And last, but not least, I would like to thank my fiancée Claire for her continual support and patience, while I spent evenings and weekends both researching and writing up this work.

Abbreviations

AOS	Active Oxygen Species
ATP	Adenosine Tri Phosphate
Da	Dalton
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra acetate
EST	Expressed sequence Tag
FAA	Fumaryl Acetoacetate
FFT	Fast Fourier Transform
GSH	Reduced glutathione
GSSG	Oxidised glutathione
GST	Glutathione S-transferase
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
IAA	Indole acetic acids
IPTG	Isopropyl-β-D-thiogalactoside
LB	Luria-Bertani medium
MAA	Maleylacetoacetate
MAAI	Maleylacetoacetate Isomerase
MAD	Multiple Anomalous Dispersion
MES	2-[N-Morpholino] ethanesulfonic acid
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement with Anomalous Scattering
MOPS	3-[N-Morpholino] propanesulfonic acid
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PEG	Polyethylene Glycol
SIRAS	Single isomorphous replacement with anomalous scattering
SDS	Sodium dodecyl sulphate
TEMED	N, N, N', N'-tetramethyl-ethylenediamine
Tris	Tris (hydroxymethyl)aminomethane
ε	Extinction coefficient

TABLE OF CONTENTS

List of Figures

List of Tables

List of Equations

Chapter 1: The Roles of Glutathione S-Transferase

1.1 Introduction

In 1798 Thomas Malthus postulated that population, when unchecked, would increase at a rate such that food supplies would fall below the level where life could be sustained (Malthus, 1798). Despite the huge population increases since the 1960's, several factors including plant breeding, agrochemical discovery and farm management techniques known as the 'green revolution' have prevented Malthus prediction from being realised. The 'green revolution' has enabled cereal grain harvests to more than double in the last forty years (Mann, 1999). Although the world now produces enough food to feed everyone if it were evenly distributed, this technological achievement will have to be repeated as we enter the 21st century to keep in line with population growth.

Year	World Grain Area (million hectares)	World Grain Production (million tons)	World Grain Production per person (kilograms)
1955	639	759	273
1965	653	904	270
1975	708	1237	303
1985	715	1646	339
1995	679	1703	299

Table 1-1 Global grain production over last 40 years with respect to area of cultivation and population (http://newmedia.avs.uakron.edu/geology/ge/ch/pop/natsys.htm)

The world population is projected to grow by 73 million people a year, equivalent to the current population of the Philippines, (United Nations, 1999) between 1995 and 2020 causing the global demand for rice, wheat and maize to increase by 40% overall or 1.35% per year (Rosegrant, 1995; Rosegrant, 1997). Despite striking increases in crop yields over the last forty years, the population growth over the last 20 years has meant there has been little increase in food production per capita. In recent years increases in cereal yields per hectare have begun to decline from 2.2% per year during the period 1967-82 to only 1.5% per year in the period 1982-94 (Abelson, 1999). With the exception of Africa the production gains over the last forty years have been almost entirely due to

improvements in crop yields, as most of the world's arable land is already under cultivation. Previously Africa was able to show increased crop production through cultivation of new land, but recently the point has been reached where increases can only be gained by cultivation of environmentally sensitive areas or through improved farming methods (Pinstrup-Andersen, 1994).

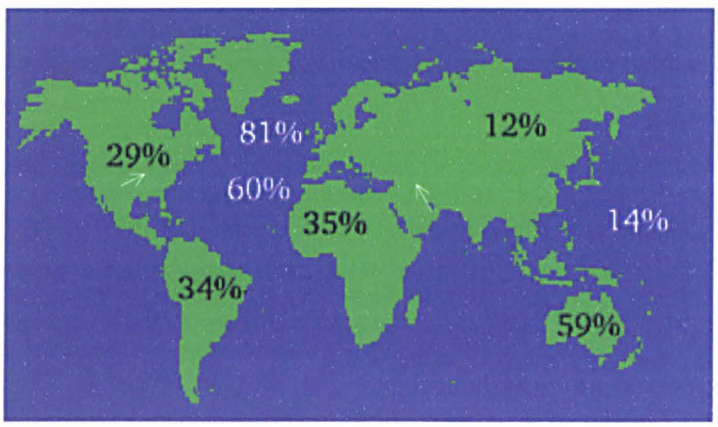


Figure 1-1: Approximately 38% of the earth's landmass has been converted into cropland, pasture and urban uses. This figure shows the domesticated land area as a proportion of land available in Ireland 81%, Spain 60%, Australia 59%, Africa 35%, Brazil 34%, United States 29%, Japan 14% and CIS 12%
(http://newmedia.avs.uakron.edu/geology/ge/ch/pop/natsys.htm)

Additional factors, which make feeding the global population more difficult, are increasing urbanisation, which has resulted in an annual loss of nearly half a million hectares (467,000), and a demand for "Western-style" diets. Demand for western style diets, i.e. for more meat and dairy products, is greatest in those developing countries with improved living standards such as China. The adoption of western style diets means these countries have an increased need of grain to enable them to feed domestic cattle and livestock not previously required to sustain their more traditional diets.

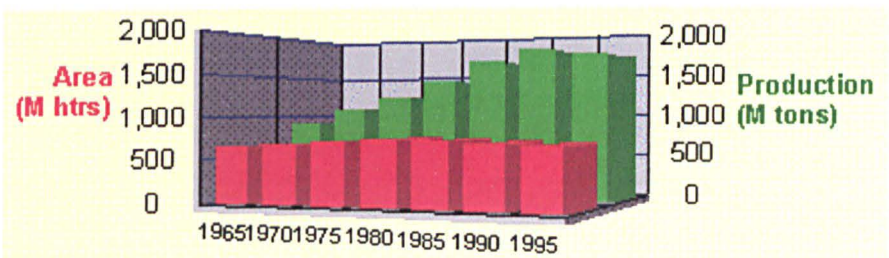


Figure 1-2: Grain production during 1965-1995. The area of land farmed between the years 1965 to 1995 has been relatively stable, however in the same period of time the total amount of grain produced has shown significant increases due to increased yields
(http://newmedia.avs.uakron.edu/geology/ge/ch/pop/natsys.htm)

Original in colour

One of the most significant factors, regardless of the farming system used that has enabled an increase in crop yields was the development of a range of herbicides. The many herbicides produced by the agrochemical industry, control weeds using various mechanisms and have provided a system of weed control, that has proven to be an essential component of profitable crop production (Table 1-2). These herbicides are typically organic molecules which interfere with vital plant specific processes such as; photosynthetic electron transport and the biosynthesis of essential amino acids, porphyrins, carotenoids, fatty acids or cellulose (Cole, 1994; Oxtoby & Hughes, 1990).

Type of Herbicide	Mechanism of action	Examples
Root mitotic inhibitors	Block cell division	trifluralin, pendimethalin, ethalfluralin
Pigment inhibitors	Contribute to chlorophyll destruction	clomazone, norflurazon
Fatty acid inhibitors	Block the synthesis of very long chain fatty acids Affect cell growth and division	EPTC, butylate alachlor, metolachlor, acetochlor, dimethenamid
Photosynthetic inhibitors	Block electron transfer, damage plant membranes	atrazine, cyanazine, simazine, metribuzin, bentazon, bromoxynil, paraquat, pyridate
Hormone (auxin)-type	Affect protein synthesis and cell division	2,4-D, 2,4-DB, dicamba, picloram,triclopyr, clopyralid, MCPA
ALS enzyme inhibitors	Block acetolactate synthase (ALS), Blocks cell division and metabolism	sulfonylureas & imidazolinones - iamzethapyr,clorimuron, imazaquin, nicosulfuron, thifensulfuron, halosulfuron, primisulfuron + prosulfuron, rimisulfuron + thifensulfuron
Acetyl CoA carboxylase inhibitors	Block lipid formation	sethoxydim, fluazifop, quizalofop, fenoxaprop, diclofop, clethodim
Protoporphyrinogen oxidase inhibitors	Disrupt internal cell membranes, prevent electron transfer	acifluorfen, lactofen, fomesafen, flumiclorac-pentyl
Amino acid biosynthesis Inhibitors	Inhibitors of amino acid biosynthesis via a number of targets	glyphosate, phosphinothricin

Table 1-2: Mechanisms by which herbicides kill weeds and injure crops.

Depending on the crop and weeds present, weeds are significant pests that can:-

- reduce crop yields by up to 100% depending of the crop system and geography, by competing with the crops for nutrients, water and sunlight.
- impair crop quality due to contamination with weed material (some of which may be toxic or unpalatable).

- increase the likelihood that the crop will be attacked by insects or disease by acting as pest or disease reservoirs.
- prevent the use of mechanical harvesting equipment.

Although weeds can be controlled mechanically (tilling and hoeing), chemically (herbicides) or culturally (crop rotation), the effectiveness and economical benefit of herbicides means that mechanical and cultural weed control practices are rarely used alone in commercial farming (Cobb, 1992; Parry, 1989). A key requirement of all herbicides is that they must be able to kill weeds with minimum damage to the crop. In view of this herbicides are divided into two broad types, non-selective or selective to describe the way that the damage they might cause to the crop is minimised.

Non-selective herbicide products are often used where all vegetation is to be destroyed, for example; roadsides, railroad tracks and industrial sites. Until relatively recently non selective herbicides used in an agricultural setting could only be applied to fields when no crop was growing (before planting or following harvest) which led to weed control problems during the growing season. Genetically enhanced crops (Table 1.3) have changed this. The engineering of crops by either altering the level and sensitivity of the target enzyme for the herbicide (LeBarron & Gressel, 1982) or the introduction of detoxifing resistance genes (Gressel, 1985) has enabled crop varieties to be produced which show artificial resistance to previously phytotoxic non-selective herbicides. An example of this is the Roundup Ready Soybean produced by Monsanto which is resistant to the environmentally favourable non-selective herbicide, glyphosate. The use of herbicide resistant crop varieties allows the use of a non-selective herbicide during the growing season to control a wide spectrum of weed species and offers several potential advantages over using a selective herbicide, for example; reduced cost, fewer applications of herbicide and the use of only one herbicide. A potential disadvantage of herbicide resistant crops is the possible evolution of herbicide resistant weeds or a shift in the spectrum of troublesome weeds, due to an over reliance on a single class of herbicide.

The suggestion that weeds, previously susceptible to herbicide control might become resistant was first proposed in 1956 (Harper, 1956). Herbicide resistance was first discovered in the United States, during the late 1960s, in a pine nursery, where triazine herbicides had been used repeatedly (Ryan, 1970). It has since been determined that weeds typically become resistant to a particular herbicide when said herbicide is used consecutively for 4-10 years. In addition to becoming resistant to a particular herbicide, the plant may become resistant to a class of herbicides or indeed to chemically unrelated herbicides. This more serious consequence can arise due to mutations occurring in the herbicides target enzyme, or the plant developing mechanisms by which to metabolise the herbicide before it can act.

Herbicide/ active ingredient	Chemical Type	Resistance gene product	Resistant Crops	References
Pursuit, Contour, Resolve, Lightning (imazethapyr)	Imidazolinone	Mutant acetolactate synthase (ALS) genes	Imidazolinone tolerant Corn & Tobacco hybrids	(Lee et al., 1999)
Poast, Poast Plus, Headline (sethoxydim)	Cyclohexendione	(Mutant acetyl-coenzyme A carboxylase – ACCase)	Sethoxydim resistant Corn hybrids	(Young & Hart, 1997)
Liberty (glufosinate)	Amino acid phenoxy	N-Acetyl transferase (bacterial)*	Liberty Link or Glufosinate Corn, Wheat, Rice, Cotton & Soybean hybrids	(Deblock et al., 1987)
Roundup Ultra (glyphosate)	Amino acid phenoxy	EPSPS gene amplification, AroA gene (bacterial)*	Roundup Ready Soybean & Corn Tobacco and Tomato hybrids	(Comai et al., 1985)
Synchrony, Reliance	Sulfonylurea	Mutant Acetolactate synthase (ALS) genes	Soybean, Tobacco	(Lee et al., 1988)
Bromoxynil	Nitrile	Nitrilase*	Oilseed Rape, Cotton, Potato	(Stalker et al., 1988)
2,4-D	Amino acid phenoxy	Monooxygenase from gene *tfd*A*	Cotton	(Bayley et al., 1992)
Dalapon	Chlorinated aliphatic acids	Dehalogenase *	Tobacco	(Buchananwollaston et al., 1992)
Phosphinothricin	Glutamine analog	Gluatamine synthase gene amplification bar gene (bacterial)*	Tobacco	(Deblock et al., 1987)

Table 1-3: Resistant crop varieties now on the market.

* Genes introduced to plant from another source.

The selection of naturally herbicide tolerant variants which become dominant with time is now recognised as a problem of modern farming methods and techniques such as crop rotation and herbicide mixtures are currently being used to try and minimise selection pressures (Cole, 1994).

Selective herbicides, unlike non-selective herbicides, are able to distinguish between species of plant due to differences in the uptake, movement or metabolism of the herbicide allowing their use during the growing season without causing crop damage.

Figure 1-3: Isoxaflutole plus chloracetanilide pre-emergence treatment showing selective grass weed control in maize. Untreated controls are shown at either side of the picture, while treated plants are shown in the two middle rows of the picture. The treated plants can be observed to be taller and stronger than the control plants.

A range of selective herbicides, working against different targets, is commonly applied to a crop, minimising the generation of resistant weeds. Although the basis for selectivity can be complex, with differences in herbicide uptake and transport being factors (Devine, 1993), the single most important factor in herbicide resistance is the ability of the plant to detoxify the herbicide rapidly. Many enzymes involved in xenobiotic metabolism appear to exist in multiple isoforms and commonly each isoform exhibits stringent substrate specificities with little overlap. The existence of particular enzyme isoforms, expressed at differing levels between species, allows the exploitation of subtle differences between the crop and weed metabolism to select compounds which damage the weed, but not the crop. To date, random compound screening has been used to

determine compounds which show both herbicide activity and are metabolised at different rates between the crop and the weed. Although this method, along with imitative chemistry, has led to the discovery of all the herbicides currently in use, it is hoped that a greater understanding of plant metabolism will allow an extension of this screening approach to allow targeting of specific enzymes involved in metabolism. Until recently the chemical tailoring of herbicides to achieve optimum selectivity has been impossible, as the plant specificity and substrate affinities of enzymes involved in herbicide metabolism have been unknown. However, as the genomes of more crop plants are elucidated and the genes involved in herbicide metabolism are cloned, enzyme kinetics and knowledge of the three dimensional structures of herbicide metabolising enzymes will provide the necessary information to rationally design herbicides with improved selectivity.

1.2 The Detoxification System

Selective herbicides rely on differences in uptake, movement or metabolism between the crop and weed to function. In some cases these differences can be marked, for instance when compared to upright narrow-leaved crop plants, broad-leaf weeds have more leaf surface area exposed to herbicide compounds and light and thus it may be postulated that they would be easier to damage by agents which cause bleaching of the leaves (Cobb, 1992). The design of herbicides able to distinguish between two similar species, is more challenging and knowledge of the subtle biochemical differences between different plants is required.

Research over the last few decades has revealed that plants naturally tolerant to certain herbicides often have very efficient means of herbicide inactivation or removal from the target. Agrochemical companies have therefore focussed much of their research into understanding the metabolism of foreign compounds with the aim of detecting differences in the metabolism of these compounds between species, which might be usefully exploited in the design of new herbicides. Details of plant physiology and herbicide metabolism have been described in a number of texts (Cobb, 1992; Dodge, 1982; LeBarron et al, 1982) and articles

(Cole & Edwards, 2000; Sandermann, 1994), therefore only a brief description of plant herbicide metabolism is included. This brief description summarises the various roles particular enzymes play in the different phases of metabolism and describes the points at which the glutathione S-transferase family of enzymes act.

In both plants and animals the reactions which occur during the metabolism of toxic chemicals, can be categorised in one of three phases:

- Phase I - Transformation,
- Phase II - Conjugation
- Phase III - Compartmentation

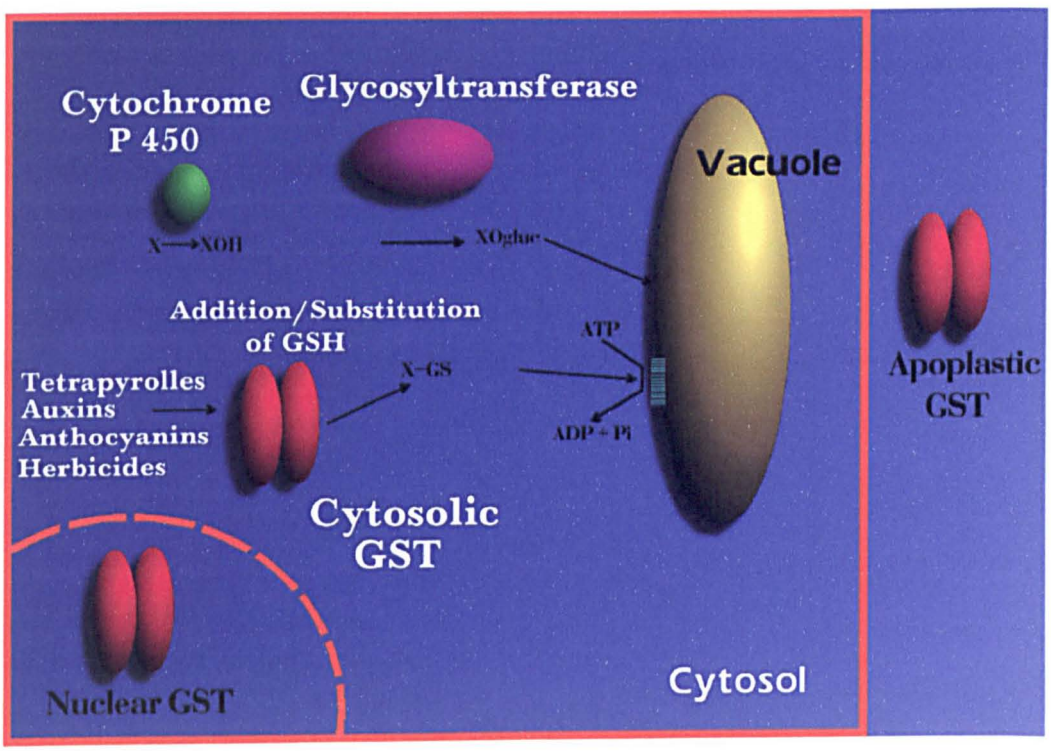


Figure 1-4: Detoxification in the plant cell. Cytochrome P450 enzymes perform phase I reactions while glycosyltransferases and glutathione s-transferases (GSTs) act in phase 2. Sugar and glutathione conjugated molecules are targeted to the vacuole where the are compartmentalised. GSTs are also observed in the nucleus and apoplasm, however their function has not been described.

1.2.1 Phase One: Transformation

1.2.1.1 Cytochrome P450

Oxidation, reduction or hydrolysis via enzymes of the cytochrome P450 superfamily are commonly the primary steps by which plants attack agrochemicals. These reactions form activated compounds with the functional

groups created during this phase acting as "handles" to enable the enzymes of phase II metabolism to act.

In contrast to their detoxifying role, cytochrome P450 molecules may also hydroxylate pro-herbicides, inactive esters that may possess more favourable physiochemical properties for uptake and transport by the plant, to yield the functionally active form of the herbicide.

$$XH + O_2 + NADPH \longrightarrow XOH + H_2O + NADP^+$$

Figure 1-5 Hydroxylation of xenobiotic by Cytochrome P450 enzyme which also requires molecular oxygen, NADPH and NADPH-Cyt-P450 reductase to perform the reaction

Although many herbicide classes contain aromatic ring structures that can be hydroxylated by cytochrome P450 monooygenases, many of the P450 enzymes present in plants show distinct substrate specificity. The distinct substrate specificity between species shown by cytochrome P450 enzymes has been exploited to produce a number of selective herbicides, such as Bentazon (3-isopropyl(1-H)2,1,3,benzothiadiasin-4(3H)1,2,2-dioxide).

Specific cytochrome P450 enzymes in rice convert Bentazon to 6-hydroxy Bentazon which can then be conjugated to glucose. This water soluble glucosidic product can then be transported out of the cytosol, preventing injury to the plant (Mine et al, 1975). *Cyperus serotinus* can hydroxylate Bentazon around 200 times less effectively rice. This slower rate of hydroxylation means that *Cyperus serotinus* is not resistant to Bentazon thus this herbicide can be used to selectively injure *Cyperus serotinus* in preference to rice.

Diclofop-methyl, a selective herbicide in wheat, is also detoxified by aryl hydroxylation. Resistant wheat is able to hydroxylate the diclofop molecule in the dichlorophenyl ring and thereby inactivate the herbicide. In contrast susceptible weeds, such as wild oat (*Avena fatua L.*) are unable to hydroxylate the diclofop molecule in this position and thus are unable to inactivate the herbicide. Instead these susceptible plants form a glycosyl ester of unaltered diclofop. This glycosyl reaction is readily reversible and allows the production of the active herbicide again (Cole, 1994).

A further instance of resistance through rapid hydroxylation has been observed for the sulfonylurea class of herbicides, with the herbicide Chlorsulfuron. This herbicide is often hydroxylated at the phenyl ring in susceptible grass species, (Sweetser et al., 1982) whereas tolerant broad leaved plants generally hydroxylate the methyl group present on the triazine ring (Hutchison et al., 1984) (figure 1-6).

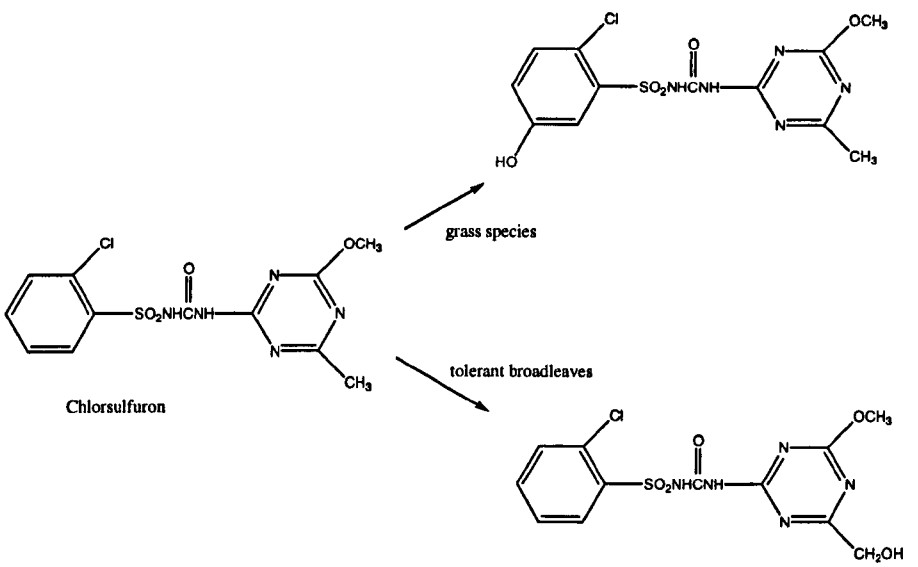


Figure 1-6: The sulfonylurea class shows differential hydroxylation in susceptible grasses to that observed in tolerant broad-leaved plants.

In some cases, the action of cytochrome P450 enzymes on a xenobiotic does not remove its toxicity *per se*, but does generate electrophilic centres within the compound. The electrophilic centres created by the cytochrome P450 enzymes may allow enzymes of phase II metabolism to act further on the xenobiotics to remove their toxicity or to generate more water-soluble compounds, which may be removed from the cytoplasm.

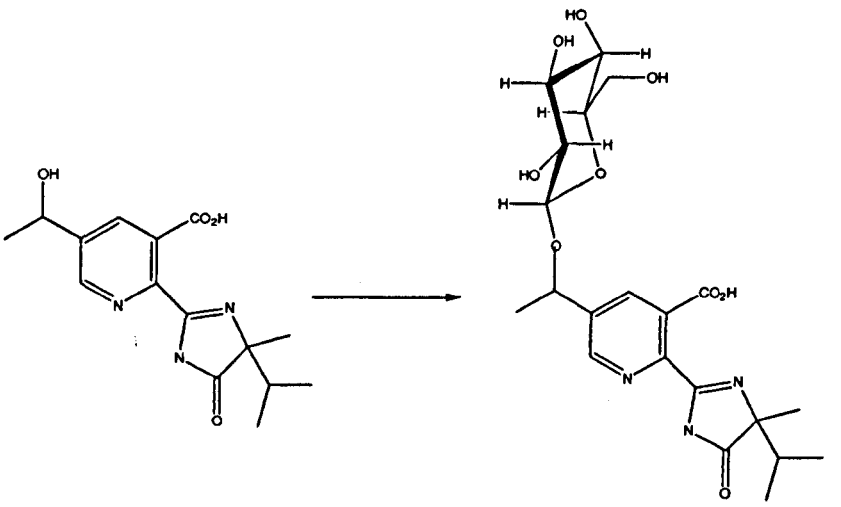
1.2.2 Phase Two: Conjugation

Following attack by phase I enzymes the activated compounds often undergo a number of different conjugation reactions. These reactions may enable the plant to increase the water solubility of potentially harmful compounds, and thus reduce their reactivity, toxicity and mobility within the plant symplast. In

addition, conjugation allows targeting of the molecules for removal to the cell vacuole, apoplast, or plant matrix (Sandermann, 1994). Plant secondary metabolism has been shown to be an important factor in herbicide resistance for maize, rice, soybean and wheat (Cole et al., 1997). The diversity in both the types of enzymes involved in phase two of metabolism and their differential expression between species, provides scope for the production of selective herbicides or tolerant crops. Alteration of either the substrate specificity of particular detoxifying enzymes, such that they are more able to detoxify particular compounds, or the over expression of particular detoxifying enzymes, which are able to metabolise compounds with herbicidal action, are two methods which may allow the development of new selective herbicides (Cole & Edwards, 2000).

1.2.2.1 Glycosylation

Oxidation or hydrolysis to create a hydroxy, amino or carboxylic acid functional group or groups is often followed by rapid glycosylation. Although hydroxylated herbicide metabolites generally show little phytotoxicity, glycosylation is often required to complete the detoxification. In soybean, hydroxylation of the herbicide imazethapyr causes it to be 2-3 times less reactive. However following the addition of a glucosyl moiety, the herbicide becomes completely inactive against its target, acetolate synthase (ALS) (Tecle et al., 1993). Plants, like mammalian systems, can form negatively charged O-malonyl -(O-glucosyl) double conjugates. They also contain O and N-glucosyl and O and N-malonyl transferases enabling them to form O-(-D glucosides, N-glucosides and 1-O-Glc esters (Sandermann, 1992a). A drawback to glycosylation as a detoxification route is that generally the sugar conjugated xenobiotics are not stable and may convert back to the active herbicide.



OH-Imazethapyr O-Glc-Imazethapyr

Figure 1-7: Soybean is tolerant to the imidazoline herbicide imazethapyr. Following hydroxylation of the 5-ethyl substituent on the pyridine ring this hydroxylated herbicide metabolite is *O*-glucosylated causing its complete inactivation.

1.2.2.2 Amino Acid Conjugation

Amino acid conjugation is a well-characterised form of detoxification within animals, (Huckle et al., 1981) however it appears to be of only limited importance with respect to agrochemical detoxification. One of the few examples of amino acid conjugation mentioned in the literature, is the conjugation of a variety of amino acids, mainly aspartate and glutamate, to the phenoxyacetic acid herbicide, 2,4-D. This reaction has been described in soybean and has been postulated to be analogous to the metabolism of the auxin molecules that the herbicide 2,4-D mimic (Cohen & Bandurski, 1982). In addition to 2,4-D, amide conjugation has also been described for the pyrinyloxyacetic acid herbicide triclopyr in cereals, and the conjugation of 4-hydroxy chloproham in chickweed and oat. Typically amide conjugates are biologically labile and herbicidally active, therefore they do not represent an ideal target on which to rationally improve herbicide selectivity.

1.2.2.3 Glutathione Conjugation

One of the most important conjugation reactions that occurs in plant herbicide metabolism is with the major non-protein cellular thiol, glutathione ((-L-glutamyl-L-cysteinyl-glycine), GSH). Glutathione has a high number of hydrophilic functional groups which along with its low molecular weight leads it to have a high water solubility (Kosower, 1976). Several members of the

glutathione S-transferase enzyme family irreversibly conjugate glutathione to a wide range of herbicides to form less reactive and more polar compounds. In addition, the glutathione tag also acts as a signal for the removal of the conjugated herbicide molecules from the cell. Glutathione S-transferase (GST) enzymes attack the electrophilic centre of the substrate (X-Z) with the GSH thiolate anion to form the conjugate with associated displacement of a nucleophile (Z: e.g. a halogen, phenolate or alkyl sulfoxide) (figure 1-8). Many herbicide classes such as; chloroacetanilides, triazines, triazinone, sulfoxides, aryloxyphenoxy-propionates, thiocarbamate sulfoxides, diphenyl ethers and sulfonyl ureas are directly conjugated by GSTs (Shimabukuro et al., 1970), however electrophilic centres formed by enzyme phase I of metabolism allow many additional xenobiotics to be detoxified by GSTs.

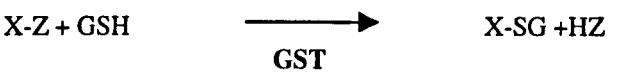


Figure 1-8 Conjugation of glutathione to a xenobiotic by glutathione S -transferase

Plant GSTs have also been described in the normal metabolism of plant secondary products such as anthocyanins and maleylacetoacetate, as discussed in section 1.3.2.

1.2.3 Phase Three: Compartmentation

1.2.3.1 Glutathione Pump

Phase three in plant xenobiotic metabolism differs from the mammalian system in the destination of the conjugated compounds exported from the cytoplasm. In mammals the GSH conjugated compounds are excreted out of the cell, whereas in plants they are stored in the vacuole, (Coleman, 1997; Schmitt & Sandermann, H., 1982). In either case the removal of glucosides and GSH conjugates from the cytosol has important consequences. GSTs show significant binding to GSH conjugates, with many GSH-conjugates bound more avidly *in vitro* than either GSH or the respective second substrate. *In vitro* data suggests the K_i values of the major human cytosolic GST for GSH-CDNB lies between 15 and 90μM, whereas K_m values for the same enzyme with CDNB and GSH alone lie between 450-910μM and 80-160μM respectively. Although the validity of this data in *vivo* is unknown, it suggests GSTs might sequester conjugates within the cell

(Meyer, 1993). Sequestration of conjugates by GST would prevent toxic compounds interacting with target molecules in the cytosol and stabilise reactive conjugates, however it would also give rise to end product inhibition of the enzyme. Export of the GSH-conjugates would prevent end product inhibition and thus sustain enzyme activity. Conjugate export plays a second important role in that it removes any conjugated compounds which might still exert biological activity from the cytosol. The importance of this second effect is exemplified by the herbicide synergist tridiphane, which when conjugated to GSH inhibits GST mediated atrazine metabolism (Lamoureux & Rusness, 1986).

Transport of the GSH-conjugates across the vacuolar membrane occurs in plants as in mammals via members of the ABC (ATP-binding cassette) superfamily. These large membrane bound proteins contain two membrane spanning domains and an ATP binding region. They are divided into two major subclasses, the multidrug resistance-associated proteins (MRPs) and the multidrug resistance proteins (MDRs). Several GS-X pumps belonging to the MRP class exist in *Arabidopsis thaliana* (Lu et al., 1998; Lu et al., 1997; Martinola et al., 1993; SanchezFernandez et al., 1998). Contrary to the prevailing chemiosmotic theory for energy dependent transport in plants, characterisation of these pumps has found that they are insensitive to the transmembrane H^+- electrochemical potential difference and can be inhibited by micromolar concentrations of the phosphoryl transition state analog, vanadate. These GS-X pumps can be driven by MgATP, but not free ATP or non hydrolysable analogues such as 5'-(ß, y-imino) triphosphate (AMP-PNP). The pumps are further described as being able to transport a variety of glutathionated compounds, including the glutathionated chloroacetanilide herbicide metalochlor and also GSSG, however they are unable to transport free GSH.

Three major pathways have so far been detected for phase three metabolism in plants, but the mode of action of each pathway is unclear. Xenobiotics are either exported into the plant cell vacuoles, exported into the extracellular space or deposited into cell wall in association with pectin, hemicellulose, or liginins and targeted into the apoplast as an insoluble conjugate termed a "bound residue" (Sandermann, 1992b). Due to current fractionation techniques having an inherent

bias in favour of the isolation of sealed vacuole membrane vesicles in comparison with transport-competent plasma vesicles, it is not known if the GS-X pump exclusively localises to the vacuole membrane of plant cells. The original physiological function of the GS-X pump is also currently unknown. As has been postulated for GSTs it seems untenable that the primary function of the GSH-pump is the transport of synthetic compounds whose manufacture and application has only been widespread for around 50 years. The true physiological function of GS-X pumps might be to remove conjugated secondary metabolites such as anthocyanin (Marrs et al., 1995), or they might be important in the efficient export of GSSH from the cytosol during oxidative stress (Kreuz et al., 1996).

1.3 The Glutathione S-Transferase Family

The irreversible conjugation of glutathione to a wide range of xenobiotics by GSTs, together with their species specific expression as various isoforms, make GSTs good targets for rational herbicide design. Knowledge of the substrate specificity of a specific GST isoform selectively expressed in crops, but not in weeds, would enable chemicals identified as having herbicidic properties in screening programs, to be modified to allow their rapid detoxification from the crop, but not the weed, while retaining their herbicidic activity.

Glutathione S-transferases were first identified in 1961, (Booth, et al., 1961; Combes & Stakelum, 1961) as an enzymatic activity in the cytosolic fraction of homogenised liver that catalysed the addition of glutathione (GSH) to 1,2, dichloro-4-nitrobenzene (CDNB).

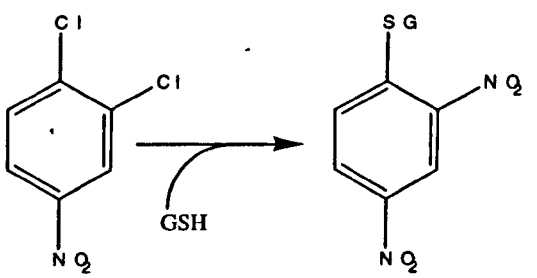


Figure 1-9: 1,2,dichloro-4-nitrobenzene (CDNB) reacts with glutathione to form S-(2,4-dinitrophenyl) glutathione. This reaction shows a significant change in absorption at 340 nm providing a simple spectrophotometric assay to detect GST activity (Habig & Jakoby, 1981).

During the next decade several Glutathione S-transferases were purified to homogeneity. It was elucidated that these cytosolic enzymes comprised a superfamily which could be subdivided into particular classes.

Further characterisation of distinct isoenzymes determined that particular isoforms were able to catalyse the addition of glutathione to distinct substrates. During the 1970's many GST isoenzymes were sequenced. The growth of sequence databases enabled cDNA searching revealing that GSTs exist ubiquitously in vertebrates, plants, insects and aerobic microorganisms. Mammalian GSTs have been extensively studied and are important in the detoxification of both carcinogens and the drugs. Plant GSTs have received

considerably less attention, but it is clear that they play a similar detoxifying role and are crucial in the removal of certain herbicides. In addition to forming glutathione conjugates, GSTs have also been shown to serve as peroxidases and isomerases (Dixon et al., 1998; Hayes & Pulford, 1995).

1.3.1 Herbicide detoxification mediated by Glutathione S-transferases

The first known function of plant GSTs was the metabolism of atrazine and the metabolism of herbicides has since provided much of the driving force behind GST research for the last 30 years (Frear & Swanson, 1970; Shimabukuro et al, 1970; Timmerman, 1989). Plant GSTs have been shown to be responsible for the tolerance of maize to atrazine (Shimabukuro et al, 1971), the chloroacetanilides, alachlor (Rossini et al., 1996), and metalachlor (Irzyk & Fuerst, 1993). In addition they have been shown to metabolise examples in herbicide classes such as; nitrodiphenyl esters, flurodifen (Frear, 1973) the oxirane derivative tridiphane (Lamoureux & Rusness, 1986) and the sulfonylureas (Brown & Neighbours, 1987) (figure 1-10).

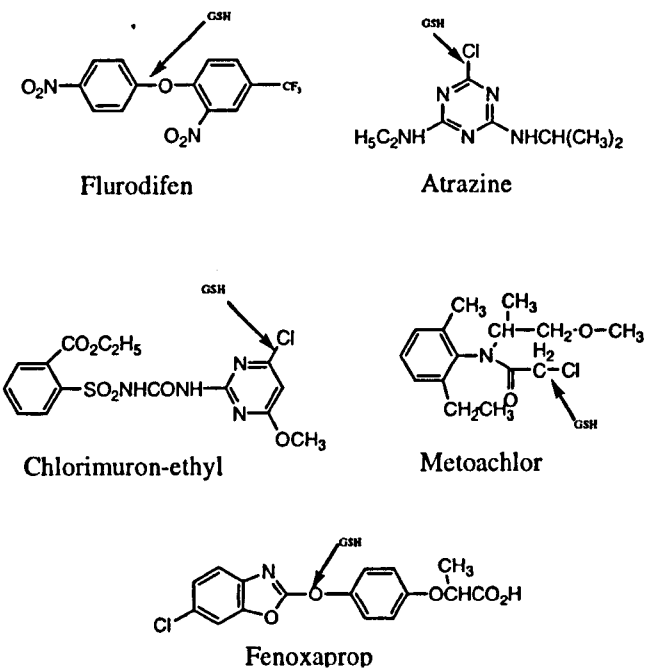


Figure 1-10: Conjugation of herbicide substrates by glutathione; Fenoxaprop (Aryl-oxy-phenoxy), Flurodifen (Diphenylether), Metolachlor (Chlroacetamide), Chlorimuron-ethyl (Sulfonylurea) & Atrazine (Traizine)

Studies of herbicide tolerance in maize and millet first suggested different isoenzymatic forms of GST existed. These different enzymatic forms were revealed because some species such as; cotton, maize, pea, peanut, okra and

soybean were tolerant to the herbicide flurodifen, while others species, such as; tomato, cucumber and melon were susceptible (Frear & Swanson, 1973). The discovery that maize was tolerant to atrazine and EPTC while *Panicum miliaceum*, (Proso Millet) was tolerant only to atrazine, (Ezra & Stephenson, 1985) further suggested that differences in isoenzymatic distribution, expression level and substrate selectivity of the GST complement within a plant was responsible for the differential ability of plant species to metabolise herbicide (Cole, 1994; Lamoureux et al., 1991).

Herbicide studies in pea have also shown that GSTs display high substrate selectivity. The GSTs present in pea are able to conjugate glutathione to flurodifen, but not to the structurally related herbicide nitrofen (Frear & Swanson, 1973).

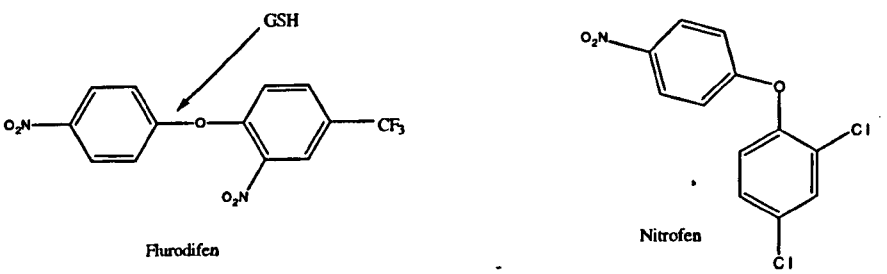


Figure 1-11: The herbicides flurodifen and nitrofen are structurally very similar, however.GST enzymes in pea are only able to conjugate GST to Flurodifen, showing that GSTs can show a high degree of substrate specificity.

These herbicide studies suggest that if a plant species contained higher levels of GSTs with appropriate substrate specificities, it would be more able to withstand exposure to specific herbicides.

Multiple GSTs are encoded by each plant species, with at least seven isoforms being identified in maize (Table 1-5) (Dixon et al., 1998), and greater than 35 genes encoding GSTs identified in the completed *A. thaliana* genome (Dixon et al., 1998). As previously discussed, the widespread use of herbicides has only been practised for around fifty years and it is not clear why *A. thaliana* or other plant species should have developed multiple GSTs. One possible explanation

for the existence of these multiple isoforms is that glutathione S-transferases may be performing roles other than that of detoxification in the organism (Edwards et al., 2000; Marrs, 1996).

Subunit Composition	Classification	Activity
Two 29kDa subunits	GST I / I	Higher activity towards atrazine than other isoenzymes, low activity with other herbicides and appreciable activity with other xenobiotic substrates CDNB
29kDa and 27kDa subunit	GST I / II	High activities towards choloroacetanilides and fluorodifen, CDNB Shows activity as a glutathione peroxidase
29kDa and 26kDa subunit	GST I / III	High activities towards choloroacetanilides and fluorodifen and alachlor, metolachlor and CDNB
Two 27 kDa subunits	GST II / II	Choloroacetamide and S-triazines, metalachlor, CDNB (weak)
Two 26kDa subunits	GST III/III	High activities towards choloroacetanilides and fluorodifen and alachlor metalachlor and CDNB
Two subunits 28.5kDa	GST V /V	Activities towards metolachlor higher than seen in GST I Activities against fluorodifen highest of all GSTs, but not against atrazine Lower activity towards CDNB, NBC and NPB than GST I/I and GST I/II Conjugates toxic alkenal derivatives; crotonaldehyde, vinyl pyridine and ethancrynic acid Shows some activity as a glutathione peroxidase it is able to reduce cumene hydroperoxide, but not linoleic acid hydroperoxide to corresponding non-toxic alcohols 2,4-D was found to inhibit CNDB activity competitively (cf *Arabadopsis* GST 5 CNDB, IAA and 2,4 -D competitively inhibit enzyme activity and salicylic Acid non competitively
28.5kDa and 25.3kDa subunit	GST V / VI	Intermediate activities of GST V/ V and GST VI / VI as two subunits appear kinetically independent
Two subunits 25.3kDa	GST VI / VI	Lower Km for CDNB than (0.12 mM) than GST V/V (1.01 mM)
Two subunits 25.4kDa	GSTVII/ VII	Not biochemically characterised

Table 1-4: Biochemical characterisation of the multiple maize isoenzymes has identified that they are show distinct substrate specificity and are taken together able to conjugate a wide range of substrates (Dixon et al 1998; Dixon et al. 1997; and personal communication with David Dixon)

1.3.2 Natural substrates

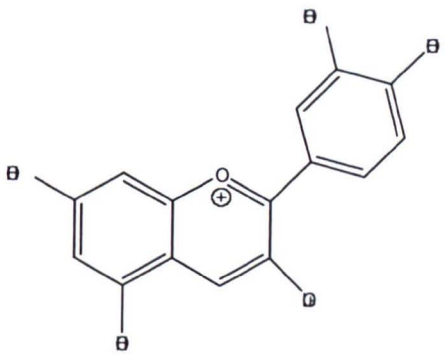
Plants are thought to have evolved a complex array of detoxification enzymes, not to protect against man-made herbicides, but to combat their continual exposure to naturally synthesised toxic chemicals. The sedentary nature of plants means that the ability of a particular plant to detoxify harmful compounds by

various means (cytochrome P450, glycosylation, glutathione transferase activity) provides it with a survival advantage. GSTs are thought to protect the plant cell against exogenous chemicals and also potentially harmful endogenous compounds which can form as a consequence of aerobic respiration and ionising radiation (reactive oxygen species, ROS) or as products of secondary metabolism.

1.3.2.1 Anthocyanins

In maize, studies of gene expression using mutator transposons (the most aggressive eukaryotic mobile elements) produced a mutant of the gene Bronze-2 (Bz-2) (Marrs et al, 1995). These Bz2 mutants accumulate the anthocyanin molecule cyanidin-3-glucoside in the cytoplasm causing the tissues to turn bronze in colour. Normally anthocyanin molecules, responsible for the red, blue and purple pigmentation in fruits and flowers, are synthesised in the cytoplasm and then transported to the vacuole. However in the Bz2 mutants the cyanidin-3-glucoside is not transported across the vacuolar membrane, but instead remains in the cytoplasm leading to localised cell death, poor vigour and possible plant death. Sequence analysis showed the Bz-2 gene shared similarity to previously identified plant GSTs and enzymatic assays with the substrate CDNB confirmed the assignment that Bz2 was related to GSTs. The GST encoded by the gene Bz2 was first thought to conjugate glutathione to cyanidin-3-glucoside. This was proposed to target the removal of the anthocyanin from the cytosol to the vacuole via a glutathione pump (Marrs et al., 1995), however, more recently it has been suggested that Bz-2 transports the cyanidin molecule to the GSH pump. It is proposed that the Bz-2 GST presents the cyanidin to the pump and enables it to be removed to the vacuole (Dixon et al., 1998; Edwards et al., 2000).

Subsequently the anthocyanin biosynthetic gene *An9* present in Petunia, which also shows similarity to plant GSTs (30% identity with ZmGST1-1 & 7% identity to Bronze2), has been cloned. This gene is able to functionally complement Bz-2 mutants, restoring the purple pigmentation in Bz-2 tissues. The reciprocal is also true and the maize GST Bz-2 gene can restore the red pigmentation seen in the flower sectors of *Petunia* An9 mutants (Alfenito et al., 1998).

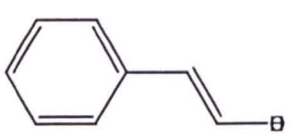


Cyanidin-3-glucoside

Figure 1-12: Petunia petal showing AN9 mutation in around 50% of the cells. These are observed in the sector of the plant with no pigment. The AN9 mutation prevents the movement of the anthocyanin pigment, Cyanidin-3-glucoside, into vacuole and this prevents the red pigmentation from being observed

1.3.2.2 Phytoalexins & Cinnamic acid

Plants synthesise and release phytoalexins when elicitors such as cell wall fragments of plants, fungi or microorganisms are detected. This defence response against microbial and fungal attack promotes cell death and thus inhibits pathogen attack. Cinnamic acid (CA) is a precursor in phytoalexin synthesis which can inhibit phytoalexin synthesis. As discussed in the review by Marrs (Marrs, 1996) Glutathione transferases were initially believed to conjugate glutathione to CA and remove inhibitory cinnamic acid from the cytosol to the vacuole during the intial steps of phytoalexin synthesis. This was based on the description of a monomeric 30Kda protein from maize which catalysed the addition of GSH to the olefinic bond of CA (Dean et al., 1995). However further work in this area has shown that this 30kDa enzyme is an ascorbate peroxidase, probably acting by indirectly generating thiyl free radicals of GSH. These free radicals may then react with the alkyl double bond of CA resulting in the formation of the observed GSH conjugate (Dean & Devarenne, 1997).



Cinnamic acid

Figure 1-13: Structure of cinnamic acid

Original in colour

1.3.2.3 Auxins

Auxins, often thought of as plant hormones, are a generic term for compounds that are not required to resemble one another chemically, but which are able to induce shoot elongation. Several GSTs with high homology to the tobacco type I GST, par B have been described as binding tritiated 5-azido-IAA during photoaffinity labelling studies performed to detect auxin binding proteins (Bilang & Sturm, 1995). Further studies using the natural auxin, indole acetic acid (IAA), and synthetic auxins α-napthyloxacetic acid and 2,4-D (figure 1-14) have shown that these different auxin molecules bind GSTs at different distinct sites. It has been proposed that the site of binding is auxin dependent, with IAA and α-NAA binding to a noncatalytic site on the GST, whereas 2,3-D and 2,4,6-T bind to a catalytic site. The binding of IAA by GSTs may serve to modulate IAA activity or allow its temporary storage. *In vivo*, the conjugation of IAA by amino acids or sugars prevents its degradation and it has been suggested that these IAA conjugates escape inactivation and that through decarboxylation they act as slow release forms of the hormone. Conjugation of auxin molecules with GSH may provide a similar mechanism.

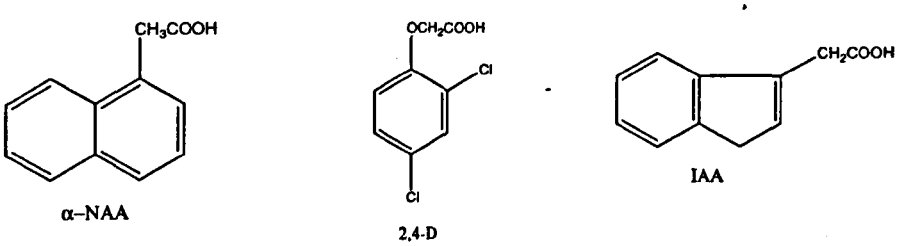


Figure 1-14: IAA present in the plant tissues.

1.3.3 Oxidative stress

Rapid generation of superoxide and accumulation of the H_2O_2 is a characteristic early feature of the hypersensitive response following pathogen attack and this small diffusable molecule has been found to selectively induce several GSTs.

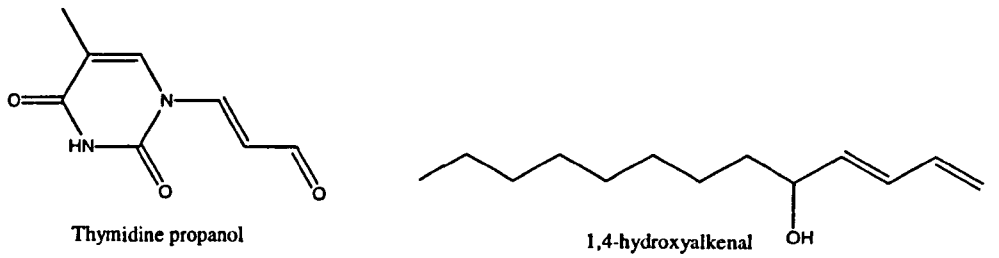


Figure 1-15: GSTs are able to 'mop up' the harmful products of oxidative damage such as; the lipid peroxide, 1,4-hydroxyalkenals and the DNA gradation product thymidine propanol.

A significant number of GSTs have been described as having glutathione peroxidase activity (Board et al., 1997) and these are thought to play a key role in the oxidative stress response. Glutathione peroxidases (GPOXs) catalyse the reduction of cytotoxic hydroperoxides, generated by lipid peroxidation and oxidative damage to DNA. Attack by glutathione on the electrophilic oxygen causes the hydroperoxides to be reduced to their corresponding alcohols in a two step process (Habig, 1983).

1.	ROOH +GSH ⟶ ROH + [GSOH]	enzymatic reaction
2.	[GSOH] + GSH ⟶ GSSG + H2O	spontaneous reaction
	ROOH +2GSH ⟶ ROH + GSSH +H2O	overall reaction

Glutathione transferases which show peroxidase activity (GST-GPOX) are a distinct group of glutathione peroxidases described in plants (Eshdat et al., 1997). Unlike selenium-dependent glutathione peroxidases (Sabeh et al., 1993) and non-selenium dependent phospholipid hydroperoxides glutathione peroxidases (PHGPXs), GST-GPOXes show no activity toward phospholipid hydroperoxides and hydrogen peroxide, however they are able to reduce a range of hydroperoxides (Eshdat et al., 1997).

Active oxygen species appear to enhance the production of GST-GPOXs in soybean, maize and wheat suggesting these enzymes may be responding to oxidative stress. Further, transgenic tobacco in which GST-GPOX Nt107 was overexpressed were found to be more tolerant of chilling and salt stress (Roxas et al., 1997a; Roxas et al., 1997b). These results, along with the finding of decreased levels of lipid peroxidation following GST-GPOX expression, suggests GST-GPOX directly protects against oxidative stress. GST-GPOX has

also been shown to increase levels of GSSG and this increase in concentration of GSSH may act as a second signal to further activate protective responses (Dixon et al., 1998).

1.3.4 Isomerisation of Maleylacetoacetate

Cloning of a gene which encoded an enzyme with Maleylacetoacetate Isomerase (MAAI) activity from *Aspergillus nidulans,* by a group working on tyrosine metabolism, led to the discovery that MAAI showed strong amino acid sequence identity to GSTs previously identified in *Caenorhabditis elegans*, *Dianthus caryophyllus* (carnation) and *Homo sapiens* (human) (Fernández-Cañón & Penalva, 1998). This led to the proposition that a subdivision of the superfamily of GST related sequences (Board et al., 1997) might catalyse the penultimate step in the phenylalanine/tyrosine pathway, the *cis-trans* conversion of maleylacetoacetate (MAA) to fumarylacetoacetate (FAA) utilising glutathione as a coenzyme which had been identified in the 1950's (Knox, 1955; Knox & Edwards, 1956) (figure 1-16).

The Zeta class GSTs are unusual in comparison to other GST classes in that sequences of this class are well conserved over a considerable period of evolution, being found in vertebrates, plants, insects and fungi. Indeed sequence identity of 38% exists between human and carnation Zeta sequences and 45% between Human and *Aspergillus nidulans.* In particular sequence similarity among the Zeta class GSTs is particularly striking within the N-terminal region which contains the signature motif, SSCX[WH]RVIAL, identified by Board and coworkers as being unique to Zeta class GSTs (Board et al., 1997).

Studies on plant Zeta class GSTs (Meyer et al., 1991b; Subramaniam et al., 1999; Dixon et al., 2000) has shown that this class of enzyme demonstrates little activity toward the traditional GST substrates, but does show isomerase activity for malelyacetone, converting it to fumarylacetone.

The human and *Arabidopsis thaliana* Zeta class GSTs are able to convert the pollutant compound dichloroacetic acid (DCA) to glyoxylic acid (Dixon et al.,

2000; Tong et al., 1998b). This DCA oxygenase activity is of interest not only because of its possible use in reducing DCA in the environment, but also because dichloroacetic acid is used therapeutically to treat lactic acidosis and has been proposed for use as a neuroprotective agent (Tong et al., 1998a).

The *cis-trans* conversion of Maleylacetoacetate is one of the first endogenous roles assigned to GSTs. In addition, this reaction is different from those previously observed for GSTs in that the glutathione is not irreversibly bound to the substrate, but is released following the isomerisation.

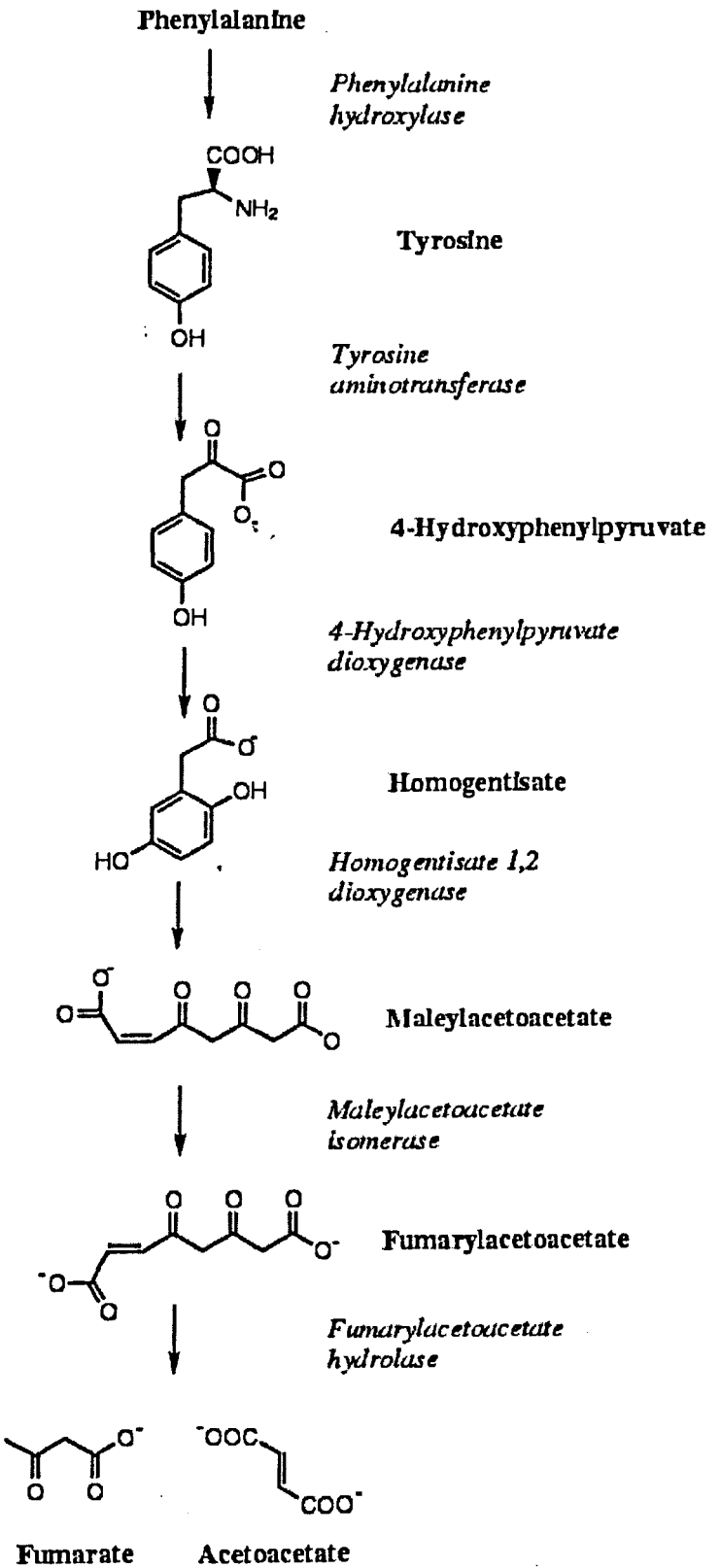


Figure 1-16: The phenylalanine catabolic pathway. Maleylacetoacetate isomerase is a member of the Zeta class of GSTs. Database searches have also identified homologues of Homogentisate dioxygenase and Fumarylacetoacetate hydrolase in the *A. thaliana* genome.

1.3.5 Non catalytic binding activities

In addition to the catalytic binding site, GSTs also possess a non-catalytic binding site, allowing them to serve as non-enzymic carrier proteins or ligandins. Reactive metabolites formed from carcinogens such as dimethylaminoazobenzene and 3-methylcholanthrene (3-MC) have been shown to bind convalently to some GSTs (Ketterer et al, 1967). The covalent binding of these genotoxic compounds to GST provides a method to prevent these carcinogenic molecules from reacting with DNA. Mammalian GSTs have also been shown to non-covalently bind; steroid hormones, bilirubin, heme, bile salts and penicillin (Hayes & Mantle, 1986; Hayes & Chalmer, 1983; Ishigaki et al, 1989; Kirsch et al, 1975). This ligandin binding is thought to occur at a different site to that of either the glutathione or hydrophobic binding site that most conjugated xenobiotics have to date been shown to bind.

Affinity labelling experiments suggest this non-catalytic site may lie within the cleft between the subunits of the dimeric enzyme (Barycki & Colman, 1997) and that binding is of moderate affinity with K_d values of between 18^{-8} and 10^{-5}M. As discussed in section 1.3.2.3, some plant GSTs can bind to the natural auxin indole-3-acetic acid (IAA) at a position distinct from the active site (Bilang & Sturm, 1995). Enzyme activity studies have also described ligandin binding of the open and closed chain tetrapyrroles; bilrubin, biliverdin, chlorophyllin and hemin (Dixon, 1998; Singh & Shaw, 1988) in GST isoforms from both oats and maize. The oat GST isoform shows non-competitive inhibition of 15-19% for open chain tetrapyrroles and 83-86% inhibition with closed chain tetrapyrroles, while the maize isoform GST V/V is 66% non-competively inhibited by the closed chain tetrapyrrole chlorophyllin.

The biological significance of these non enzymatic activities is unknown, but it has been proposed that as GSTs exist in most tissues in the range of 5-100μM that they constitute a high capacity intracellular binding pool for hormones (Listowski et al., 1988). It is also thought GSTs could play a role in transport and direct hormones such as IAA to sites of action or that GSTs may act as a buffer to moderate fluxes in hormone levels within the cell.

1.4 MultiGene family of GSH Transferase

1.4.1 The Glutathione S-transferase Superfamily

The GST superfamily is a large and sequence diverse family of enzymes with a wide phylogenetic distribution. Currently over a hundred plant GST sequences have been determined and at least 35 different GSTs have been detected in plant species *Arabidopsis thaliana* (Edwards et al., 2000). To date, GSTs, have been classified according to sequence similarities and immunological cross reactivity into the classes; Alpha, Mu, Pi, Theta, Zeta, Omega, Phi, Tau, Delta and Sigma, (Chelvanayagam et al, 2000; Dixon et al., 1998; Droog, 1997; Jakoby et al., 1984; Mannervik et al., 1985), however, it has been proposed from sequence comparisons of known GST sequences, that at least 25 classes exist, as different from each other as the currently recognised classes (Snyder & Maddison, 1997). GSTs like cytochrome P450's (Nebert et al., 1991) show clear evolutionary divergence between animals and plants with the classes Alpha, Mu and Pi only observed in mammals, while the classes Tau and Phi are only observed in plants.

With the exception of the Theta class, intraclass identity between GSTs is generally in the range of 50-80%, while interclass sequence identity is significantly lower ~25-35%. The Theta class has the lowest sequence identity and widest species distribution of all the classes, being found in mammals (Meyer et al., 1991a), plants (Frear & Swanson, 1970), insects (Fournier et al, 1992) and also in mitochondria (Harris et al., 1991). However the sequence diversity assigned to the Theta class is somewhat false, as during the classification of GSTs, many of the GST isoenzymes not fitting into the mammalian classes Alpha, Mu or Pi have been classified Theta. Recent work has further classified enzymes of the Theta class, particularly in insects and plants and assigned many of the previously designated Theta class GSTs to distinct classes. The Theta class does appear to be one of the oldest classes of GST. This has led to the postulation that it is the progenitor GST class and that duplication events throughout evolution have led to the development of new classes (figure 1-17) (Pemble & Taylor, 1992). The discovery of GSTs in the mitochondrial matrix led to the suggestion that the initial function of GST was to protect purple and cyanobacteria against oxygen toxicity. If this is true GSTs were perhaps

crucial in enabling bacteria to survive in the cytoplasmic environment of its symbiotic host (Fahey & Sundquist, 1991).

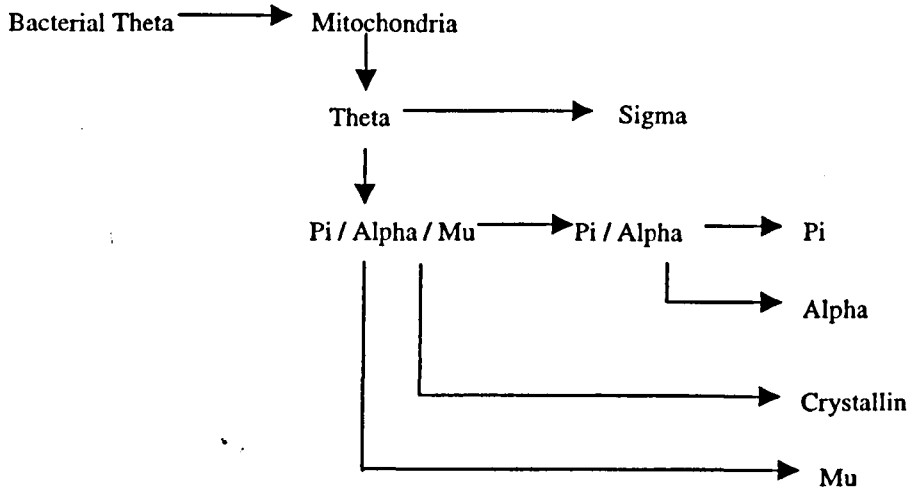


Figure 1-17: Evolutionary relationships of major classes of GST. The animal classes Alpha, Mu and Pi are very similar to the squid lens crystallin and probably arose due to a duplication event of class Theta. Class Zeta is found in a large range of organisms including mammals and plants. Conservation suggests it too arose from class Theta, but before the evolutionary split between plants and animals. Adapted from (Pemble & Taylor, 1992).

In addition to the cytosolic classes of GST, two distinct microsomal forms of glutathione transferases also exist. These show little sequence similarity to their cytosolic counterparts with one catalysing the addition of a broad range of substrates (microsomal glutathione transferase, (Morgenstern et al, 1983) while the other conjugates leukotrieneA_4 to glutathione (LTC_4 synthase, (Nicholson et al, 1993)). Microsomal glutathione transferase which belongs to the MAPEG superfamily (Membrane Associated Proteins in Eicosanoid and Glutathione metabolism) exists as a trimer of 17 kDa subunits. Currently 2D crystals have yielded structural data suggesting each subunit is formed by a four helix bundle (Schmidt-Krey et al, 1999). The second enzyme Leukotriene C4 synthase, is 18 kDa in size and is functionally active as a dimer. It is responsible for the production of leukotriene C_4 (LTC_4) which following cleavage yields LTD_4 and LTE_4. These products are smooth muscle contractile lipid mediators that elicit bronchial smooth muscle contraction and increase vascular permeability.

Although primary sequence analysis of the cytosolic GSTs cannot suggest the exact substrate conjugated by a GST, sequence comparison can aid elucidation

of important catalytic or structural residues within each class and suggest the possible size and character of substrates that may be conjugated by that class.

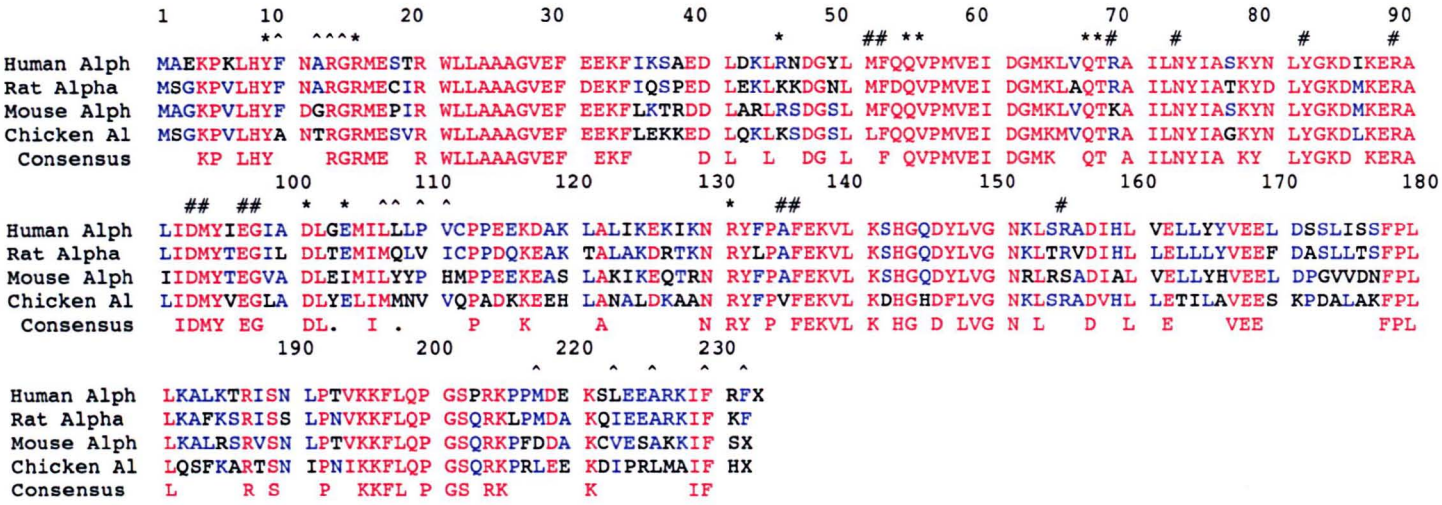


Figure 1-18: Alignment of Alpha class GSTs-Rat (P00502), Human (P08263), Mouse (M73483) &Chicken (L15386). G-site residues marked with(*), H-site residue (^) & residues involved in subunit dimerisation (#) based on Rat Alpha structure (Sinning *et al.*, 1993) (Red = similar/identical residues, blue = conservation)

Original in colour

1.3.1.1 The Alpha Class

All the Alpha class genes isolated from rats, mice and humans are composed of seven exons and share at least 55% sequence identity with each other. Three distinct subclasses form the Alpha class. These have been defined as subunits; GSTA1, GSTA2; GSTA3, GSTA4 and GSTA5. Subunits rGSTA1 and rGSTA2 are encoded on a different chromosome from subunits rGSTA3 and rGSTA5, however all four of these subunits are able to form both homo and heterodimers. Examination of the dimer interface of these subunits indicates that they are conserved with only a Lys to Arg substitution occurring in the rGSTA5 subunit in this region. GSTA4 is only able to form homodimers as it differs at four of the nine conserved residues observed in the other Alpha class GSTs at the dimer interface. The active site residues of the Alpha class enzymes which bind substrate are hypervariable and to date it remains to be established the exact residues that confer substrate specificity.

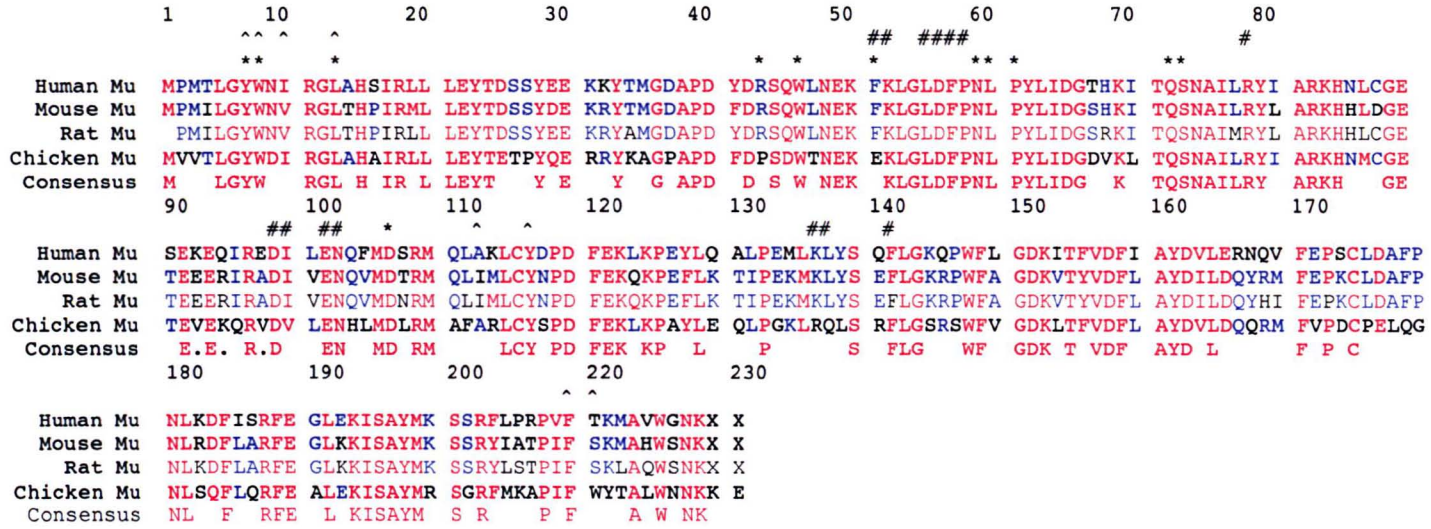


Figure 1-19: Alignment of Mu class GST from Human (P28161), Mouse (J03952), Rat (P04905) &Chicken (X58248). G-site residues marked with(*), H-site residue (^) & residues involved in subunit dimerisation (#), based on Mouse structure (Ji *et al.*, 1992) (Red = similar/identical residues, blue = conservation).

1.3.1.2 The Mu Class

The class Mu genes from rats, mice and humans are made up of eight exons and are about 5 kb in length. Genes of the Mu class typically share at least 65% sequence identity with each other and can be grouped into two distinct subfamilies. It is unknown whether these two subfamilies are able to form heterodimers with each other.

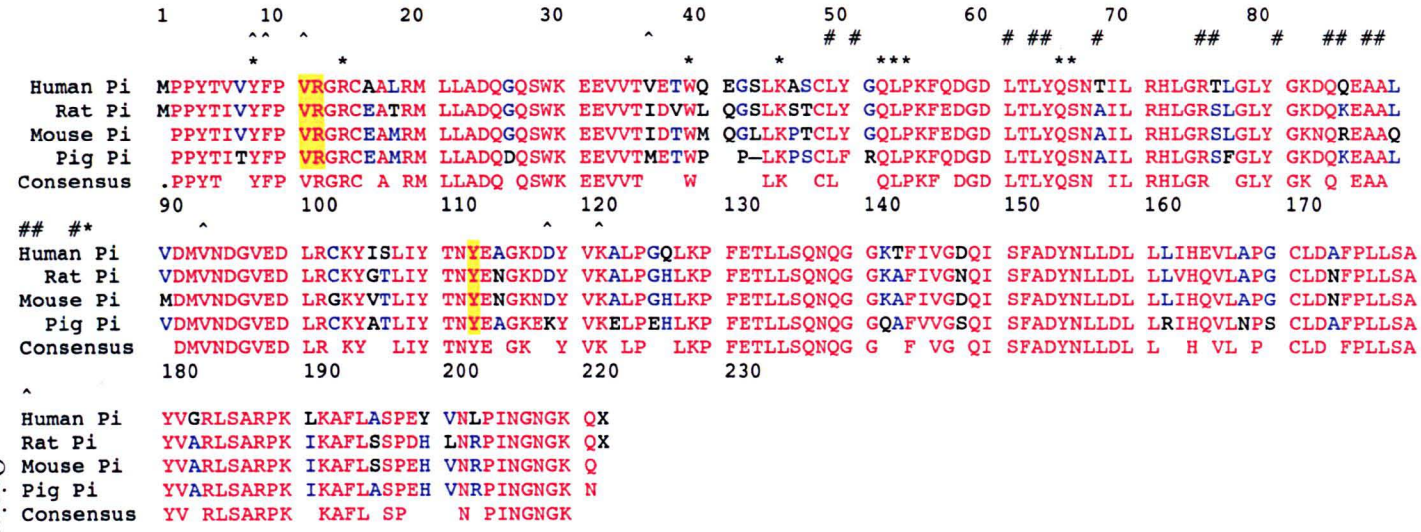


Figure 1-20: Alignment of Pi class GST from Human (P28161), Rat (L29427), Mouse (P19157) & Pig (P80031) G-site residues marked with(*), H-site residue (^) & residues involved in subunit dimerisation (#) (Reinemer *et al.*, 1991) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.3 The Pi Class

The Pi class GSTs from rats, mice and humans are about 3kb long and contain seven exons. They are highly conserved and unlike Alpha, Mu and Theta only one functional Pi class GST gene is found in rats (Okuda et al., 1987). Humans also have one functional class Pi gene encoded at a single locus that shows some allelic variation. Residues Val11 and Arg11 are thought to cause the substrate stereo-selectivity of Pi class enzymes while Tyr113 is essential for activity towards epoxides (Hiratsuka et al., 1989).

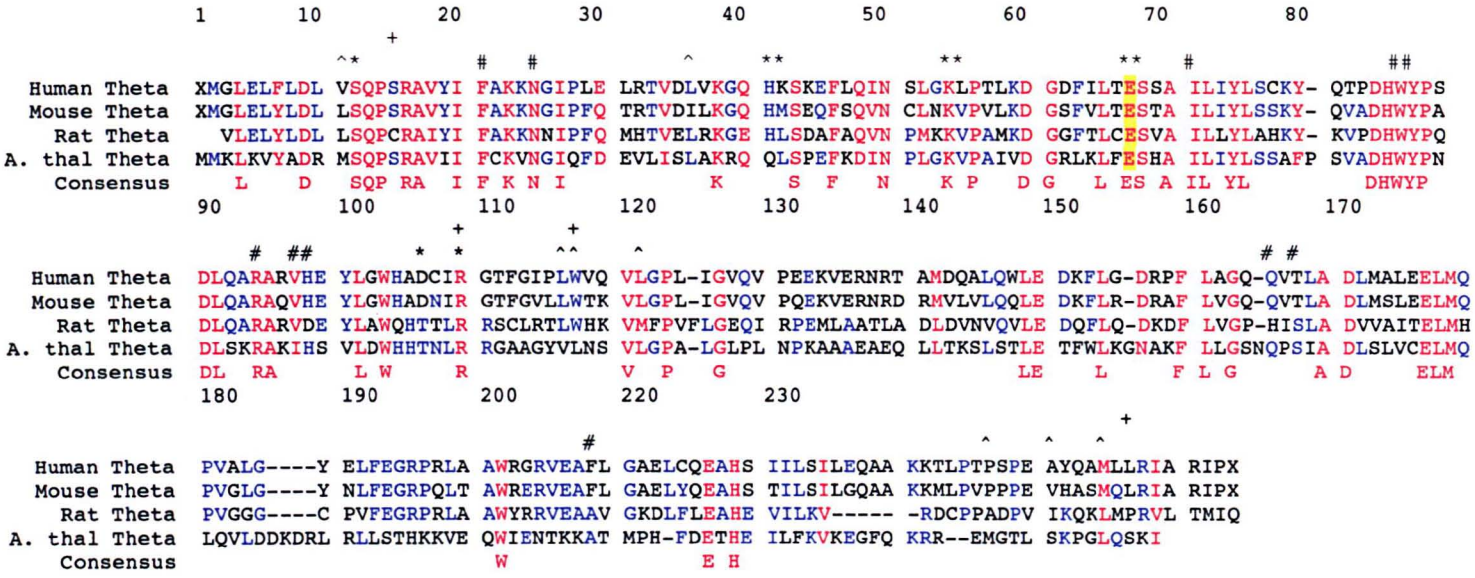


Figure 1-21: Alignment of Theta class GST from Human (P30712), Mouse (U48419), Rat (Q01579) & *Arabidopsis thaliana* (AJ131580) G-site residues marked with (*), H-site residue (^), residues involved in subunit dimerisation (#) & those binding sulphate (+) (Rossjohn et al., 1998) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.4 The Theta class

Theta is the most diverse class of GST and includes enzymes from bacteria, insects, plants, fish and mammals. Although considered the default class, it has several strict sequence conservations, which distinguish it from other classes. Gln67 which interacts with the γ-glutamic acid moiety of GSH in Alpha, Mu and Pi is a conserved Glu in the Theta class (Reinemer *et al.*, 1996; Wilce *et al.*, 1995), while a conserved serine replaces tyrosine to stabilise the GSH thiolate ion.

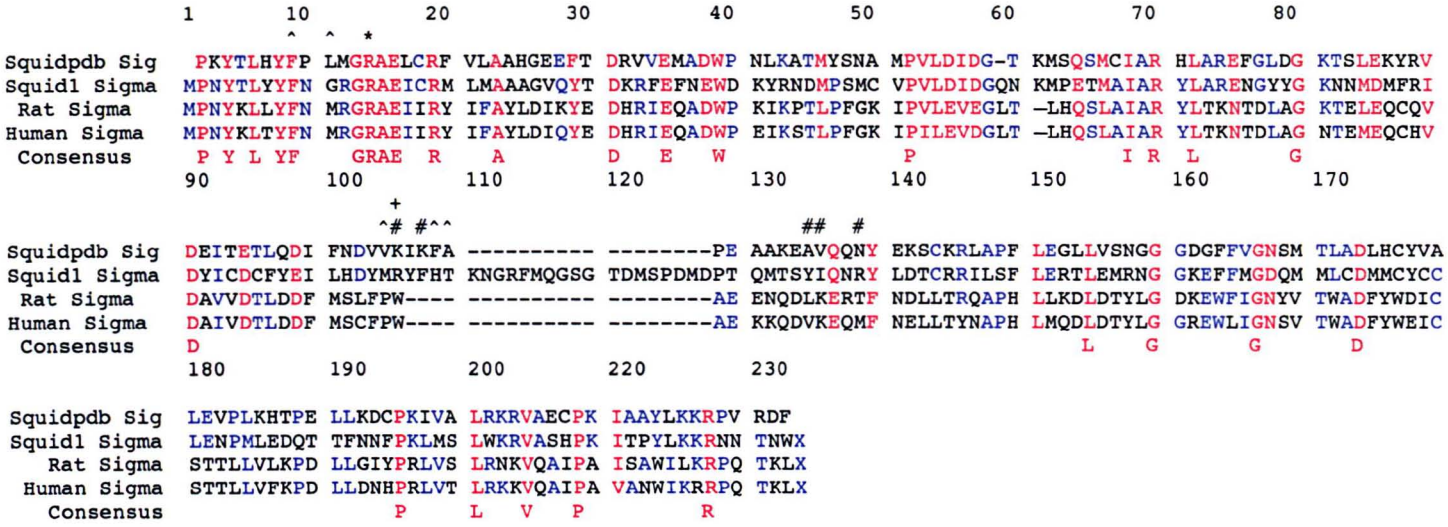


Figure 1-22: Alignment of Sigma class GST from Squidpdb (P46088), Squid1 (M36937), Rat (D82071), Human (D82073) G-site residues marked with (*), H-site residue (^), residues involved in subunit dimerisation (#) & those binding sulphate(+) (Ji et al., 1995) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.5 The Sigma Class

The Sigma class was first suggested from sequence alignments of the S-crystallins from mollusc lens (Buetler & Eaton, 1992) and shares 20% identity with the Alpha, Mu and Pi class GSTs. It most closely resembles the Pi class in both sequence and intron/exon boundaries (Tomarev *et al.*, 1993), with greatest similarity at the N-terminal region (residues 1-27) rather that the G-site. Sigma class GSTs show activity toward prostoglandins and CDNB (Ji *et al.*, 1995) and are thought to have been recruited to cephalopod lenses due to their high thermodynamic stability and ability to accumulate to high concentrations without causing opacification (Tomarev & Zinovieva, 1988).

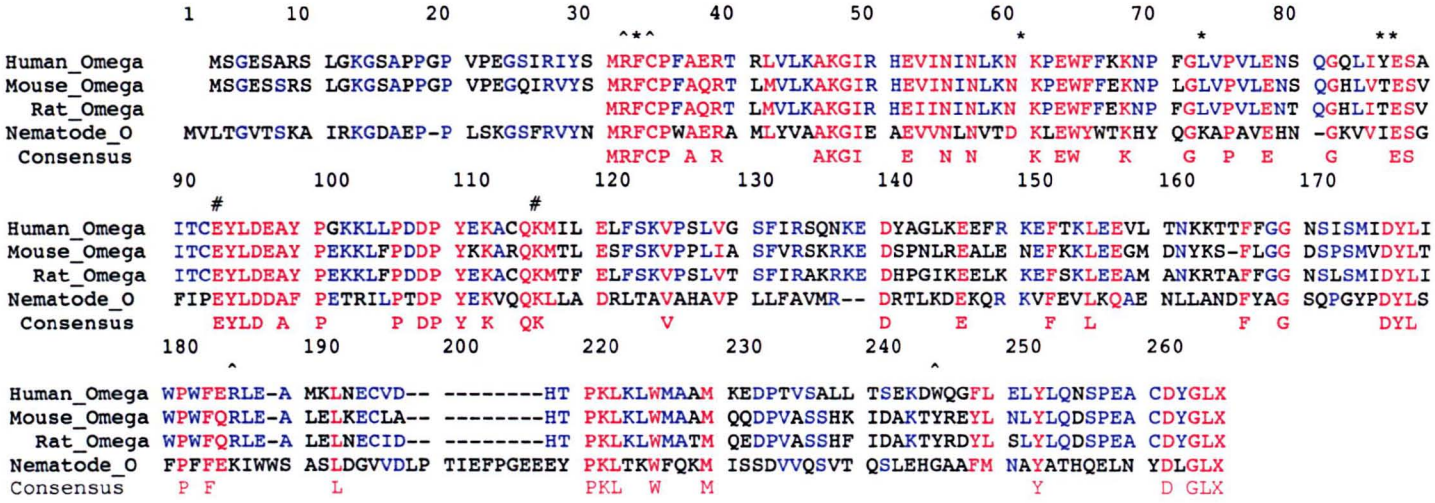


Figure 1-23: Alignment of Omega class GST from Human (P78417), Mouse (U80891), Rat (AB008807), Nematode *C. elegans* (L23651), G-site residues marked with (*), H-site residue (^), residues involved in subunit dimerisation (#) & those binding sulphate(+) (Board *et al.*, 2000) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.6 The Omega Class

The Omega class was detected by searching and subsequent alignment of sequences from expressed sequence tag databases. Omega GSTs are expressed in most tissues in several mammalian species, as well as *Caenorhabditis elegans*, insects and plants. They exhibit GSH-dependent thiol transferase and dehydroascorbate reductase activities. Although structurally similar to other GST classes they have an invariant cys residue in substituted for a ser or tyr catalytic residue and a novel nineteen residue proline rich N-terminal extension of unknown function. Due to its presence in most mammalian tissues and conservation in a range of species, it has been suggested that like the Zeta class enzymes, Omega class GSTs might act in endogenous metabolism, possibly helping protect organisms from oxidative stress (Board *et al*, 2000).

```
           1       10        20        30        40        50        60        70        80

Blow Fly     MDFYYLPGS APCRSVLMTA KALGIELNKK LLNLQAGEHL KPEFLKINPQ HTIPTLVDGD FALWESRAIM VYLVEKYGKND SLFPKCPK
House Fly    MDFYYLPGS APCRSVLMTA KALGIELNKK LLNLQAGEHL KPEFLKINPQ HTIPTLVDGD FALWESRAIM VYLVEKYGKTD SLFPKCPK
Fruit Fly    MVDFYYLPGS SPCRSVIMTA KAVGVELNKK LLNLQAGEHL KPEFLKINPQ HTIPTLVDNG FALWESRAIQ VYLVEKYGKTD SLYPKCPK
 Consensus     DFYYLPGS  PCRSV MTA KA G ELNKK LLNLQAGEHL KPEFLKINPQ HTIPTLVD   FALWESRAI  VYLVEKYGK D SL PKCPK
           90         100       110       120       130       140       150       160       170

Blow Fly     KRAVINQRLY FDMGTLYKSF ADYYYPQIFA KAPADPELYK KMEAAFDFLN TFLEGHQYVA GDSLTVADLA LLASVSTFEV AGFDFSKYAN
House Fly    KRAVINQRLY FDMGTLYKSF ADYYYPQIFA KAPADPELFK KIETAFDFLN TFLKGHEYAA GDSLTVADLA LLASVSTFEV ASFDFSKYPN
Fruit Fly    KRAVINQRLY FDMGTLYQSF ANYYYPQVFA KAPADPEAFK KIEAAFEFLN TFLEGQDYAA GDSLTVADIA LVATVSTFEV AKFEISKYAN
 Consensus   KRAVINQRLY FDMGTLY SF A YYYPQ FA KAPADPE  K K E AF FLN TFL G  Y A GDSLTVAD A L A VSTFEV A F  SKY N
180        190       200       220       230

Blow Fly     VAKWYANAKT VAPGFDENWE GCLX
House Fly    VAKWYANLKT VAPGWEENWA GCLEFKKYFG X
Fruit Fly    VNRWYENAKK VTPGWEENWA GCLEFKKYFE X
 Consensus   V WY N  K  V PG  ENW  GCL
```

Figure 1-24: Alignment of Delta class GST from Blow fly (L23126), House fly (X61302), Fruit Fly (X14233). G-site residues marked with (*), H-site residue (^), residues involved in subunit dimerisation (#) & those binding sulphate (+) (Wilce *et al.*, 1995) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.7 The Delta Class

Two distinct classes of GST were identified in insects through biochemical studies (Fournier et al., 1992) and these were designated as class I and class II. Following further characterisation of the GSTs in insects and other organisms it has become apparent that those GSTs of class II can be grouped into the mammalian classes Sigma, Theta, Omega and Zeta while the class I enzymes form an insect specific class, Delta (Chelvanayagam et al., 2000).

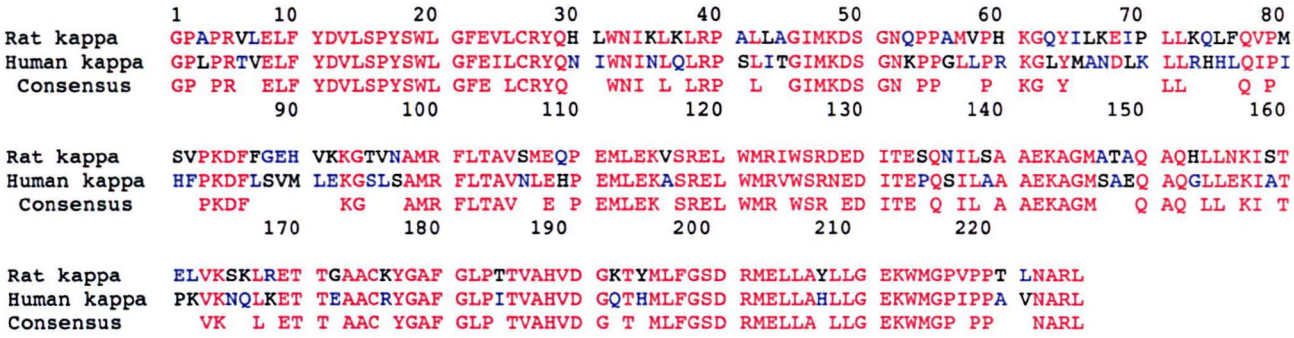


Figure 1-25: Kappa class GST from, Rat (P24473) and Human (AF070656) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.8 The Kappa class

A soluble mitochondrial specific GST, from rat termed Kappa, has also been described (Pemble *et al.*, 1996), however the mechanism of catalysis and the residues involved have yet to be described. Using the Rat GST as a probe further Kappa GSTs were searched forusing the program Psi-Blast (Altschul *et al.*, 1997), revealing a human GST which shows strong similarity to the Rat.

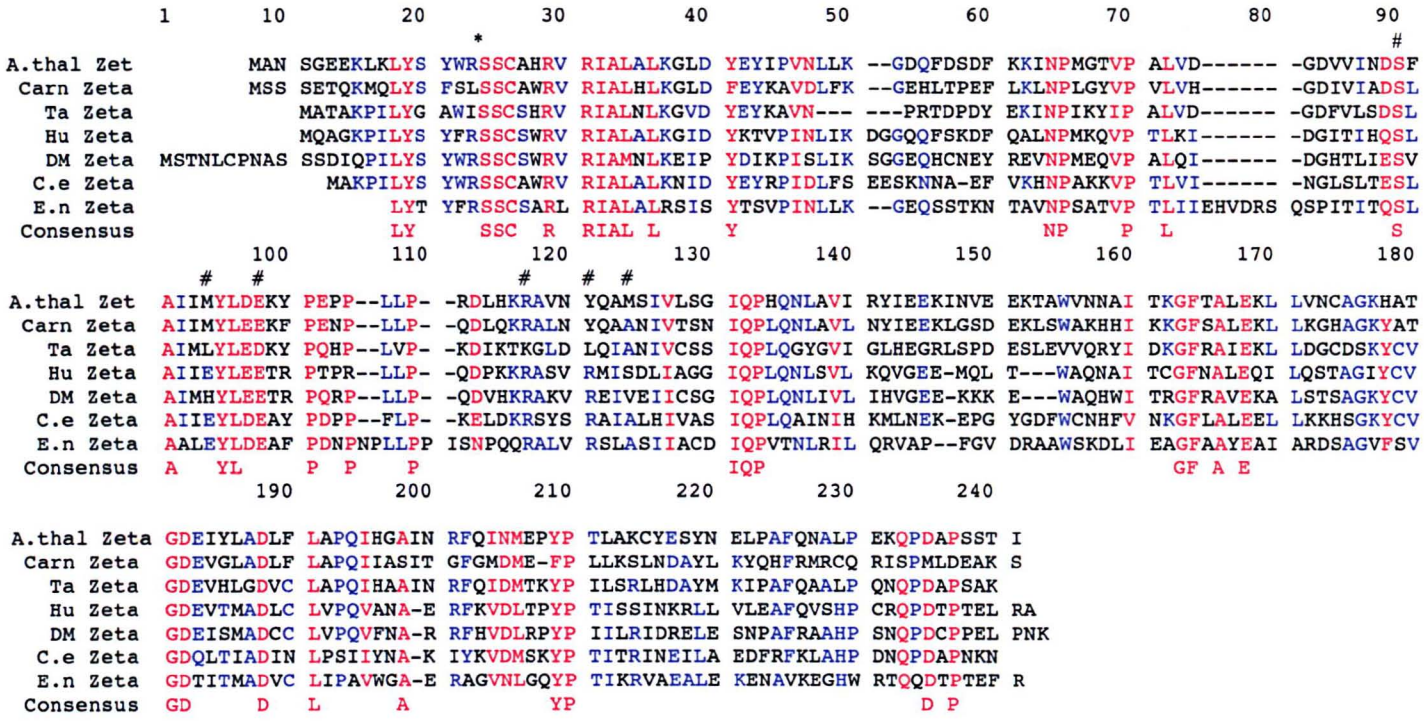


Figure 1-26: Alignment of Zeta class GST from *A. thaliana* (AC005312), Carnation (L05915), Wheat (AF002211), Human (O43708), *Drosophilia Meglanostar* (AAF54382) , *C. elegans* (AAF36004C) & *Emericella nidulans* (AJ001837). G-site residues marked with (*), H-site residue (^), residues involved in subunit dimerisation (#) (Red = similar/identical residues, blue = conservation).

Original in colour

1.4.1.9 The Zeta Class

This class of GST was found by database searching and sequence comparison in a wide range of species from mammals to plants (Board et al., 1997) and contain ten exons and nine introns. This class of GST shows high similarity between the different species with 49% identity found between *C. elegans* and the human sequence. In plants the Zeta class was first identified in carnation, as an ethylene responsive enzyme (Meyer et al., 1991b) and only tentatively classified as a plant subclass (Droog, 1997). A wheat Zeta sequence was recently detected supporting the hypothesis that these GSTs belonged to a discrete class (Subramaniam et al., 1999) while searches of the EST database of higher plant sequences determined that this class of GST was widely spread among plant species (Dixon et al, 2000). Zeta class GSTs are thought to act in tyrosine degradation pathway, to catalyse the *cis-trans* isomerisation of maleylacetacetate to fumarylacetoacetate. This class appears to be similar to class Theta in that it contains a serine in the active site, this serine appearing to be used to stabilise the thiolate ion of glutathione. However a cysteine residue which forms part of the active site is also highly conserved. These residues lie in the motif, SSCXWRVIAL, which distinguishes the Zeta class GSTs from other classes (Board et al., 1997).

1.4.2 Classification of Plant Glutathione Transferases

Plant GSTs were originally classified as part of the Theta class, but as more plant sequences were discovered it became clear that this distinction was too simplistic. Subdivision of the plant GSTs was first proposed by Droog on the basis of polypeptide sequence similarity and exon structure (Droog, 1997). Three distinct classes of plant GST were proposed; type I or Theta which comprised GSTs containing two introns, a tentative type II class containing GSTs with nine introns (Meyer et al., 1991b) and type III or Tau which contained one intron.

Mammalian homologues with similar intron structure and high sequence similarity for type II GSTs were also identified by Board and co-workers (Board et al., 1997) and this group of GSTs has since been termed the Zeta class. Following Droog's classification, further sequences have been detected for all the plant GST families. In addition, a putative A. *thaliania* genomic sequence with closer similarity to mammalian type Theta GSTs than the type I plant GSTs has been detected (Dixon et al., 1998) and this has been proposed to belong to a new class termed type IV.

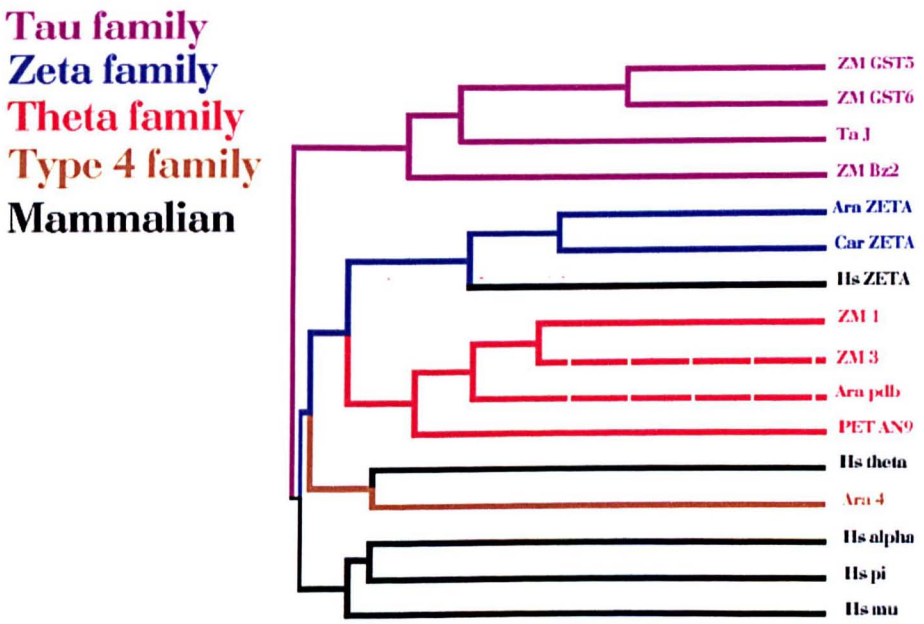


Figure 1-27: Classification of GST sequences. This diagram shows the relationship between the GSTs currently undergoing crystallographic investigation at Glasgow and also the relationship between these GSTs and structurally characterised GSTs from plant and mammals. Black lines indicate mammalian classes, dashed lines indicate plant GSTs, not worked on in Glasgow.

Original in colour

Recently a new system of nomenclature has been proposed for plant GSTs (Edwards et al., 2000) This new nomenclature allows plant GSTs to be categorised in the same manner as mammalian GSTs. using greek letter designations (Hayes & McLellan, 1999). Under the new classification scheme the following classes have been recognised;

- Phi (F) - a plant specific class replaces type I
- Zeta (Z) – a non specific plant class that replaces type II
- Tau (U) - a plant specific class replaces type III
- Theta (T) – a non specific plant class that replaces type IV

Using the same system as that for mammalian GSTs, plant GSTs are given a name structured to indicate the species and the class. This nomenclature is in the form XyGSTZ1, indicating that the GST was isolated from species Xy (such as *At* for *Arabidopsis thaliania*), Z denotes the GST belongs to the Zeta class and the 1 indicates it is the first gene of the Zeta class from that species. This classification will be adopted to describe the GSTs mentioned in the remainder of this work.

Species	Old name	Proposed name
Zea mays (Maize)	*Zm*GST I	*Zm*GSTF1
Zea mays	*Zm*GST II	*Zm*GSTF2
Zea mays	*Zm*GST III	*Zm*GSTF3
Zea mays	*Zm*GSTV	*Zm*GSTU1
Zea mays	*Zm*GSTVI	*Zm*GSTU2
Zea mays	*Zm*GSTVII	*Zm*GSTU3
Zea mays	*Bronze2*	*Zm*GSTU4
Arabidopsis thaliania	*At*GST10	*AtGSTT1*
Arabidopsis thaliania	GSTI, PM239x14	*At*GSTF1
Arabidopsis thaliania	GST2, PM24.1	*At*GSTF2
Arabidopsis thaliania	GST3, ERD13	*At*GSTF3
Arabidopsis thaliania	GST4, ERD13	*At*GSTF4
Arabidopsis thaliania	GST6	*At*GST5
Arabidopsis thaliania	GST7	*At*GST6
Arabidopsis thaliania	GST8	*At*GST7
Arabidopsis thaliania	GST11	*At*GSTF8
Arabidopsis thaliania	GST5, 103-1a	*At*GSTU1
Aegilops tauschii	wheat J,	-
Petunia	AN9	-

Table 1-5: New nomenclature for GSTs under crystallographic investigation together with other GSTs present in *A. thaliana* and *Z. mays* subsequently mentioned in this text.

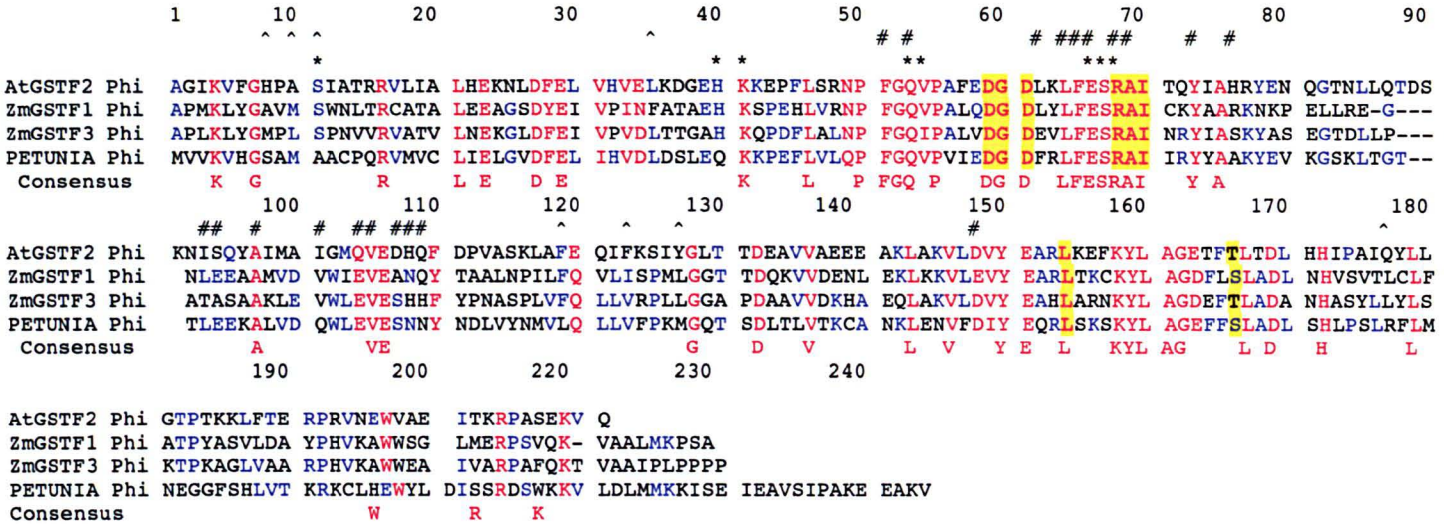


Figure 1-28: Alignment of Phi class GSTs - *Zm*GSTF1 (P12653) & AN9 (Y07721), *Zm*GSTF3 (AJ010295) & *At*GSTF2 (P46422). Conserved residues of this class as detailed in text are highlighted in yellow. G-site residues marked with (*), H-site residue (^) & residues involved in subunit dimerisation (#) (Reinemer et al., 1996) (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.1 Type Phi (F) Plant Glutathione Transferases

This plant specific class of GSTs have been shown to have ligandin, isomerase and peroxidase activity (Edwards et al., 2000). Conserved motifs within the Phi class are found at positions 59-61 Asp -Gly-Asp, 68-70 Arg-Ala-Ile and Leu154. In addition, three residues upstream of Asp169 (strictly conserved in all GSTs) Phi class GSTs contain either a Ser or a Thr residue (Droog, 1997). Three Phi class GSTs, AtGSTF2, ZmGSTF1 and ZmGSTF3 have been structurally determined (Neuefeind et al., 1997a; Neuefeind et al., 1997b; Reinemer et al., 1996) and these show greatest structural similarity to the Theta class mammalian GST. As with the class Theta GSTs a serine residue (Ser11) and not a tyrosine residue is responsible for the stabilisation of the glutathione thiolate ion.

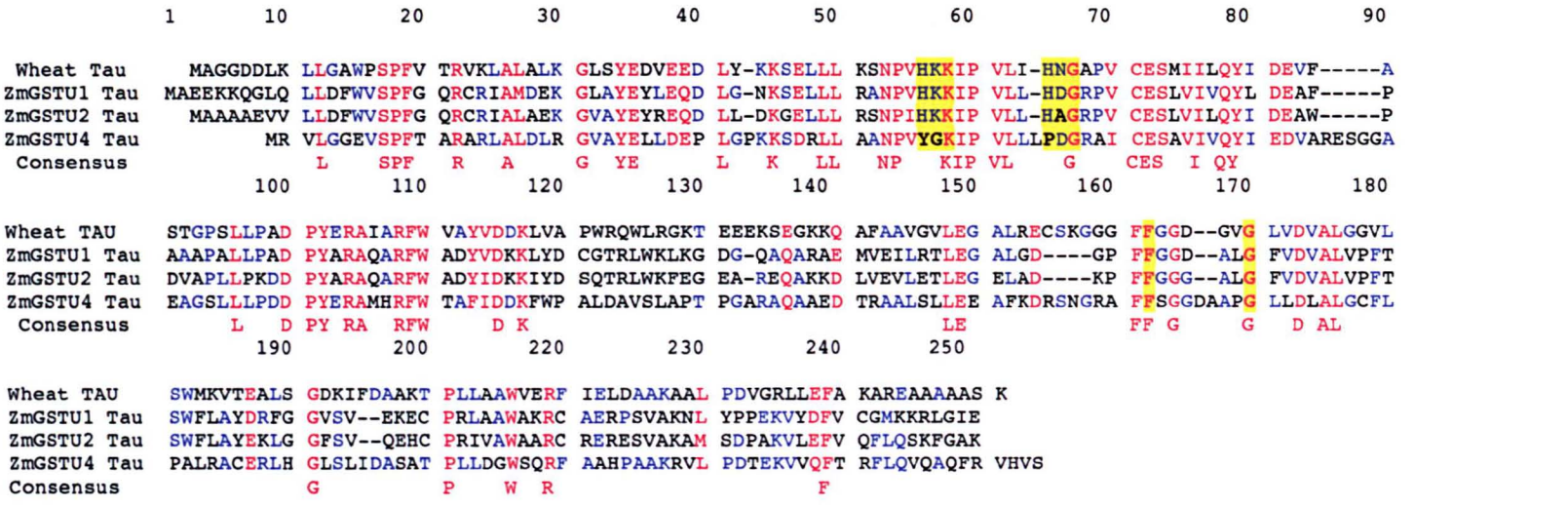


Figure 1-29: Alignment of Tau class GSTs. Wheat Tau (AF004358), ZmGSTU1 Tau (Y12862), ZmGSTU2 (AJ010439), & ZmGSTU4 (JQ0987) are worked on in Glasgow Conserved residues detailed in text are highlighted in yellow (Red = similar/identical residues, blue = conservation).

Original in colour

1.3.1.2 Type Tau (U) Plant Glutathione Transferases

The Tau class was originally identified as a set of auxin related gene products, later found to be GSTs. In contrast to the Phi class GSTs which predominate in maize (Dixon et al., 1997), Tau class GSTs are more common in wheat (Cummins et al., 1997; Riechers et al., 1997) and most dicotyledenous species (Marrs, 1996). Although not expressed constitutively, a number of Tau class GSTs have been characterised from maize following stimulation by the herbicide safener Dichloromid. These have been shown to catalyse the conjugation of glutathione to diphenyl ether herbicides.

Droog classified Tau class GSTs as those with His-Lys-Lys at position 56-58, His-Asp-Gly at position 65-67. In addition, he indicated that Tau class GSTs, in contrast to Phi class enzymes, have a phenylalanine residue instead of a leucine residue at position 162. (Droog, 1997).

1.5 Structural information

1.5.1 Gross Structural Features

In 1991 the first three-dimensional structure of a GST was described, a Pi class GST in complex with the GSH analogue glutathione sulfonate (Reinemer et al., 1992). Since this first structure ~80 GST structures have been deposited in the PDB (Sussman et al., 1998) (August 2000). Including the structures presented in this text, examples of structures for all the GST classes, with the exception of the Kappa class, have been elucidated. These three dimensional structures have confirmed the dimeric nature of GSTs and shown that each monomer is composed of two domains joined by a covalent linker element. In addition, they have shown that despite the low sequence identity between the different classes, all the GSTs described share a similar overall topology (figure 1-31). However although the overall topology is similar, these structures also illustrate that significant differences, including which residues catalyse the conjugation of substrate to GSH do exist between members of different classes.

Domain I (residues 1-~80), or the N-terminal domain, is formed by a four strand mixed β-sheet as part of a βαβαββα structural motif. This motif shares significant homology to that found in the glutaredoxin fold (figure 1-30), and it has been postulated that GSTs may be evolutionarily related to glutaredoxins.

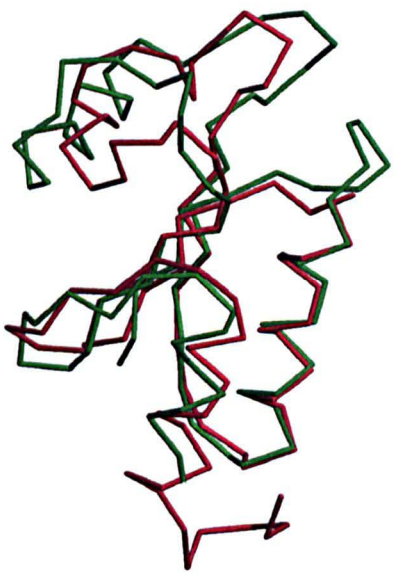
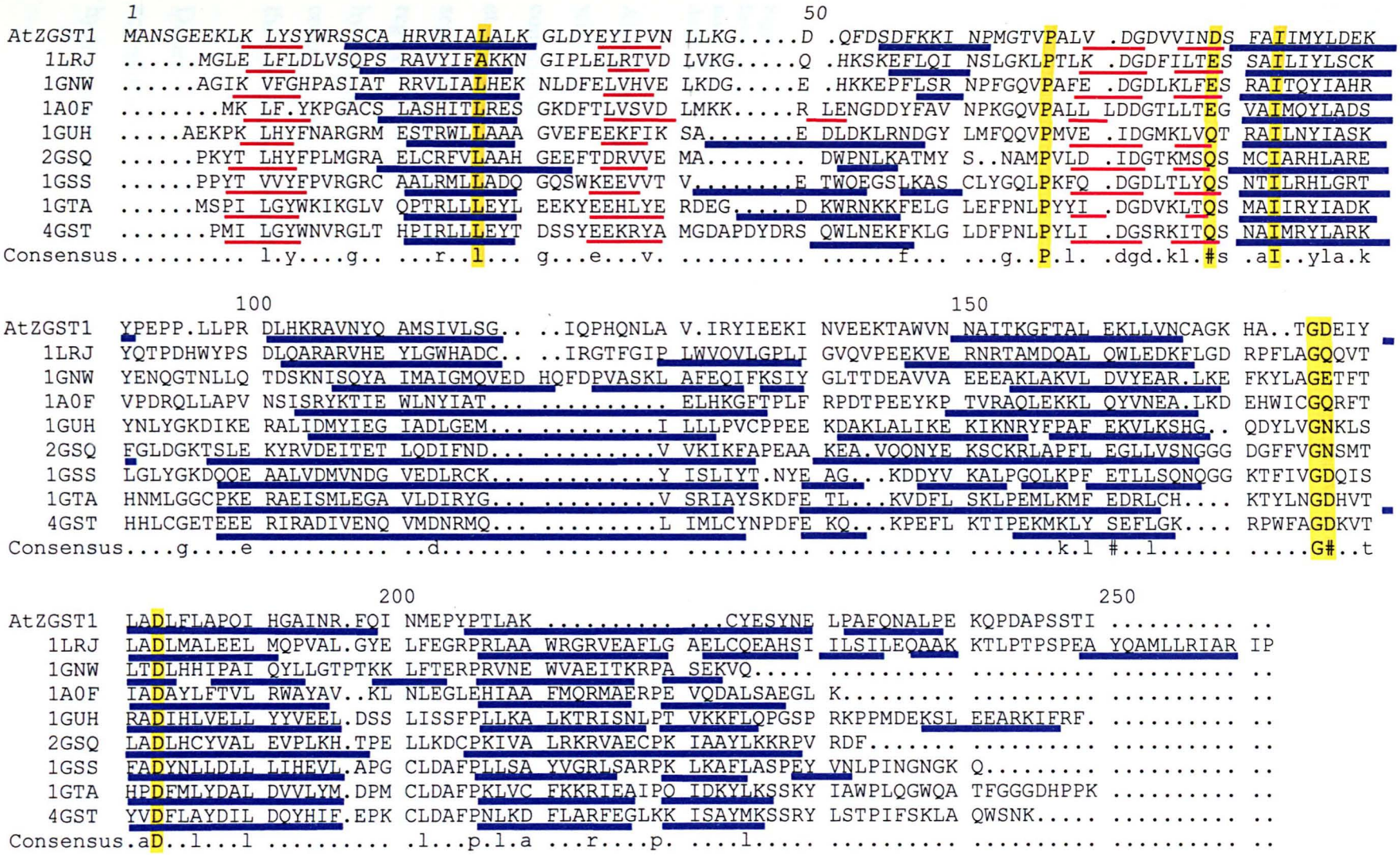


Figure 1-30: Domain I containing βαβαββα structural motif which is conserved between glutaredoxin (shown as green) and N-terminal domain (truncated at residue Leu85) of glutathione S-tranferase 1GNW (Reinemer et al., 1996) (red)

Original in colour

Figure 1-31: Sequence alignment of glutathione *S*-transferases. The GST sequences are listed for *At*GSTZ1, 1LRJ-Theta from human hGSTT2-2 (Rossjohn et al, 1998), 1GNW-Phi from *A. thaliana* (Reinemer et al., 1996), 1A0F-Beta from *E. coli* (Nishida et al., 1998), 1GUH-Alpha from Human (Sinning et al., 1993), 2GSQ- Sigma from Squid (Ji et al., 1995), 1GSS-Pi from Human (Reinemer et al., 1991), 1GTA- *S. Japonicum* (McTigue et al., 1995) and 4GST-Mu from Rat (Ji et al., 1992). The residues in the α-helices are underlined by thick blue lines and the β-strands by thin red lines. The different classes of GST show similar overall secondary structure despite low sequence similarity. 1A0F displays a β-sheet instead of the more commonly observed α-helix 2. Both human Theta and Alpha GST show a longer C-terminus with an additional helix. Those residues highly conserved throughout all the classes are highlighted in yellow.



Original in colour

Although domain II (C-terminal domain) is typically formed by eight or nine helices in the different GST classes, it is less conserved than the smaller N-terminal domain, differing in both the length and folding of the loop regions and in the orientation of the secondary structure elements between isoforms. The active site of the GSTs structurally characterised has always been shown to exist between the two monomer subunits. In particular it is within the large hydrophilic and solvent accessible furrow between the two N and C terminal domains as indicated in figure 1-32.

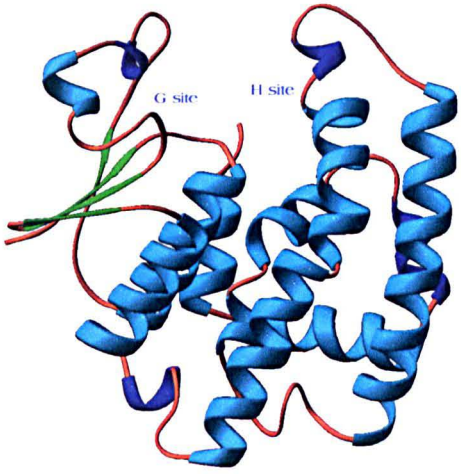


Figure 1-32: Monomer structure of GST displaying N-terminal domain consisting of sheet and helices and the larger C-terminal domain, which is entirely helical. The active site cleft is marked by its two components, the glutathione binding site (G-site) within the N-terminal domain and the substrate binding region marked (H-site), of the C-terminal domain.

All structurally characterised GSTs to date are active as either hetero or homodimers, with the dimer exclusively formed from GSTs from the same class, although considerable efforts are currently being made to produce active monomers (Stenberg et al, 2000). Each dimer contains two kinetically independent active sites typically separated by around 20 Å from each other. Each active site is comprised of two regions: a glutathione binding site (G-site) formed by the N-terminal domain and a hydrophobic binding site (H-site) formed by the C-terminal domain. Together these create the complete active site with the substrate being brought into close proximity of the cysteine moiety of the GSH.

Dimerisation of the monomer subunits results in around 13%-20% of the solvent accessible area being buried. The majority of the dimer interface is comprised of hydrophobic residues and in addition to this hydrophobic component, salt bridges and

Original in colour

hydrogen bonds form, particular to the GST class, stretch across the interface stabilising the dimer.

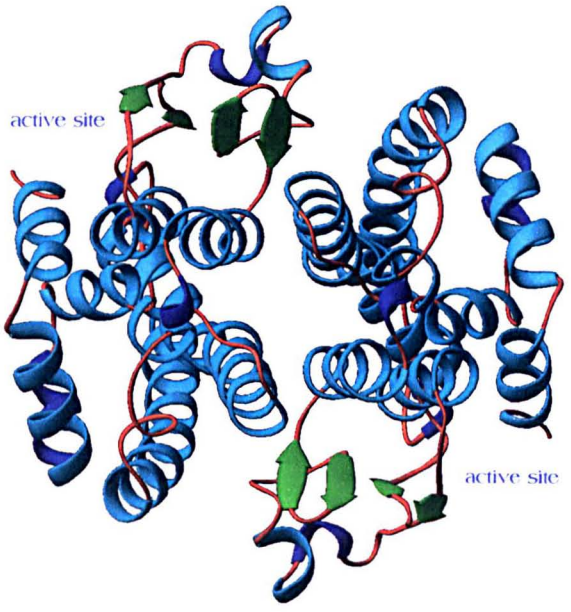


Figure 1-33: Dimeric structure of glutathione S-transferase shown annotated with active site positions of each monomer.

1.5.2 Structural Features of the Different Classes

The smaller N-terminal domain is typically of ~80 residues in length (Alpha 1-78, Mu 1-82, Pi 1-74, Zeta 1-84 and Theta 1-78) and is formed by two alpha helices and four beta-sheets (Armstrong, 1993). The N-terminal domain contains the canonical glutathione binding domain (Reinemer et al., 1991; Sinning et al., 1993) which allows the GST to bind glutathione and stabilise the formation of the GS^- thiolate ion. In the Alpha, Mu, Pi and Sigma classes stabilisation of the GS^- ion is via a conserved tyrosine residue (Armstrong, 1993), while in the Theta, Phi, Tau and Zeta classes this role is performed by a conserved serine (Board et al., 1995; Wilce et al., 1995). Other important residues within the active site are; a conserved *cis* Pro responsible for hydrogen bonding to the backbone of the GSH substrate and a conservatively replaced acidic residue (Gln67 of Alpha, Gln71 of Mu, Gln62 of Pi, Asp72 of Zeta and Glu66 of Theta). This conservatively replaced acidic residue shows unfavourable Phi and Psi angles and is proposed to interact with the γ-glutamate of GSH (Dirr *et al.*, 1994; Wilce & Parker, 1994). The N-terminal domain is relatively well conserved between different classes of GST, although structural differences are observed in the loop region between β-strands 2-3 (figure 1-23). In particular the Mu class displays a

Original in colour

unique structural feature termed the "Mu loop", a seven to eight residue section in the GSH-binding domain which forms one wall of the approach to the active site. The recently solved Omega class GST also contains a unique proline rich N-terminal extension. No function for this extension has yet been postulated for this Omega class enzyme, but perhaps when more is known about the preferred substrates of this enzyme, its role will become clear.

Among the different classes of GST, helices α4 and α6 of the C-terminal domain typically form the hydrophobic binding region that binds substrates to be conjugated to glutathione. Variations in the length and orientation of the C-terminal helices thus affect the size and shape of the substrate binding site. In addition the hydrophobic binding region, identified by either X-ray determined structures or photoaffinity-labelling experiments, (Hoesch & Boyer, 1989) shows considerable sequence diversity among different classes of GST. This region is particularly hydrophobic in the Phi class structures, whilst showing hydrophilic character in the Zeta class enzymes. In the Alpha and Theta classes an additional C-terminal helix (α9) sits on top of the active site and forming a lid to the substrate binding site, and influencing the entry and egress of substrate (Armstrong, 1993).

Although the three dimensional structures of the class Alpha, Mu, Pi, Theta and Sigma enzymes are topologically similar, the RMS deviations on the α-carbon backbone core, as determined by the program Lock (Singh & Brutlag, 1997) is around 2Å. Although part of this deviation can be attributed to difficulties in superimposition of the molecules, visual inspection clearly shows that isoenzymes from different classes of GSTs are structurally distinct (figure 1-34). Further, superimposition of dimers from different classes indicates that due to the different relative positions of the secondary structure elements within each respective monomer, the dimer interface is particular to each class.

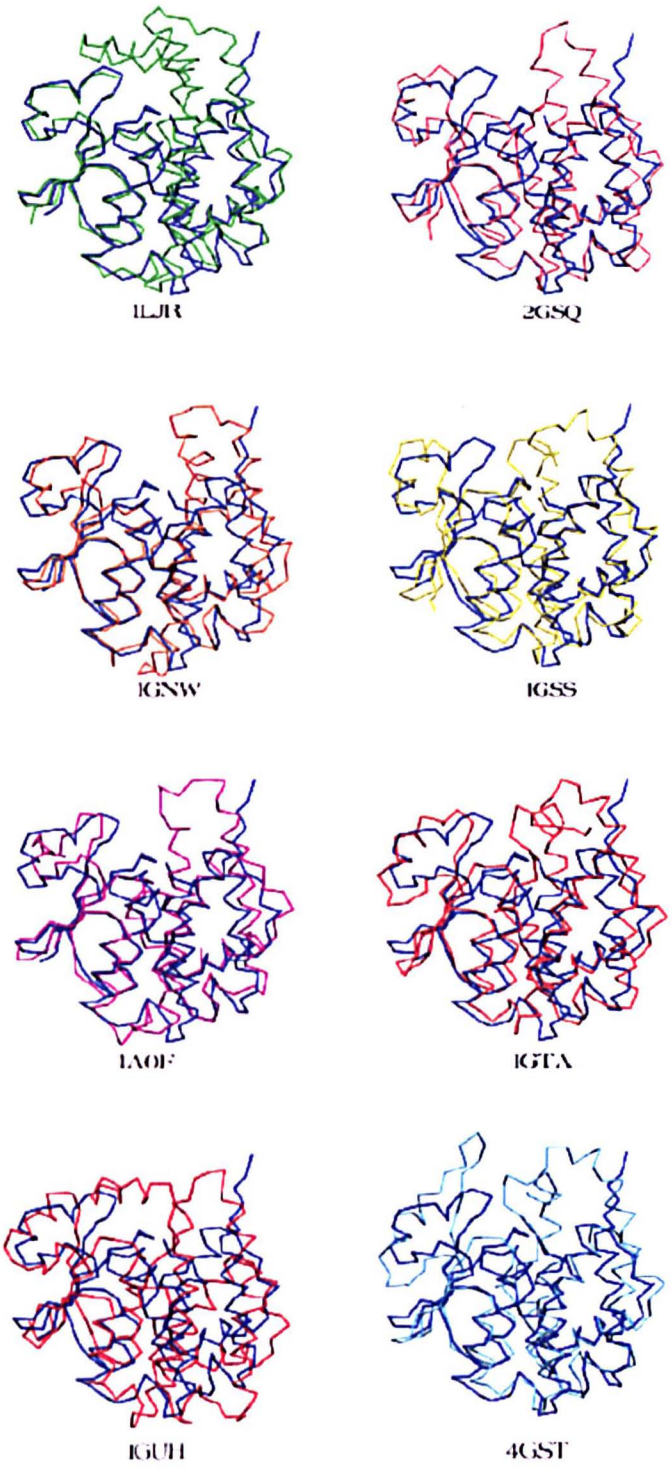


Figure 1-34: Superimposition of *At*GSTZ1 (blue) and other GST structures from different classes. Structures used are 1LJR (Human Theta) green, 1GNW (A. *thaliana* phi) gold, 1A0F (E. *coli* Beta) magenta, 1GUH (human Alpha) red, 2GSQ (Squid Sigma) pink, 1GSS (Human Pi) yellow, 1GTA (S. *Japonicum*) orange, 4GST (Rat Mu) light blue. Images in this figure were drawn using Setor (Evans, 1993).

Original in colour

1.5.3 Catalytic Activity

When complexed with GST, glutathione exists as the thiolate (GS^-) anion at neutral pH (Chen et al, 1988; Graminski et al, 1989a; Graminski et al, 1989b; Huskey et al, 1991). By lowering the pKa of the reduced glutathione sulphydryl group from 9.0 in aqueous solution to around 6.5 when bound in the active site (Armstrong, 1993) it is proposed that the GS^- ion is capable of reacting spontaneously with electrophilic compounds by nucleophilic attack (Jakoby, 1978). In order for this reaction to occur the electrophilic compounds must be in close proximity to the thiolate ion. X-ray crystallography has confirmed that this is the case, describing a highly specific glutathione binding site (G-site) present in the conserved N-terminal region of the enzyme, and a hydrophobic binding site (H-site) in the C-terminal domain, which is able to orientate substrate, bringing them close to the glutathione moiety.

1.6 Induction of Glutathione S-transferases

Glutathione S-transferases in plants are selectively expressed in response to a range of conditions including; cell division, senescence, exposure to pathogens, changes in environmental conditions and in response to chemical exposure (Droog, 1997). Induction of individual genes by a specific stimulus or in a specific tissue has been observed (Meyer et al., 1991b), however more frequently multiple stimuli have been found to cause induction of the same GSTs. In tobacco, *Nt*103, *Nt*107 and *Nt*114 genes encoding GSTs were induced by differing extents following treatment by a range of agents including auxins, auxin transport inhibitors, cytokinins, absicissic acid (ABA), heavy metals, salicylic acid (SA), H_2O_2, GSH, ethanol, ethancrynic acid and pathogen infection (Boot et al., 1993; Droog et al., 1995). The soybean gene Gmhsp26-A was also found to be induced by a variety of agents including; auxins, nonauxin analogues, heavy metals, H_2O_2, SA, dithiothreitol (DTT), GSH, and jasmonic acid (Ulmasov et al., 1995). The induction of GSTs in response to a range of stimuli is not solely observed in plants. Phytochemicals, indole-3-carbinol, indole-3-acetonitrile and flavone have been found to induce GSTs in plant eating insects (Yu, 1984) and over 100 structurally diverse chemicals have been found to induce GST activity in rats and mice (Hayes & Pulford, 1995).

The wide spectrum of chemicals which induce both plant (Marrs, 1996) and mammalian GSTs (Hayes & Pulford, 1995) cannot be grouped by any single chemical feature, suggesting GST induction is part of an adaptive response mechanism to chemical stress. This response is thought to up regulate the expression of GSTs and other detoxifying enzymes to provide protection against a wide range of harmful substances (Talalay et al, 1988).

Identification of chemicals which specifically upregulate GSTs is often hampered by the biotransformation of substrates within the organism. These biotransformation reactions hinder the determination as to whether the parental compound or its metabolites cause induction of GST expression.

1.6.1 Induction through oxidative stress

It has been postulated that structurally different chemicals could induce GSTs if the multiple regulatory elements present upstream of most GST genes were activated through a common step in signal transduction pathways. This step has been proposed to be a change in the oxidative state of the cell, as different treatments such as wounding and pathogen attack, or heavy metal, ethylene or ozone treatment, which have been shown to induce GSTs, all promote an oxidative burst within the cell (Lamb, 1997).

During pathogen or fungal attack, plants initiate the hypersensitive response (HR), through activation of a NADPH oxidase (Marrs, 1996). Activation of this NADPH oxidase leads to a 2-3 minute transient formation of H_2O_2 reinforcing cell walls, restricting fungal growth and in high concentrations (>4mM) promoting cell death. The hypersensitive response also forms Active Oxygen species (AOS) such as superoxide radicals (O_2^-) and hydroxyl radicals (OH^-), which cause lipid peroxidation and form reactive electrophiles. These reactive electrophiles inhibit DNA and protein synthesis and inactivate enzymes involved in cellular signalling (Esterbauer & Cheeseman, 1987; Pickett & Lu, 1989).

At concentrations of 1-2mM, H_2O_2 has been found to selectively induce GSTs and glutathione peroxidases. These enzymes are able to bind and detoxify lipid peroxides and catalyse the formation of glutathione disulphide respectively. This dual response

may appear to be a paradox, with H_2O_2 causing both cell death and initiating transcription of cellular defence mechanisms. However, this dual effect means those cells closest to the site of H_2O_2 release are selectively killed, while cells further away from the site of damage or infection are protected from the effects of the toxic molecules produced.

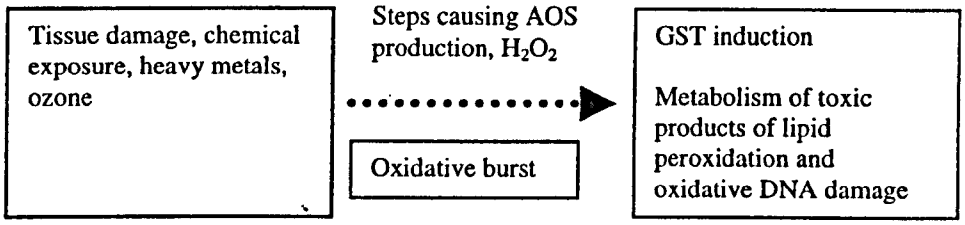


Figure 1-35: Pathogen attack, tissue damage or stress inducing agents generate AOS which are thought to induce GSTs. Heavy metals such as copper and cadmium also generate AOS through a series of redox reactions and it has been suggested that these give rise to the observed induction of GSTs. Ozone and ethylene treatment, also shown to induce GSTs, is thought to promote induction in a similar manner. The induction of GSTs in these circumstances is thought to be part of an adaptive to conjugation of harmful AOS compounds and remove them from the cytosol.

1.6.2 Safeners

Herbicide safeners (also known as herbicide antidotes) were discovered in 1947 following a chance observation that tomato plants treated with 2,4,6-T did not suffer damage when accidentally exposed to vapour of the herbicide, 2,4-D. Further 1,8 napthathalic anhydride (NA), was found to protect maize from injury by thiocarbamate herbicides (Hoffman, 1953; Hoffman, 1969). An array of safeners are now marketed as herbicide mixtures as detailed in figure 1-9 (Davies & Caseley, 1999).

Safener	Crop	Herbicide
Benoxacor	Maize	Metolachlor
Cloquintocetmexyl	Wheat	Clodinafoppropargyl
Cyometrinil	Sorgum	Metolachlor
Dichlormid	Maize	EPTC, butylate, vernolate
Fenchlorazole-ethyl	Wheat	Fenoxaprop-ethyl
Fenclorim	Rice	Pretilachlor
Flurazole	Sorghum	Alachlor
Fluxofenim	Sorghum	Metoachlor
Furilazole	Cereals	Halosulfuron-methyl
Mefenpyr-diethyl	Wheat, Rye, Triticale, Barley	Fenoxaprop-ethyl
MG 191	Maize	Thiocarbamates
Naphthalic anhydride (NA)	Maize	EPTC, butylate, vernolate
Oxabetrinil	Sorghum	Metolachlor

Table 1-6: Table Herbicide safeners available as commercial products (Davies & Caseley, 1999)

Safeners can act by interacting with the herbicide target or receptor proteins involved in herbicide activity (Ebert & Ramsteiner, 1984; Gorog et al., 1982), binding the herbicide to lower its activity, (Hoagland, 1989) blocking herbicide uptake and translocation (Walton & Casida, 1995) or enhancing herbicide metabolism (Fuerst & Gronwald, 1986). Examples of safener enhanced detoxification by P450 enzymes and elevation of reduced glutathione have been observed following safener treatment (Adams et al., 1983; Farago & Brunold, 1990; Frear et al., 1991; Moreland et al., 1993; Zimmerlin & Durst, 1992). In addition, increased tolerance to chloroacetanilide injury by varieties of maize, sorghum, rice and wheat following treatment with the safeners, Dichlormid, Benoxacor, Benzenesulfonamide and Flurazole (table 1.10), have been shown to be due to increased GST activity (Cole et al., 1997; Gronwald et al., 1989; Marrs, 1996).

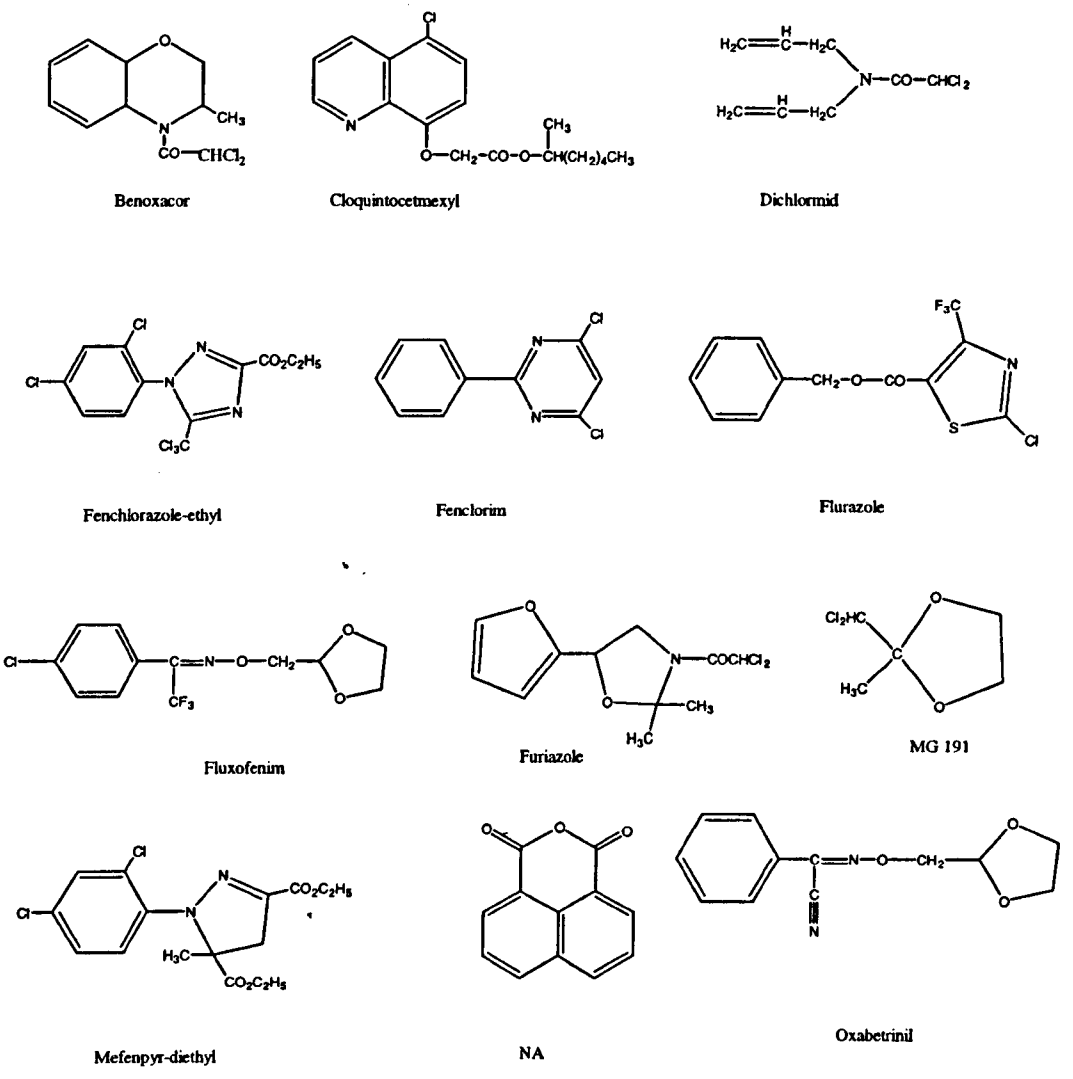


Figure 1-36: Structure of Safeners presented in Table 1-9

Safener	Species	Reference
AD 67	Maize	(Ekler et al., 1993)
BAS 145138	Maize	(Ekler et al., 1993)
Benoxacor	Maize	(Cottingham & Hatzios, 1991)
2-chlorobenzenesulfonamide	Maize	(Hershey & Stoner, 1991)
Cloquintocet-mexyl	Wheat	(Riechers et al., 1997)
Dichlormid	Maize	(Dixon et al., 1997)
Flurazole	Maize	(Wiegand et al, 1986)
MG191	Maize	(Ekler et al., 1993)
NA	Maize	(Mozer et al, 1983)
Oxabetrinil	Sorghum	(Gronwald et al., 1987)

Table 1-7: Studies showing safener enhancement of GST activities in crop plants.

Although safeners may structurally resemble the herbicides they provide protection against, they appear to be non toxic and not to directly activate GSTs *in vitro* (Lay et al, 1976; Gronwald et al., 1987). It has been suggested that safeners enhance herbicide

metabolism by altering the expression of particular genes. This has meant that they have become powerful tools to study the molecular mechanisms and signal transduction events, which lead to upregulated expression of detoxification enzymes.

1.7 Molecular Basis of Regulation of GST Expression

Development, tissue location, stress, pathogen attack and chemical stimulus effect GST expression in plants. There has been relatively little study of the molecular basis of GST induction in plants, although some parallels can be drawn from studies of mammalian systems (Lamoureux & Rusness, 1992; Lamoureux et al., 1991). In animals a number of elements in the 5' flanking region have been found which respond to xenobiotics, `Xenobiotic Regulatory Elements (XREs), Antioxidant Response elements (AREs), Electrophile Responsive elements (EpREs), Barbie box, and NF-κB binding sites (Hayes & Pulford, 1995). These elements are not found upstream of plant GST sequences (Marrs, 1996), however the well characterised plant enhancer, *ocs* element has been found (Chen et al., 1996; Ellis et al., 1993; Takahashi et al., 1995; vanderZaal et al., 1996) upstream of several GSTs. These *ocs* elements are similar to EpREs in that they also contain tandemly duplicated binding sites and are able to bind the dimeric b-ZIP transcription factors, Fos/Jun proteins (Daniel, 1993; Friling et al., 1992; Zhang & Singh, 1994). *Ocs* elements have been described as playing a role in both basal and induced expression of plant GSTs conferring induction by auxins, inactive auxins, salicylic acid, GSH, H_2O_2 and methyl jasmonate (Ulmasov et al., 1994; Ulmasov et al., 1995). The level of induction conferred by *ocs* elements appears to be tissue specific with greatest induction occurring in the root tissue (Benfey et al., 1989) (Ellis et al., 1993; Fromm et al., 1989; Zhang & Singh, 1994; Chen & Singh, 1999).

1.7.1 Postranscriptional control of GST

There is also some evidence that the levels of GST activity can be controlled by modulation of the mRNA and protein. Several plant GST genes contain one or more copies of the sequence ATTTA in the 3' untranslated region, a sequence shown to be an instability element that targets mRNAs for degradation by RNases (Takahashi et al., 1995). At the protein level alteration of GST activity has been observed where Reactive Oxygen Species (ROS) and methylation has been shown to modulate the activity of class Mu enzymes (Murata et al, 1990).

It is clear that many different GSTs exist in plants and little is known regarding the specific functions that each particular GST plays within the plant. The work of this thesis was to better structurally characterise a range of GSTs in plants to gain an improved understanding of the differences in overall structure and particularly in substrate binding between different GSTs in a range of species.

In order to try and gain the most information from any structural data that might be collected it was decided to focus on GSTs that were of commercial interest, in that they were known to detoxify relevant herbicides and had been biochemically characterised. In addition, GSTs were included in crystallisation trials if they were thought to be novel or show different substrate specificity to previously characterised GSTs.

Using the above criteria GSTs from every known plant class were eventually included in the trials. Multiple GSTs of the Tau class and Zeta class were included in the crystallisation trials. Each of the Tau class GSTs under study were commercially significant and it was hoped that if a number of GSTs of the same class were studied, differences in their structure may enable one to crystallise more easily than others.

At the end of the three years of study, ten GSTs had been included in the crystallisation trials from a range of species including *Arabidopsis thaliana*, Carnation, Maize, Petunia and Wheat.

Chapter 2: Materials and Methods

2.1 Expression

To enable large quantities of protein to be produced and easily purified the individual GST isoenzymes were cloned and transformed into *Escherichia coli.* Dr David Dixon, (Dr Robert Edward's group, Crop Protection Unit, Durham University) provided the constructs of *Zm*GSTF1, *Zm*GSTU1, *Zm*GSTU1-2, *Zm*GSTU2, Wheat J, CAR SR8, *At*GSTZ1 and *At*GSTT1, while the constructs AN9 and *Zm*GSTU4, were provided by Lukas Müeller (Dr Virginia Walbot's group, Stanford University, USA). Details of the production and purification of the recombinant proteins *Zm*GSTF1, *Zm*GSTU1 and *At*GSTZ1 have been previously published (Dixon et al., 1997a; Dixon et al., 1997b; Dixon et al, 2000). The constructs *Zm*GSTU1-2 and *Zm*GSTU2 were cloned as described as for *Zm*GSTU1 (Dixon et al., 1997b) while the carnation Zeta class GST (CAR SR8), was cloned as described for *At*GSTZ1 into pET24 adding a hexa-histidine tag at the C-terminus. The enzyme *At*GSTT1 was cloned into the pET-11 expression system, which also added a histidine tag, but in this case at the N-terminus of the protein. The Durham constructs were transfected into the E. *coli* strain BL21 DE3 while the Stanford constructs (*Zm*GSTU4 and AN9) were transfected into JM105 and M15pREP4 strains respectively.

2.1.1 Expression of Glutathione S-Transferase isoenzymes - *Zm*GSTF1, *Zm*GSTU1, *Zm*GSTU1-2, *Zm*GSTU2, Wheat J & *Zm*GSTU4

A fresh single colony was picked into 10ml Luria Broth (LB) containing Ampicillin (100µg/ml) and incubated at 37°C with vigorous shaking overnight. This overnight culture was used to inoculate 500ml LB supplemented with Ampicillin (100µg/ml) in a 2L cut flask to give an initial OD_{600} of 0.1. This 500ml culture was then incubated at 37°C with vigorous shaking for two hours until the OD_{600} reached approximately 0.5. At this point 1ml of the culture was transferred to a universal and incubated at 37°C (section 2.1.5). IPTG, at a final concentration of 1mM, was added to the remainder of the culture and it was further incubated at 37°C for three hours to allow induction of gene expression to

occur. Following the three hour induction a second 1ml sample was removed from the flask and this was used to check induction of gene expression had occurred (section 2.1.5). The remainder of the induced sample was centrifuged at 3000×g for 45min (4°C) and the resulting pellet stored at -80°C. Prior to purification the pellets were re-suspended in a 1/50 of the original culture volume of 10mM phosphate buffer pH 7. The bacterial pellet was lysed using a Philip Harris, US 200 sonicator set at half power (six 30 second bursts with 30 second rest between bursts). The crude lysate was then centrifuged at 14000×g for 30min at 4°C. The resultant supernatant was removed and used in the subsequent purification steps (section 2.3).

2.1.2 Expression of GST isoenzymes - CAR SR8, *At*GSTZ1 & *At*GSTT1

These GST isoenzymes were expressed as described above with the substitution of Kanamycin (50µg/ml) for Ampicillin (100µg/ml) as the selective antibiotic. The pellets were lysed and resuspended as before, but in a different buffer, 20mM Imidazole, 50mM Phosphate pH8.0, 300mM NaCl which allowed loading of the sample ontó a nickel charged metal chelate column, the first step in the subsequent purification.

2.1.3 Expression of GST isoenzyme - AN9

AN9 was expressed using Carbenicillin (100µg/ml) and Kanamycin (50µg/ml) as the selective antibiotics. Following addition of the overnight culture to the larger 500ml cultures these larger cultures were incubated until OD_{600} was 0.6. IPTG was then added, to a final concentration of 0.2mM to the cultures, which were then further incubated for 4 hours at 30°C to allow induction of gene expression to occur. Following induction the cultures were centrifuged as described above and the resultant pellet was resuspended in the same buffer described in section 2.1.2. It was found that the addition of glycerol (10%) into the resuspension buffer increased the amount of soluble protein which could be recovered during the subsequent purification steps.

2.1.4 Expression of Selenomethionine protein

Use of a recombinant expression system for protein crystallographic studies allows for the manipulation of the overexpression system such that methionine

residues in the protein can be substituted with selenomethionine. This substitution of methionine with selenomethionine allows phase information to be gained using Multiwavelength Anomolous Dispersion (MAD), (section 4.3.3) which has been discussed in a large number of publications (Hendrickson & Ogata, 1991; Hendrickson, 1997; Wilson et al, 1997).

In order to gain phase information for the GST isoenzymes wheat J and *At*GSTT1 (thought to be too dissimilar to previously solved GSTs for molecular replacement) the respective plasmids for each enzyme were obtained using a Qiagen Mini Prep kit, as detailed in the manufacturers instructions. Following transformation the *E. coli* methionine auxotroph strain B834 by the respective plasmids the transformed E. *coli* were grown and expressed in minimal media containing selenomethionine (further details are given in table 2.1). The growth of the *E. coli* in this selenomethionine containing media was significantly slower and although IPTG was not added until OD_{600} was 0.5 only around 1/4 of the normal yield of protein was found to be expressed.

Solution 1	
Chemical	Quantity (g)
1. alanine	0.5
2. arginine HCL	0.58
3. aspartic acid	0.4
4. cystine	0.03
5. glutamic acid	0.67
6. glutatmine	0.33
7. glycine	0.54
8. histidine	0.06
9. isoleucine	0.23
9. leucine	0.23
10. lysine HCL	0.42
11. phenylalanine	0.13
12. proline	0.10
13. serine	2.08
14. threonine	0.23
15. cytosine	0.17
16. valine	0.23
17. adenine	0.5
18. guanosine	0.67
19. thymine	0.17
20. uracil	0.5
21.sodium acetate	1.5
22. succinic acid	1.5
23. ammonium	0.75
24. sodium hydroxide	1.08
25. dibasic K_2HPO_4	10.5

Solution 2	
Chemical	Quantity (g)
glucose	10g
MgSO4	0.25g
Fe SO4	4.2mg
conc H_2SO_4	8µl
thiamine -HCl	5mg

Table 2.1: Components of solution 1 were mixed together and made up to one litre with degassed dH_2O. This solution was then autoclaved for 20 minutes at 15lb pressure and the pH of the subsequent solution was adjusted to 7.5. Components of solution 2 were mixed and made up to 100ml with degassed dH_2O and the pH of this solution adjusted to pH 7.0. Solution 2 was through a 0.2µM filter into solution 1. 50mg of DL-SeMet (Sigma) was added to the resulting mixture.

2.1.5 Testing for Protein Induction

To test protein induction occurred following addition of IPTG, 1 ml samples of the bacterial culture were taken before and after IPTG addition. These samples were pelleted by centrifugation in a microfuge (13000×g for 1min), the supernatant removed and the pellet resuspended in 100µl sample buffer (2×sample buffer; 0.125M Tris-Cl, pH 6.8, 4% SDS, 20% glycerol and 10%

mercaptoethanol). These samples were then boiled for 2 min prior to loading on a 12.5% SDS-PAGE gel. Electrophoresis was performed as described in section 2.2. The gels were stained using Coomassie Brilliant Blue and destained until the background was clear.

2.2 Electrophoresis

2.2.1 SDS-PAGE

SDS-PAGE is a method by which the polypeptides of a sample can be separated based on their mass, allowing quick analysis of protein solutions to determine if the target protein is present and if protein contaminants exist. The BioRad Mini Protean II system was used for all gel electrophoresis. It was assembled according to the manufacturers instructions and the 12.5% separating gel solution and stacking gel solution were prepared (Sambrook et al, 1989). The apparatus, the 12% separating gel solution and stacking gel solution were prepared according to the manufacturers instructions.

2.2.2 Staining of SDS-PAGE gels

Analysis of electrophoretic profiles was performed using either Coomassie Brilliant Blue R-250 (CBB R-250) or silver staining.

The non-polar, sulphated triphenylamine dye Coomassie Brilliant Blue R-250 was used to detect if protein induction had occurred following addition of IPTG to bacteria. Staining was performed for about 2 hours. Destaining was carried out using 4 washes with water heated to around 80°C.

As silver staining is 20-200 times more sensitive than CBB R-250 at detecting protein, it was used to monitor the protein purification steps and to check the purity of the final purified protein product. In this method gels were first washed with 20% w/v trichloroacetic acid (TCA) to precipitate and immobilise the separated proteins and to remove non-protein components which might subsequently interfere with staining procedures. Gels were subsequently washed with 8% glutaraldehyde for 20 minutes before 2×1 minute washes with water. Following incubation with 0.25% silver nitrate solution for 20 minutes, gels were

again washed with water for 2×1min. The bands within the gels were then visualised by incubation with 2.5% sodium carbonate; 0.1% formaldehyde. The staining reaction being terminated through the addition of solution containing 10% glycerol and 5% acetic acid. Once destained, gels were preserved by air drying between sheets of porous polyethene.

2.3 Purification

Purification of all the GST isoenzymes was performed using affinity purification as the first step. This was typically followed by ion exchange and gel filtration. In all cases the final purity of the protein was high with only trace impurities observed when the sample was run on a 12.5% SDS PAGE gel subsequently silver stained.

2.3.1 *Zm*GSTF1 purification

Purification of the *Zm*GSTF1 isoenzyme was based on a previously determined protocol (Dixon et al., 1997), using an Orange A DyeMatrex gel column (Amicon - triazinyl dye immobilized on cross linked agarose) as a rapid and highly specific affinity purification step. A 5ml column of Orange A DyeMatrix was equilibrated in buffer A (10 mM potassium phosphate buffer, pH 6.0), before loading the crude extract. The column was then washed with 5 column volumes of Buffer A followed by Buffer B (50 mM potassium phosphate buffer, pH 7.0) to remove non specifically bound proteins. The *Zm*GSTF1 isoenzyme was eluted using Buffer B containing 5 mM glutathione and the fractions pooled before overnight dialysis in 10mM Tris-HCl (pH 7.8). Buffer exchange into a low ionic strength buffer enabled ion exchange chromatography to be performed as a next purification step (section 2.3.4).

The Orange A DyeMatrix column was cleaned using 1 M NaCl and stored in 50 mM Tris pH 8.0 containing 0.02M sodium azide to prevent bacterial growth.

2.3.2 *Zm*GSTU1, *Zm*GSTU1-2, *Zm*GSTU2 & Wheat J

These Tau class GSTs were purified using an S-hexyl glutathione affinity column, prepared as described in sections 2.3.2.1 and 2.3.2.2, using a previously determined protocol (Dixon et al., 1997b). S-hexyl glutathione is a substrate-conjugate analogue and this affinity step allowed the rapid purification of almost pure enzyme. Crude supernatant was loaded onto the 6 ml S-hexyl glutathione column pre-equilibrated in 20 mM Tris (pH7.8). Following washing with equilibration buffer the column was further washed with a solution of 20 mM Tris (pH7.8) + 50 mM NaCl. The GST was eluted using 20 mM Tris (pH 7.8) containing 5 mM S-hexyl glutathione. The column was regenerated using 20 mM Tris (pH 7.8) + 3 M NaCl and stored in 20mM Tris (pH 7.8) + 0.02% sodium azide. Eluted fractions were pooled and further purified using Q-Sepharose anion exchange chromatography.

2.3.2.1 S-Hexyl glutathione synthesis

S-hexyl glutathione was prepared as previously described (Mannervik & Guthenberg, 1981), Three grams of glutathione were dissolved in 10mls distilled water then 10mls 2 M NaOH was added. Ethanol was added dropwise until the cloud point of the solution was reached. Two grams of iodohexane was added with stirring and the solution left to stand overnight at room temperature. The pH of the solution was then dropped to 3.5 using 47% hydriodic acid and the resultant suspension of precipitated S-hexyl glutathione stored at 4°C overnight. The precipitate was collected by centrifugation, washed with cold water and dried over silica gel in a vacuum desiccator.

2.3.2.2 Preparation of S-hexyl glutathione-Sepharose

S-hexyl glutathione-Sepharose was prepared by adding two grams of epoxy-activated Sepharose CL-6B (Sigma) to 20ml of distilled water. This mixture was left to swell for 25min at room temperature, washed on a sintered glass funnel with 200 ml of distilled water, then adjusted to pH 12 with 5 M NaOH. Synthesised S-hexyl glutathione was added to the washed Sepharose then mixed gently for 16 hours at room temperature, allowing the S-hexyl glutathione to covalently bind to the activated Sepharose. The conjugated Sepharose was then washed twice with 200 ml of distilled water. Following conjugation, any

unreacted groups on the Sepharose were blocked by incubation with 10 mls of 2 M ethanolamine for four hours at 30°C. The conjugated Sepharose was washed twice with distilled water and three times with 20 mls of 0.1 M sodium acetate (pH 4.0) + 0.5 M NaCl. Finally, the conjugated Sepharose was washed with 0.1 M Tris-HCl pH 8.0 containing 0.02% w/v sodium azide and stored at 4°C. The resultant gel, with an approximate volume of 6 ml, was packed into a 10 mm diameter chromatography column.

2.3.3 His tagged GSTs

Clones of the GSTs *Zm*GSTU4, AN9, CAR SR8, *At*GSTZ1 & *At*GSTT1 were supplied in vectors which incorporate a hexa-histidine tag onto the protein, as these isoenzymes were found not to bind well to glutathione based matrices (Dixon et al., 2000). The poly-histidine tag has a very high affinity for nickel ions thus affinity purification using a nickel charged metal chelate column can be performed. The histidine tagged proteins were initially purified using a Nickel charged POROS Metal Chelate column attached to a BioCad Sprint machine. The column was equilibrated with wash buffer, (20 mM imidazole, 50 mM Phosphate buffer (pH 8.0) and 300 mM NaCl (with addition of 10% glycerol to buffers in AN9)) and supernatant from the lysed pellet loaded onto the column. Following washing with ten column volumes of equilibration buffer, the GST was eluted by increasing the concentration of imidazole in the buffer, (250 mM imidazole, 50 mM Phosphate buffer, 300 mM NaCl).

The small POROS Metal Chelate column had only limited binding capacity and was also found to develop high back pressures with repeated use. A second 10ml column was poured using Qiagen Ni-NTA Superflow matrix which allowed a greater quantity of protein to be purified using similar conditions as established for the POROS matrix with no apparent loss in resolution.

2.3.4 Ion Exchange

2.3.4.1 Q-Sepharose ion exchange purification

The GST isoenzymes (with the exception of *At*GSTT1) were further purified after the initial affinity step using Q-Sepharose fast flow matrix from Pharmacia

Biotech. A 10ml column of this matrix was poured and packed following manufacturer's instructions. The column was equilibrated with Buffer A, (Buffer A – 20 mM Tris-HCl pH 7.8, 14 mM mercaptoethanol), and the dialysed elutant from the previous affinity step loaded. Following washing of the column for 5 column volumes with buffer A, a linear gradient was applied using a second buffer (Buffer D – 20 mM Tris-HCl pH 7.8, 1 M NaCl). The GSTs typically eluted from the column at a concentration of 0.25 M NaCl. Five column volumes of Buffer D were used to clean the Q-Sepharose column and it was then stored in 20 mM Tris + 0.02 M sodium azide following each purification run to prevent bacterial growth.

2.3.4.2 Cation exchange

While the other GST isoenzymes under study had theoretical isoelectric points (calculated by Prot Param: http://www.expasy.ch/tools/protparam.html) in the range of 5.3-6.6, the isoenzyme *At*GSTT1 had a theoretical pI of around 9.5. The higher pI of this protein meant it was unlikely that this protein would bind the Q sepharose column used for the other GSTs. Ion exchange was thus performed using a 5ml Hi-Trap S column. Following metal chelate chromatography fractions containing *At*GSTT1 were dialysed overnight in 50mM phosphate (pH 7.1). The sample was then loaded onto a 5 ml Hi-Trap S column which had been equilibrated with 50 mM phosphate (pH 7.1). The column was then washed with 5 column volumes of equilibration buffer and a linear gradient was applied using 50 mM phosphate pH 7.1 + 1 M NaCl to elute *At*GSTT1.

2.3.5 Gel filtration

Gel filtration using a 100ml Superdex 200 column was employed as final purification step for all the GSTs. The column was equilibrated using 150 ml of 50 mM Tris-HCl, pH 7.5, 200 mM NaCl (the high salt preventing non specific interactions with the matrix), before loading of a concentrated 1ml sample of the purified protein. The column was then washed with 150 mls of equilibration buffer at 0.5 ml per minute. A sharp peak was observed for all the isoenzymes and the middle fractions of this sharp peak were used in crystallisation trials.

2.3.6 Protein concentration using Absorbance (280nm)

Measurement of the absorbance of proteins in the near UV (280nm) due to their tyrosine and tryptophan content (and to a small extent the amount of Phe and Cys) allows quantitative determination of protein concentration in solution. Although this method is not the most accurate at determining protein concentration due to interference from other chromophores, it is simple to perform, the sample can be recovered and, if similar conditions are used, it is reproducible. In order to determine the specific absorption value the following formula was used for each isoenzyme studied.

$$E = \frac{[1280nY + 5960nW + 120nC]}{Mwt}$$

Equation 2-1: The molar extinction coefficient (M^{-1} cm^{-1}) for each protein is calculated using the above equation and by substituting the Y, W and C with the number of tyrosine, tryptophan and cysteine residues present in the sequence respectively.

Using the appropriate molar extinction coefficient value as calculated for each protein along with the absorbance value (A) gained at 280nm, protein concentration (c) can be determined using equation 2-2.

$$A = Ecl$$

Equation 2-2: Rearranging the equation allows determination of the protein concentration (c) in mg / ml where (A) is the absorbance measured in OD units at 280 nm and (l) is the path length of the spectrophotometer in cm.

2.4 Crystallisation

2.4.1 Theory of Crystallisation

As more synchrotron beamlines become available with greater flux and faster detectors for use in protein crystallography, it has become the crystallisation of biological macromolecules which is the rate limiting step in the structure determination process. Many texts have been published devoted solely to the theory and practical techniques of protein crystallisation, therefore only a brief account of the theory surrounding the crystallisation methods used in this study is presented (Bergfors, 1999; Ducruix & Giege, 1992; McRee, 1999).

X-rays scattered by a single molecule are too weak to detect. To allow detection of macromolecular X-ray diffraction crystals, thermodynamically stable aggregates with a repeating lattice, are grown.

Crystallisation is one of several means (including non-specific aggregation / precipitation) by which a metastable supersaturated solution can reach a stable lower energy state by reduction of solute concentration. If crystallisation is to occur in preference to non-specific aggregation or precipitation, three processes must occur; nucleation, growth, and termination.

2.4.1.1 Nucleation

Nucleation is the first step in crystallisation, in which free molecules or non crystalline aggregates (dimers, trimers) existing in a supersaturated solution come together to form a stable aggregate lattice typically of a free energy ~3-6 kcal/mole less than the relative solution state. This stable aggregate must exceed a critical size (thought to be between 10-200 molecules) to form a macroscopic crystal, otherwise the free energy of formation will be positive, spontaneous dissolution will occur and non-specific aggregates / crystalline precipitates will form. The formation of nuclei seems to increase with the degree of supersaturation, thus to achieve few nuclei and larger crystals in preference to showers of small crystals, supersaturation must be approached slowly.

2.4.1.2 Growth

Once critical nuclei have formed, the crystal must grow to a size suitable for diffraction studies, typically around 0.2-0.3mm. Crystal growth begins at solute concentrations sufficient for nucleation to occur and continues at concentrations beneath the nucleation threshold. The growth rate is determined by the nature of the growing crystal and the diffusional rate of the molecules in the crystallising solution.

2.4.1.3 Cessation of Growth

Cessation of crystal growth typically occurs due to a decrease in the concentration of the crystallising protein to a point where the solid and solution

phases reach exchange equilibrium. Additional factors which prevent crystal growth irrespective of solute concentration are; cumulative lattice strain effects (Feher & Kam, 1985) and poisoning of the crystal growth surface (Sato et al, 1992). In the second case impurities or damaged crystallising protein is incorporated into the crystal, resulting in defects which prevent further packing of molecules. To remove surface poisoning of the crystal and allow further growth of the crystal, macroseeding can be attempted (Thaller et al, 1981).

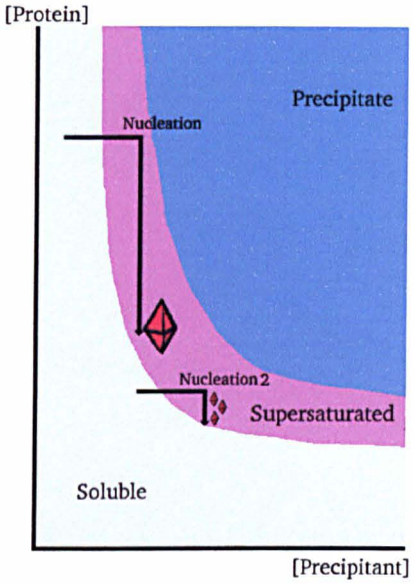


Figure 2-1: Schematic solubility diagram describing the protein state at varying levels of protein precipitant concentration during a crystallisation experiment. Within the supersaturated region crystallisation can occur. Ideally only limited nuclei form which then spend longer in the supersaturated region allowing larger crystals to grow (nucleation 1). Nucleation 2 in the above diagram describes the situation where supersaturation is achieved quickly with the formation of many nuclei. This production of multiple nuclei quickly reduces the available protein concentration which in turn leads to more rapid cessation of growth and thus smaller crystals.

2.4.2 **Method of Crystallisation**

A number of factors including precipitant type, precipitant and protein concentration, temperature, pH and molecule specific factors (such as prosthetic groups and ligands) affect the rate at which nucleation and subsequent crystal growth occurs.

Original in colour

2.4.2.1 Precipitants

Precipitant compounds can be divided into six categories; salt, high molecular weight straight chain polymers (e.g. PEG), organic solvents, sulphonic dyes and deionised water. These molecules are used at a range of concentrations when screening for crystals in initial sparse matrix trials (Jancarik & Kim, 1991). If a particular precipitant and concentration is found that yields crystals, fine linear grid screens surrounding this condition are used to optimise this initial condition. In vapour diffusion experiments the difference in the concentration of precipitant in the well and that in the mother liquor often affects the speed of crystal growth. In addition different ratios of precipitant to protein can also influence the rate of nucleation and crystal growth.

2.4.2.2 Temperature

Many factors related to solubility of the protein are temperature dependent, thus this is an important parameter in crystallisation. Crystallisation at lower temperatures often results in slower growth of crystals and may mean that a condition which yields many small rapidly formed crystals at room temperature, yields fewer larger crystals at 4°C.

2.4.2.3 pH

Proteins are charged species and the net charge of a protein often determines its solubility. Alteration of pH and the salts present in the crystallisation solution modifies the net charge of the proteins in the solution and thus effect their solubility in the precipitant solution. Buffers are always included in the crystallisation solution to control the pH of the crystallisation solution and hence the solubility of the protein.

2.4.2.4 Molecule specific factors

Factors such as ligands and small molecules have been shown to affect the crystallisation of proteins. Small molecule interactions may fix a protein in one conformation or may alter interactions between solvent or other molecules in the lattice and the protein. The small interactions effected by these molecules can perturb the sample in such a way that larger or better diffracting crystals may be grown.

2.4.2.5 Vapour diffusion

Several different techniques have been described which allow a supersaturated state to be attained, as required for crystallisation (Bergfors, 1999). One such method, the vapour diffusion technique, utilises evaporation and diffusion of water between solutions of different concentration, as a means of approaching and achieving supersaturation of the macromolecules. Typically, in the vapour diffusion experiment the solution containing the macromolecule is mixed 1:1 with the crystallising well solution to form the drop solution. The drop containing the 1:1 mixture of protein and precipitant (both of which have thus been diluted to 1/2 their original concentration) is then sealed over the crystallising well solution, which contains the precipitant at the target concentration.

Differences in the precipitant concentration of the drop and the well solutions means a non equilibrium condition is initially present in the sitting drop cell (figure 2-1). This non equilibrium condition means that the rate of evaporation of water from the drop is be greater than that from the well solution allowing the slow approach of the precipitant concentration in the drop towards that of the crystallising well solution. Eventually equilibrium will be reached when the concentration of the precipitant in the drop equals that of the well solution.

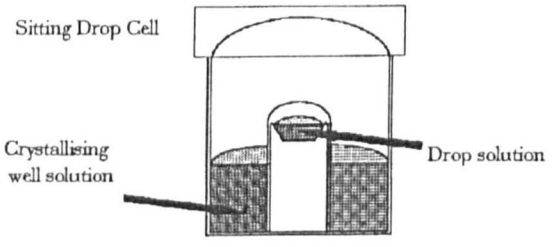


Figure 2-2: Sitting drop apparatus, typically well solution is around 800µl and drop solution is around 2-4µl.

Original in colour

Chapter 3: Wet Lab Results

3.1 Expression

In order to maximise the amount of glutathione S-transferase produced in each prep. (typically 5-10mg/litre), expression tests were performed to determine at which time point during induction the greatest ratio of target protein over background was obtained. An optimal induction time of between 3-4 hours incubation following addition of IPTG was found to maximise expression for all the GST isoenzymes under study.

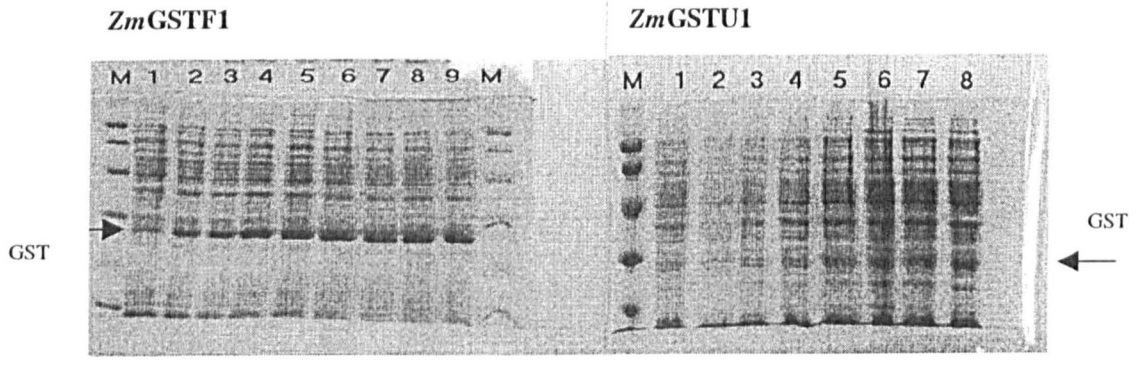
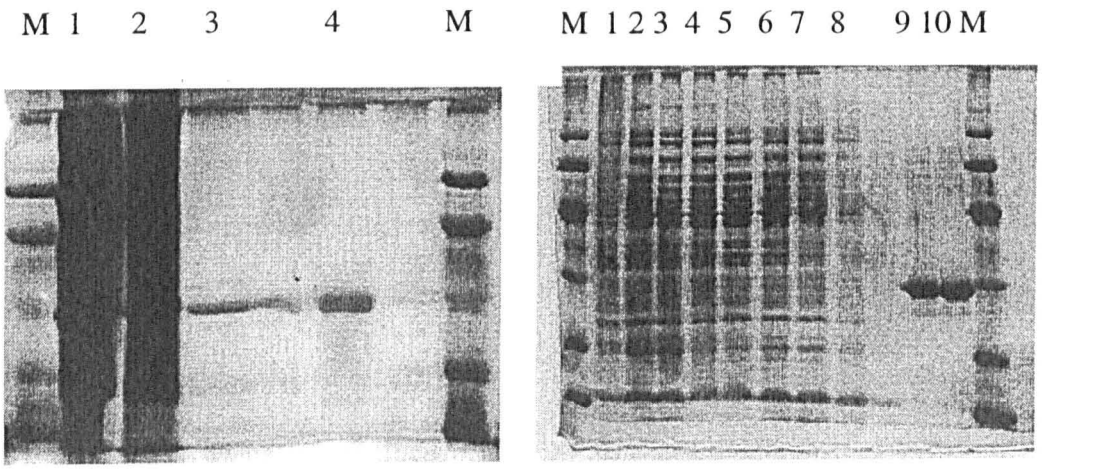


Figure 3-1 Example of Typical time course of induction with 1mM IPTG for maize isoenzymes *Zm*GSTF1 and *Zm*GSTU1. M, Markers (Phosphorylase 94kDa, Albumin 67kDa, Ovalbumin 43kDa, Carbonic Anhydrase 30kDa, Trypsin Inhibitor 20.1kDa and Lactalbumin 14.4kDa) Time point 1 was taken before IPTG addition, then samples were then taken every 30 minutes following IPTG addition. Maximal expression between 3-4 hours (timepoint 7).

3.2 Purification

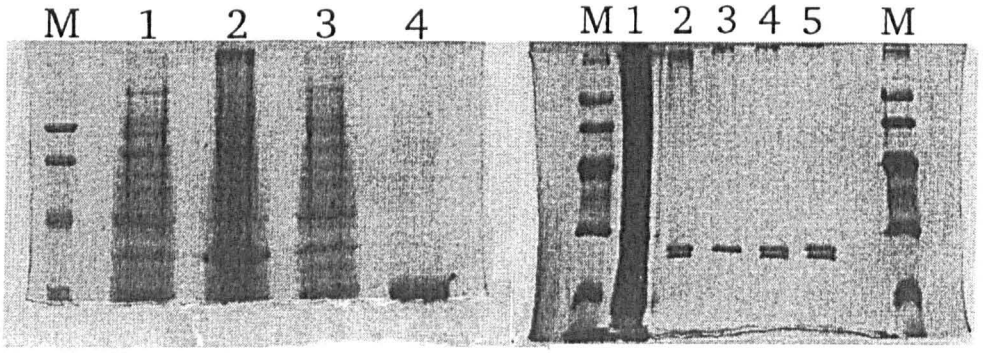
Following over expression, each isoenzyme was purified as described in section 2.3 to yield protein typically greater than 95% pure (figures 3-2 to 3-5)



ZmGSTF1: M = markers, 1= pellet, 2 = crude lysate, 3= Orange A affinity step, 4 = Hi-Trap Q

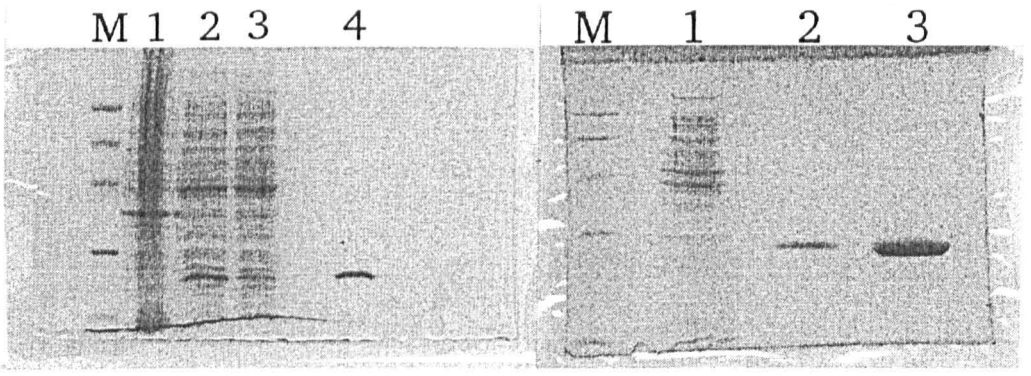
An9: M = markers, 1-8 = flow through, and wash of metal chelate column, 9-10 eluted GST from metal chelate column

Figure 3-2: Purification of Phi Class GST, *Zm*GSTF1 utilised an Orange A affinity step followed by anion exchange. The Phi class enzyme AN9 was purified via a Nickel charged metal chelate column, using its His -Tag.



ZmGSTU1: M = markers, 1 = pellet, 2= lysate, 3 = flow through, 4 = S-hexyl affinity column elutant

ZmGSTU1-2: M = markers, 1 = lysate, 2 = S-hexyl elutant, 3 = Hi-Trap Q peak 1 (Zm GSTU1), 4 = Hi-Trap Q peak 2 (Zm GSTU1 and ZmGSTU2), 5 = Hi-Trap Q peak 3 (Zm GSTU1 and ZmGSTU2)

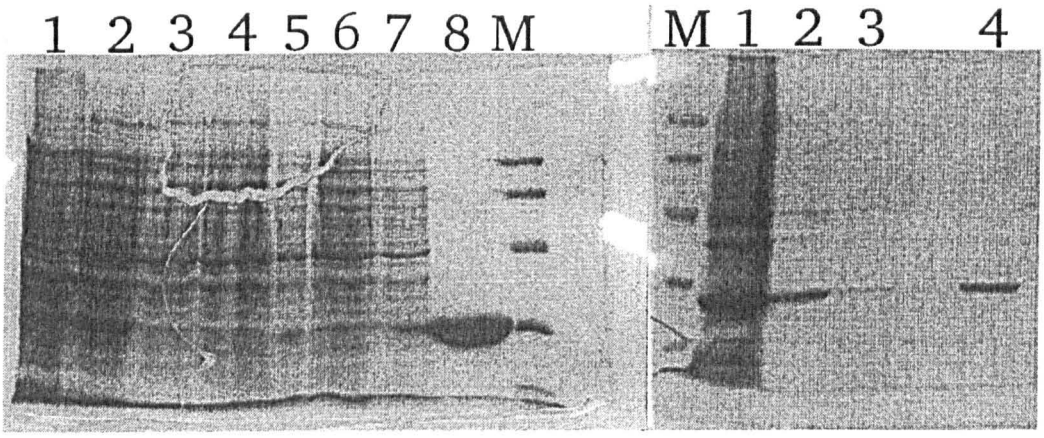


ZmGSTU2: M = markers, 1 = pellet, 2= lysate, 3 = flow through, 4 = S-hexyl affinity column elutant

Wheat J : M = markers, 1 = Lysate, 2 = S-hexyl affinity column elutant, 3 = Hi –Trap Q peak

Figure 3-3: Purification of Tau class GSTs, *Zm*GSTU1, *Zm*GSTU1-2, *Zm*GSTU2 & Wheat J which utilised a S-hexyl affinity column

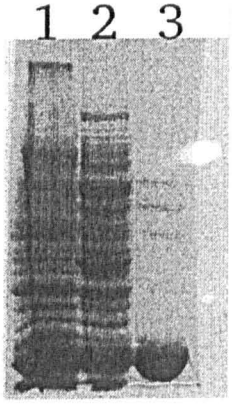
A number of attempts were made to express the maize GST, *Zm*GSTF4, suggested to transport anthocyanin from the cytosol to the vacuole, however expression of this protein was found to lead to insoluble protein which accumulated in inclusion bodies. Despite trying to optimise factors known to increase the expression of soluble protein such as; the growth temperature of the *E. coli*, concentration of IPTG used for induction and use of different types of media, soluble *Zm*GSTF4 could not be produced. No attempt was made to refold the protein from inclusion bodies.



*At*GSTZ1: 1 = pellet, 2= lysate, 3-7 = flow through from metal chelate, 8 = metal chelate elutant, M = markers

Carnation GST SR8 M = markers, 1 = lysate, 2 = metal chelate elutant, 3 = wash through from metal chelate, 4 = protein from gel filtration

Figure 3-4: Purification of Zeta Class GSTs, *At*GSTZ1 and SR8, which utilised a nickel, charged metal chelate column.



*At*GSTT1: 1 = Pellet, 2 = lysate, 3 = elutant from metal chelate purification

Figure 3-5: Purification of Theta Class GST, *At*GST1 that utilised a nickel charged metal chelate column.

3.3 Crystallisation

All the crystallisation conditions performed were set up using the sitting drop method of vapour diffusion. Initial screening of the protein for appropriate crystallisation conditions was performed using sparse matrix screens made in house, based on a number of commercial screens (Jancarik & Kim, 1991). Optimisation of these initial conditions through variation of protein and precipitant concentration, drop size and pH, as described in section 2.5, enabled crystals of sufficient size and quality to be grown for a number of the GST isoenzymes.

3.3.1 **Crystallisation of Phi class GSTs**

3.3.1.1 Crystallisation of ZmGSTF1

Sparse matrix screening by a previous student (Richard Grant) suggested conditions in which small crystals could be grown. Further optimisation of these conditions enabled diffraction size crystals to be grown using 1µl of protein solution (3mg/ml protein in 10mM Tris-HCl buffer, 0.4mM dithiothreitol (pH 7.8)) and 1µl precipitating buffer (20-25% (w/v)) PEG 2000 mono-methylether (Me2KPEG) in 0.1M BTP buffer (pH 6.8 to 7.4)). These 2µl droplets were then equilibrated against 800µl of precipitating buffer at 17°C. The crystals, which grew in approximately 14 days, were thin plates of maximum dimension 0.2 mm (figure 3-6). These thin plates were found to diffract to 2.8Å.

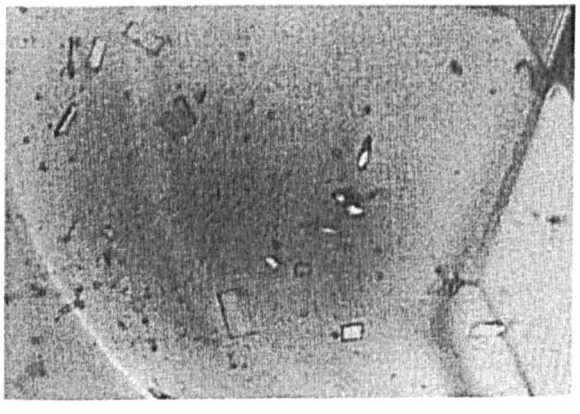


Figure 3-6: Plate like crystals of *Zm*GSTF1, which were found to belong to the monoclinic lattice system $P2_1$

Using similar conditions as above, (19% Me2KPEG, BTP pH 7.2) but with 5 mM chlorophyllin also present in the crystallisation mixture, more crystals of *Zm*GSTF1 were grown that displayed different morphology to those described above. Unfortunately despite screening a number of these larger cubic crystals (~ 0.3mm), which were light green in colour, these cubic crystals were found to diffract only to 2.9Å.

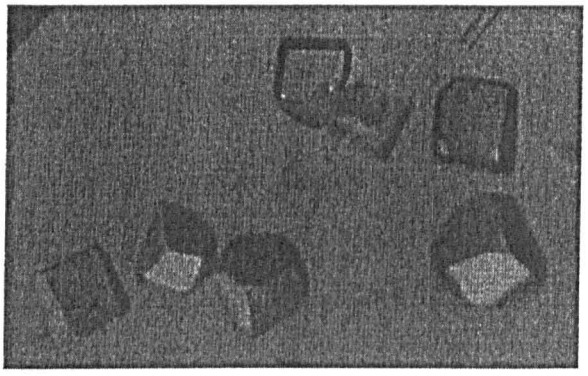


Figure 3-7: Cubic crystals of *Zm*GSTF1 grown following the addition of chlorophyllin (a noncompetitive inhibitor of *Zm*GSTF1) to the crystallisation mixture.

3.3.1.2 Petunia AN9

Optimisation of initial conditions found by sparse matrix screening enabled small needle like crystals of this GST to be grown in PEG 8000 (12-17%), 0.1M cacodylate (pH 6.5) & 0.2M Magnesium acetate. These crystals were found to diffract to 2Å at the microfocus beamline (ID13) at the ESRF, Grenoble, however this data was clearly twinned and further attempts to optimise these crystals was unsuccessful.

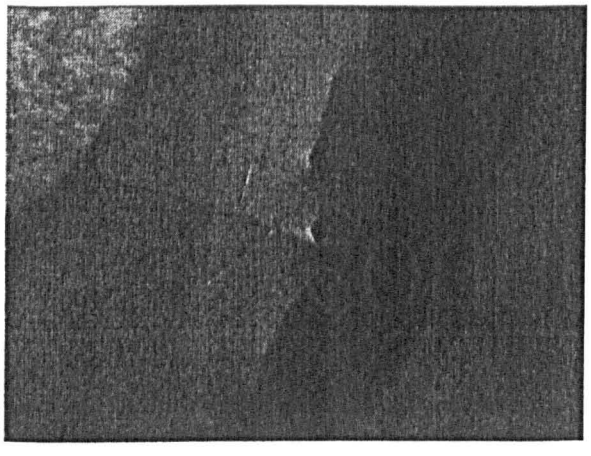


Figure 3-8: Small needle like crystals of the Phi class GST proposed to be used in the transport of anthocyanin from the cytosol to the vacuole.

3.3.2 Crystallisation of Tau Class GSTs

3.3.2.1 ZmGSTU1

Optimisation of initial conditions found by sparse matrix screens for the GST enzyme *Zm*GSTU1 determined that the largest crystals could be grown using a precipitant solution of PEG 8000 (19-24%), 0.2 M NaCl in 0.1 M MES buffer (pH 6.0-6.6). Unfortunately these small needle like crystals could not be grown to a sufficient size to be used for diffraction.

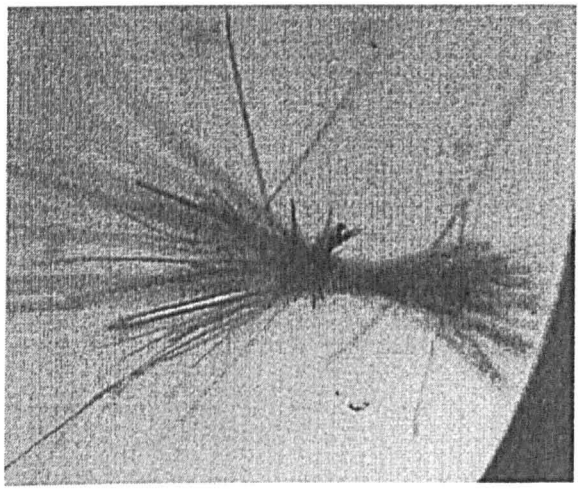


Figure 3-9: Needle like crystals of *Zm*GSTU1

3.3.2.2 ZmGSTU1-2

Conditions for crystallisation of the maize isoenzyme *Zm*GSTU1-2, the only heterodimeric GST included in the crystallisation trials, were investigated using a number of sparse matrix screens (Magic 50, Foot Print Screen and M screen (Jancarik & Kim, 1991; Stura et al., 1991; McPherson, 1990)). Different protein concentrations of 5mg/ml and 10mg/ml were also used in these initial screens, but no crystals were observed in any of the conditions tried.

3.3.2.3 ZmGSTU2

Small crystals of the homodimeric Tau class GST, *Zm*GSTU2, were observed in a number of conditions of the M screen following sparse matrix screening (Table 3-1).

Condition number	**Precipitant concentration, buffer and ions present**	**Crystal morphology**
M48	20% PEG 8K, 0.1M HEPES pH 7.5	Fine Needles
M60	20% PEG 8k, 0.1M TRIS pH 8.7, 0.2M NaKPhos	Single Rod
M88	10% Me2KPEG, 0.1M MOPS pH 6.3, 0.2M $MgCl_2$	Fine Needles
M104	15%% Me2KPEG, 0.1M TRIS pH 8.5, 0.2M $MgCl_2$	Fine Needles
M112	20% PEG 8K, 0.1M MOPS pH 6.5, 0.2M $MgCl_2$	Fine Needles
M114	20% PEG 8K, 0.1M MOPS pH 6.5, 0.2M $CaCl_2$	Cluster of Needles

Table 3-1: Initial crystallisation conditions observed for *Zm*GSTU2

Optimisation of these initial conditions produced the largest crystals in a condition containing 12% Me2KPEG, 0.1 M MES (pH 6.5) and 0.2 M $CaCl_2$ using a protein concentration of 10mg/ml and temperature of 20°C. The long thin needles produced in this condition were however not large enough to be used to collect diffraction data.

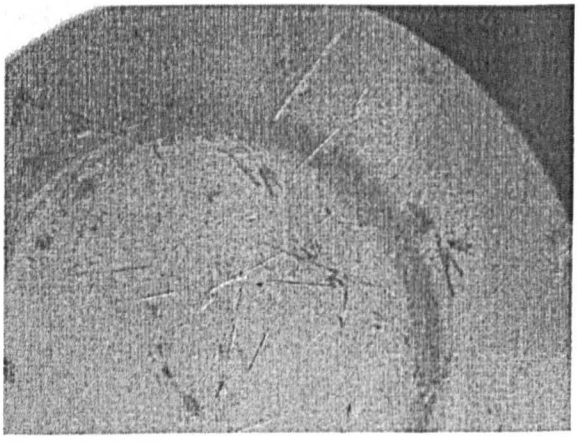


Figure 3-10: Crystals of the homodimer *Zm*GSTU2

3.3.2.4 Wheat J

Crystals of the wheat J enzyme were grown by equilibrating a mixture of 1 µl protein solution (10 mg/ml in 20 mM Tris pH 7.5) and 1 µl precipitant solution (1.1-1.5 M ammonium sulphate, 0.2M lithium sulphate, 0.1 M Tris (pH 7.5) containing 5 mM S-hexyl glutathione) against 0.8 ml of the precipitant solution. Initial crystals of 0.1 mm x 0.05 mm in size were tested at the microfocus beamline, ESRF, Grenoble and were found to diffract to 2.1Å, however the data collected proved very difficult to index (section 4.2.5). Additional optimisation of this condition allowed larger crystals to be grown, (figure 3-11) and more data to be collected on this isoform. Further problems were experienced in processing data gained from crystals of this type. A 3Å data set was eventually collected which was successfully processed in the hexagonal spacegroup $P6_122$. The phase problem for this data set was solved using molecular replacement.

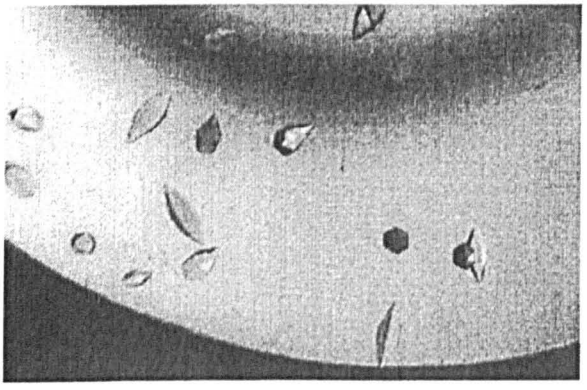


Figure 3-11: Crystals of Wheat J that appear to be hexagonal

3.3.3 Crystallisation of Zeta Class GSTs

3.3.3.1 AtGSTZ1

Using the human Zeta class sequence as a probe, plant sequence databases were searched to determine if this class of GST was widespread among plants. Following the discovery of a number of plant Zeta class GSTs, David Dixon cloned a Zeta class GST from *Arabidopsis thaliana* for use in crystallisation studies. Hexagonal crystals of *At*GSTZ1 with dimensions 0.4×0.4×0.5 mm (figure 3-12) were grown following optimisation of condition 18 from the Magic 50 sparse matrix screen (condition 18, 20% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate). These crystals, grown by

equilibrating a mixture of 2 µl protein solution (20 mg/ml in 20 mM Tris pH 7.5) and 2 µl precipitant solution (11% PEG 8000, 0.2 M magnesium acetate, 0.1 M sodium cacodylate pH 6.5) against a 0.8ml reservoir of the precipitant solution, were found to diffract to 1.65Å at station 7.2 SRS, Daresbury (section 4.2.4).

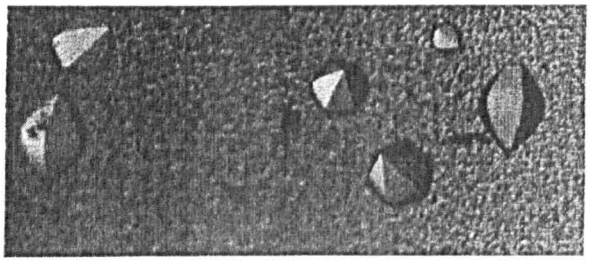


Figure 3-12:Hexagonal crystals of *At*GSTZ1

3.3.3.2 DCSR8

In addition to the Zeta class GST from *Arabidopsis thaliana*, David Dixon also supplied a clone of a Zeta class GST from *Dianthus caryophyllus* (carnation). Attempts to crystallise this GST were made using a range of sparse matrix screens, however no crystals were found to grow in any of the conditions tested. This GST was found to show significant levels of precipitation at a protein concentration of 5 mg/ml in more than 50% of the conditions tested, markedly contrasting the solubility of *At*GSTZ1.

3.3.4 Crystallisation of type Theta GSTs

3.3.4.1 Arabidopsis Theta Class GST ZmGSTT1

Thin plate like crystals of this GST were initially grown based on condition 90 of the M screen (condition 90, 30% Me2KPEG, 0.1 M MOPS pH 6.5). Optimisation of this condition enabled crystals large enough for diffraction purposes (figure 3-13) to be grown. These were produced by mixing 2 µl of protein solution (20 mg/ml in 20 mM Tris pH 7.5) with 2 µl of precipitant solution (16-22% Me2KPEG, 0.2 M ammonium sulphate, 0.1 M MOPS, pH 6.5) and equilibrating it against 0.8mls of the precipitant solution for 2-3 days.

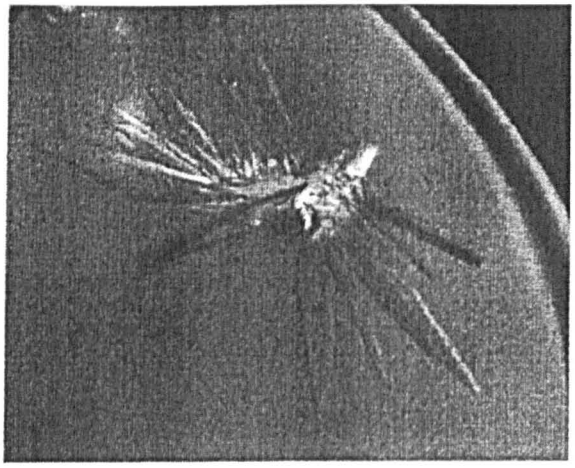


Figure 3-13: Crystals of *At*GSTT1 that were determined as growing in the orthorhombic spacegroup $P2_12_12_1$.

Chapter 4: Data Processing

4.1 Protein Crystallography

Since the X-ray diffraction pattern of pepsin was observed by J. D. Bernal and Dorothy Hodgkin (Bernal & Crowfoot, 1934), there has been a steady increase in the number of proteins whose structures have been determined. Although with the development of crystallography there has been an increase in automation and in the use of computers to facilitate the solving of structures, the underlying theory of structure determination remains relatively unchanged since these early days. A variety of excellent textbooks have dealt with the crystallographic theory of protein crystallography (Blundell & Johnson, 1976; Drenth, 1994; McRee, 1999) and so only the basic concepts are introduced in this text in order to explain the crystallographic terms used.

4.1.1 Relationship of diffraction to the crystal and the model

When atoms are placed in a beam of electromagnetic radiation like X-rays, the atom particles (neutrons, protons and electrons), begin to oscillate and become a source of radiation themselves, scattering the incident wave. Electrons within a molecule oscillate at the same frequency as the incident radiation and thus emit radiation of the same frequency as the incident wave. In a periodic arrangement of atoms (crystal) the scattered waves interfere with each other, cancelling each other out when they are out of phase and reinforcing each other when they are in phase (figure 4-1). This interference gives rise to a characteristic pattern of intensities dependent of the location of atoms within the molecule (figure 4-2).

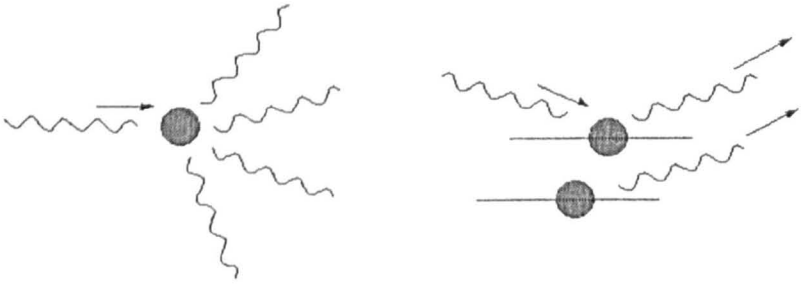


Figure 4-1: In a periodic arrangement of atoms the emitted radiation of waves becomes diffraction.

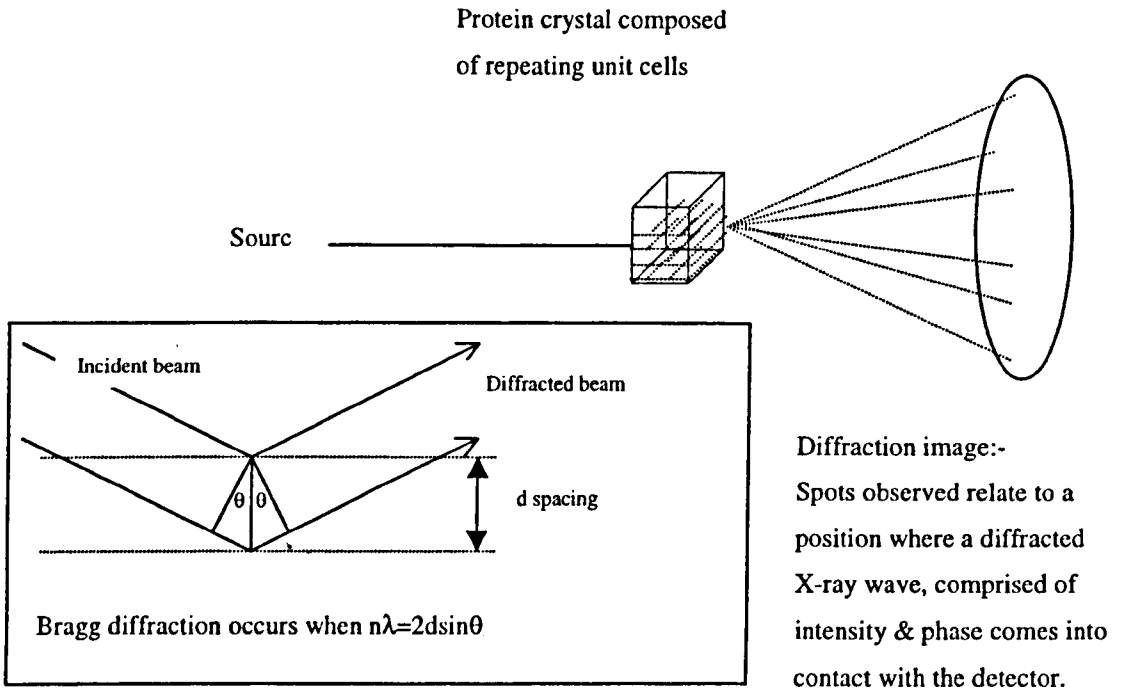


Figure 4-2: Generation of diffraction pattern from a protein crystal

The diffraction pattern observed in X-ray crystallography is the convolution of the molecular transform of the protein and the diffraction pattern of the crystal lattice (Blundell & Johnson, 1976). The spacing of the spots allows calculation of the unit cell dimensions, while the intensities of the spots depend on the molecular structure of the protein molecules within the unit cell.

The relationship between the diffraction pattern of a protein and the electron density, which describes its molecular structure, (equation 4-1) is a Fourier series of the structure factors F_{hkl}, where $|F_{hkl}|$ is a description of the amplitude and exp $i\alpha_{(hkl)}$ is a description of the phase of a diffracted ray. In order to generate an electron density map each term of the diffracted ray must be gained. The amplitude of the structure factor F_{hkl} is proportional to the square of the intensity of a diffracted beam hkl, but the phase of F_{hkl} cannot be directly gained from the diffraction pattern, (the phase problem). Indirect methods such as; Isomorphous Replacement, Anomalous Scattering or Molecular Replacement are used to determine the phase of the diffracted X-ray.

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}^{\infty}\sum_{-\infty}^{\infty}\sum_{k}^{\infty}\sum_{-\infty}^{\infty}\sum_{l}^{\infty} F_{(hkl)} \exp -2\pi i(hx+ky+lz)$$

Equation 4-1: Electron density ρ(x,y,z) can be calculated from diffraction data using the relationship below. $F_{(hkl)}$ is the structure factor known for all reflections, hkl and it contains both amplitude and phase information. The equation describes a Fourier series for a three dimensional wave with frequencies h in the x direction, k in the y direction and l in the z direction.

The aim of data processing is to reduce the graphical diffraction images to an ordered list of reflection indices with properly measured and corrected intensities. This allows calculation of the amplitudes of the diffracted X-rays and these can then be used to solve the phase problem using indirect methods.

4.2 Data processing

4.2.1 Theory

Macromolecular data processing has developed considerably over recent years due to the increased use of synchrotron radiation and faster detectors which require rapid identification of a crystals space group to enable optimal data collection (Minor et al., 2000; Rossmann et al, 1999). In addition, the growing popularity of techniques such as Multiwavelength Anomalous Dispersion (MAD) has required the accurate processing of large quantities of data. A number of data processing programs have been developed, XDS, Mosflm, DENZO, (Kabsch, 1993; Leslie, 1992; Otwinowski & Minor, 1997) however during the course of this study the program DENZO (Otwinowski & Minor, 1997) has been used to process all the data collected. Generally all the programs perform the following major steps;

- Account for detector characteristics
- Determination and refinement of crystal orientation and unit cell parameters
- Prediction of reflection positions
- Reflection integration, profile fitting
- Scaling and refinement

All the data processing programs perform these steps in a relatively similar manner, however in this work they have been expanded upon in relation to the data processing program DENZO in particular.

4.2.1.1 Detector Characteristics

Currently diffraction data for proteins is typically measured using the oscillation technique and multi-wire proportional counters, image plate (IP) scanners or charge coupled detectors (CCD). In order to correctly index and integrate the diffraction images produced from these detectors the data processing program must be able to account for the characteristics of each detector.

4.2.1.2 Determination of crystal orientation and unit cell parameters

One of the major advances which has been made in data processing in recent years is the ability of data processing programs to autoindex diffraction patterns where the orientation of the reciprocal lattice is not known in relation to the X-ray beam and the detector. The novel algorithim which enables autoindexing in Denzo has not been published, however it is suggested in the manual that it utilises a Fast Fourier Transform (FFT) which allows the independent identification of h, k & l. The FFT was suggested by Bricogne (Bricogne, 1986) to be a powerful indexing tool, however excessive amounts of memory and time were found to be required for large unit cells (Campbell, 1997).

In other processing programs, autoindexing is typically performed by the calculation of vectors between adjacent lattice points (difference vectors) based on the selection of a few hundred bright, well-defined spots from one image (Kabsch, 1993).

In DENZO, following identification of the unit cell which indexes all observed spots, the program attempts to find the best cell for all fourteen bravais lattices and calculates the distortion required to fit this basic cell to a higher symmetry cell.

4.2.1.3 Prediction of reflection positions

Once the user has defined a lattice and space group the chosen cell dimensions are used to predict the expected reflection positions. The crystal orientation is refined throughout data processing to account for any small change in the crystal orientation.

4.2.1.4 Reflection integration and profile fitting

The integration of a reflection is one of the most important steps in data processing, as it generates the raw measurement of I (hkl) used in further steps of the structure determination process. To calculate the intensity of the spot, the detector background (present due to non-Bragg scattering of the sample or electronic events in the x-ray detector such as, dark current) has to be estimated and then subtracted from the reflection profile. To minimise the error in the background calculation DENZO applies a series of statistical tests to remove pixels in the background with unreasonable values or which overlap other spots.

In addition Denzo also uses two-dimensional profile fitting to obtain a more accurate estimate of the intensity in the integrated peak. By predicting the profile of a reflection based on the profiles of strong, well measured, local reflections a more accurate estimate of the integrated intensity can be made for weaker reflections. Profile fitting also provides relatively reliable intensities for overloaded reflections in which some pixels of the spot are greater that the detector threshold. Particular care must be taken that the spot positions are accurately predicted when profile fitting is applied to ensure noise is not added to the reduced data (Leslie 1987), or in the case of barely separated spots, adjacent peaks do not intrude upon areas used for background or peak measurement.

4.2.1.5 Scaling and refinement

The intensity of diffraction is governed by a number of factors as shown in equation 4.2.

$$I(hkl) = I_o \times \frac{\lambda^3}{\omega} \times \frac{V_{cr} \times L \times P \times A}{V^2} \times |F(hkl)|^2$$

Equation 4-2: The intensity diffracted by a crystal rotated with angular velocity ω through the reflection position is dependent on the incident beam intensity I_o, the scattering of the atoms it contains, $|Fhkl|^2$, the wavelength or radiation (λ), the volume of the unit cell *V* and the volume of the crystal *Vcr*. Additionally a number of other corrections L (lorentz), P (polarisation) and A (adsorbtion factor) are made to the raw diffraction intensities as explained in the text.

Equation 4-2 describes that the intensity of the diffracted spot is dependent on the incident beam intensity I_0 and the size of the crystal, with larger crystals generally producing stronger diffraction. However not all crystals of equal size diffract to the same extent, as other factors also influence the diffraction intensity. $|F_{(hkl)}|^2$ is a measure of the scattering by atoms in the unit cell, with the light atoms which comprise proteins (carbon, nitrogen, oxygen) causing protein crystals to diffract relatively weakly. In a perfect crystal diffraction spots would be observed as sharp points, however typically they show a small angular spread of around 0.25-0.5°. This is due to mosaicity, small imperfections in the crystal lattice as shown in figure 4-3. This increase in the width of the spot affects the intensity and therefore is taken into account by including V_{cr}.

Figure 4-3: Mosaicity is a measure of small imperfections in the crystal. These small imperfections in the crystal can be thought of as small blocks, which cause a breakdown in the reflection condition of infinite unit cells and cause a small angular spread in the diffraction spots

The volume of the unit cell (V) is also a factor that influences the diffraction intensity. If crystals of fixed size are considered, then those crystals with bigger unit cells diffract more weakly. This can be explained by observing that intensity varies with $f^2 / V^2 \times n$ (where f is the scattering factor and n is the number of molecules per unit cell). A doubling of the unit cell causes $n \rightarrow 2n$ and the intensity of diffraction to half.

The choice of wavelength (λ) for a diffraction experiment is often determined by either the source available or the need of a particular wavelength to optimise anomalous scattering for phase determination. However the wavelength also has a significant effect on the intensity of the diffracted beam. Strongest diffraction intensities are gained using longer wavelengths, however these also have the disadvantage of increased absorption by the crystal, resulting in increased radiation damage. The choice of CuKα (1.54Å) for home sources represents a good compromise between scattering and absorption.

Corrections to the raw intensities are also made during data processing for the effects of Lorentz (L), Polarisation (P) and Absorption (A). The Lorentz correction is a measure of the velocity of the Bragg spot through the Ewald Sphere. A polarisation correction is applied where the beam striking the crystal is non-polarised (does not apply to synchrotron radiation or x-rays collimated using a monochromator). This correction is applied, because on reflecting from the plane of the crystal unpolarised x-rays are split into distinct components, which effect the intensity of the scattered beam. Corrections for absorption are made to account for X-ray scattering due to the crystal. Further absorption effects occur due to the absorption of the X-ray by the air between the sample and the detector. and the sample holder which results in increased background.

The corrected diffraction intensities produced by DENZO for each data set were scaled using the program SCALEPACK (Otwinowski & Minor, 1997). This program provides a method of merging symmetry-related observations and a method by which outlying reflections can be statistically rejected. Scaling also provides a method by which global refinement of the unit cell dimensions,

mosaicity and crystal orientation can be performed and produces a series of statistics that allow assessment of the data quality. Traditionally R_{merge} is employed to provide some idea of data quality, but R_{merge} can be influenced by a number of factors. These factors include removal of weak data and decreasing the overall redundancy, both of which artificially decrease this value.

$$R_{merge} = \sum_h \sum_j \frac{|I(h) - I(h)j|}{\sum_h \sum_j I(h)j}$$

Equation 4-3: R_{merge} (Where I (h) is the mean intensity) is typically less than 10% for most protein structures.

Data was collected and processed for a number of GST enzymes included in the crystallisation study. Details of the data collected are provided and the methods used are provided in the rest of this chapter.

4.2.2 ZmGSTF1 ($P2_1$)

Thin plates of *Zm*GSTF1 apoenzyme were grown in a solution of 25% Me2KPEG, BTP pH 6.8-7.4 (figure 3-6). Following soaking of the crystal for ~30s in the mother liquor from the well, which was employed as a cryoprotectant, the crystal was cryocooled using an Oxford cryostreams cryosystem. Data to 2.8Å was collected at station 9.5 SRS, Daresbury and processing in Denzo (Otwinowski & Minor, 1997) indicated that the protein had crystallised in the monoclinic spacegroup $P2_1$, with unit cell dimensions a=52.94 b=75.46 c=106.98, $\alpha=\gamma=90.0°$, $\beta=100.04°$. Data collection statistics are shown in Table 4.1.

4.2.3 ZmGSTF1 (P432)

Based on work that demonstrated tetrapyrroles were potent inhibitors of GSTs in oat (*Avena sativa*) (Singh & Shaw, 1988), our collaborator David Dixon tested tetrapyrrole inhibition of CDNB conjugation in maize GSTs. Enzyme assays using these maize GSTs and tetrapyrrole molecules demonstrated that chlorophyllin was able to inhibit the Phi class enzyme *Zm*GSTF1 by 36%. Further studies using the Tau class enzyme *Zm*GSTU1 enzyme (inhibited by

66% by chlorophyllin) suggested that tetrapyrroles inhibited this GST noncompetitively. This noncompetitive binding suggests that the tetrapyrrole molecules bind GSTs in a region distinct from the active site, which is able to promote a decrease in GSH conjugating activity, as a result of conformational change in the enzyme (Dixon, 1998). In order to determine where the tetrapyrrole binds to the GST and to determine any structural effects of tetrapyrrole binding, crystallisation trials were set up which included chlorophyllin in the crystallisation drop.

In crystallisation trials of *Zm*GSTF1, using similar conditions as described above but including 5 mM chlorophyllin, crystals of different morphology to those apoenzyme crystals of spacegroup $P2_1$ were grown (figure 3.5). In addition these crystals with different morphology also appeared to be light green in colour which suggested that the green pigment chlorphyllin had bound. These crystals were frozen as before using 30% Me2KPEG as a cryoprotectant and data collected to 2.9Å at station BM30, Grenoble, ESRF. Processing in DENZO suggested the spacegroup was primitive cubic and following scaling in SCALEPACK it became clear that the spacegroup was P432 with unit cell dimensions of a=116.5. Data collection statistics are shown in Table 4.1.

Data collection	Apoenzyme *Zm*GSTF1	Chlorophyllin *Zm*GSTF1
Source	DL9.5	ESRF BM30
Detector	Mar 300	Mar 345
Wavelength (Å)	1.192	1.2
Space Group	$P2_1$	P432
Resolution (Å)	2.8	2.9
Total number of reflections	21345	137408
Completeness (%)	99.4	99.7
Multiplicity[a]	4-5	13-19
Solvent content (%)	36	49
No. molecules in a.u.	4	1
R_{merge} (%)[b]	5.9	8.8
Wilson B	34.2	32.1

[a] Redundancy of data, defined as the ratio of the number of measured and the number of unique reflections.

[b] $R_{merge} = \Sigma_h\Sigma_i | I(h,i) - \langle I(h)\rangle | / \Sigma_h I(h,i)$, where $I(h,i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of h for all i measurements of h; is the summation over all measurements.

Table 4-1: Data collection statistics for *Zm*GSTF1 data

4.2.4 *Arabidopsis thaliana* Zeta class GST

Hexagonal crystals of *At*GSTZ1 were grown by equilibrating a mixture of 2µl protein solution (20mg/ml in 20mM Tris pH 7.5) and 2µl precipitant solution (11% PEG 8000, 0.2M magnesium acetate, 0.1M sodium cacodylate pH 6.5) against a 0.8ml reservoir of the precipitant solution. The crystals were loop-mounted in a drop of artificial mother liquor containing up to 25% (v / v) glycerol and flash frozen at 100K using an Oxford cryostream cryosystem for all data collections. The first native dataset to 2.2Å was obtained at BM30 at the ESRF, Grenoble with a second higher resolution dataset to 1.65Å collected at station 7.2, SRS, Daresbury. The X-ray reflection data for the heavy atom derivative was collected in house using a Nonius FR591 rotating anode with a Macscience DIP2000 image plate (CuKα - λ=1.54Å). All the diffraction data were processed using DENZO and scaled using SCALEPACK (Otwinowski & Minor, 1993). The crystals were shown to belong to a hexagonal crystal system with cell constants $a = b = 75.54$Å, c = 140.76 Å. Analysis of the systematic absences in the data revealed absences at l=6n along the (00l) axes, consistent

with the space group $P6_522$ or its enantiomorph $P6_122$. The *Vm* value (Matthews, 1968) was calculated to be 2.43 $Å^3$ / dalton and the solvent content estimated to be 43%, consistent with a monomer in the asymmetric unit. Table 4.2 summarises the data collection statistics.

Data collection	Native 1	Native 2	$Hg(CN)_2$
Source	ESRF BM 30	SRS 7.2	Rotating Anode
Detector	Mar 345	Mar 345	DIP2000
Wavelength (Å)	1.2	1.488	1.54
Space Group	$P6_522$	$P6_522$	$P6_522$
Resolution (Å)	2.2	1.65	3.0
Total number of reflections	95908	150341	54765
Completeness (%)	95.3	94.7	89.9
Multiplicity[a]	5-6	5-6	7-8
Solvent content (%)	43	43	43
No. of molecules per a.u.	1	1	1
R_{merge} (%)[b]	6.2	4.0	5.7
Wilson B	34.2	32.1	-

[a]Redundancy of data, defined as the ratio of the number of measured and the number of unique reflections.

[b]$R_{merge} = \Sigma_h\Sigma_i | I(h,i) - \langle I(h)\rangle | / \Sigma_h I(h,i)$, where $I(h,i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of h for all i measurements of h; is the summation over all measurements.

Table 4-2: Data collection statistics for *At*GSTZ1 data

4.2.5 Wheat J

Hexagonal crystals of wheat J were grown by equilibrating a mixture of 1 µl protein solution (10 mg/ml in 20 mM Tris pH 7.5) and 1 µl precipitant solution (1.1-1.5 M ammonium sulphate solution, 0.2 M lithium sulphate, 0.1 M Tris pH 7.5) against a 0.8 ml reservoir of the precipitant solution. Data was collected at 100 K using crystals which had been loop-mounted following soaking in a drop of artificial mother liquor containing 20% (v / v) glycerol. Initial data collection at station 7.2, SRS Daresbury yielded diffraction to only 4.0Å, however this resolution was significantly improved upon using similar crystals at the microfocus beamline ID13, ESRF, Grenoble. Unfortunately the data collected from this beamline was very difficult to process and the diffraction spots did not lie on the predicted positions for spacegroups of the hexagonal, trigonal, c

orthorhombic or monoclinic lattice systems for a significant proportion of the data. A second data set was collected at station 7.2 using bigger crystals, which had been soaked in mercury cyanide for 1min. These were found to diffract to 3Å, but again this data could not be processed due to similar problems as experienced for the ID13 data. A native data set that could be processed was finally collected on station X31 in Hamburg. This data extended to 3Å and was found to both index and scale in the hexagonal crystal system with cell constants $a = b = 87.98$Å, c = 146.77Å. Analysis of the systematic absences in the data revealed absences at l=6n along the (00l) axes, consistent with the space group $P6_122$ or its enantiomorph $P6_522$. The *Vm* value (Matthews, 1968) was calculated to be 3.37Å^3 / Dalton and the solvent content estimated to be 63%, consistent with 1 molecule in the asymmetric unit (Matthews, 1968). Table 4.3 summarises the data collection statistics.

Data collection	DESY X31	ID13	ID13 data indexed in $C222_1$
Detector	Mar 345	CCD	CCD
Wavelength (Å)	1.1	0.782	0.782
Space Group	$P6_122$	$P6_122$	$C222_1$
Resolution (Å)	2.9	2.3	2.3
Total number of reflections	127875	268134	315731
Completeness (%)	99.4	99.9	96.9
Multiplicity	13-19	9-12	5-6
Solvent content (%)	63	63	61
No. of molecules per a.u.	1	1	3
R_{merge} (%)[b]	26.9	11.1	9.5
Wilson B	43	25	

[a]Redundacy of data, defined as the ratio of the number of measured and the number of unique reflections.

[b]$R_{merge} = \Sigma_h\Sigma_i | I(h,i) - \langle I(h)\rangle | / \Sigma_h I(h,i)$, where $I(h,i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of h for all i measurements of h; is the summation over all measurements.

Table 4-3: Data collection statistics for Wheat J data

4.2.6 *Arabidopsis thaliania* Type 4

Crystals of *At*GSTT1 were grown by equilibrating a mixture of 2µl protein solution (20mg/ml in 20mM Tris pH 7.5) and 2µl precipitant solution (16-22% Me2KPEG, 0.2M ammonium sulphate, 0.1M MOPS pH 6.5) against a 0.8ml reservoir of the precipitant solution. Using loop-mounted crystals which had been soaked in a drop of artificial mother liquor containing 20% (v / v) glycerol and flash frozen using an Oxford cryostream cryosystem, data were collected at 100K. The first data set of *At*GSTT1 was collected at station BW7B, DESY, Hamburg. Although in some orientations this data extended to 2.1Å the data at this resolution was very weak and incomplete, however diffraction to 2.3Å was both complete and > 50% of the data had a ratio of intensity to the error of the intensity (I/σ) greater than two. This data was indexed in the primitive orthorhombic spacegroups $P2_12_12_1$ or $P2_12_12$. The final determination of the correct spacegroup is discussed in chapter 7. The unit cell dimensions were determined to be a=50.70 b=132.19 c=157.38. The *Vm* value (Matthews, 1968) was calculated to be 2.37Å^3 / Dalton and the solvent content estimated to be 47.5%, consistent with four molecules in the asymmetric unit. Molecular replacement using AmoRe (Navaza, 1994) was attempted with various models for this data, but no solution was gained. A second data set was collected on beamline X31, DESY, Hamburg also indexed and scaled in the spacegroup $P2_12_12_1$ and $P2_12_12$ with similar cell dimensions of a=50.55 b=131.79 c=157.11, this lower resolution data was used successfully to find a molecular replacement solution. Table 4-4 summarises the data collection statistics of the *At*GSTT1 data.

Data collection	Native1	Native2
Source	BW7A	DESY X31
Detector	Mar 345	Mar 345
Wavelength (Å)	0.8424	1.1
Space Group	$P2_12_12_1$	$P2_12_12_1$
Resolution (Å)	2.3	2.6
Total number of reflections	971097	264175
Completeness (%)	89.7.0	99.0
Average Multiplicity	5-6	7-8
Solvent content (%)	41%	41%
No. of molecules per a.u.	4	4
R_{merge} (%)[b]	8.3	12.3
Wilson B	28.5	51.8

[a]Redundacy of data, defined as the ratio of the number of measured and the number of unique reflections.

[b]$R_{merge} = \Sigma_h\Sigma_i |I(h,i) - \langle I(h)\rangle| / \Sigma_h I(h,i)$, where $I(h,i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of h for all i measurements of h; is the summation over all measurements.

Table 4-4: Data collection statistics for *At*GSTT1 data

4.3 Intensities to electron density

The resulting merged intensity data (sca) files from SCALEPACK were converted to mtz format using scalepack2mtz [CCP4, 1994]. After calculation of each cells respective solvent content and the number of molecules present in the unit cell (Matthews, 1968) the intensities were converted to amplitudes using Truncate [CCP4, 1994]. The overall B factor was calculated using Wilson statistics and twinning of the data was checked for by examining the distribution of centric and acentric reflections in the cumulative intensity distributions generated by Truncate. Solution of the phase problem for each structure is detailed within each respective section.

Chapter 5: Structure solution of ZmGSTF1

*Zm*GSTF1 was one of the first GSTs to be selected for use in crystallisation trials at Glasgow, because it was found to be the major constitutively expressed GST in both roots and shoots of maize (Holt et al., 1995; Mozer et al, 1983), and it showed activity against a broad range of xenobiotics. In particular *Zm*GSTF1 had significantly higher activity toward atrazine than other GST isoenzymes isolated from maize (Dixon et al., 1997). At the outset of the project GST structures from the mammalian classes Alpha, Mu, Pi and Theta and one plant GST from *Arabidopsis thaliana* had been described. It was hoped that one of these structures could be used as a homologous probe in the molecular replacement of *Zm*GSTF1 to provide initial phases.

5.1 Molecular replacement

Molecular replacement methods were first developed by Rossmann and Blow (Rossmann et al, 1962) to exploit the presence of non-crystallographic symmetry allowing phase information to be obtained and thus reduce phase uncertainties. However, as the number of structures in the Protein Data Bank has grown, molecular replacement has been increasingly used to gain initial phases for proteins whose structures are homologous to those already solved. In order to gain phase information for homologous proteins molecular replacement techniques aim to find the transformation of a known protein structure from its crystalline arrangement to that of the crystal of the unknown structure. Correct orientation and positioning of the probe in the new unit cell is performed using the Patterson function (equation 5.1) which does not require knowledge of phases.

$$P_{(u,v,w)} = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}\right|^2 \cos 2\pi(hu+kv+lw)$$

Equation 5-1: Patterson function does not require knowledge of phases. P (u,v,w) is a contour map in which the peaks correspond to the vectors between atoms. Like the electron density function it is a Fourier series of simple sine and cosine functions, however in this case Fhkl contains only amplitudes and not phase information

The Patterson function produces a map as described in figure 5-1 in which the vectors of the Patterson map can be separated into two forms;

- Self-Patterson vectors - These are short intramolecular vectors between atoms of the same molecule.
- Cross-Patterson vectors - These vectors are generally longer than self-vectors and arise between intermolecular atoms.

To determine both the orientation and the position of the molecule simultaneously would require a six dimensional search which until recently was not computationally feasible (Sheriff et al, 1999) and even with present computers is very slow. Splitting the search into two three dimensional searches to determine first the orientation then the position of the molecule (rotational then translation search) the calculation is more tractable.

5.1.1 Rotation function

Although many software packages now use the more elegant Crowther fast rotation function (Crowther, 1972) the original Rossmann and Blow method will be used to discuss the principles of molecular replacement. Using the Patterson function a vector map can be calculated which describes the relationship between each atom to every other atom present in the unit cell. To increase the chances of a strong rotation solution being found, only the intramolecular vectors (chosen by defining a maximum Patterson radius) are used in the rotation function as these are independent of relationships between molecules in the unit cell. The rotation function, which can be performed in real space or reciprocal space, is monitored by computing the product of the probe Patterson and the Patterson of the unknown structure. When the correct orientation is found the peaks of both Patterson maps will be superimposed and the product will be large (figure 5-1).

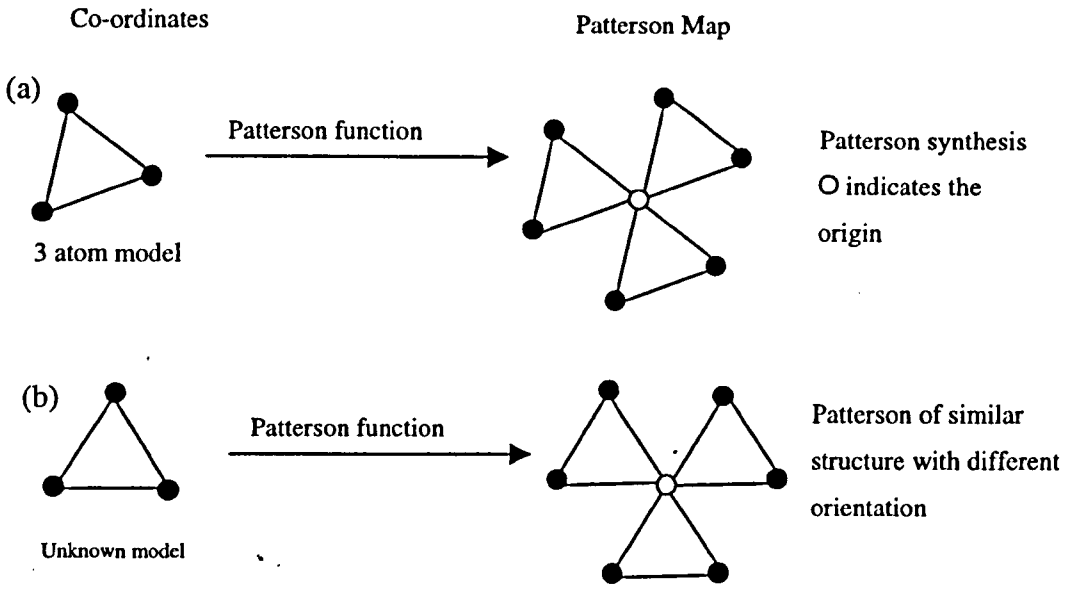


Figure 5-1: A Patterson map may be calculated by placing each atom at the origin and drawing vectors from this atom to all the other atoms of the unit cell. The magnitude of the vector is proportional to the product of the atomic numbers of the atoms at each end of the vector. Application of the corresponding rotation necessary to enable Patterson map (b) to match that of (a) onto the atomic co-ordinates of the known molecule determines the orientation of the unknown molecule within the unit cell.

5.1.2 Translation methods

Following the rotation function to correctly orientate the molecule, its position in the unit cell must be found. Two methods have been used to position the molecule in the unit cell. The first method, the translation function, is similar to the rotation function, however in this case, the product maximum of intermolecular vectors indicates a correct solution. The second method, a translational search, is a trial and error based search in which the structure factor amplitudes of the probe are calculated at various positions within the asymmetric unit and compared to those of the unknown structure. The translational search is monitored using both R-factor (equation 5-2) and a correlation coefficient. The correlation coefficient is a more robust monitor as it does not suffer from scaling errors.

$$R = \frac{\sum \| F_{obs} | - | F_{calc} \|}{\sum | F_{obs} |}$$

Equation 5-2: Crystallographic R-factor examines the agreement between the observed structure factors (Fobs) and the calculated structure factors (Fcalc). If the observed and calculated amplitudes are similar then the sum of the differences when compared to the sum of observed structure factors will be small.

5.1.3 AMoRe

The program AMoRe (Navaza, 1994) was used to perform molecular replacement for *Zm*GSTF1. This program shows a high degree of automation and allows many top potential orientation and translation solutions to be explored. This enables detection of solutions not among the first peaks of the rotation and translation functions. Improved algorithms allowing rapid calculation of structure factors and a fast least squares minimisation rigid body refinement of the top translation hits allows solutions to be found in ~20-40 minutes (Silcon Graphics, Inc. O_2), with potential solutions scored by both R factor and correlation coefficient.

5.2 Solution of phases of ZmGSTF1 in $P2_1$

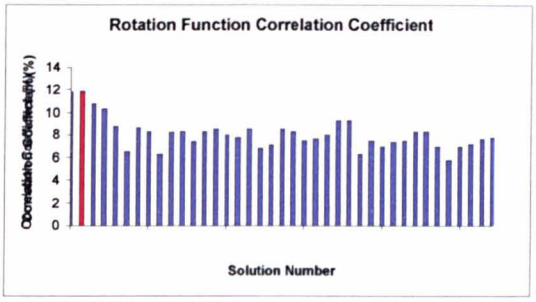
Following the collection of a 2.8Å data set for *Zm*GSTF1 (section 4.2.2) its sequence was compared with sequences of the GST structures in the Protein Data Bank to indicate which structures might be most similar to *Zm*GSTF1 for use as homologous probes in molecular replacement. The most similar sequence was that of the *Arabidopsis thaliana* structure (PDB entry-1GNW) which shared 35% sequence identity with *Zm*GSTF1 (figure 5.2). To improve the *A. thaliana* model for use as a probe to search for a molecular replacement solution in AMoRe, side chains not conserved between *A. thaliana* and *Zm*GSTF1 were trimmed back to the Cβ atom.

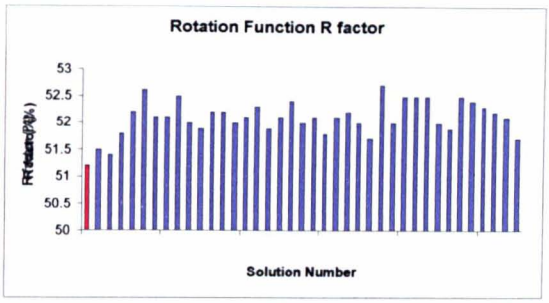
```
  ZmGSTF1 APMKLYGAVM SWNLTRCATA LEEAGSDYEI VPINFATAEH KSPEHLVRNP
     1GNW AGIKVFGHPA SIATRRVLIA LHEKNLDFEL VHVELKDGEH KKEPFLSRNP
Consensus A  K  G    S    R   A L E   D E  V       EH K    L RNP

  ZmGSTF1 FGQVPALQDG DLYLFESRAI CKYAARKNKP E---LLREG-- NLEEAAMVD
     1GNW FGQVPAFEDG DLKLFESRAI TQYIAHRYEN QGTNLLQTDSK NISQYAIMA
Consensus FGQVPA  DG DL LFESRAI   Y A          LL      N    A

  ZmGSTF1 VWIEVEANQY TAALNPILFQ VLISPMLGGT TDQKVVDENL EKLKKVLEVY
     1GNW IGMQVEDHQF DPVASKLAFE QIFKSIYGLT TDEAVVAEEE AKLAKVLDVY
Consensus     VE  Q          F         G T TD  VV E    KL KVL VY

  ZmGSTF1 EARLTKCKYL AGDFLSLADL NHVSVTLCLF ATPYASVLDA YPHVKAWWSG
     1GNW EARLKEFKYL AGETFTLTDL HHIPAIQYLL GTPTKKLFTE RPRVNEWVAE
Consensus EARL   KYL AG    L DL  H      L   TP         P V  W

  ZmGSTF1  LMERPSVQKV AALMKPSA
     1GNW  ITKRPASEKV Q
Consensus     RP   KV
```

Figure 5-2: Alignment of the *Zm*GSTF1 sequence and the sequence of the probe structure 1GNW (Reinemer et al., 1996). Those residues in red are similar or identical. The sequence alignment was performed by Multialign (Corpet, 1988).

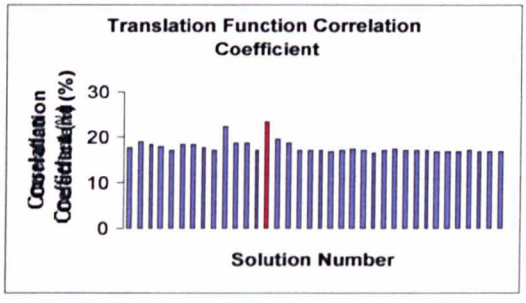
Using the trimmed *A. thaliana* monomer model as a probe for molecular replacement, a search for four molecules in the asymmetric unit (as determined by Matthews coefficient (Matthews, 1968)) was performed using a resolution range of 10Å-4Å in AMoRe. A clear solution for the *Zm*GSTF1 structure was found which had a higher correlation coefficient and lower R factor than the other orientations tested (figure 5.3).

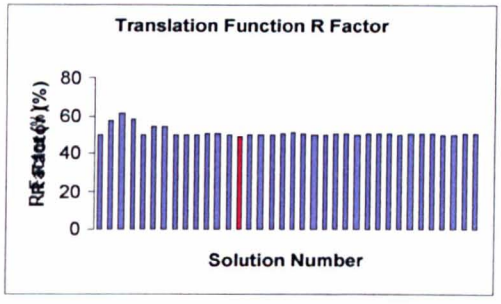
Original in colour



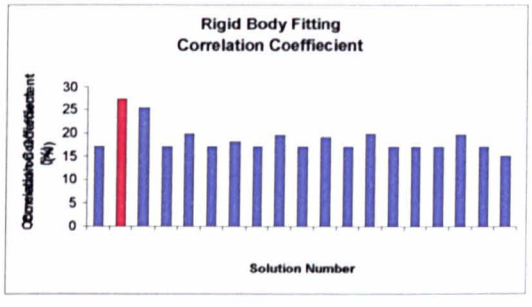


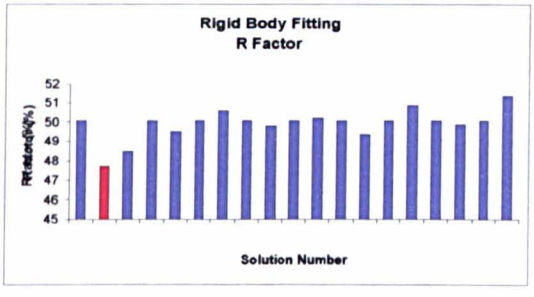
Ranking of Rotation Function Solutions for Apoenzyme *Zm*GSTF1. Top solution is shown in red. Top 50% of these solutions were used subsequently in the Translation function





Ranking of Translation Function Solutions for Apoenzyme *Zm*GSTF1. Top translation function is shown in red. Translation function solution then undergo rigid body refinement.





Ranking of Translation Function solutions following Rigid-Body fitting Solutions for Apoenzyme *Zm*GSTF1. Top solution is shown in red.

Figure 5-3: Solutions for *Zm*GSTF1 found by AMoRe

Using the matrices supplied by AmoRe, the probe was positioned in the new unit cell using LSKQAB (CCP4, 1994) and the model visually checked using SETOR. The molecule packed in the unit cell (figure 5.4) with no clashes between crystallographic or non-crystallographic symmetry equivalent molecules.

Original in colour

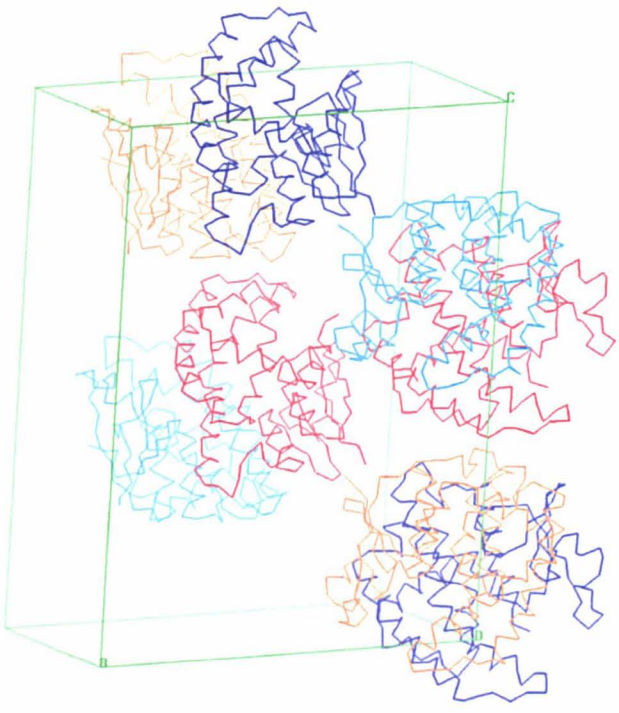


Figure 5-4: Crystal packing of molecules related to each other by non-crystallographic symmetry in the unit cell (Picture made using SETOR (Evans, 1993)

The presence of four molecules in the asymmetric unit allowed the use of fourfold molecular averaging to improve the electron density maps. The non-crystallographic symmetry matrices shown below for the molecules in relation to monomer A were calculated using LSQKAB. Matrices are shown in the O format.

Matrices A to A not shown

Matrices relating B to A are

-0.609893	-0.021294	0.792196
-0.018272	-0.998999	-0.040921
0.792266	-0.039434	0.608896
-0.149770	0.699037	0.141890

Matrices relating C to A are

-0.382810	0.018158	0.923654
-0.005582	-0.999830	0.017337
0.923813	0.001478	0.382842
-9.120162	0.575860	52.957893

Matrices relating D to A are

0.964548	0.006734	-0.263830
-0.021200	0.998420	-0.052027
0.263062	0.055776	0.963166
-8.958243	-0.115023	52.941040

Original in colour

Using these matrices NCSMASK was used to calculate the mask of the monomer and MAPROT was used to create an averaged map through interpolation of the map for the four molecules over monomer A [CCP4, 1994]. The resultant electron density showed considerable phase bias, however with the exception of the loop regions the observed electron density was fairly continuous and the $2F_o$-F_c map indicated residues suggested by the *Zm*GSTF1 sequence. Using the graphics program QUANTA (Molecular Structures, Inc.) an initial model of a single molecule comprising of 211 residues was built into the map. Co-ordinates for the second, third and fourth molecules were generated using non-crystallographic symmetry. Refinement was performed initially in the maximum likelihood refinement program REFMAC (Murshudov et al, 1996), however when an R factor of 30% was reached, REFMAC seemed unable to improve the structure any further. Positional and B factor refinement in the program XPLOR enabled further improvement in both $R_{(cryst)}$ and $R_{(free)}$. REFMAC containing an imported bulk solvent correction, in the form of partial structure factors from XPLOR (Brunger, 1992), was used to generate the electron density maps. Solvent molecules were added manually in the latter stages of refinement based on whether density was present and hydrogen bonds to the protein or ordered solvent molecules made chemical sense. After several cycles of rebuilding and refinement the R_{cryst} and R_{free} remained at 23.6 and 27.8 respectively. Refinement statistics are shown in Table 5.1.

Resolution Range (Å)	25.0-2.8
No of unique reflections in resolution range	21345
R_{Work} (%)[a]	23.6
R_{Free} (%)	27.8
Model	
Total number of amino acids	211
Total number of Solvent atoms (excl. H)	444
No. of Heteroatoms	0
Stereochemistry[b]	
Ramachandran quality, % in	
Most favoured regions	91.5
Allowed regions	8.5
R.M.S. deviation[c]	
Bond length (Å)	0.008
Bond angle (degrees)	1.6

[a] $R = \Sigma_{hkl}(|F_o - F_c|) / \Sigma|F_o|$.
[b] Calculated using the program PROCHECK (Laskowski et al., 1993)
[c] (Cruikshank, 1999)

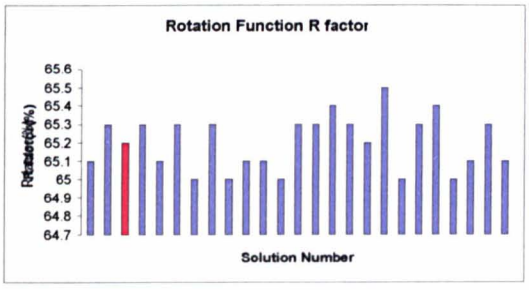
Table 5-1: Refinement statistics for *Zm*GSTF1 solved in $P2_1$

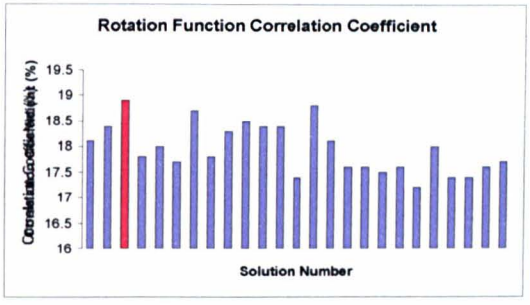
5.3 Solution and subsequent refinement of ZmGSTF1 in P432

As described in section 3.3.1.1 addition of chlorophyllin to the crystallisation mixture used to gain crystals of *Zm*GSTF1 in the space group $P2_1$, produced crystals of significantly different morphology. These crystals despite, growing to a size of around 0.3mm in all directions, only diffracted to 2.9Å resolution. Data processing and scaling using the HKL suite of programs found the crystals belonged to the cubic space group, P432 with cell constants a=116.50Å (section 4.2.3). Using AmoRe to determine a molecular replacement solution, the $P2_1$ apo *Zm*GSTF1 model was used to gain phases for the P432 data. As expected, a clear solution was found (figure 5.5) for the cubic data set.

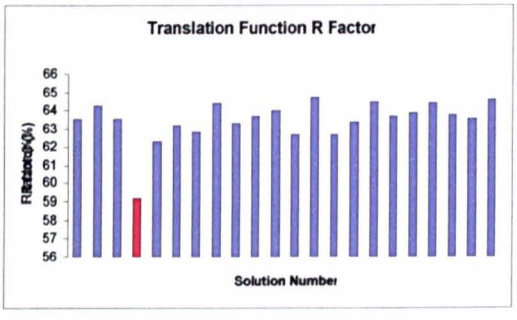
A monomer was found to be present in the asymmetric unit and this packed together with the symmetry related molecules with no clashes. Several cycles of rebuilding and refinement using QUANTA (Molecular Structure, Inc.) and CNS (Brunger et al, 1998) led to a final R free and R cryst of 29.9% and 23.4% respectively. Despite the different crystal morphology and light green colour of

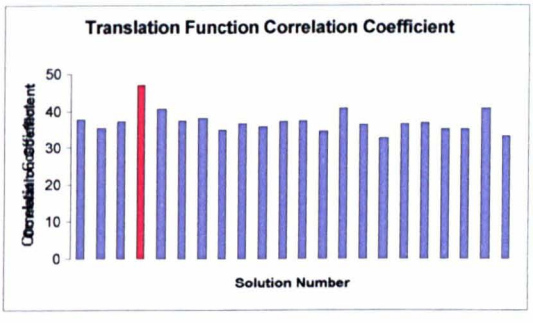
the crystals, no molecules of chlorophyllin could be observed in the electron density maps.



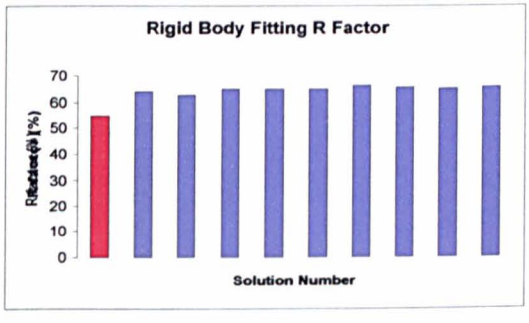


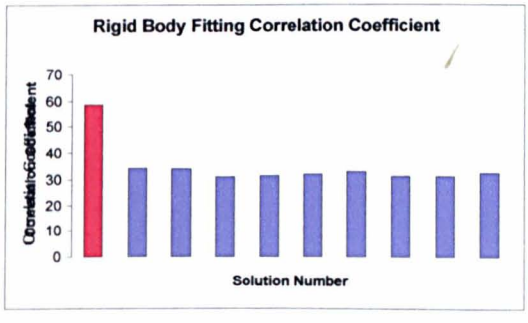
Ranking of Rotation Function Solutions for Apoenzyme P432 *Zm*GSTF1. Top solution is shown in red. Top 50% of these solutions were used subsequently in the Translation function





Ranking of Translation Function Solutions for P432 *Zm*GSTF1. Top solution shown in red





Ranking of translation function solutions following Rigid-Body fitting for P432 *Zm*GSTF1. Top solution shown in red.

Figure 5-5: Solutions for *Zm*GSTF1 P432 crystals as found by AmoRe (Top solution shown in red

Original in colour

The structure of *Zm*GSTF1 solved in the space group P432 is very similar to that of the model solved in $P2_1$, with a root mean square deviation of 0.6Å over 210 residues. Only a small number of valine and leucine side chains differ in conformation between the structures, due to selection of different rotamers during building of the model into the relatively low resolution electron density map. Refinement statistics for the P432 model are shown in Table 5-2.

Resolution Range (Å)	25.0-2.9
No of unique reflections	7136
R_{Work} (%)[a]	23.4
R_{Free} (%)	29.9
Model	
Total number of amino acids	211
No. of Solvent atoms (excl. H)	3
No. of Heteroatom	0
Stereochemistry[b]	
Ramachandran quality, % in	
Most favoured regions	79.3
Allowed regions	20.2
R.M.S. deviation[c]	
Bond length (Å)	0.008
Bond angle (degrees)	1.4

[a] $R = \Sigma_{hkl}(|F_o - F_c|) / \Sigma|F_o|$.

[b] Calculated using the program PROCHECK (Laskowski et al., 1993)

[c] (Cruikshank, 1999)

Table 5-2: Final refinement statistics for *Zm*GSTF1 in spacegroup P432

5.3.1 Structure of ZmGSTF1

The maize *Zm*GSTF1 enzyme exists as a dimer of identical subunits with each subunit consisting of two distinct domains (figure 5.6). The smaller N-terminal domain (residues 1-78) is formed by a four stranded β-sheet flanked by two helices on one side and a section of irregular loop structure on the other. A short nine residue linker (residues 79-87) joins the N-terminal domain to the larger and entirely helical C-terminal domain (residues 87-214).

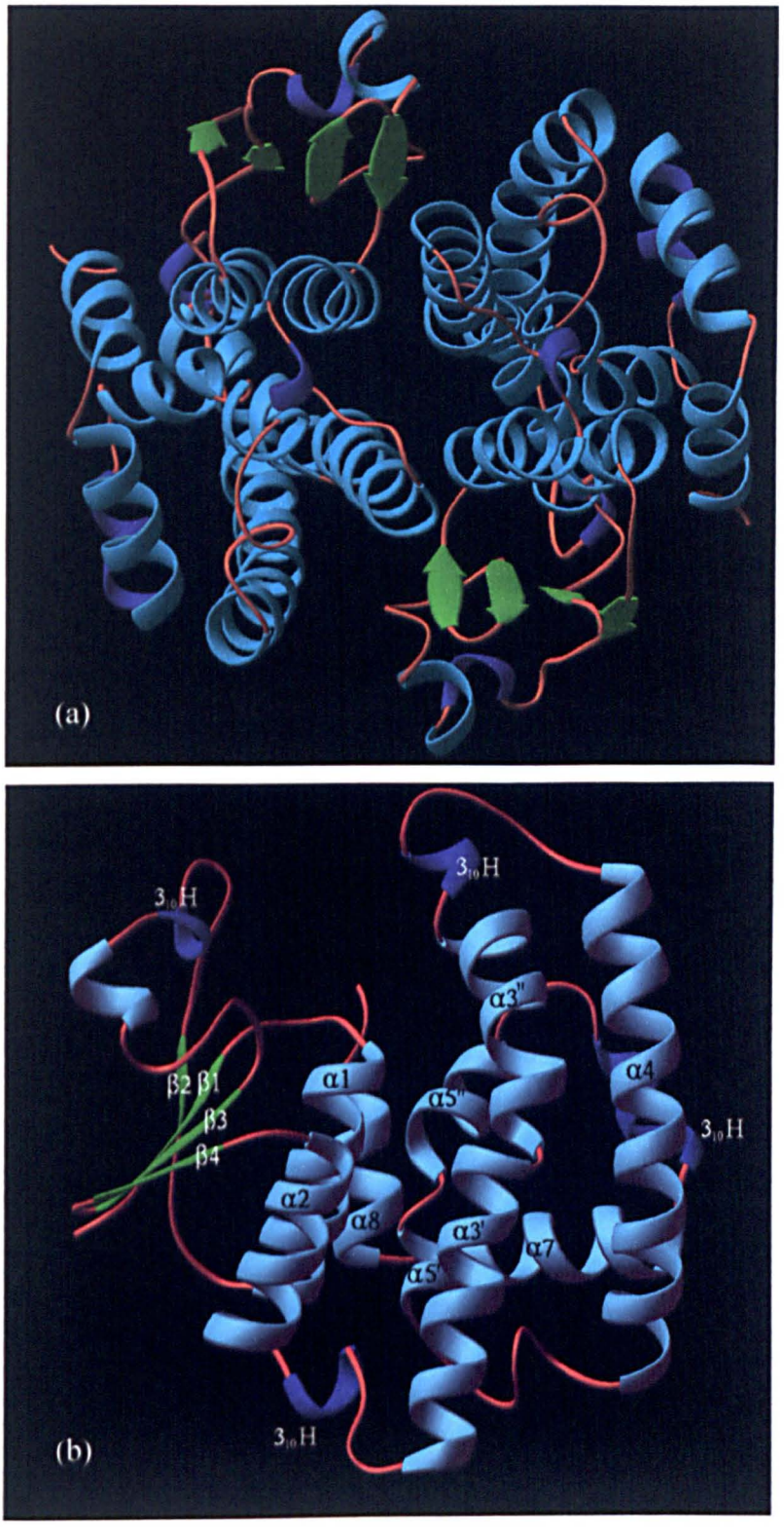


Figure 5-6: Diagram (a) shows the dimeric structure of the *Zm*GST1 enzyme, consisting of two identical monomers, while diagram (b) displays the monomer structure with annotated secondary structure elements.

Original in colour

5.3.1.1 N-terminal domain

The N-terminal region consists of a four stranded β-sheet typically observed in GSTs, with strand three (β3, 56-59) orientated in an antiparallel direction to the others (β1, 3-6; β2, 28-31; β4, 62-64). Strand β1 is connected to β2 via an α-helix (α1, 11-25), the top of which forms the base of the active site. Following this helix is an irregular loop region containing a 3_{10} helical element (residues 38-41) and an α-helix (residues 42-47), which pack against the β-sheet and link strand β2 to strand β3. Strand β3 and strand β4 form a type II β-turn before forming α2 (87-104) which is orientated similarly to α1. A short nine residue linker region containing a 3_{10} helix (80-83) covalently links the N-terminal domain to the C-terminal domain.

5.3.1.2 C-terminal domain

The C terminal domain is formed by six partially interrupted α-helices. The first helix, α3 (87-118) is partially interrupted into two regions (α3`, 87-104 and α3``, 104-118) before it forms a $3_{10}$ helix ($3_{10}$ H3, 119-122). Two glycine residues form the loop region after these helical elements enabling a sharp turn before helix α4 (127-152) which runs almost parallel to α3. A series of β-turns links α4 to the interrupted helix α5 (α5` 161-166 and α5`` 167-176), which in turn leads into the short 3_{10} helix (3_{10} H6 179-185). The penultimate helix α7 (186-198) runs almost perpendicular to α3 and α4 to form the back of the C-terminal domain along with α8 (199-200) which lies along the rear of the N and C-terminal domain interface.

5.3.1.3 Interactions between domains

The N and C-terminal domain interface is predominately formed by aliphatic and aromatic residues. These form multiple hydrophobic interactions. A particularly significant hydrophobic interaction is that within the β2-β3 loop element of the N-terminal domain which packs against residues Leu113, Leu117 and Ile118 of the C-terminal domain, and forms a hydrogen bond across the dimer interface with Asn110. Additionally two polar interactions between residues Glu23-Arg119 and Arg68-Glu101 are formed between the domains.

5.3.1.4 Interactions between subunits

Residues on helices α3 and α4 and sheets β3 and β4 form the 1397Å^2 dimer interface, which like the domain interface, is predominately mediated by hydrophobic interactions. As depicted in figure 5-7, Phe51 plays a crucial role in dimer formation interacting with three valine residues (Val100, Val141 and Val114) and a glutamine residue (Gln104). Gln104 also forms a polar interaction with Gln53 and Glu66 across the dimer interface as does Glu66 with Asn103.

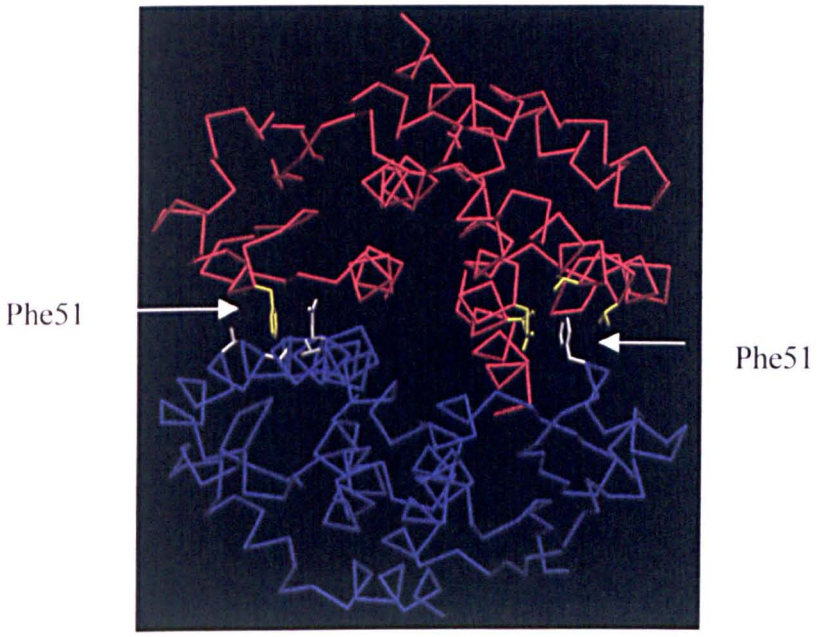


Figure 5-7: Phe51 interacts across the dimer interface with residues Val100, Val141 and Val114 in the opposite monomer subunit. These residues are conserved in the Phi class sequences

5.3.2 Comparison of primary sequence of ZmGSTF1

*Zm*GSTF1 belongs to the plant specific Phi class (Droog, 1997) and shares pairwise identities of 50% and 36% to the plant GSTs *Zm*GST3 and *At*GSTF2 whose structures have been determined (figure 5-8). Although the primary sequence of *Zm*GSTF1 shares only 20% sequence identity to the mammalian isoenzymes, *Zm*GSTF1 like the other structurally characterised members of the Phi class is topologically similar to the mammalian GSTs (Reinemer et al., 1996).

Original in colour

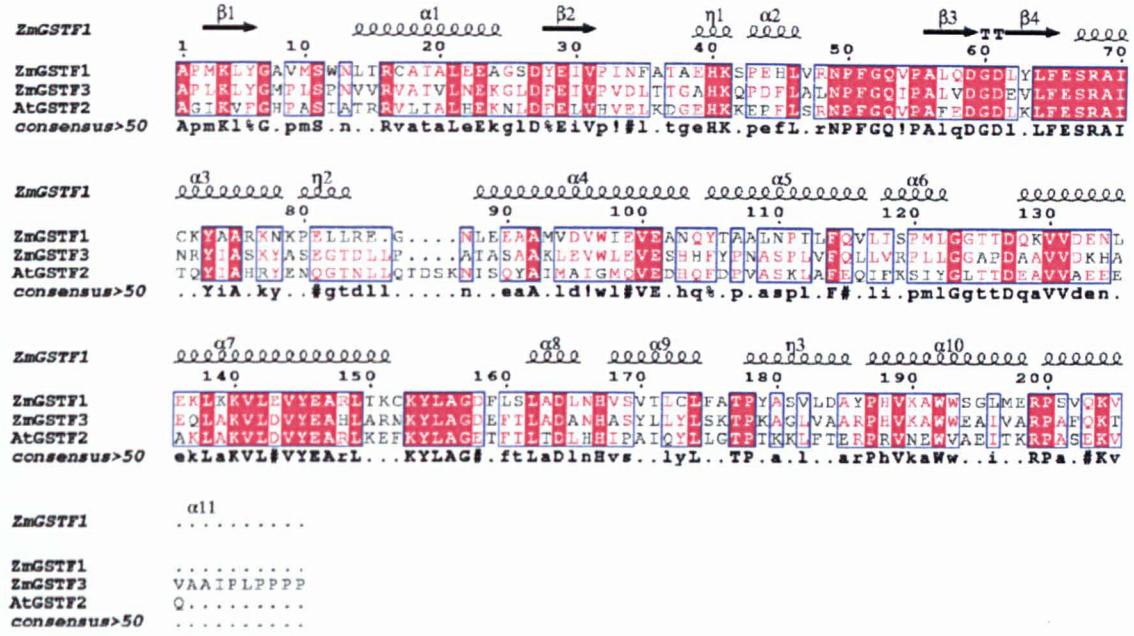


Figure 5-8: Sequence comparison of Phi class GST structures which have been determined. The secondary structure of *Zm*GSTF1 (PDB - 1AXD) is shown. Sequences from *Zm*GSTF1 (P12653), *Zm*GSTF3(AJ01095) and *At*GSTF2 (P46433) have been aligned using Multialign (Corpet, 1988). This figure was produced using ESPRIT (Gouet et al, 1999).

Many of the residues conserved between the three Phi class models are structurally important. The hydrophobic core of the C-terminal domain is formed by the conserved residues Val131, Leu138, Val141, Tyr145, Tyr154, Leu155, Ala156 and Trp192. Glu146 forms hydrogen bonds with the amide groups of Val189 and Pro188 promoting packing of helices α6 and α7 to α3 and α4. Asp16, buried in the hydrophobic core, forms hydrogen bonds with the amide groups of Leu155 and Ser161. Other structurally important residues are; the conserved charged residues Glu23 and Arg68 of the N-terminal domain which form salt bridges to Arg199 and Glu101 of the C-terminal domain, *cis*-Pro55 which forms the characteristic turn of the glutathione binding site and Gly123 that forms the sharp loop structure linking α3 to α4.

5.3.3 Comparison of tertiary structure of ZmGSTF1

During the structural characterisation of the apo *Zm*GSTF1 structure, two ligand bound structures (PDB codes 1AXD and 1BYE (Neuefeind et al., 1997a; Prade et al., 1998)) containing lactoylglutathione and a glutathione conjugate of atrazine were deposited in the Protein Data Bank. These two ligand bound

structures share close similarity to the apo structure and can be superimposed with a maximum root mean square deviation of 0.79Å over 206 residues. The loop region between helices α3 and α4 (residues 114-134) shows the greatest difference among the *Zm*GSTF1 structures with the lactoylglutathione bound structure differing by ~2Å in this region when superimposed on the apo structure. In the apo, P432 and two ligand bound structures this β2-β3 loop region shows higher B factors than the rest of the model. This suggests that this region shows some flexibility. Two glycine residues, one of which is strongly conserved among members of the Phi class GSTs, are found within this loop region and these may provide flexibility around the active site enabling larger aromatic molecules to be accommodated. It has been suggested that *Zm*GSTF1 may undergo an induced conformational change on substrate binding, with the loop region between β2-β3 proposed to act as a lid to the active site providing a hydrophobic environment (Ketterer et al, 1988; Neuefeind et al., 1997b). However, comparison of the apo enzyme and ligand bound *Zm*GSTF1 models suggests that no induced movement occurs in this region. The lack of induced conformational change of the β2-β3 loop region, contrasts the induced fit mechanism proposed for Pi class GSTs, on the basis of crystallographic temperature factors, spectroscopic and proteolysis data, (Lo Bello et al, 1993; Oakley et al, 1998). However it agrees with data for Alpha class (Cameron et al., 1995; Sinning et al., 1993) and the *S. japonicum* (McTigue et al., 1995) GSTs, which show no significant conformational changes on glutathione binding.

In addition to the differences observed in the α3-α4 loop region between the apo and ligand bound *Zm*GSTF1 structures, slightly different hydrogen bonding patterns between residues 42-47 and 19-122 are observed. The different hydrogen bonding results in the α-helical structure observed in the apo structures between β2-β3 forming a 3_{10} helix in the ligand bound structures and the 3_{10} helix described at the top of α3`` in the apo structures forming a third α-helical component in the ligand bound structures.

Comparison of *Zm*GSTF1 with the other plant GST structures, *Zm*GSTF3 (Neuefeind et al., 1997b) and *At*GSTF2 (Reinemer et al., 1996) indicates that

there is strong structural homology within the Phi class, with many of the secondary structure elements being of similar length and orientation among the three structures. *Zm*GSTF1 can be superimposed on the *Zm*GSTF3 and *At*GSTF2 structures with a root mean square deviation (r.m.s.d.) of 1.25Å over 196 residues and 1.02Å over 190 residues respectively. The greatest differences between these plant GSTs occurs within the N-terminal loop joining β2 to β3, the linker region and at the C-terminal tail as shown in figure 5-9.

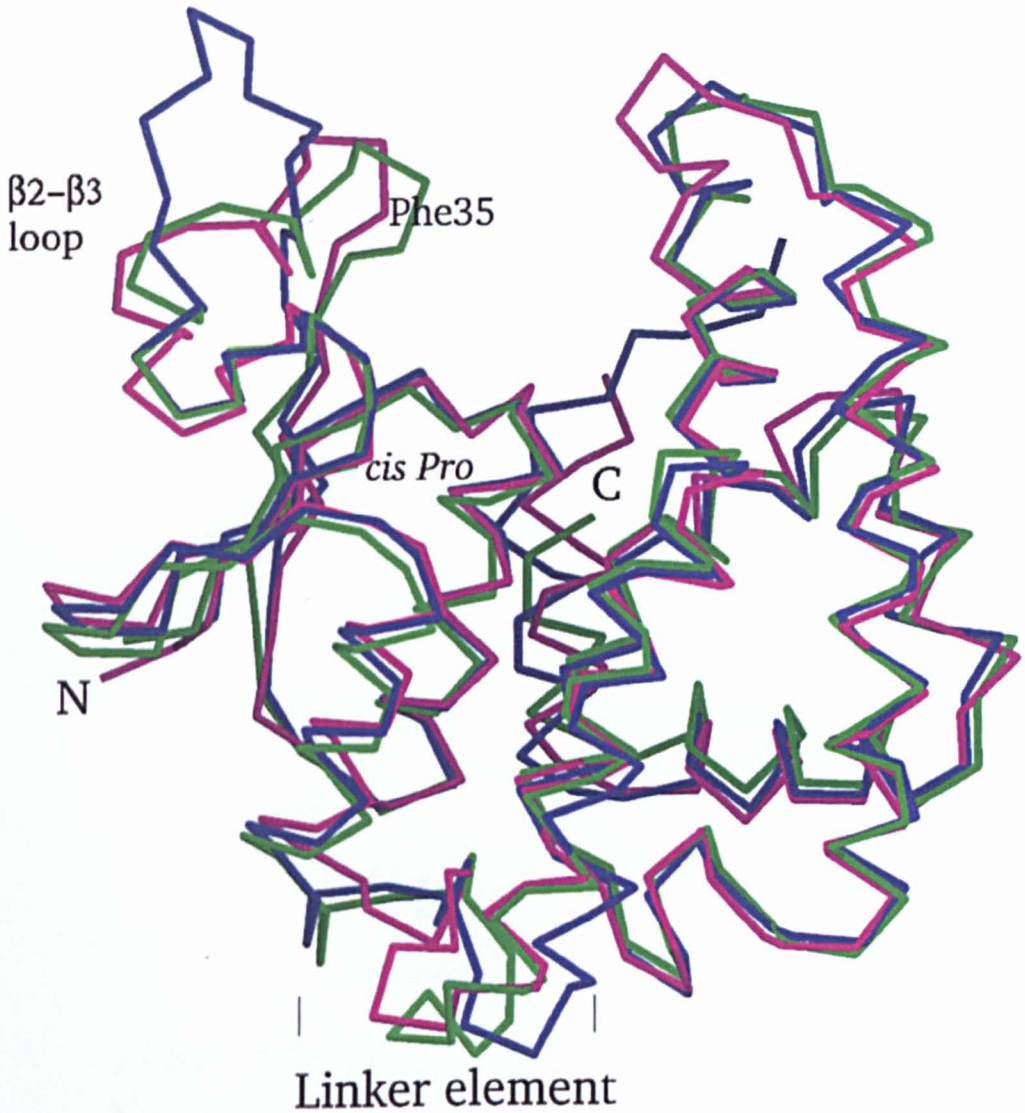


Figure 5-9; Superposition of three Phi class structures, *Zm*GSTF1, *Zm*GSTF3 and *At*GSTF2 determined to date. Structures displayed in Magenta (Apo *Zm*GSTF1), Blue (*Zm*GSTF3 - 1AW9, (Neuefeind et al., 1997b)) and Green (*At*GSTF2 -1GNW, (Reinemer et al., 1996). This figure was created in SETOR (Evans, 1993).

Original in colour

*Zm*GSTF1 contains two structural elements within the β2-β3 loop region, (residues 36-50) a short 3_{10} helix (residues 38-41) and an α-helix (α2, 42-47). This is unlike the irregular structure described for *Zm*GSTF3, but similar to the three short 3_{10} helices observed in the *At*GSTF2 structure (residues 35 to 37, 39-41 and 45-47).

Following this loop region the three Phi class structures become very similar again, the conserved *cis*-pro bend superimposing in each of the three structures. The linker region between the N and C-terminal domains of *Zm*GSTF1 is only nine residues in length and is therefore similar to the short linker region described in mammalian structures. This is in contrast to the longer linker regions observed in the *Zm*GSTF3 and *At*GSTF2 structures, which are 14 and 15 residues in length respectively. In addition unlike the other plant GST structures the *Zm*GSTF1 structure also contains a 3_{10} helix structural element within this linker element.

The C-terminal domain of all three plant GST structures is comprised of six partially interrupted amphipathic helices. Helix α3, formed by three shorter helices, is significantly longer in the Phi class structures than in the mammalian enzymes and is characteristically bent due to distortions in the main chain hydrogen bonding pattern. The loop region between α3-α4, which allows the helices to lie in the same plane, but run in opposite directions is similar in all three plant structures, and adopts a more extended structure in the plant enzymes than in the mammalian structures. This more extended structure causes it to protrude over the active site to a greater extent. The three Phi class structures C-terminal domains are generally similar to each other with the exception of the C-terminal helix. This helix in *Zm*GSTF3 when compared to *Zm*GSTF1 and *At*GSTF2 structures, shows additional residues that form an 11 residue coil following helix α8 (Neuefeind et al., 1997b).

5.3.4 Comparison of active sites of Phi class structures

The glutathione binding site (G-site) residues (Arg16, Ser11, Lys41, His40, Glu66, Ser67, Arg68) are structurally well conserved between the Phi class

structures. The catalytic residues Ser11, that stabilises the glutathione anion, (Armstrong, 1993; Dirr et al, 1994) and Glu66, thought to interact with the GSH-γ-glutamate, superimpose closely in the structures of the these three enzymes. In the paper describing the *Zm*GSTF3 model it is indicated that residues 33-44 of the β2-β3 loop could not be traced due to weak electron density. It was postulated that this region might be involved in glutathione binding through an induced fit mechanism, based on comparison of the ligand bound *Zm*GSTF1 and apo *Zm*GSTF3 structures (Neuefeind et al., 1997b). Comparison of the apo and ligand bound *Zm*GSTF1 structures does not agree with this hypothesis, as no structural differences in this region are observed between the apo and ligand bound structures. However this irregular loop region might be important in influencing substrate approach or the egress of product in a similar manner as proposed for the Mu loop structure which has also been shown to display high B factors (Armstrong, 1993).

Although several key residues are conservatively replaced in the hydrophobic site (H-site) of the Phi class GSTs, each plant GST exhibits unique structural features which determine the enzymes substrate specificity. Ligand bound complexes of *Zm*GSTF1 (Neuefeind et al., 1997a; Prade et al., 1998) and *At*GSTF2 (Prade et al., 1998; Reinemer et al., 1996) provide a description of the residues involved in binding the conjugated substrate (figure 5.11).

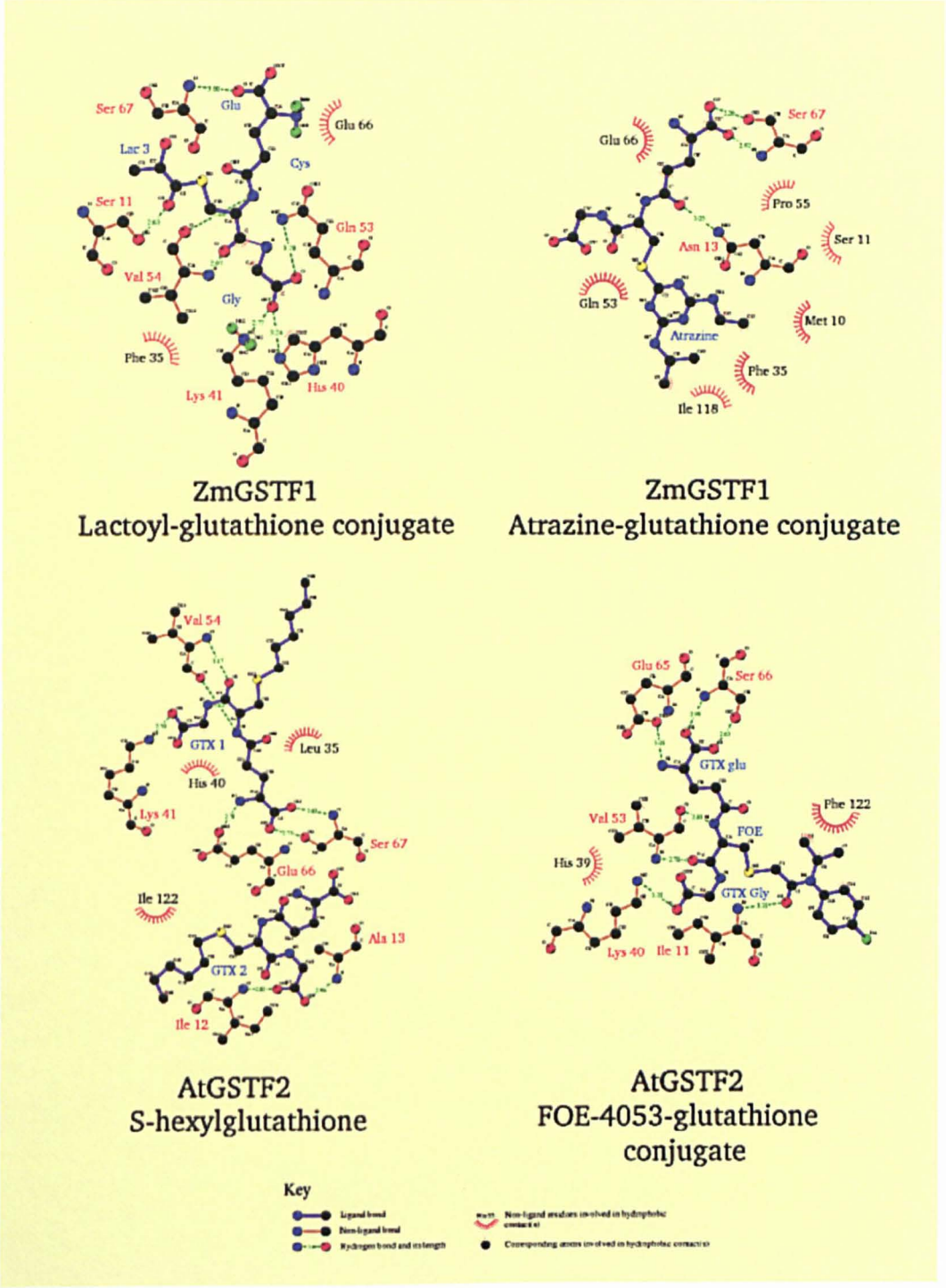


Figure 5-10: Ligplot schematics to describe interactions within the active site of ligand bound structures 1AXD (*Zm*GSTF1 bound to lactoyl-GSH), 1BYE (*Zm*GSTF1 bound to Atrazine-GSH), 1GNW (*At*GSTF2 bound to S-hexyl GSH) and 1BX9 (*At*GSTF2 bound to FOE-GSH) (Prade et al., 1998).

Analysis of the ligand bound *Zm*GSTF1 and *At*GSTF2 structures indicates that the C-terminal domain forms one face of the H-site with hydrophobic residues

Original in colour

being contributed from helices α3 and α5 (figure 5-11). The other face of the H-site of *Zm*GSTF1 is formed by residues of the N-terminal domain with Met10, Trp12 and Phe35, which also create a hydrophobic environment.

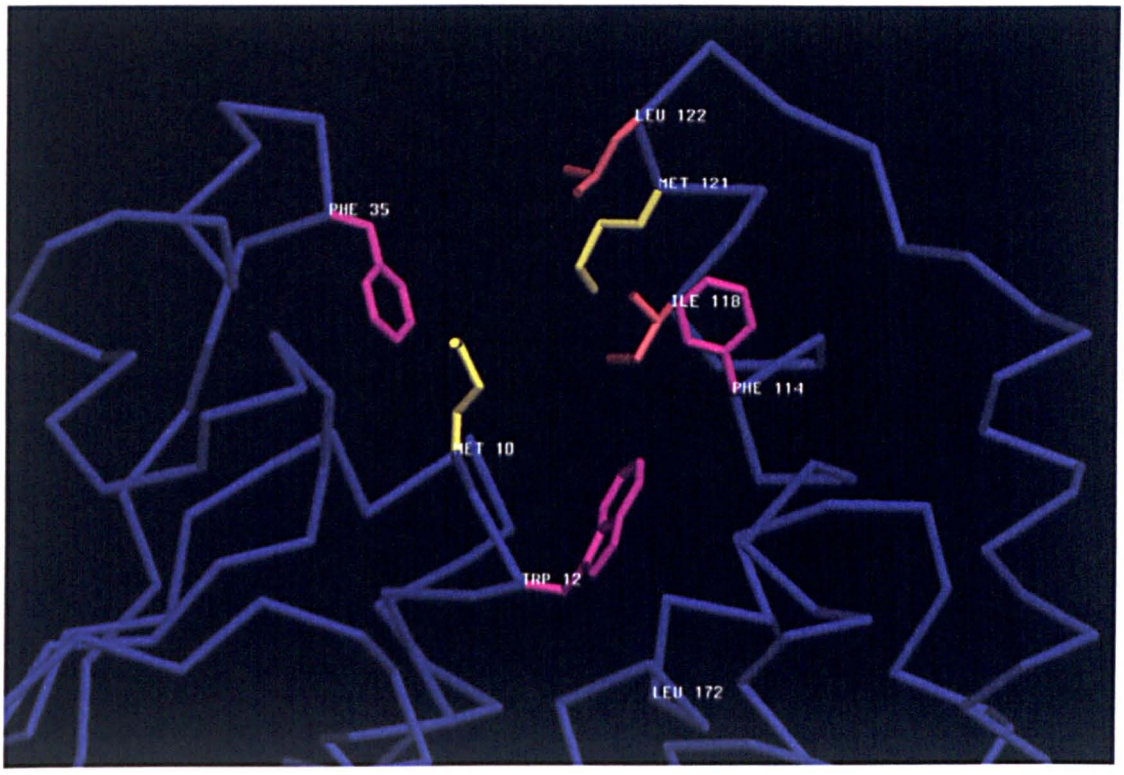


Figure 5-11: The H-site of the Phi class enzymes is predominately composed of hydrophobic residues: This diagram depicts the position of these hydrophobic residues in *Zm*GSTF1 with Leu172 marked to indicate the position where Tyr172 of *Zm*GSTF3 protrudes into the active site. Residues of *Zm*GSTF1 shown are Met10, Trp12, Phe35, Phe114, Ile118, Met121 & Leu122. These are conservatively replaced in *Zm*GSTF3 by; Met9, Phe116, Val120, Leu123, Leu124, Tyr172 and in *At*GSTF2 by; His8, Ile12, Phe119, Phe123, Ile126, and Tyr127 (Neuefeind et al., 1997a).

Phe35 of *Zm*GSTF1 appears to be important in the fixing of the substrate into the active site and is observed to form a stacking interaction with the aromatic ring of atrazine (Neuefeind et al., 1997a). In the *Zm*GSTF1 structure Trp12 enables the interaction of N-terminal residues with substrate by blocking the hydrophobic pocket observed between the two domains in the *At*GSTF2 structure (Prade et al., 1998) and enabling the C-terminal domain to move towards the N-terminal domain on substrate binding. Although the bulky side group of Trp12 causes the active site of *Zm*GSTF1 to be shallower in comparison to *At*GSTF2, the flexibility of the loop between α4-α5 and the interaction of substrate with hydrophobic residues of the N-terminus enables *Zm*GSTF1 to bind large

Original in colour

aromatic substrates such as CDNB and atrazine. This distribution of hydrophobic residues on the N and C terminal domains may also explain the broad substrate specificity of the *Zm*GSTF1 enzyme toward hydrophobic compounds.

The substitution of Trp12 by Ile in the *At*GSTF2 structure results in the formation of a hydrophobic pocket. This pocket appears to be crucial in determining substrate specificity with the isopropyl group of the herbicide FOE-4043 found to be completely buried within this pocket in the ligand bound *At*GSTF2 structure (Prade et al., 1998). In addition the N terminal domain of the *At*GSTF2 structure contrasts that of *Zm*GSTF1 in that it contains less bulky hydrophobic residues with the residues Phe35, Met10 and Trp12 of *Zm*GSTF1 replaced by Leu35, Ala10, Ile12 respectively. The substitution of Trp12 with Ile12 has been postulated to prevent movement of the C-terminal domain towards the N-terminal domain and thus substrate molecules conjugated to glutathione do not interact with residues of the N-terminal domain. The lack of binding of substrates with the N terminal domain may act as a limiting factor on the size of substrates able to bind the *At*GSTF2 active site.

Neither glutathione or substrate bound structures for *Zm*GSTF3 have been determined, therefore the interactions which occur for this molecule are unknown. However, as *Zm*GSTF3 shares the two glycine molecules (residues 125 and 126) found in *Zm*GSTF1, it is likely that the active site will show some degree of flexibility. The replacement of Trp12 with a proline residue would also suggest that *Zm*GSTF3 may have a hydrophobic pocket similar to that observed for *At*GSTF2 and that this might be important in determining substrate specificity for *Zm*GSTF3.

Chapter 6: Structure solution of *At*GSTZ1

The *Arabidopsis thaliania* Zeta class GST (*At*GSTZ1) was selected for use in crystallisation trial for two reasons. Firstly no members of the Zeta class GSTs had previously been structurally characterised and secondly genetic studies had identified this class of GSTs as functioning as an isomerase in the phenylalanine/tyrosine catabolic pathway (Fernández-Cañón & Penalva, 1998). The isomerase activity, first identified by Knox, (Knox & Edwards, 1956) is necessary for the penultimate step in this phenylalanine/tryosine catabolic pathway causing the *cis-trans* conversion of maleylacetoacetate to fumarylacetoacetate (figure 6-1).

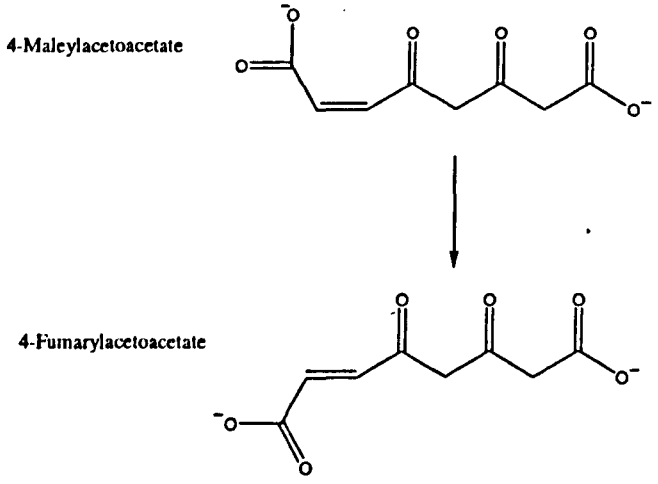


Figure 6-1: *Cis–trans* isomerisation of maleylacetoacetate to fumarylacetoacetate.

Using a single cryocooled crystal, data for *At*GSTZ1 was collected on beamline BM 30, ESRF, Grenoble to 2.2Å. Sequence comparison of *At*GSTZ1 with GSTs in the PDB revealed the maize GST, *Zm*GSTF3 shared greatest sequence identity (21%) and it was initially hoped that this structure might be used as a search model for molecular replacement. Several other models (shown in grey in table 6-1) suggested by the program SSS-align (Sturrock & Dryden, 1997) which aligned the predicted secondary structure (obtained from PHD - http://www.embl-heidelberg.de/predictprotein/ (Rost et al, 1994)) were also used as probes in AMoRe (Navaza, 1994).

Protein	PDB	Sequence conservation (%)	Structural similarity (%)	residue range
Toxin	1DLC	24.8	87.5	66-208
Transferase	2GTU	27.5	87.19	9-221
Transferase	6GSX	27.5	87.1	9-221
Transferase	1GTU	27.0	86.2	9-221
Transferase	1GSU	26.11	85.7	9-221
Signal protein	1B8X	29.6	84.9	11-221
Hydrolase	1CP9	26.8	84.9	30-221
Glycosidase	1A47	18.4	83.7	6-201
Transferase	1LJR	40.3	83.08	7-207
Transferase	1AW9	41.23	81.9	8-218

Table 6-1: Top 9 results from SSS-Align and the score for the plant GST (1AW9) that share greatest sequence identity with AtGSTZ1. SSS-Align utilises both a Smith/Waterman best local similarity algorithm and a secondary structure prediction of *At*GSTZ1, (PhD, (Rost, 1994)) to align sequences.

Despite trying various models, different resolution ranges and changing a variety of other parameters in AMoRE, molecular replacement proved unsuccessful, therefore isomorphous replacement was used to gain phase information.

6.1 Isomorphous Replacement

6.1.1 The heavy atom method

In the isomorphous replacement method the aim is to add heavy atoms (atoms with high atomic numbers) to specific points on the protein surface displacing solvent without distorting the structure of the protein in the crystal. Although the preparation of heavy atom derivatives is usually a 'trial and error' process, the amino acid residues cysteine, histidine and methionine, are known to bind to particular heavy atoms, therefore some rational selection of heavy atoms most likely to bind the protein can be made. As each atom in the unit cell contributes to every reflection in the diffraction pattern, the introduction of heavy atoms changes the intensities observed in the derivative diffraction pattern in comparison to those observed for the native protein. Subtraction of the native diffraction intensities from the derivative intensities yields a simpler diffraction pattern produced by the crystal lattice and the heavy atoms present in the unit cell. As only a few heavy atoms are present in this simplified diffraction pattern,

Patterson methods allow detection of their positions in the unit and enable the phases of these heavy atoms to be calculated. Given that the structure factor (F_{HP}) for the heavy atom derivative is the sum of the vectors from the native structure F_P and the heavy atom F_H, then if F_H is determined and the length of F_P is known, (intensity measurement) $F_P=F_{HP}-F_H$ allows calculation of the phase angle for the structure factor F_P.

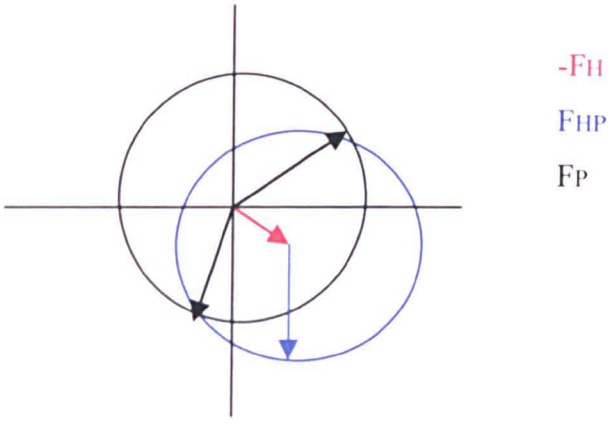


Figure 6-2: Vector solution of F_{HP}-F_H generates all phase possibilities shown by blue circle. Knowledge of the magnitude of F_P generates phase possibilities shown by the black circle. Points of intersection of the two circles indicate phase solutions for the structure factor of F_P.

One heavy atom derivative produces two possible solutions for the phase (shown by the black arrows in figure 6-2). If the two intersections lie near each other then an average phase might be reasonably used to generate initial maps (Single Isomorphous Replacement-SIR), however, typically additional derivatives are used (Multiple Isomorphous Replacement-MIR) to obtain an unambiguous solution.

6.1.2 Anomalous scattering

Just below their characteristic emission wavelengths elements display a sharp change in the absorption of X-rays known as absorption edges. At these absorption edges a proportion of the diffracted beam is absorbed and then re-emitted with altered phase. This absorption effect (anomalous scattering or anomalous dispersion) is negligible for the light atoms in protein structures; carbon, nitrogen and oxygen, at the wavelengths used in protein crystallography, but is measurable for heavier atoms (the anomalous effect of sulphur has been used to phase lysozyme (Dauter et al, 1999)). Where anomalous scattering

conditions are met, the atomic scattering factor can be described by two perpendicular contributions, one real (ΔF_r) and the other imaginary (ΔF_i). The opposite direction of the imaginary anomalous component of bijovet pairs breaks Friedels law (-hkl = -h-k-l and $\alpha_{hkl} = -\alpha_{-h-k-l}$.) (Figure 6-3).

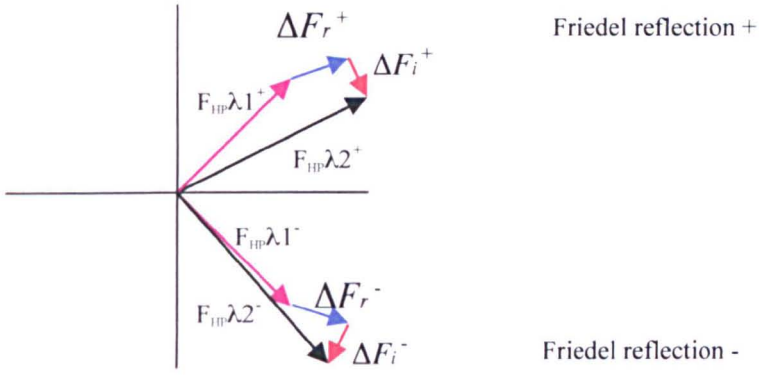
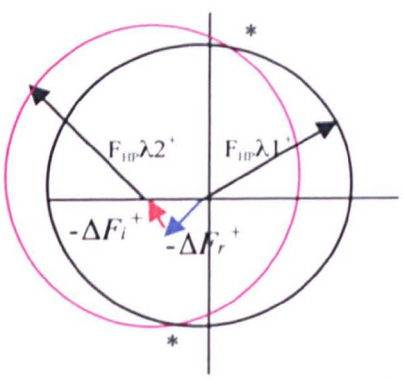
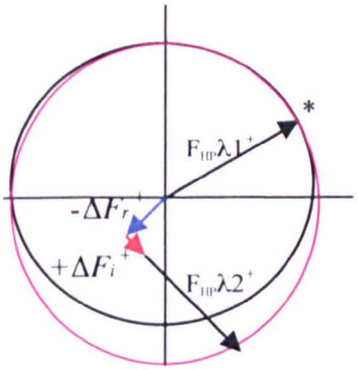


Figure 6-3:$F_{HP}\lambda 2^+$ is no longer the mirror image of $F_{HP}\lambda 2^-$ when anomalous scattering is present $F_{hkl} \neq F_{-h-k-l}$.

As the anomalous scattering contributions of (ΔF_r) and (ΔF_i) are constants for a given element, location of the heavy atom using Patterson methods allows the respective phases of (ΔF_r) and (ΔF_i) to be calculated. Used in conjunction with isomorphous replacement, anomalous scattering information can yield unambiguous phase information for FP as shown in figure 6-4.



Vector solution using one Friedel reflection and additional structure factor with no anomalous component. $F_{HP}\lambda 1^+ = F_{HP}\lambda 2^+ - \Delta F_r^+ - \Delta F_i^+$. Possible phase angles *



Vector solution using second Friedel reflection with $+\Delta F_i^+$ component. $F_{HP}\lambda 1^+ = F_{HP}\lambda 2^+ - \Delta F_r^+ + \Delta F_i^+$ which when taken with (a) allows determination of an unambiguous solution *

Figure 6-4: Determination of unambiguous phases using anomalous scattering.

Original in Colour

6.2 Solution of phases of ZmGSTZ1

6.2.1 Preparation of Heavy Atom bound crystals

Hexagonal crystals of maximum size 0.4 mm × 0.4 mm × 0.5 mm were grown by equilibrating a mixture of 2 µl of protein solution (20 mg/ml in 20 mM Tris pH 7.5) and 2 µl precipitant solution (11% PEG 8000, 0.2 M magnesium acetate, 0.1 M sodium cacodylate pH 6.5) against a 0.8 ml reservoir of the precipitant solution. Examination of the *At*GSTZ1 sequence indicated two free cysteines might be present in the *At*GSTZ1 structure, thus mercury derivatives were screened first. Native crystals were soaked in native mother liquor containing 1 mM $Hg(CN)_2$ for two minutes. The crystals which showed no visual signs of deterioration in the $Hg(CN)_2$ solution were then transferred to the cryo-solution (mother liquor containing 25% glycerol) for 60s before being flash cooled in an Oxford Cryostream. A 3.0Å data set for a mercury cyanide soaked crystal was collected in house using a Nonius FR591 rotating anode with a Macscience DIP2000 image plate (CuKα - λ=1.54Å). This $Hg(CN)_2$ data was processed and scaled using the HKL suite of programs as discussed in section 4.2.4.

6.2.2 SIRAS phasing, model building and refinement

To determine if the mercury cyanide soaked crystal was a derivative the $Hg(CN)_2$ derivative data and native data was scaled together using SCALIT from the CCP4 suite of programs (CCP4, 1994). SHELX-90 (Sheldrick, 1990) was then used to search the Patterson function and this indicated that a single heavy atom was present. The positional parameters (*x*, *y*, *z*) derived from SHELX-90 were then refined along with the temperature factor and relative real and anomalous occupancies in the space groups $P6_122$ and $P6_522$ using MLPHARE (Otwinowski, 1991). The anomalous occupancy suggested that the space group was $P6_522$ (A summary of phasing statistics is given in Table 6-2) and phases for this spacegroup were calculated by MLPHARE using single isomorphous replacement with anomalous scattering (SIRAS). These initial phases were then used as the starting point for density modification procedures in DM (Cowtan, 1994) which used a combination of histogram matching and solvent flattening to refine and extend the phases to 2.2Å.

SIR phasing (25.0-1.65Å)	20-2.9
No. of sites	1
Occupancy (%)	32.4
Anomalous Occupancy (%) $P6_522$	31.4
Anomalous Occupancy (%) $P6_122$	3.4
Temperature factor ($Å^3$)	32.8
Phasing Power[a]	0.87
R_{iso} (%)[b]	13.9
R_{cullis}[c]	0.76
Mean figure of merit[d]	0.36
Fractional coordinates of heavy atom (x, y, z)	-0.726, -0.906, -0.731

[a] Phasing Power, mean heavy-atom contribution over lack of closure, defined as $\sum|F_H^c| / \sum||F_{PH}^o| - |F_P^o|\exp(i\phi_c)+F_H^c||$, where F_H^c is the calculated heavy atom structure factor, $|F_H^c|$ the corresponding amplitude, $|F_P^o|$ and $|F_{PH}^o|$ are the observed amplitudes for the protein and heavy atom derivatives and ϕ_c is the calculated phase.

[b] $R_{iso} = \sum_{hkl}||F_{PH}| - |F_P|| / \sum_{hkl}|F_{PH}|$, where $|F_P|$ and $|F_{PH}|$ are the structure factor amplitudes for the native and the heavy-atom derivatized crystal, respectively.

[c] R_{cullis}, defined as lack of closure over isomorphous difference; data were used in the range of 25 to 2.9Å resolution.

[d] Figure of merit, defined as $\int P(\phi)\exp(i\phi)\, d\phi / \int P(\phi)\, d\phi$, where P is the probability distribution of ϕ, the phase angle.

Table 6-2: Phasing statistics for *At*GSTZ1 $Hg(CN)_2$ derivative. In alternative hand $P6_122$ the anomalous occupancy was only 3.4% clearly indicating that $P6_522$ was correct spacegroup.

The resultant electron density map generated from the phases calculated by DM was interpretable and secondary structure elements were quickly built using the program QUANTA (Molecular Simulations Inc.). The *Zm*GSTF3 structure (Neuefeind et al., 1997) was visually aligned with the helical secondary structure elements built for the *At*GSTZ1 model to provide an orientation of each of the subunits within the dimer. The alignment of the *Zm*GSTF3 structure with the partially built AtGSTZ1 structure aided building of the β-sheet secondary structure elements into the poorer electron density present in this region. During the first round of building it was possible to trace approximately ~63% of the sidechains. This initial model was then refined using the maximum likelihood refinement program REFMAC (Murshudov et al, 1996). At this stage a higher resolution (1.65Å) native data set was collected on station 7.2, Daresbury SRS and refinement was continued using this data, maintaining the same Free R set as before and extending it where the resolution allowed. Following corrections to

the Cα backbone and the addition of further sidechains, the program ARP (Lamzin & Wilson, 1993) was used to add waters. After several cycles of refinement and rebuilding the R_{cryst} and R_{free} remained at 21.5% and 24.5%. Given the high resolution of the data individual restrained aniosotropic B factors were refined (Murshudov et al., 1999) causing an almost immediate reduction in the R_{cryst} and R_{free} to 19.6% and 23.4% respectively. The refined co-ordinates of the model and the structure factors were deposited with the Protein Data Bank under the accession number 1E6B.

6.3 *At*GSTZ1 model

6.3.1 Quality of the model

The asymmetric unit of the *At*GSTZ1 model contains one molecule with the GST dimer lying on a crystallographic dyad. The final model contains one protein molecule of 194 amino acids, a molecule of mercaptoethanol attached to residue Cys154 and 137 water molecules. The polypeptide chain was traced with the exception of seven residues at the N-terminus, one residue at the C-terminus and a nineteen residue region joining helix α4 to helix α5 (residues 114-132). The quality of the refined structure was assessed with PROCHECK (Laskowski et al., 1993) with the Ramachandran plot indicating that 94% of the residues lie within the most favoured regions. Only residue Asp72 was found to lie in the non-favoured region. This is typical of GST structures, which exhibit this residue adopting a strained conformation to allow it to bind γ-glutamate GSH. The model shows good stereochemistry with an estimated co-ordinate error (r.m.s.) of 0.15Å (Cruikshank, 1999) and a final r.m.s. deviation from ideal bond lengths and angles of 0.011Å and 1.6° respectively. With the exception of a few residues, both sides of the unobserved loop region, between α4–α5, the averaged temperature factors for both main chain and side chain atoms agree with the value estimated by the Wilson plot. The final $2F_o - F_c$ electron density map contoured at 1σ is of good quality and shows continuous, well-defined density. The parameters for the refined model are summarised in Table 6.3.

Resolution Range (Å)	25.0-1.65
No of unique reflections in resolution range	29026
R_{Work} (%)[a]	19.6
R_{Free} (%)	23.4
Model	
Total number of amino acids	194
No. of Solvent atoms (excl. H)	137
No. of Heteroatoms	4
Stereochemistry[b]	
Ramachandran quality, % in	
Most favoured regions	94
Allowed regions	5.4
R.M.S. deviation[c]	
Bond length (Å)	0.011
Bond angle (degrees)	1.6

[a] $R = \Sigma_{hkl}(|F_o - F_c|) / \Sigma|F_o|$.
[b] Calculated using the program PROCHECK (Laskowski et al., 1993)
[c] (Cruikshank, 1999)

Table 6-3: Final Refinement statistics for *At*GSTZ1

6.3.2 Overall structure

AtGSTZ1 adopts a characteristic GST fold, despite the low pairwise sequence identity of around 15% to the mammalian classes; Alpha, Mu, Pi and 20% to the Theta and Phi classes. The crystal structure of the apo *At*GSTZ1 is a monomer (figure 6-5) formed by two spatially distinct domains, a smaller N-terminal domain, (domain I) 84 residues in length, and a larger C-terminal domain (domain II) of 127 residues. Domain I contains both helical and β-strand secondary structure elements while domain II is entirely helical. These two domains are covalently joined by an eight residue linker region which adopts an extended structure, lacking any secondary structure elements.

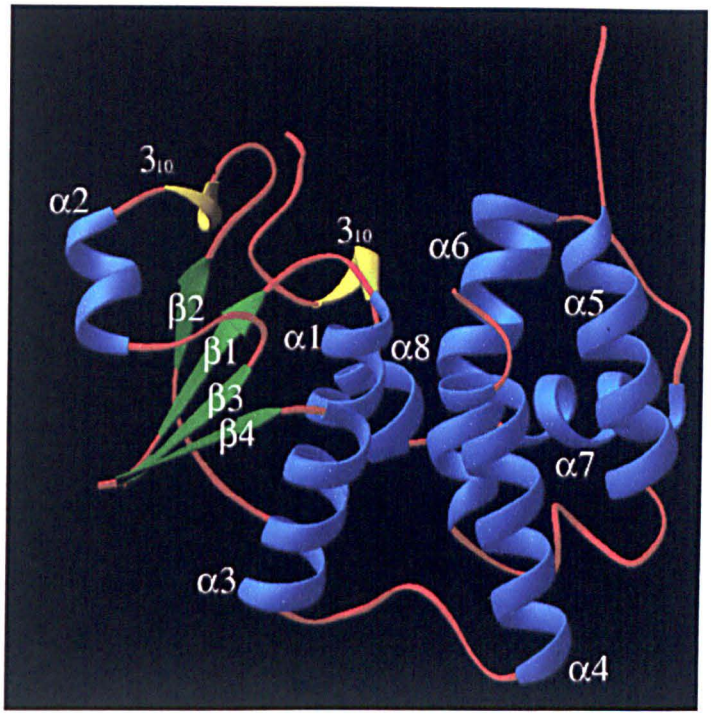


Figure 6-5: Ribbon representation of *At*GSTZ1 monomer. α-Helices are drawn as blue spirals, β-strands are in green and 3_{10} helices shown in yellow. The figure was drawn using RIBBONS (Carson, 1997).

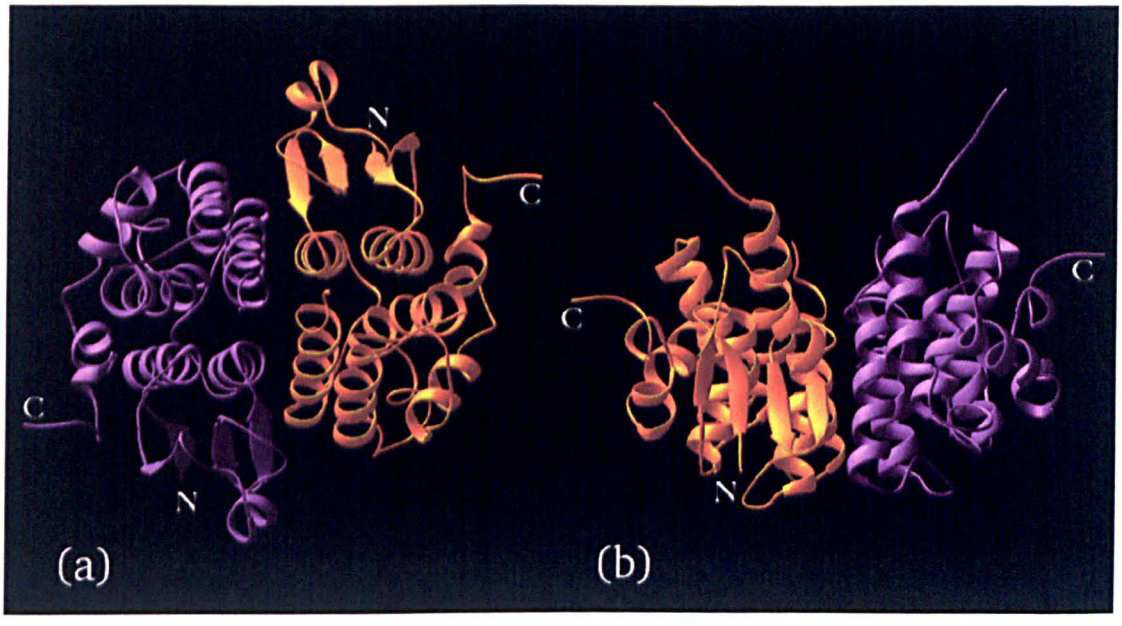


Figure 6-6: Ribbon diagram of the dimeric *At*GSTZ1 structure. The 2 fold axis relating to the dimer subunits normal to the plane in (a) and on the plane of the drawing in (b). Subunit A is coloured lilac and subunit B is gold. The N and C termini are indicated with N and C respectively. The figure was drawn using RIBBONS (Carson, 1997)

Original in Colour

6.3.2.1 N-terminal Domain

Domain I (residues 8-84) comprises a four stranded β-sheet flanked on one side by two α-helices (α1 and α3) and on the other by an irregular loop region. The β-strands are arranged in the order β2 (35-38), β1 (10-13), β3 (62-65) and β4 (68-71) with β3 in an antiparallel direction to the others. A type I β-turn joins β1 to helix α1 (17-30) the top of which forms the base of the active site. Strand 2 is connected to β3 via a mostly irregular part of structure containing two β-turns (40-43, 55-58), a 3_{10} Helix (44-47) and a short section of helix α2 (48-55). At the beginning of β3, *cis* Pro61 creates the characteristic turn essential for glutathione binding. A hairpin turn connects β3 to β4 before forming helix α3 (72-84), which is orientated almost parallel with helix α1. Helix α3 is broken by Pro85 to form an extended part of structure, eight residues in length containing four proline residues that covalently connect the N and C-terminal domains.

6.3.2.2 C-terminal domain

The larger C-terminal domain (93-220) is composed entirely of helical elements with the first two helices, helix α4 (93-110) and helix α5 (137-152), forming an up down arrangement. Of the 27 amino acid residues which link helices α5 and α4, nineteen residues (114-132) could not be traced due to weak or no electron density. Residues 133-138 are stabilised by crystal contacts with residues 32-36 of a symmetry related molecule and form a strand like structure. This flexible region may form an extended loop structure, which could close over the active site on substrate binding. At the base of helix α5 a 13 residue loop region connects this helix to helix α6 (165-183) which runs parallel to helix α4. Helix α6 is connected via a series of β-turns (186-189, 187-190) to helix α7 (189-200) which lies perpendicular to both helix α6 and helix α8 (202-209). This orientates α8 to lie along the rear of the domain interface before adopting a short 3_{10} helical structure, (210-213).

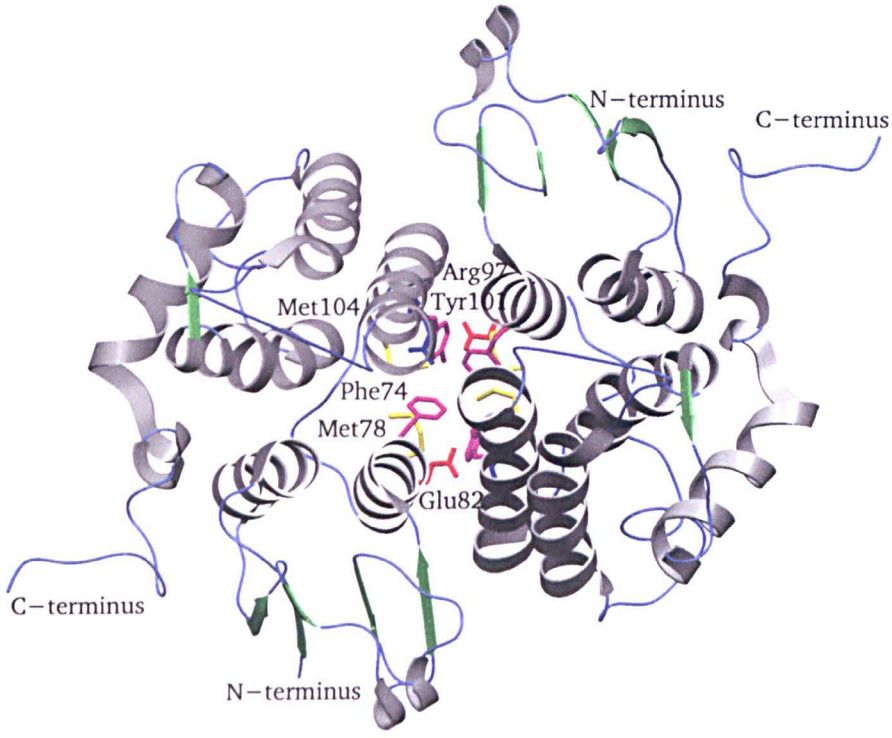
6.3.2.3 Interactions between domains

The interactions between the N and C-terminal domains are predominately hydrophobic in nature with residues from helices α1 and α3 packing against those of helices α4 and α6. The side chains Ile25, Leu29, Phe74 and Ile76 of

constitute one hydrophobic surface, helices α1 and α3, while the other is formed by residues Tyr101, Met104, Leu108 from α4 along with Leu166, Leu169, Phe170 and Pro173 from α6. The buried surface area between domains in *At*GSTZ1 (1125Å^2) is smaller than that observed in the Alpha, Mu and Pi class GST structures, (~1400Å^2) but more similar in size to the Theta and non-mammalian GSTs. The C-terminal residue Asp215 of the conserved QPD motif, present in the Zeta class sequence forms interdomain hydrogen bonds with the hydroxyl of the conserved residue Tyr36 and the amide nitrogen of Trp14 in the N-terminal domain. The domain interface is also significantly less polar than observed in other GST structures with only 28% being formed by polar residues in comparison to 40-45% typically seen in mammalian structures. *At*GSTZ1 also possesses none of the characteristic interdomain salt bridges found in the Alpha, Beta, Pi and Mu classes of GST.

6.3.2.4 Interactions between subunits

Two α-helices, α3 and α4 from each subunit, form a bundle around the 2-fold axis. The formation of the dimer is driven by the hydrophobic interaction between helices α3 of one subunit with α4 of the other. These helices form an interface of 1162Å^2 which comprises 11% of the molecules surface area and is predominately hydrophobic in nature.



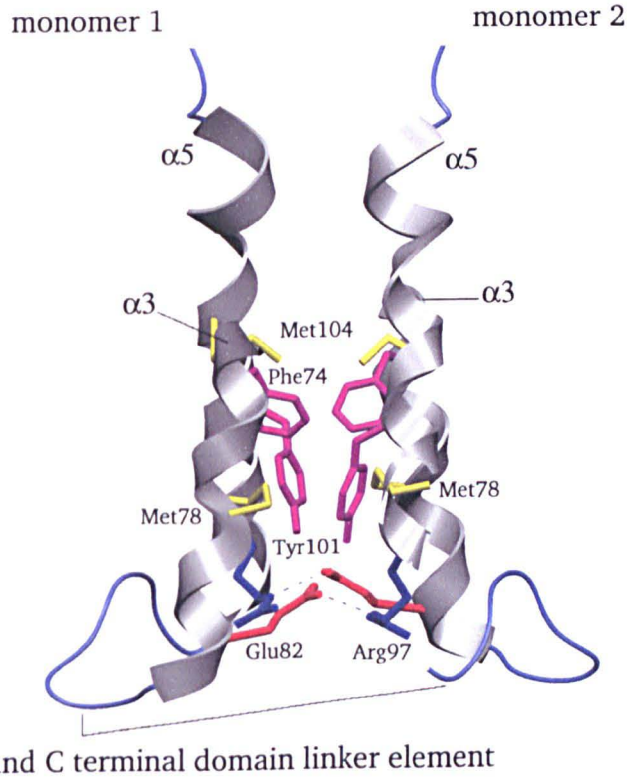


Figure 6-7: Interactions between subunits of AtGSTZ1 viewed from top and side of the *molecule to display interleaving of non polar residues and salt bridge formed between the* subunits.

Original in Colour

When viewed down the two fold axis, a number of residues; Phe74, Tyr101, Met78 and Met104 interleaving with their two fold equivalent across the interface. In addition to these hydrophobic contacts two salt bridges are formed at the base of the interface between residues Glu82 and Arg97 (figure 6-7). The interactions found at the *At*GSTZ1 dimer form a unique hydrophobic stacking of the residues across the interface. This would prevent Zeta class molecules forming heterodimers with GSTs of other classes.

6.3.3 Active Site

The active site is formed by helices α1 and α3 from the N-terminal domain and α4 and α6 from the C-terminal domain. The N-terminal region contains the residues for glutathione binding, including *cis*-proline at residue 61, which is conserved in all GST structures and is responsible for backbone hydrogen bonding to the glutathione molecule. Asp72 is thought to interact with the GSH-γ-glutamate and shows unfavourable Phi and Psi angles as described for the equivalent residues Glu67 of Alpha, Gln71 of Mu, Gln62 of Pi and Glu66 of Theta. In addition the structural residue Val60 which corresponds to Val55 of Alpha, Leu52 of Pi, Val59 of Mu and Val54 of Theta is also conserved (Dirr et al., 1994; Wilce & Parker, 1994).

It has been proposed that glutathione is active predominately as a thiolate anion and that its formation is facilitated by an active site tyrosine or serine residue, either through a general base catalytic mechanism or by donation of a hydrogen bond (Armstrong, 1993; Dirr et al., 1994). In the Beta (Nishida et al., 1998) and Omega (Board et al, 2000) structures, a cysteine residue has been suggested to be catalytically important. Sequence analysis of Zeta class GSTs shows that there are two serine residues, Ser17 and Ser18, a cysteine, Cys19 and a tyrosine, Tyr12 that are strongly conserved and could be catalytically important. Analysis of the *At*GSTZ1 structure indicates that Tyr12 is structurally rather than catalytically important (figure 6-8).

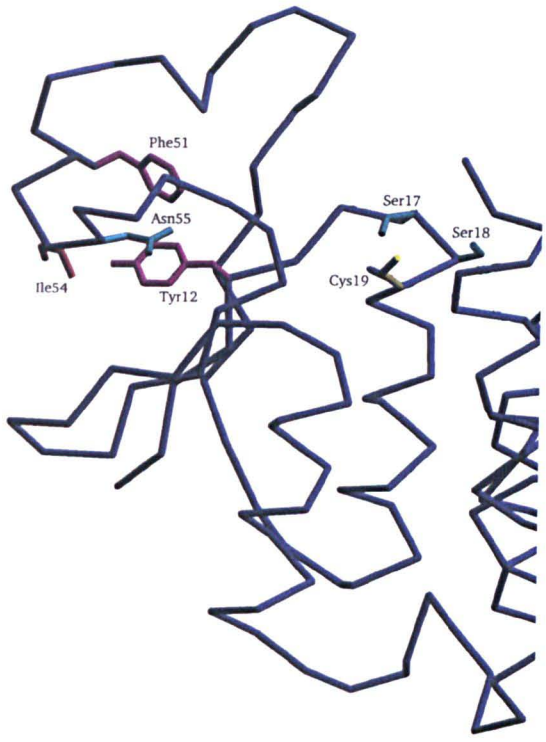


Figure 6-8: Diagram showing gross positions of strongly conserved residues Tyr12, Ser17, Ser18 and Cys19. Tyr12 is clearly not involved in catalysis at the active site, but is structurally important in preserving the structure of the 1-2 loop region, packing against Phe51, Ile54 and Asn55.

Ser17 at the N-terminus of α1 superimposes well on the conserved catalytic serine of the Theta and Phi classes of GST and like the catalytic serine in these two classes is thought to stabilise the GSH thiolate ion in the Zeta class. This serine residue was previously thought to distinguish the Theta class from all other classes, but this study shows it is also a key feature of the Zeta class enzymes. Both Ser18 and Cys19 are also orientated into the active site although their roles in substrate binding or in the isomerisation reaction are unknown.

The proposed substrate binding region (H-site) of the active site is formed by residues from α4, α6 of the C-terminal and residues of the N-terminal domain. In *At*GSTZ1 the H-site is significantly less hydrophobic than that observed for other classes of GST. It contains a number of basic residues, Arg16, and Arg22 from the N-terminal domain, along with Arg181 from the C-terminal domain. These basic residues form a positively charged region in the active site enabling it to bind the negative carboxyl moieties of the maleylacetoacetate substrate (figure 6-9).

Original in Colour

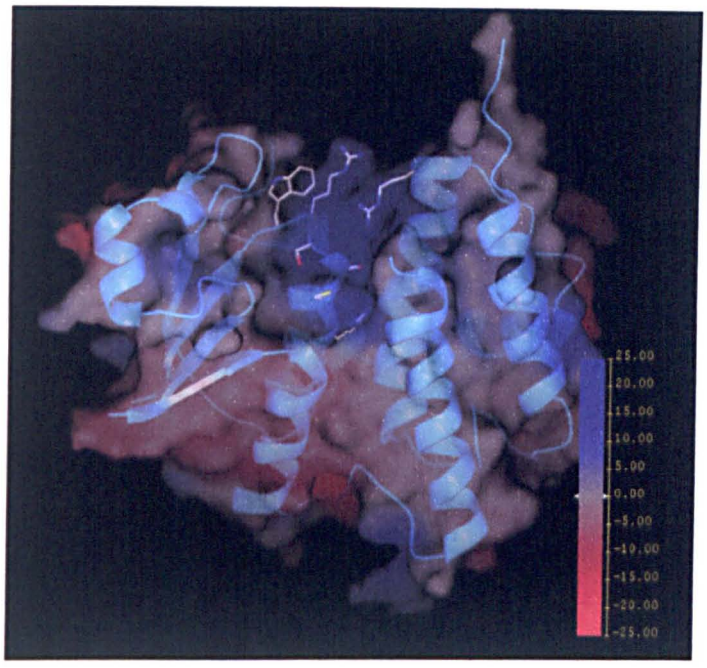


Figure 6-9: Electrostatic potential coloured surface representation of *At*GSTZ1. The proposed substrate binding region of the active site has a positively charged face formed by the basic residues Arg16, Arg22 and Arg181. Ser17, proposed to stabilise the thiolate ion of glutathione, is displayed as are Ser18 and Cys19. Trp15 which interacts with the C-terminal region is also shown. This figure was drawn using the program SPOCK (Christopher & Baldwin, 1998). This figure was further edited using GIMP (http://www.gimp.org/)

6.4 Discussion

6.4.1 Analysis of primary sequence of *At*GSTZ1

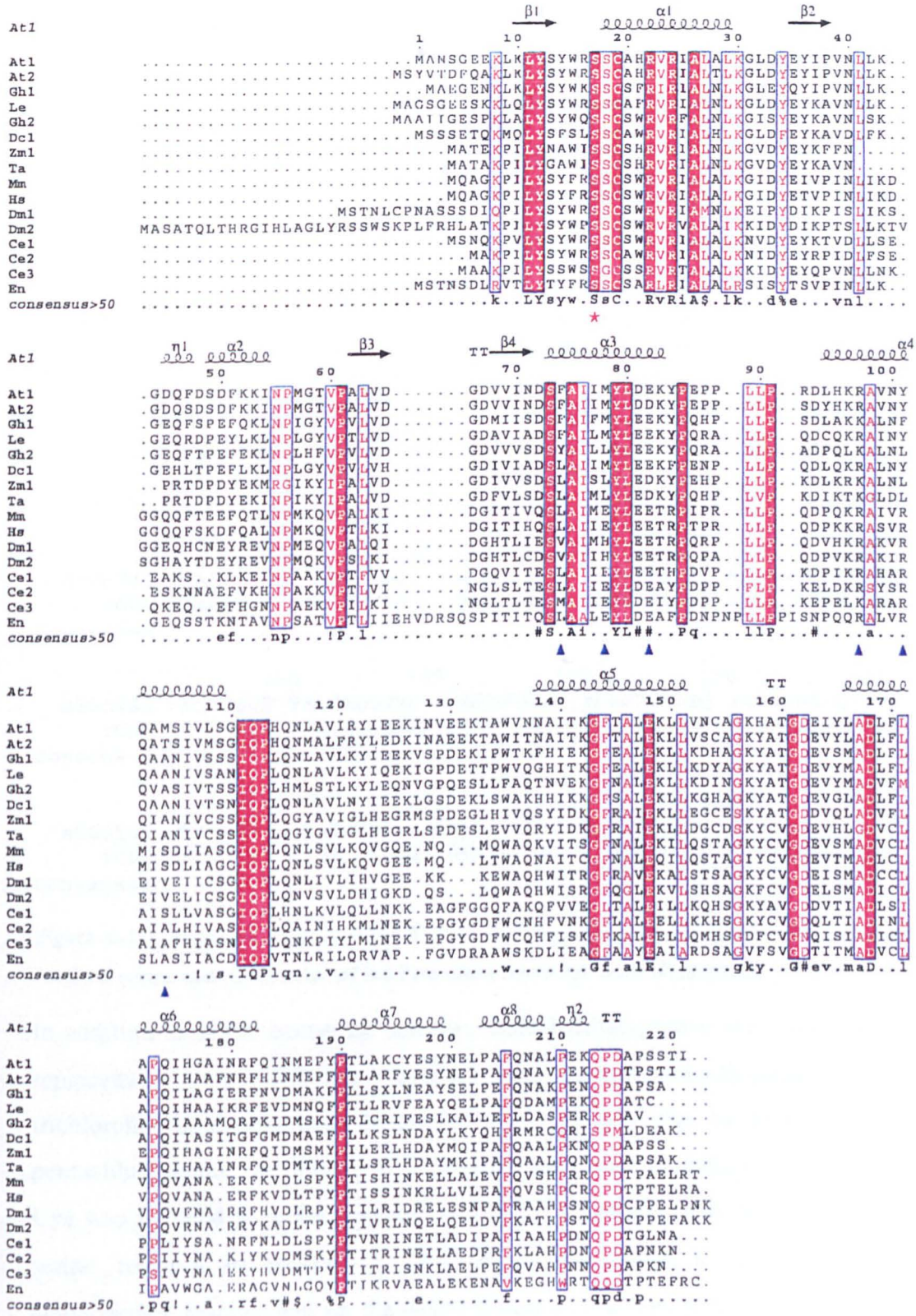
Multiple sequence alignment of the Zeta class GSTs clearly shows that the class is well conserved across a range of species with 45% pairwise identity within the group as shown in Figure 6-10. As observed in other classes of GSTs, domain I shows significantly greater percentage identity across a range of species in comparison to domain II. In domain I 51% of residues are conserved between *At*GSTZ1 and the human isoenzyme, while domain II only shares 40% identity. Those residues of domain I suggested to be significant in the binding of glutathione are almost all strictly conserved and the majority of the differences in sequence occur at the loop regions between β2-β3 and α4-α5. A few single amino acid changes namely Phe74 to Leu and Met78 to Glu, found in the human, mouse and *C. elegans* sequences are associated with helix α3 and are significant as they will effect the nature of the dimer interface. In domain II residues of α4

Original in Colour

of the non-plant Zeta sequences associated with the dimer interface are also different in size and charge, with Tyr101 significantly changed to Arg (figure 6-10). These amino acid changes will cause the dimer interface of the non plant isoenzymes to be considerably less hydrophobic. The Glu and Arg residues present in the animal sequences, may form a salt bridge which would stabilise the intersubunit interaction and the interaction between the N and C-terminal domains. In domain II a conserved motif IQP (111-113) is observed at the C-terminal end of α4. This motif forms part of the active site and may be required for substrate binding or catalytic activity. Furthermore residues Gly161 and Asp168 which are conserved in all classes of GST are present in the α5-α6 loop region and helix α6 respectively. Asp168 is structurally important, buried within the hydrophobic core and forms four hydrogen bonds to the backbone amide groups of Ala159, Tyr165, the hydroxyl of Tyr199 and a structural water 125.

In addition to the Zeta class GSTs identified in eukaryotes, a BLAST search of the NCBI nonredundant database (Altschul et al, 1997) detects a number of bacterial sequences (AF109131, *Sinorhizobium meliloti;* (Milcamps & deBruijn, 1999), Y13118 *Sphingomonas sp;* AF081362, *Pseudomonas sp*) which show significant homology to the Zeta class GSTs, but lack the conserved cysteine residue present in the Zeta class signature motif SSCX[WHI]RVIAL. Analysis of the flanking DNA sequences of these genes reveals that they lie within clusters of genes involved in the metabolism of tyrosine, strongly suggesting that these enzymes are also maleylacetoacetate isomerases. The lack of the conserved cysteine in these bacterial sequences indicates that this residue is not required for isomerase activity.

Figure 6-10: Sequence alignment of members of the Zeta family of GSTs. The secondary structure of AtGSTZ1 is also shown. The proposed catalytic residue, Ser18 which begins the conserved region SSCX[WH]RVIAL is marked *. Residues at the dimer interface are marked with_. This sequence alignment was created using the following sequences (Organism, Genbank / SWISSprot accession numbers or other source in brackets) At1 *At*GSTZ1, *A. thaliana*, AC005312 T16F16.18); At2 (A. thaliana, AC005312 T16F16.18) Gh1 (*Gossypium hirsutum*, GI7628319); Le (*Lycopersicon esculentum*, AI781796+AW218977); Gh2 (*Gossypium hirsutum*, AI7318676); Dc1 (*Dianthus caryophyllus*, S33628) *Zm* (*Zea Mays*, AW067573+AW01781); Ta (Triticum aestivum, AF002211); Mm (*Mus musculus*, NP_034493); Hs (*Homo sapiens*, U86529); Dm1 (*Drosphila melanogaster*, AAF54382); Dm2 (*Drosphila melanogaster*, AAF54381); Ce1 (C. *elegans*, T20294); Ce2 (C. *elegans*, AC024200); Ce3 (*C. elegans*, AF60754);En (*Emericella nidulans*, O43123)


At1
β1
α1
β2
1 10 20 30 40
At1 .....MANSGEEKLKLYSYWRSSCAHRVRIALALKGLDYEYIPVNLLK..
At2 .....MSYVTDFQAKLKLYSYWRSSCAHRVRIALTLKGLDYEYIPVNLLK..
Gh1 .....MAEGENKLKLYSYWKSSCSFRIRIALNLKGLEYQYIPVNLLK..
Le .....MAGSGEESKKLQLYSYWRSSCAFRVRIALNLKGLDYEYKAVNLLK..
Gh2 .....MAAIIGESPKLALYSYWQSSCSWRVRFALNLKGISYEYKAVNLSK..
Dc1 .....MSSSETQKMQLYSFSLSSCAWRVRIALHLKGLDFEYKAVDLFK..
Zm1 .....MATEKPILYNAWISSCSHRVRIALNLKGVDYEYKFFN.....
Ta .....MATAKPILYGAWISSCSHRVRIALNLKGVDYEYKAVN.....
Mm .....MQAGKPILYSYFRSSCSWRVRIALALKGIDYEIVPINLIKD.
Hs .....MQAGKPILYSYFRSSCSWRVRIALALKGIDYETVPINLIKD.
Dm1 .....MSTNLCPNASSSDIQPILYSYWRSSCSWRVRIAMNLKEIPYDIKPISLIKS.
Dm2 MASATQLTHRGIHLAGLYRSSWSKPLFRHLATKPILYSYWPSSCSWRVRVALAIKKIDYDIKPTSLLKTV
Ce1 .....MSNQKPVLYSYWRSSCSWRVRIALALKNVDYEYKTVDLLSE.
Ce2 .....MAKPILYSYWRSSCAWRVRIALALKNIDYEYRPIDLFSE.
Ce3 .....MAAKPILYSSWSSCGSSRVRTALALKKIDYEYQPVNLLNK.
En .....MSTNSDLRVTLYTYFRSSCSARLRIALALRSISYTSVPINLLK..
consensus>50 .....k..LYsyw.SsC..RvRiA$.lk..d%e...vnl....
η1 α2 β3 β4 α3 α4
At1
50 60 70 80 90 100
At1 .GDQFDSDFKKINPMGTVPALVD.....GDVVINDSFAIIMYLDEKYPEPP..LLP..RDLHKRAVNY
At2 .GDQSDSDFKKINPMGTVPALVD.....GDVVINDSFAIIMYLDDKYPEPP..LLP..SDYHKRAVNY
Gh1 .GEQFSPEFQKLNPIGYVPVLVD.....GDMIISDSFAIFMYLEEKYPQHP..LLP..SDLAKKALNF
Le .GEQRDPEYLKLNPLGYVPTLVD.....GDAVIADSFAILMYLEEKYPQRA..LLP..QDCQKRAINY
Gh2 .GEQFTPEFEKLNPLHFVPVLVD.....GDVVVSDSYAILLYLEEKYPQRA..LLP..ADPQLKALNL
Dc1 .GEHLTPEFLKLNPLGYVPVLVH.....GDIVIADGLAIIMYLEEKFPENP..LLP..QDLQKRALNY
Zm1 ..PRTDPDYEKMRGIKYIPALVD.....GDIVVSDGLAISLYLEDKYPEHP..LLP..KDLKRKALNL
Ta ..PRTDPDYEKINPIKYIPALVD.....GDFVLSDGLAIMLYLEDKYPQHP..LVP..KDIKTKGLDL
Mm GGQQFTEEFQTLNPMKQVPALKI.....DGITIVQGLAIMEYLEETRPIPR..LLP..QDPQKRAIVR
Hs GGQQFSKDFQALNPMKQVPTLKI.....DGITIHQGLAIIEYLEETRPTPR..LLP..QDPKKRASVR
Dm1 GGEQHCNEYREVNPMEQVPALQI.....DGHTLIESVAIMHYLEETRPQRP..LLP..QDVHKRAKVR
Dm2 SGHAYTDEYREVNPMQKVPSLKI.....DGHTLCDSVAIIHYIEETRPQPA..LLP..QDPVKRAKIR
Ce1 .EAKS..KLKEINPAAKVPTFVV.....DGQVITESLAIIEYLEETHPDVP..LLP..KDPIKRAHAR
Ce2 .ESKNNAEFVKHNPAKKVPTLVI.....NGLSLTESLAIIEYLDEAYPDPP..FLP..KELDKRSYSR
Ce3 .QKEQ..EFHGNNPAEKVPILKI.....NGLTLTESMAIIEYLDEIYPDPP..LLP..KEPELKARAR
En .GEQSSTKNTAVNPSATVPTLIIEHVDRSQSPITITQSLAALEYLDEAFPDNPNPLLPPISNPQQRALVR
consensus>50 .....ef...np...!P.l.....#S.Ai..YL##..Pq....llP...#....a...
α5 TT
At1
110 120 130 140 150 160 170
At1 QAMSIVLSGIQPHQNLAVIRYIEEKINVEEKTAWVNNAITKGFTALEKLLVNCAGKHATGDEIYLADLFL
At2 QATSIVMSGIQPHQNMALFRYLEDKINAEEKTAWITNAITKGFTALEKLLVSCAGKYATGDEVYLADLFL
Gh1 QAANIVSSSIQPLQNLAVLKYIEEKVSPDEKIPWTKFHIEKGFEALEKLLKDHAGKYATGDEVSMADLFL
Le QAANIVSANIQPLQNLAVLKYIQEKIGPDETTPWVQGHITKGFEALEKLLKDYAGKYATGDEVYMADLFL
Gh2 QVASIVTSSIQPLHMLSTLKYLEQNVGPQESLLFAQTNVEKGFNALEKLLKDINGKYATGDEVYMADVFM
Dc1 QAANIVTSNIQPLQNLAVLNYIEEKLGSDEKLSWAKHHIKKGFSALEKLLKGHAGKYATGDEVGLADLFL
Zm1 QIANIVCSSIQPLQGYAVIGLHEGRMSPDEGLHIVQSYIDKGFRAIEKLLEGCESKYATGDDVQLADVFL
Ta QIANIVCSSIQPLQGYGVIGLHEGRLSPDESLEVVQRYIDKGFRAIEKLLDGCDSKYCVGDEVHLGDVCL
Mm MISDLIASGIQPLQNLSVLKQVGQE.NQ...MQWAQKVITSGFNALEKILQSTAGKYCVGDEVSMADVCL
Hs MISDLIAGGIQPLQNLSVLKQVGEE.MQ...LTWAQNAITCGFNALEQILQSTAGIYCVGDEVTMADLCL
Dm1 EIVEIICSGIQPLQNLIVLIHVGEE.KK...KEWAQHWITRGFRAVEKALSTSAGKYCVGDEISMADCCL
Dm2 EIVELICSGIQPLQNVSVLDHIGKD.QS...LQWAQHWISRGFQQLEKVLSHSAGKFCVGDELSMADICL
Ce1 AISLLVASGIQPLHNLKVLQLLNKK.EAGFGGQPAKQPVVEGLTALEILLKQHSGKYAVGDDVTIADLSI
Ce2 AIALHIVASIQPLQAINIHKMLNEK.EPGYGDFWCNHFVNKGFLALEELLKKHSGKYCVGDQLTIADINL
Ce3 AIAFHIASNIQPLQNKPIYLMLNEK.EPGYGDFWCQHFISKGFKALEELLQMHSGDFCVGNQISIADICL
En SLASIIACDIQPVTNLRILQRVAP..FGVDRAAWSKDLIEAGFAAYLEAIARDSAGVFSVGDTITMADVCL
consensus>50 .....!.s.IQPlqn..v.....w....!..Gf.alE..l....gky..G#ev.maD..l
α6 α7 α8 η2 TT
At1
180 190 200 210 220
At1 APQIHGAINRFQINMEPYPTLAKCYESYNELPAFQNALPEKQPDAPSSTI..
At2 APQIHAAFNRFHINMEPFPTLARFYESYNELPAFQNAVPEKQPDTPSTI...
Gh1 APQILAGIERFNVDMAKFPLLSXLNEAYSELPEFQNAKPENQPDAPSA....
Le APQIHAAIKRFEVDMNQFPTLLRVFEAYQELPAFQDAMPEKQPDATC.....
Gh2 APQIAAAMQRFKIDMSKYPRLCRIPESLKALLEFLDASPERKPDAV.....
Dc1 APQIIASITGFGMDMAEFPLLKSLNDAYLKYQHFRMRCQRISPMLDEAK...
Zm1 EPQIHAGINRFQIDMSMYPILERLHDAYMQIPAFQAALPKNQPDAPSS....
Ta APQIHAAINRFQIDMTKYPILSRLHDAYMKIPAFQAALPQNQPDAPSAK...
Mm VPQVANA.ERFKVDLSPYPTISHINKELLALEVFQVSHPRRQPDTPAELRT.
Hs VPQVANA.ERFKVDLTPYPTISSINKRLLVLEAFQVSHPCRQPDTPTELRA.
Dm1 VPQVFNA.RRFHVDLRPYPIILRIDRELESNPAFRAAHPSNQPDCPPELPNK
Dm2 VPQVRNA.RRYKADLTPYPTIVRLNQELQELDVFKATHPSTQPDCPPEFAKK
Ce1 PPLIYSA.NRFNLDLSPYPTVNRINETLADIPAFIAAHPDNQPDTGLNA...
Ce2 PSIIYNA.KIYKVDMSKYPTITRINEILAEDPRFKLAHPDNQPDAPNKN...
Ce3 PSIVYNAIEKYHVDMTPYPIITRISNKLAELPEFQVAHPNNQPDAPKN....
En IPAVWGA.ERAGVNLGQYPTIKRVAEALEKENAVKEGHWRTQQQTPTEPRC.
consensus>50 .pq!..a..rf.v#$..%p.....e.....f....p..qpd.p.....

Tetrachlorohydroquinone (TCHQ) dehalogenase from *Sphingomonas clorophenolica* has MAA isomerase activity (Anandarajah et al., 2000). Although TCHQ shares only limited sequence similarity with enzymes of the glutathione S-transferase superfamily. (TCHQ dehalogenase shares only 19% pairwise identity with *At*GSTZ1), it shares significant homology within the Zeta class signature motif exists with both the proposed catalytic ser and the cys residue conserved between these enzymes (figure 6-11) (Orser et al, 1993).

```
           1        10         20         30         40         50
  ATGSTZ1 MANSGEEKLK LYSYWRSSCA HRVRIALALK GLDYEYIPVN L-LKGDQFDSD
     TCHQ       MPEVS LYNYTMSICS MKTRLAMEEF GVDYDDKQVD IGFALENFEPD
Consensus             LY Y  S C     R A     G DY    V            D

                  60         70         80         90          101
  ATGSTZ1 FKKINPMGT VPALVDGDVV INDSFAIIMY LDEK----YP EPPLLPR---DL
     TCHQ YVRLNEKAV VPTLVVGDRV VTNSYNIVLE AANVGKVGIP ADPVENKAALDW
Consensus      N     VP LV GD V    S  I             P   P        D

                 110        120        130        140        150
  ATGSTZ1 HKRA--VN YQAMSIVLSG IQPHQNLAVI R------YIE EKINVEEKTAWVN
     TCHQ FQKGDQVN FQVITYGHKG VPRGDELLIA RRERAKEYAE KYPELRSIYQAAH
Consensus        VN  Q       G             R      Y E

                 160        170        180        190        200
  ATGSTZ1 NAI-TKGF TALEKLLVNC AGKHATGDEI YLADLFLAPQ IHGAINRFQINME
     TCHQ DRIVEHG  NCAYDADTVA QAEVDLQKRL DELDVHLADK PFIAGSNYSIADI
Consensus   I                                D  LA      A     I

               210        220        230        240        250
  ATGSTZ1 PYPTLAK  CYESYNE--- LPAFQNALPE KQPDAPSSTI
     TCHQ MWTVLLA  RIEMLNMTAW ISERPNLLAY YQRMKARRSF GGIETARVMPNWK
Consensus     L      E  N          N L    Q
```

Figure 6-11: Sequence alignment of TCHQ dehalogenase and *At*GSTZ1 with proposed catalytic serine and the cysteine of the Zeta Class signature motif highlighted.

In addition to MAA isomerase activity, TCHQ dehalogenase also catalyses the replacement of chlorine atoms on tetrachlorohydroquinone and trichlorohydroquinone with hydrogen atoms during the biodegradation of pentachlorophenol. A mutant of TCHQ dehalogenase in which the conserved Cys was mutated to a Ser revealed that TCHQ requires both the cysteine and serine residues to reduce the glutathione conjugates formed during the dehalogenase reaction and for the isomerisation of maleylacetoacetate.

The absence of the conserved cysteine residue in bacterial MAAI sequences is difficult to reconcile with its proposed catalytic role in TCHQ dehalogenases. It is therefore unclear whether the sequence similarity between TCHQ and MAAI

is merely a reflection of the conserved glutathione binding site, or if fundamental similarities in the structure and mechanism of these enzymes exist.

6.4.2 Comparison to the Zeta class enzyme with other GST structures

The tertiary structure of *At*GSTZ1 was compared with isoenzymes of the classes; Alpha (Sinning et al., 1993), Mu (Ji et al., 1992), Pi (Reinemer et al., 1991), Theta (Rossjohn et al., 1998), Sigma (Ji et al., 1995), Beta (Nishida et al., 1998), Phi (Reinemer et al., 1996), Omega (Board et al, 2000), Delta (Wilce et al., 1995) and the GSTs from *Schistosoma japonicum* (Lim et al., 1994; McTigue et al., 1995). When the structural core of *At*GSTZ1 is superimposed on GST structures from other classes the degree of structural similarity decreases in the order, domain 1> monomer > dimer. The core structure of *At*GSTZ1 superimposes best on the mammalian Theta class enzyme (hGSTT2-2) with a root mean square deviation (r.m.s.d.) of 1.46Å / 154 residues, (calculated using the program LOCK, (Singh & Brutlag, 1997)) the N-terminal secondary structure elements being of similar length and orientation. The loop structure between β2 and β3, which is particularly variable among different classes of GST, is structurally almost identical between the Zeta and Theta class enzymes. This loop region is stabilised by Phe51 and the conserved residue Tyr12, which hydrophobically stack in a manner similar to Phe45 with Phe6 in hGSTT2-2, Phe43 with Tyr4 in insects (Delta class) and Phe45 with Phe6 in plants (Phi class) (Reinemer et al., 1996; Rossjohn et al., 1998). Following this loop region the conserved residues Asn55 and Pro56 form a sharp turn as observed in Beta, Theta and Phi class GST structures. These residues play a critical role in the formation of the loop prior to the functionally important *cis* proline (Rossjohn et al., 1996). The residue Gly58, which is conserved in the Theta class GST, is also observed in the Zeta class. This residue allows the β2-β3 loop structure to pack with the β-sheet of the N-terminal, promoting a similar conformation in this flexible and variable region of structure between the two classes.

There are significant differences in domain II and the linker region between the structures of the different classes of GST. These differences are also reflected in the primary sequence where sequence comparison indicates domain I shares percentage sequence identity of 34%, 36% and 32% with human (1LJR), *A.*

thaliana (1GNW) and *E. coli* (1A0F) GSTs, while domain II shows percentage identities of 17%, 13% and 16% respectively. The linker section of *At*GSTZ1 has 8 residues (85-92) four of which are proline residues, causing this region to adopt an extended structure not previously observed in other GST classes. The helical elements α4 and α5 of the C-terminal domain of *At*GSTZ1 adopt a significantly different alignment, with a difference of ~8° between the orientation of α4 in the Zeta class structure to that of the α4 observed in hGSTT2-2. The loop region between α4 and α5 could not be traced, however, in the presence of substrate, it is postulated that this loop might close over the active site, in an analogous fashion to the C-terminal tail observed in the mammalian Theta and Alpha classes. Further investigation is required to determine if this hypothesis is correct. Helices α6 to α8 resemble those observed in previous Theta class structures, however *At*GSTZ1 contains fewer aromatic and more aliphatic residues in the core of the C-terminal domain. The stretch of residues which connects α5 and α6 in *At*GSTZ1 contains an additional residue in comparison to the mammalian Theta class structures causing this loop to be enlarged and pushed further back from the dimer interface. The other secondary structure elements of the C-terminus, α-helix (α8) and a short 3_{10} helix, are disrupted by three proline residues. These two helices are orientated at the rear of the domain interface and interact with the N-terminal domain via residues of the conserved QPD motif at the C-terminus which form a number of interactions with the N-terminal domain and prevent the intrusion of the C-terminal region into the active site.

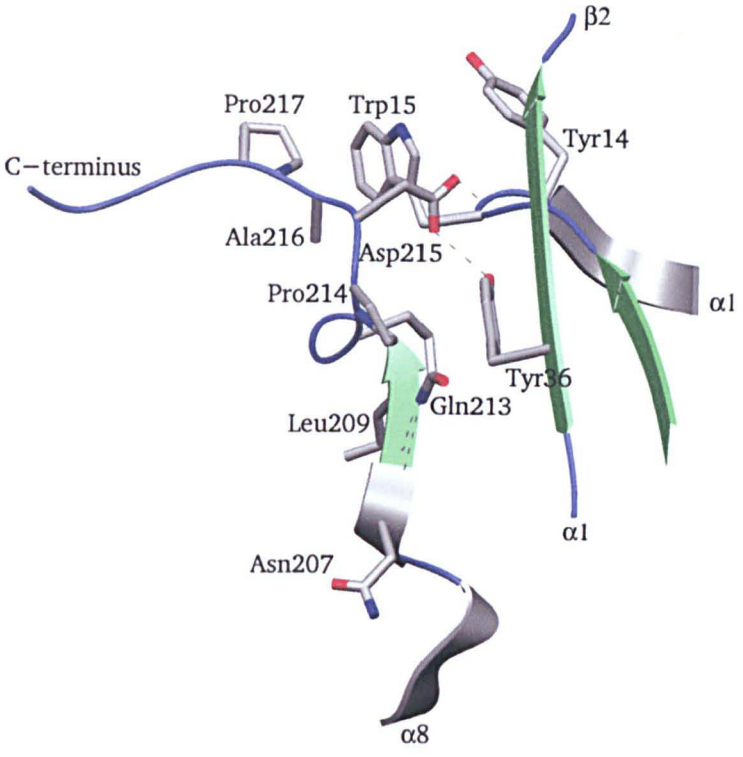


Figure 6-12: Interaction of QPD motif with N-terminal domain.

From this motif, Gln213 forms hydrogen bonds via NE2 to the carbonyl oxygens of Asn207 and Leu209, Pro214 stacks against Tyr36 and Asp215 forms hydrogen bonds with both the hydroxyl of Tyr36 and the amide nitrogen of Trp15 in the N-terminal domain. These interactions together with the packing of Ala216 against the indole ring of Trp15 orientates Pro217 such that it directs the C-terminus away from the active site (figure 6-12).

The superimposition of the dimers of the various GST classes on *At*GSTZ1 reveals large structural differences. These differences are due to the differential packing of α3 and α4 to their dimeric mates, which in turn results in the significant shifts observed in the position of the two monomer units within each dimer. The Theta, Phi and Mu classes are more similar to *At*GSTZ1 with helices α3 and α4 packing closer, than in Alpha, Pi and Omega class structures. The subunit interface of the Zeta class GST is predominately hydrophobic as observed in other classes of GST. However, it does not contain the Phe and Val interactions 'lock and key' which are strictly conserved in previous plant structures nor the stacking of the two symmetrically equivalent Arg residues observed in both maize and mammalian structures. The most significant polar

Original in Colour

interaction occurring between the monomer subunits of *At*GSTZ1 is found at the base of helices α3 and α4, between the strictly conserved residues, Glu82 and Arg97.

6.4.3 Comparisons of active site of *At*GSTZ1 and proposed mechanism

Structural studies of mammalian and squid GSTs have indicated that a tyrosine residue, located at the C-terminal end of β1, is the catalytic residue responsible for the stabilisation of the thiolate anion of enzyme bound glutathione (Dirr et al., 1994). In the GSTs from *L. cuprina* (Board et al., 1995) and plants (Reinemer et al., 1996), a serine residue at the N-terminus of α1 replaces the function of this tyrosine residue. The *At*GSTZ1 enzyme contains two tyrosine residues at the N-terminus, Tyr12 and Tyr14, however from the structure it is clear that these residues are not in a position within the active site to stabilise the glutathione thiolate ion. Analysis of the *At*GSTZ1 structure indicates that Ser17 lies at the N-terminal end of α1 and superimposes well on the catalytic serine of the Theta and Phi GSTs suggesting this serine residue is the catalytic residue in the Zeta class. In addition, the active site structure of *At*GSTZ1 shares the characteristic glutathione backbone conformation observed in the other classes of GST (Wilce & Parker, 1994) together with other residues previously identified as being important in the binding of glutathione, such as Asp72.

Although displaying some differences, the nature of the residues present in the G-site and the interactions with glutathione are comparable in all GSTs. In contrast the H-site is known to display considerable variability between structures, with the Phi class H-site known to be broader and deeper than mammalian isoforms, while the Alpha and Theta classes have been shown to contain an additional C-terminal element which is orientated over the top of the active site.

*At*GSTZ1 has a novel H-site, which is hydrophilic, with several basic residues orientated into the active site. The basic residues Arg22 and Arg181 are highly conserved throughout the Zeta class enzymes and provide a method by which the enzyme can selectively bind the negatively charged substrate maleylacetoacetate in preference to the hydrophobic xenobiotics conjugated by other classes of GST. In order to bind the MAA substrate the active site of the Zeta class enzyme is

required to have both hydrophobic and hydrophilic character. These two characteristics are observed in the active site of *At*GSTZ1, with basic residues in a position that enables the enzyme to bind the carbonyl groups of the substrate and hydrophobic residues Tyr14, Trp15, Leu41 and Leu42 at a second position that enables interaction with the aliphatic Cα backbone of the substrate. A means by which MAA might interact with the enzyme is suggested by the conformation of the unproductively bound S-hexyl glutathione molecule in the *A. thaliana* structure (Reinemer et al., 1996). Given the position of the two S-hexylglutathione molecules in the *A. thaliana* structure the method by which substrate might interact with the enzyme and the co-factor can be postulated even though these are absent in *At*GSTZ1. The unsaturated carboxyl of MAA is postulated to interact with the backbone amides of residues Ser18 and Cys19 in an orientation similar to the glycine carboxylate of the second S-hexylglutathione. In this position the double bond of MAA would be in close proximity to the thiol of the productively bound S-hexylglutathione molecule, permitting conjugation. In this orientation Cys19 would not play a catalytic role in the isomerisation reaction. This hypothesis supports the observation that treatment of *At*GSTZ1 with iodoacetate does not inhibit the isomerisation of maleylacetoacetone (David Dixon, unpublished) and also suggests why bacterial MAAI sequences lack this cysteine. In addition, recent work presented by Coggan and co-workers (Coggan et al, 2000) has shown that the human Zeta enzyme is inactivated by the mutation Ser14Ala whereas Cys16Ala increases the enzyme activity.

In contrast to other classes of GSTs that act to irreversibly conjugate glutathione to substrate, *At*GSTZ1 conjugates then removes glutathione from the substrate during the isomerisation reaction. In combination with some recent studies on TCHQ dehalogenase and the human Zeta class isoenzyme (McCarthy et al., 1997; Tong et al., 1998a) the structure of *At*GSTZ1 indicates a possible mechanism for the isomerisation of MAA. The results of this work suggest that Zeta class GSTs perform the conversion of maleylacetylacetate to fumarylacetylacetate using Ser17, to stabilise the thiolate ion of glutathione and catalyse the nucleophilic addition of the glutathione moiety to C-2 of

maleylacetoacetate. This promotes the formation of an enolate followed by rotation around the resultant single bond. The elimination of glutathione completes the tautomerization of *cis*-maleylacetoacetate to fumarylacetoacetate (Figure 6-13).

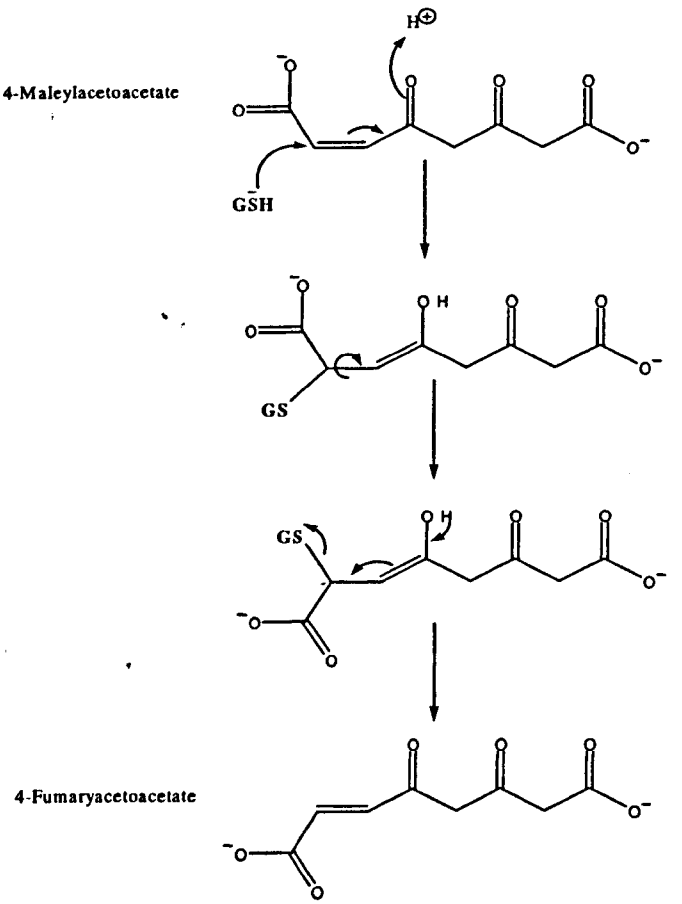


Figure 6-13: Proposed mechanism for isomerisation of Maleylacetoacetate to Fumarylacetoacetate.

Tetrachlorohydroquinone (TCHQ) dehalogenase, has been shown to perform both the two step reductive dehalogenation of tetrachlorohydroquinone to 2,6-dichlorohydroquinone and also the isomerisation of Maleylacetoacetone to Fumarylacetoacetate (Anandarajah et al., 2000). *At*GSTZ1 and TCHQ dehalogenase both only require one molecule of glutathione to perform the isomerisation reaction. In contrast the dehalogenase reaction requires two molecules of glutathione (figure 6-14).

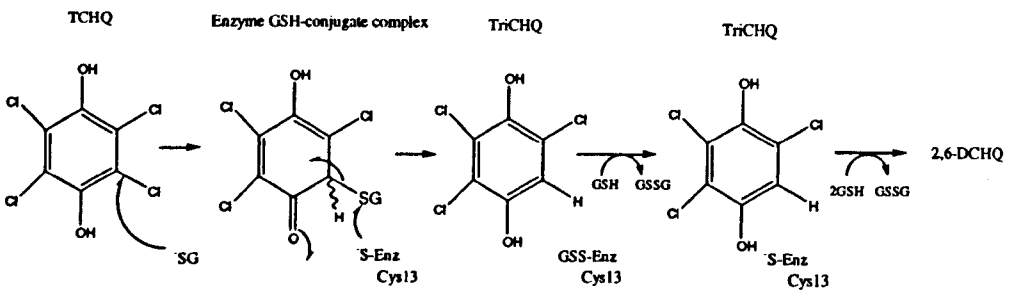


Figure 6-14: Postulated mechanism of TCHQ dehalogenase (Wilett & Copley, 1996).

The dehalogenase Cys13Ser mutant is able to conjugate glutathione to form 2,3,5-trichloro-6-*S*-glutathionyl-4-hydroxycyclohexa-2,4-dienone, however it is unable to complete the reaction to form TriCHQ. It has been proposed that the active site cysteine residue is required to attack the glutathionyl conjugate and that the second molecule of glutathione is then used to regenerate the free enzyme (McCarthy et al., 1996). The requirement of two molecules of glutathione to perform the dehalogenase reaction and the need for Cys13 to be present suggests that the dehalogenase reaction is mechanistically different from the isomerase reaction. The reason the cysteine residue is highly conserved in the Zeta class enzyme in mammals, plants, insects and fungi, but not bacteria is currently unknown, but it may be that these higher organisms have a need for this enzyme to catalyse an additional endogenous reaction.

The Zeta class glutathione S-transferases show distinct catalytic activity in comparison to other GST classes, displaying a clear endogenous function as maleylacetacterate isomerases in the tyrosine and phenylalanine degradation pathway. Despite sharing a similar overall fold to other classes of cytosolic GSTs, *At*GSTZ1 shows significantly different interactions both at the domain and dimer interface. In addition this isoenzyme has a novel active site which contains both a serine, crucial for isomerase activity, and a highly conserved cysteine residue. The active site sequence conservation of *At*GSTZ1 and TCHQ dehalogenase is remarkable given that both enzymes catalyse quite different reactions. It has been suggested that TCHQ and other dehalogenases may have arisen from MAAI enzymes (Anandarajah et al., 2000) due to evolutionary pressure on existing enzymes to optimise new catalytic functions.

Structure function studies and the generation of mutant enzymes are required to allow further the characterisation of important residues within this structure.

Chapter 7: Structure solution AtGSTT1

DNA sequence searches of the plant EST databases using mammalian GST sequences identified an *Arabidopsis thaliana* EST which showed higher homology to the mammalian Theta class GSTs than to any plant GST sequence (Edwards et al., 2000). In contrast to previously described plant GSTs shown to contain 1, 2 or 9 introns, this *A. thaliana* sequence (classified as an *Arabidopsis thaliania* Theta class GST (*At*GSTT1)) contained five introns (mammalian Theta class GSTs have four). Further, the location of two of the five introns were found to be conserved with those observed in mammalian Theta sequences suggesting a common ancestral origin. Characterisation of *At*GSTT1 revealed it to had low activity towards CDNB (2.0 nkat/mg), higher activity towards nitrobenzylchloride (21.7 nkat/mg) and no activity towards benzyl isothiocyanate or ethancrynic acid. As this GST differed significantly in sequence from the other GSTs under study, it was included in crystallisation trials to enable determination of the regions involved in dictating substrate specificity. Following optimisation of crystallisation conditions (section 3.3.4.1) crystals of *At*GSTT1 were grown which allowed data to be collected to 2.6Å at station X31, DESY, Hamburg (section 4.2.6). Processing and scaling suggested that the crystal belonged to the orthorhombic spacegroups $P2_12_12_1$ or $P2_12_12$. The Matthews coefficient (Matthews, 1968) calculated for these spacegroups suggested that four molecules existed in the asymmetric unit.

Comparison of the primary sequence of *At*GSTT1 with sequences of GST structures in the Protein Data Bank suggested it shared the highest pairwise identity (34%) with the human Theta class GST structure (1LJR) (figure 8-1). Based on this sequence identity it was thought that molecular replacement techniques would be able to provide a solution to the phase problem. Despite trying various modifications to the search models and varying the resolution and packing constraints, all attempts to molecular replace *At*GSTT1 using the program AMoRe (Navaza, 1994) were unsuccessful.

```
                                                                    50
     ARA_4  MMKLKVYADR MSQPSRAVII FCKVNGIQFD EVLISLAKRQ QLSPEFKDIN PLGKVPAIVD
      1ljr   MGLELFLDL VSQPSRAVYI FAKKNGIPLE LRTVDLVKGQ HKSKEFLQIN SLGKLPTLKD
 Consensus   M L    D  SQPSRAV I F K NGI        L K Q   S EF IN  LGK P   D
                                                   100
     ARA_4  GRLKLFESHA ILIYLSSAFP SVADHWYPND LSKRAKIHSV LDWHHTNLRR GAAGYVLNSV
      1ljr  GDFILTESSA ILIYLSCKYQ T-PDHWYPSD LQARARVHEY LGWHADCIRG TFGIPLWVQV
 Consensus  G   L ES A ILIYLS       DHWYP D L  RA  H   L WH    R          V
                                                   150
     ARA_4  LGPALGLPLN PKAAAEAEQL LTKSLSTLET FWLKGNAKFL LGSNQPSIAD LSLVCELMQL
      1ljr  LGPLIGVQVP EEKVERNRTA MDQALQWLED KFL-GDRPFL AGQ-QVTLAD LMALEELMQP
 Consensus  LGP  G                  L  LE    L G   FL  G  Q   AD L    ELMQ
                                       200
     ARA_4  QVLDDKDRLR LLSTHKKVEQ WIENTKKATM PHF-DETHEI LFKVKEGFQK RREMGTLSKP
      1ljr  VALG----YE LFEGRPRLAA WRGRVEAFLG AELCQEAHSI ILSILEQAAK KT-LPTPSPE
 Consensus    L       L          W               E H I      E   K      T S
            240
     ARA_4  GLQSKI
      1ljr  AYQAMLLRIA RIP
 Consensus    Q
```

Figure 7-1: Sequence comparison of *At*GSTT1 with the human type Theta GST (1LJR). Strongly conserved residues are shown in red. This alignment was performed using the program Multialign (Corpet, 1988).

As molecular replacement using AmoRe had proven unsuccessful it was thought that isomorphous replacement techniques or multiwavelength anomalous scattering techniques would have to be employed to gain phase information. Selenomethionine protein was prepared as described in the methods section of this text and heavy atom soaks were prepared, however it was also decided to try a second molecular replacement program EPMR (Kissinger et al, 1999), that unlike AmoRe performs a limited six dimensional search using evolutionary programming.

In contrast to AMoRe, EPMR successfully determined a solution to the molecular replacement problem which had a significantly lower R factor and higher correlation coefficient than the other orientations and translations tested. The electron density map generated using the phases calculated from molecular replacement were sufficient to enable building of the model, and as a result of this unexpected success, the other methods of phase determination were abandoned.

Original in colour

7.1 Molecular Replacement Methodologies

As discussed previously (section 5.1), molecular replacement requires the correct orientation and positioning of an approximate structural model in the asymmetric unit. This provides initial phase estimates that can be improved upon by rebuilding and refinement to allow determination of the final structure. The success of molecular replacement relies critically on the probe model being homologous to the unknown structure, as even relatively slight structural differences between the probe and the unknown structure can prevent a solution being found. Traditionally molecular replacement has been separated into two components, a rotation function and a translation function, to greatly enhance the speed of the molecular replacement calculation. However this division of the molecular replacement problem does have a number of disadvantages with the rotation search suffering from a low signal to noise ratio and the translation function being very susceptible to errors in the rotation search. Current molecular replacement programs based on the traditional rotation then translation searches have developed strategies to try and overcome these problems. AMoRe (Navaza, 1994) automatically performs translation searches on all of the top solutions from the orientation function, while the molecular replacement package in X-plor performs Patterson Correlation refinement (Brunger, 1990) to optimise the solution from the rotation function before performing the translation function.

Molecular replacement based on six dimensional searches is more sensitive and reliable than the traditional methods, but despite considerable advances in computational power, systematic searching of six dimensional space is still not generally feasible due to the large number of calculations required (Sheriff et al, 1999). Methods utilising genetic algorithms (Chang et al, 1997) or evolutionary programming (Kissinger et al, 1999), which allow limited searching of six dimensional space, have been suggested as being able to solve more difficult molecular replacement problems, while greatly reducing the computational time required to search the increased dimensions of parameter space.

7.1.1 Evolutionary programming

EPMR uses evolutionary programming to efficiently search six dimensional space, allowing it to overcome the disadvantages of the split rotational then

translational search methods. This causes it to be significantly less computationally intensive than systematic six dimensional search methods (Kissinger et al, 1999).

Evolutionary programming is a search and learning method which has proven useful in solving difficult optimisation problems where there is no differentiable objective function, many local optima or in cases where a large number of parameters exist. It is broadly similar to genetic algorithms in that it is based on Darwin's theory of natural selection, whereby more individuals are produced in every generation than can survive and reproduce. Phenotypic variation exists among individuals of each generation, some of which is heritable. If individuals with particular phenotypes or traits are able to survive and reproduce better than others, then more individuals with said particular phenotypes will be present in the next generation.

Within this context a population evolves according to rules of recombination and mutation whereby each individual within the population is evaluated by a measure of fitness that determines if it survives to pass on its traits.

Using this basis then in more detail, evolutionary programming generates an optimised solution by the following procedure. A random initial population (P) is chosen in which all individuals of that population are thought of as parents. The characteristics of these parents are mutated to produce an offspring population (2P), which differ from the parents according to a normal distribution. Each individual belonging to the parent and the children population (2P) is evaluated according to a fitness function in a tournament selection to determine the survivors (P1). These survivors undergo further mutation and the selection criteria is applied again. As the tournament selection always ensures that the best individual in terms of the fitness function is retained, an optimum once found cannot be lost. Although genetic algorithms (GA) are similar to evolutionary programming, they differ in that in GA characteristics are assigned to particular structures (part of a chromosome) and the crossover and recombination of these structures generates new offspring (figure 7-2) rather than mutation of the phenotype.

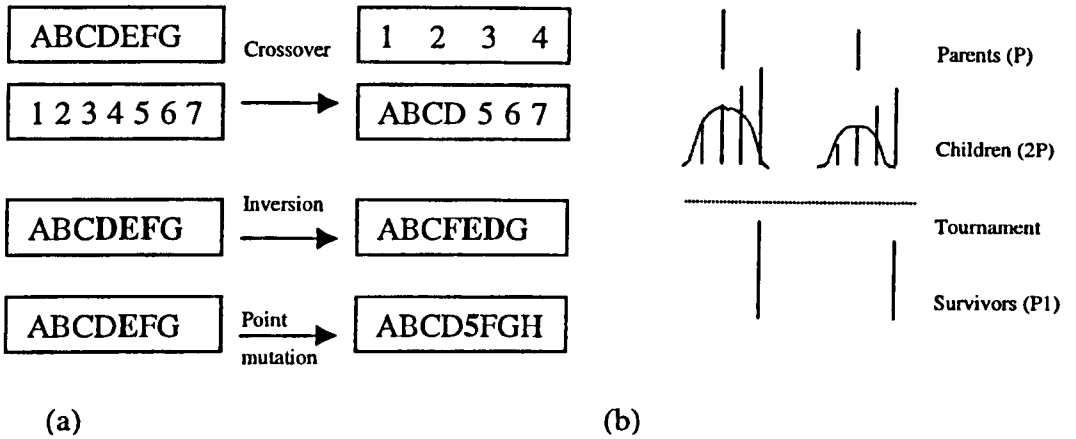


Figure 7-2: (a) describes the operators used by genetic algorithms, where each character is taken to be a heritable feature of phenotype. The heritable features forming a string or chromosome. (b) describes evolutionary strategies where each phenotype is mutated according to a normal distribution in each generation, and a measure of fitness (in this case height) selects those individuals with the most advantageous phenotype according to a scoring system.

In the molecular replacement problem evolutionary programming allows the simultaneous optimisation of the orientation and position of a search model with an aim to successively optimise initial trial solutions until a global optimum is found by repetition of three basic steps

- An initial selection of trial solutions is made.
- Addition of random noise (mutation) to each solution. This generates offspring similar to the initial population, but enables local minima to be escaped and parameter space to be explored.
- Ranking of trial solutions. In accordance with a function of its fitness, the 'best' solutions are retained while the rest are culled.

In more detail, EPMR aims to obtain a molecular replacement solution as described in figure 7-3.

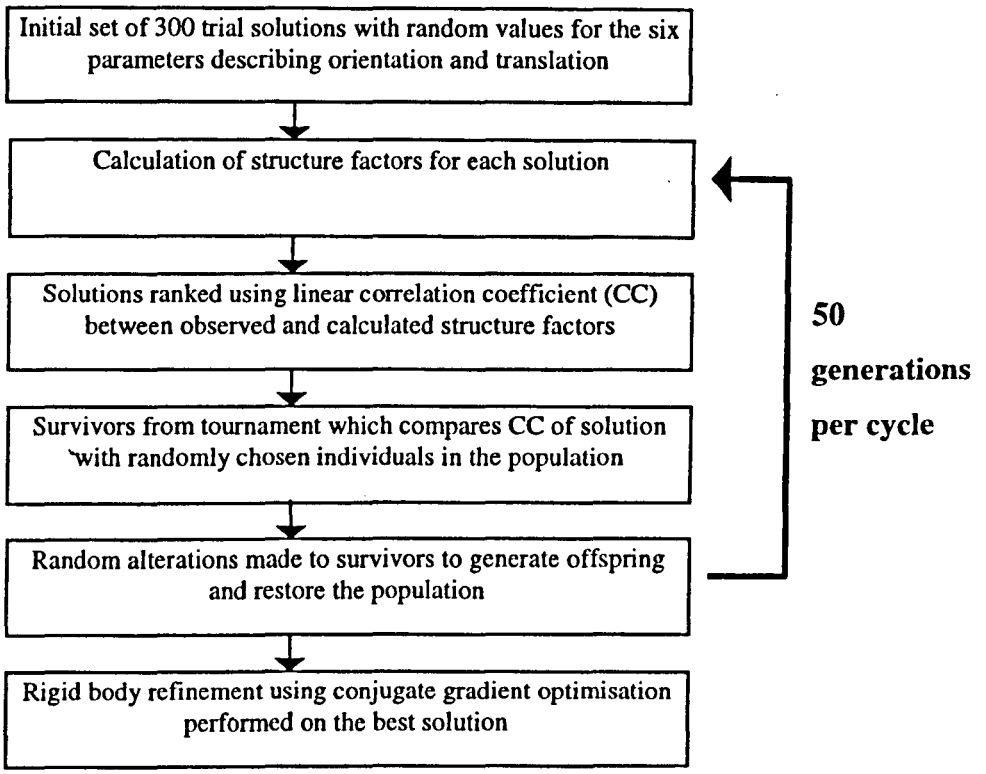


Figure 7-3: Flowchart of Molecular Replacement by EPMR adapted from (Kissinger et al, 1999)

7.2 Solution of phases using EPMR

Using a truncated model of the human Theta GST, (PDB entry 1LJR truncated to residue 180) results from EPMR appeared to favour the spacegroup $P2_12_12_1$ in comparison to $P2_12_12$, however no clear solution was found (Table 7.1). Visual examination of the models generated by the top solutions from both the $P2_12_12_1$ and $P2_12_12$ trials identified that they both suffered from poor packing, with serious clashes between symmetry related molecules in the asymmetric unit.

$P2_12_12_1$	Correlation Coefficient	R factor
Soln 1	0.228	0.621
Soln 2	0.205	0.628
Soln 3	0.209	0.627
Soln 4	0.212	0.628
Soln 5	0.206	0.627
Soln 6	0.216	0.631
Soln 7	0.198	0.631
Soln 8	0.204	0.634
Soln 9	0.203	0.635
Soln 10	0.204	0.629

$P2_12_12$	Correlation Coefficient	R factor
Soln 1	0.134	0.691
Soln 2	0.142	0.687
Soln 3	0.133	0.687
Soln 4	0.145	0.683
Soln 5	0.130	0.684
Soln 6	0.139	0.692
Soln 7	0.139	0.691
Soln 8	0.140	0.685
Soln 9	0.137	0.689
Soln 10	0.142	0.678

Positioning of second dimer in asymmetric unit

Soln 1	0.240	0.625
Soln 2	0.240	0.626
Soln 3	0.231	0.626
Soln 4	0.239	0.619
Soln 5	0.247	0.615
Soln 6	0.236	0.617
Soln 7	0.233	0.633
Soln 8	0.242	-0.618
Soln 9	0.224	0.625
Soln 10	0.242	0.622

Soln 1	0.168	0.677
Soln 2	0.163	0.686
Soln 3	0.164	0.684
Soln 4	0.161	0.682
Soln 5	0.161	0.677
Soln 6	0.163	0.681
Soln 7	0.173	0.679
Soln 8	0.164	0.680
Soln 9	0.156	0.681
Soln 10	0.161	0.681

Table 7-1: Solutions generated using a truncated model of human Theta class enzyme in spacegroups $P2_12_12_1$ and $P2_12_12$. As a dimer was used to search for four molecules in the asymmetric unit the program separates finding each dimer into a separate solution using the top solution from the first search as a partial solution during the second. Solution found in $P2_12_12_1$ shows larger correlation coefficient and lower R factors that those of $P2_12_12$, although no solution stands out with a lower correlation coefficient and R factor than rest.

A minimum bump distance (smallest distance allowed between centre of mass of a solution and that of a symmetry mate) of 30Å was included and the program rerun. The inclusion of this bump parameter generated four solutions with significantly higher correlation coefficients and lower R factors (Table 7.2) in spacegroup $P2_12_12_1$. Visual examination of the top solution identified only small clashes between the symmetry related molecules within a loop region. Following calculation of a 2Fo-Fc map, it was clear that a solution had been gained with the map clearly covering the molecule with 'extra' density present. This 'extra' density corresponded to sequence differences between *At*GSTT1 and the probe structure for example density was observed for Phe66 and His69, which replaced residues Thr65 and Ser68 respectively in the probe structure.

$P2_12_12_1$	Rotation			Translation			Correlation coefficient	R factor
Soln 1	123.37	122.58	336.64	23.82	8.78	34.10	0.207	0.630
Soln 2	77.37	79.17	79.71	72.40	117.45	79.98	0.221	0.626
Soln 3	36.39	205.66	153.69	85.51	87.59	20.30	0.255	0.607
Default threshold correlation coefficient of 0.3 was passed and so program began to position second dimer in asymmetric unit								
Soln 1	131.12	95.96	246.62	30.19	119.54	100.76	0.254	0.611
Soln 2	155.86	100.33	228.86	24.07	31.15	150.97	0.253	0.612
Soln 3	131.13	83.52	37.19	3.70	11.73	155.05	0.249	0.615
Soln 4	128.91	73.44	56.67	11.10	47.20	9.14	0.252	0.615
Soln 5	303.33	101.33	312.85	3.66	11.50	15.30	0.382	0.560
Soln 6	303.29	101.34	312.82	3.68	11.47	15.30	0.382	0.560
Soln 7	125.28	75.20	37.80	22.09	8.78	25.43	0.250	0.615
Soln 8	124.13	77.49	230.69	35.41	32.62	150.21	0.253	0.611
Soln 9	123.34	78.69	47.20	46.88	77.37	63.24	0.382	0.560
Soln 10	156.62	101.36	226.89	46.87	77.42	63.27	0.382	0.560

Table 7-2: Table of solutions gained using EPMR with bump radius of 30 Å in spacegroup $P2_12_12_1$. Top solutions displaying significantly higher correlation coefficient and lower R factor are shaded grey. The rotations are shown as Eulerian angles as defined in X-plor while the X, Y and Z translations are in orthogonal Anstroms to be applied to the search model after it has been centered at the origin.

Refinement using simulated annealing in CNS (Brunger et al, 1998) with strict non-crystallographic symmetry between the four molecules in the asymmetric unit followed by manual rebuilding into averaged maps significantly improved the structure (R_{free} 32.3%, R_{cryst} 31.5%). It was hoped that this partially refined structure could be refined against previously collected higher resolution data (2.3Å collected at the Hamburg Beamline BW7B - section 4.2.6), however this proved impossible, as the two data sets were not isomorphous. The partially refined structure was used therefore as probe in molecular replacement against the previously collected higher resolution data. Using a dimer model of the partially built structure a clear solution (figure 7-4) was found in AMoRe for the higher resolution data. The model generated from this solution displayed similar packing to that observed by the model for the 2.6Å data.

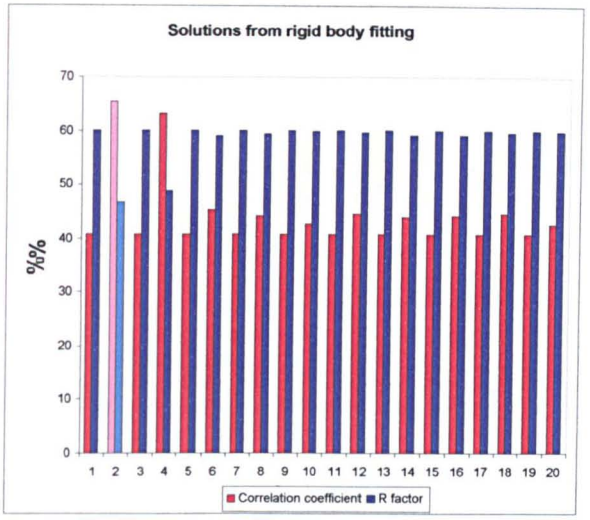


Figure 7-4: Graph depicting the correlation coefficient (red) and R factor (blue) for the solutions gained in AMoRe for the 2.3Å data using the model built with the 2.6Å data. Solution is shown in pink and light blue.

Following one round of rebuilding using the higher resolution data it became clear that the four molecules in the asymmetric unit were not identical within the α4-α5 loop region among the different chains, and the NCS restraints were released for this part of the model. Further refinement and rebuilding allowed complete tracing of the chain for molecules A, B and D (Two residues Leu125 and Gly126 in the loop region α4-α5 in chain C could not be traced due to poor density). The final model comprised four protein molecules and four sulphate ions. The quality of the refined structure was assessed with PROCHECK (Laskowski et al., 1993) which indicated that the model displayed stereochemical parameters consistent with models of 2.3Å resolution. The Ramachandran plot indicated that 90.2% of the residues are within the most favoured regions, with a further 9.2% present in favourable regions. Only residues Glu67 from all four chains and Asn107 from molecule D were found to lie in the generously allowed regions of the plot. Glu67 binds the γ-Glu moiety of GSH and has been found in the generously allowed region of the Ramachandran plot in all GST structures solved to date. Asn107 lies within a loop region of the molecule and it is difficult to refine its position due to the relatively poor density of this region in molecule D. Overall the model shows good stereochemistry with an estimated coordinate error (r.m.s.) of 0.94 (Cruikshank, 1999) and a final r.m.s. deviation from ideal bond lengths and angles of 0.007Å and 1.7° respectively. The parameters for the refined model are summarised in Table 7.3.

Original in colour

Resolution Range (Å)	30-2.3
No of unique reflections in resolution range	41718
R_{Work} (%)[a]	22.0
R_{Free} (%)	27.5
Model	
Total number of amino acids	917
No. of Solvent atoms (excl. H)	146
No. of Heteroatoms	4
Stereochemistry[b]	
Ramachandran quality, % in	
Most favoured regions	90.2
Allowed regions	9.2
R.M.S. deviation[c]	
Bond length (Å)	0.007
Bond angle (degrees)	1.3

[a] $R = \Sigma_{hkl}(|F_o - F_c|) / \Sigma|F_o|$.
[b] Calculated using the program PROCHECK (Laskowski et al., 1993)
[c] (Cruikshank, 1999)

Table 7-3: Final refinement statistics for *At*GSTT1 structure

7.3 Overall structure

The *Arabidopsis thaliana* Theta class structure (*At*GSTT1) contains two dimers in the asymmetric unit with each monomer adopting the canonical GST fold, comprised of two domains separated by a linker region. Superimposition of the four monomers shows that pairs of molecules A and C are more similar, as are molecules B and D (r.m.s.d. ~0.3Å over 228 residues for each pair). Significant differences in the α4-α5 loop region between the pairs of monomers are observed with an overall r.m.s.d. of the four molecules ~0.7Å over 214 residues. These differences are due to crystal contacts between a symmetry related region of the B and D monomer subunits (residues 126-130), similar to that observed for the atrazine-GSH conjugate bound *Zm*GSTF1 structure (Prade et al., 1998). These crystal contacts lead to stabilisation of the α4-α5 loop region in the B and

D subunits causing them to adopt a different to subunits A and C, as depicted in figure 7-5. In subunits A and C this loop region is not stabilised by crystal contacts and a more typical conformation of the protein in solution is observed. As monomer A could be fully traced and also shows no distortion due to crystal contacts this subunit will be considered alone in further descriptions and discussions of the *At*GSTT1 monomer structure.

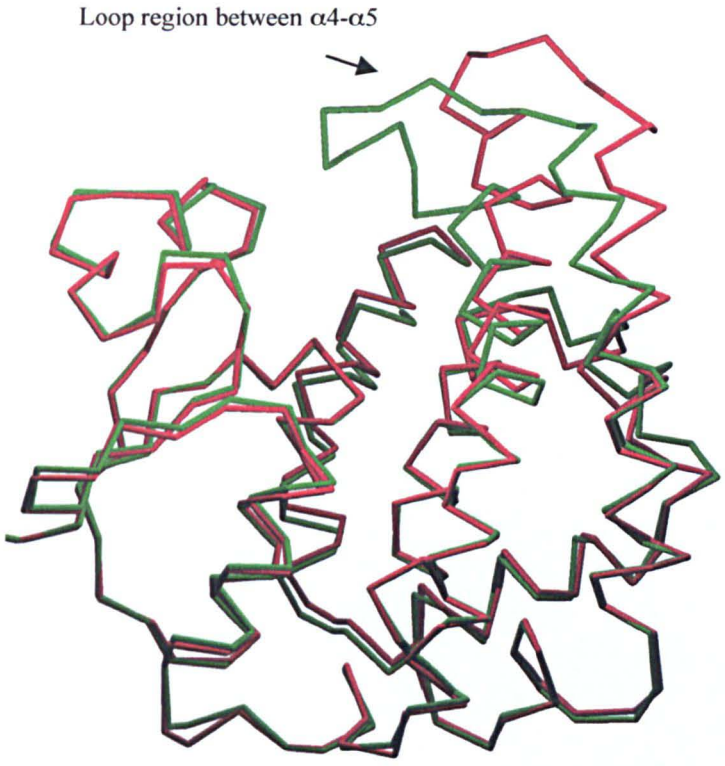


Figure 7-5: Superimposition of molecules A (shares close structural similarity to molecule C) and B (shares close structural similarity to molecule D) from *At*GSTT1 structure. Crystal contacts cause the loop region between α4-α5 to be artificially fixed in the B and D monomer subunits. Molecule A is shown in green and molecule B in red.

The smaller alpha-beta N-terminal domain (residues 1-79) consists of a four stranded β-sheet (residues β1 5-8, β2 30-33, β3 57-60, β4 63-66) flanked on one side by two α-helices (α1 12-25 and α3 67-79) and on the other by an irregular loop region. This irregular loop region contains a short 3_{10} helix (39-42) and an α-helix (43-50) (The secondary structure nomenclature has been made consistent with other published Theta class structures). A linker region of 10 residues containing a 3_{10} helical element joins the N-terminal domain to the larger entirely helical C-terminal domain. This second domain contains five partially

Original in colour

interrupted helices with helix α4, formed by two helical elements α4` (residues 90-107) and α4`` (residues 107-120), before a series of β–turns forms the loop to α5 (residues 130-153). A short 3_{10} helical element (residues 159-163) joins α5 to α6 (residues 167-179) with α6 packing on one side into the core of the C-terminal domain while on a second side it forms interactions across the domain interface via a series of van der Waals contacts and a salt bridge between Glu179 and Gln13. Helix α6b (residues 184-193, previously observed as an unstructured region in the human Theta class GSTs) and helix α7 (residues 195-208) pack against the rear of the C-terminal domain before forming the partially disrupted helix α8 (α8` 211-217, α8`` 217-233).

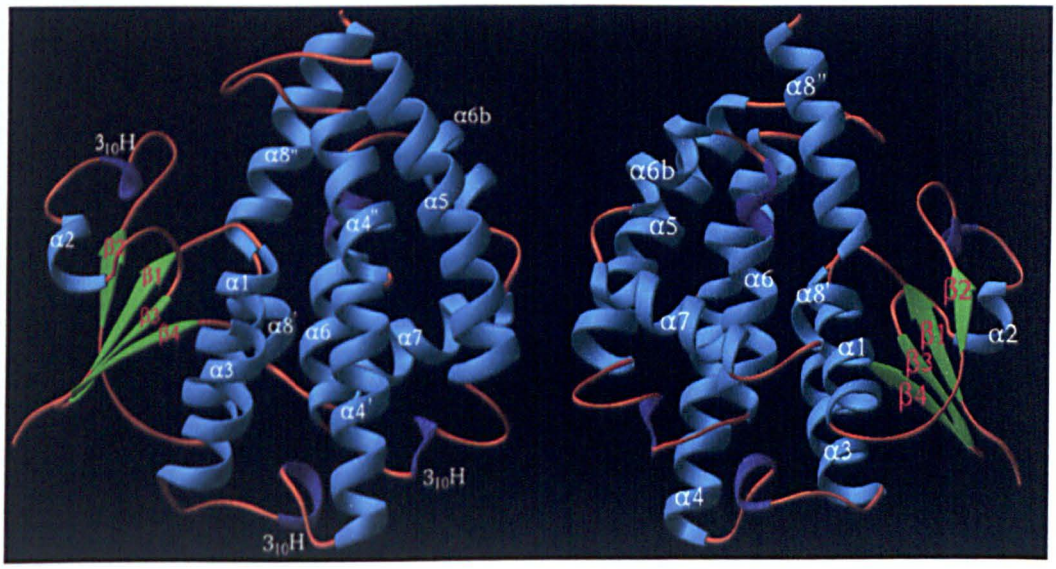


Figure 7-6: Structural elements present in *At*GSTT1. The picture on the left hand side picture shows conventional view, (from dimer interface) while that on the right shows the view from the back. Helices are shown in light blue, ribbons in green and 3_{10} helices in dark blue.

7.4 Discussion

A number of structures classified as Theta class GSTs from human, plant and insect sources have been compared with the mammalian GST classes, Alpha, Mu and Pi in the literature (Rossjohn et al., 1998; Wilce et al., 1995), therefore a detailed comparison will not be presented here. Subsequent subdivision of the Theta class now means that the insect (*Lucilia cuprina* (Wilce et al., 1995)), *Arabidopsis thaliana* (1GNW (Reinemer et al., 1996)) and *Zea mays* (1AXD, 1AW9 (Neuefeind et al., 1997a; Neuefeind et al., 1997b)) GSTs have now been

Original in colour

separated from the human Theta class GST (hGSTT2-2) (Rossjohn et al., 1998). The insect and plant GSTs have now been respectively assigned to the species specific classes of Delta and Phi. Due to the historical grouping of the Delta and Phi classes with the Theta class, both the Delta and Phi class GSTs have been included in the comparison of *At*GSTT1 with the structure of the human Theta class GST. *At*GSTT1 shows 34% identity with the human Theta class GST compared to 25% with the insect GST and around 20% with the plant GSTs (figure 7-7). In addition to sequence similarity *At*GSTT1 also shares a number of structural features with these enzymes.

Many of the key catalytic residues identified in these structures are strongly conserved, for example Ser12, *cis*-Pro56 and Glu67 of the glutathione binding domain. Further a number of structurally important residues are also conserved such as Phe45, which enables packing of the β2-β3 loop, the buried polar residue Asp170 and the aromatic residue Trp201, which are important in packing of the C-terminal helices. However, particularly within the C-terminal domain, significant variation in residues is observed between the sequences of the Theta, Phi and Delta classes with only ~30% of residues being strongly conserved in the C-terminal domain, while ~50% of residues are strongly conserved between the *At*GSTT1 and hGSTT2-2 in the N-terminal domain.

```
                  10         20         30         40         50         60
    AtGSTT1  MMKLKVYADR MSQPSRAVII FCKVNGIQFD EVLISLAKRQ QLSPEFKDIN PLGKVPAIVD
   hGSTT2-2  MGLELFLDL VSQPSRAVYI FAKKNGIPLE LRTVDLVKGQ HKSKEFLQIN SLGKLPTLKD
 L.cuprina     MDFYYLP GSAPCRSVLM TAKALGIELN KKLLNLQAGE HLKPEFLKIN PQHTIPTLVD
    ZmGSTF1  APMKLYGAV MSWNLTRCAT ALEEAGSDYE IVPINFATAE HKSPEHLVRN PFGQVPALQD
    ZmGSTF3  APLKLYGMP LSPNVVRVAT VLNEKGLDFE IVPVDLTTGA HKQPDFLALN PFGQIPALVD
    AtGSTT2  AGIKVFGHP ASIATRRVLI ALHEKNLDFE LVHVELKDGE HKKEPFLSRN PFGQVPAFED
  Consensus             S                                       N      P   D

                   70         80              90         100        110
    AtGSTT1  GRLKLFESHA ILIYLSSAFP SVADHWYP-- -NDLSKRAKI HSVLDWHHTN LRRGAAGYVL
   hGSTT2-2  GDFILTESSA ILIYLSCKYQ T-PDHWYP-- -SDLQARARV HEYLGWHADC IRGTFGIPLW
 L.cuprina   GDFALWESRA IMVYLVEKYG K-NDSLFP-- -KCPKKRAVI NQRLYFDMGT LYKSFADYYY
    ZmGSTF1  GDLYLFESRA ICKYAARKNK PELLRE-G-- --NLEEAAMV DVWIEVEANQ YTAALNPILF
    ZmGSTF3  GDEVLFESRA INRYIASKYA SEGTDLLP-- --ATASAAKL EVWLEVESHH FYPNASPLVF
    AtGSTT2  GDLKLFESRA ITQYIAHRYE NQGTNLLQTD SKNISQYAIM AIGMQVEDHQ FDPVASKLAF
  Consensus  G   L ES A I  Y                        A

                  120        130        140        150        160        170
    AtGSTT1  NSVLGPALGL PLNPKAAAEA EQLLTKSLST LETFWLKGNA KFLLGSNQPS IADLSLVCEL
   hGSTT2-2  VQVLGPLIGV QVPEEKVERN RTAMDQALQW LEDKFL-GDR PFLAGQ-QVT LADLMALEEL
 L. cuprina  PQIFAKA--- PADPELYKKM EAAFDFLNTF LE------GH QYVAGD-SLT VADLALLASV
    ZmGSTF1  QVLISPMLGG TTDQKVVDEN LEKLKKVLEV YEARLT--KC KYLAGD-FLS LADLNHVSVT
    ZmGSTF3  QLLVRPLLGG APDAAVVDKH AEQLAKVLDV YEAHLA--RN KYLAGD-EFT LADANHASYL
    AtGSTF2  EQIFKSIYGL TTDEAVVAEE EAKLAKVLDV YEARLK--EF KYLAGE-TFT LTDLHHIPAI
  Consensus                                   E                    D

                  180        190        200        210        220        230
    AtGSTT1  MQLQVLDDKD RLRLLSTHKK VEQWIENTKK ATMPHF-DET HEILFKVKEG FQKRREMGTL
   hGSTT2-2  MQPVALG--- -YELFEGRPR LAAWRGRVEA FLGAELCQEA HSIILSILEQ AAKKT-LPTP
 L. cuprina  STFEVAG--- --FDFSKYAN VAKWYANAKT -VAPGFDENW EGCL
    ZmGSTF1  LCLFATP--- YASVLDAYPH VKAWWSGLME --RPSVQK-V AALMKPSA
    ZmGSTF3  LYLSKTP--- KAGLVAARPH VKAWWEAIVA --RPAFQKTV AAIPLPPPP
    AtGSTF2  QYLLGTP--- TKKLFTERPR VNEWVAEITK --RPASEKVQ
  Consensus                        W

              240
    AtGSTT1  SKPGLQSKI
   hGSTT2-2  SPEAYQAMLL RIARIP
 L. cuprina
    ZmGSTF1
    ZmGSTF3
    AtGSTF2
  Consensus
```

Figure 7-7: Sequence comparison of structurally characterised Theta class GSTs, *At*GSTT1, human (PDB code 1LJR) (Rossjohn et al., 1998), *Lucilia cuprina* (now Delta class) (Wilce et al., 1995) (The structure for this has not been deposited in the PDB), and the plant Phi class GSTs from *Arabidopsis thaliana At*GSTF2 (PDB code 1GNW) (Reinemer et al., 1996), *Zea mays* GSTF1 (1AXD) (Neuefeind et al., 1997a) and *Zea mays* GSTF3 (1AW9) (Neuefeind et al., 1997b). Numbered according to *At*GSTT1 structure.

Original in colour

The overall r.m.s deviation on superimposition of the human Theta class GST (hGST2-2) on *At*GSTT1 is 1.21Å over 198 residues, however the greater sequence identity of the N terminal domain is reflected structurally with an r.m.s.d. of 0.76Å over the first 79 residues, while the C-terminal displays a r.m.s.d. of 1.44Å over 123 residues. No detailed comparison could be made for the insect Theta class structure as the coordinates have not been deposited in the PDB, however the most structurally similar Phi class structure (1GNW) has an overall r.m.s.d. of 1.54Å over 170 residues (figure 7-8).

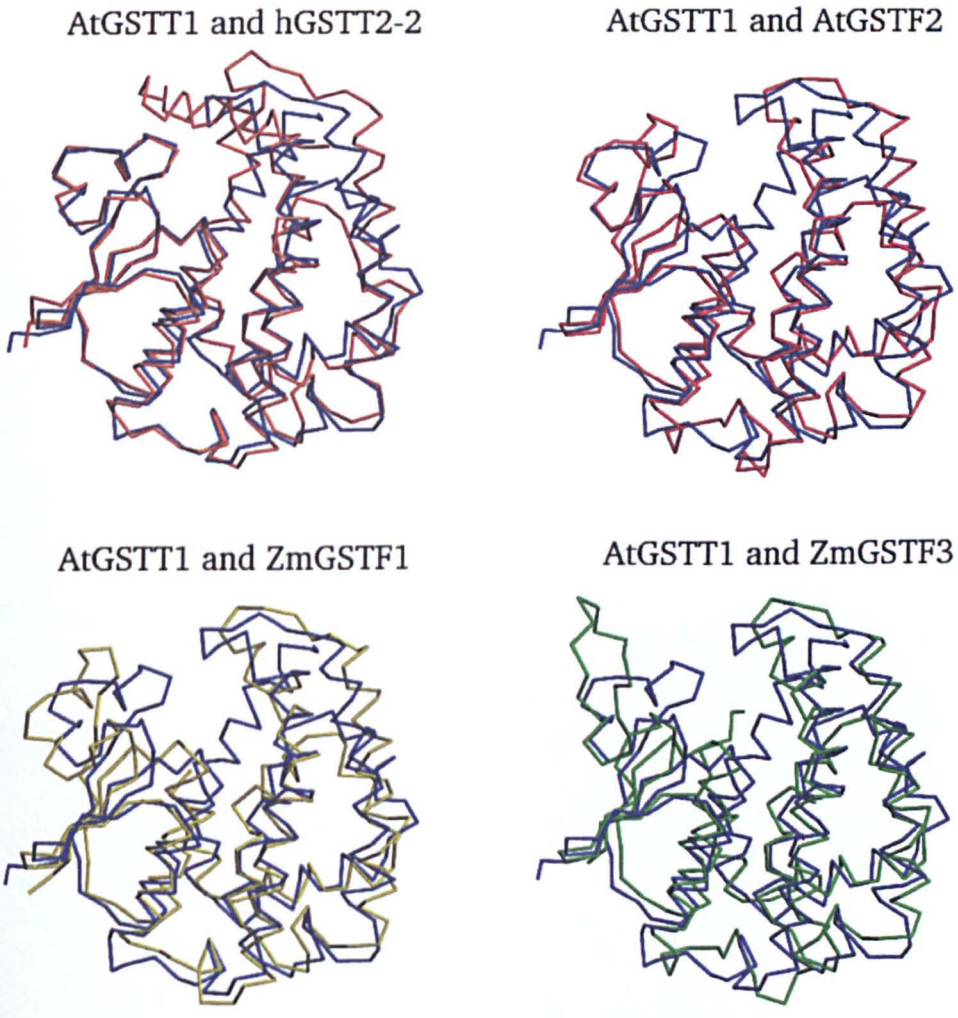


Figure 7-8: Diagram showing superimposition of *At*GSTT1 with human, *Lucilia cuprina*, *Arabidopsis thaliana* and *Zea mays* GSTs. *At*GSTT1 is shown in blue in all the comparisons. *At*GSTT1 can be seen to be most similar to the human isoform, however the long C-terminal helix of the mammalian Theta class enzyme is missing

Original in colour

In addition to the N-terminal domain, the covalent linker region between the two domains of *At*GSTT1, hGST2-2 and the *Lucilia cuprina* GST and *At*GSTT1 share close homology. As observed for the hGST2-2 and the insect GST structures the conserved residues Trp86 and Tyr87 of *At*GSTT1 act as wedges between N and C-terminal domains with Trp86 sitting in a hydrophobic pockets formed by residues Phe21, Val24, Val82, Pro166 and Ile168. Tyr87 forms a hydrogen bond with the His98 and is surrounded by Ile73, Ile168 and the aliphatic region of Arg94. The linker element in the *Arabidopsis thaliania* class Phi structure (Reinemer et al., 1996) is considerably longer linker region being composed of fifteen residues and like the *Zea mays* structures it does not contain aromatic residues in this region, causing it to adopt a different conformation.

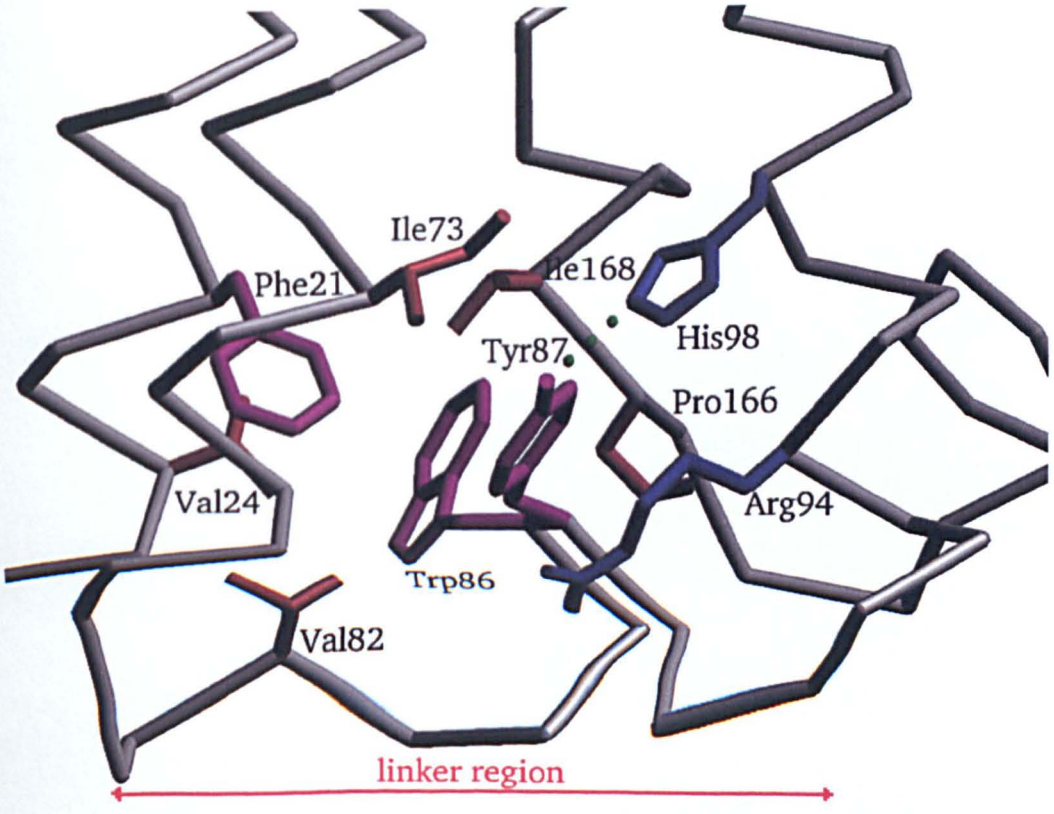


Figure 7-9: Domain interface at linker region of *At*GSTT1. Trp86 and Tyr87 of *At*GSTT1 act as wedges between N and C-terminal domains. Both residues lie in a hydrophobic pocket with Tyr87 forming a hydrogen bond to His98.

Although *At*GSTT1 is most similar to hGSTT2-2 overall, significant differences between the *At*GSTT1 and the hGSTT2-2 structure are observed in the C-terminal domain where helices α4 and α5 of *At*GSTT1 are shifted towards the N-

Original in colour

terminal domain by ~1.5Å and 2-4Å respectively. This shift together with a conformational change in the loop region of α4-α5 (figure 7-10) results in the top of these helices protruding into the active site to a much greater degree in the *At*GSTT2 structure than observed in hGSTT2-2. Following this significant shift, helix α6 of both structures superimpose closely, however the insertion of four residues Asp, Lys, Asp, Arg (residues 185-188) in the *At*GSTT1 sequence causes a rearrangement of the loop region at the top of α6, causing it to form a helical element (α6b). This is analogous to that observed in Phi class structures (this region is observed to be unstructured in both the human and insect structures). Helices α7 and α8 superimpose relatively closely for both the Theta and Phi class structures, with α7 running almost perpendicular to α4 and α5 at the back of the C-terminal domain, while α8 lies at the back of the domain interface. In contrast to the hGSTT2-2 which has a ninth helix (α9 was observed in both apo and ligand bound hGSTT2-2 structures) no long C-terminal extension is observed to wrap over the top of the *At*GSTT1 monomer and point toward the twofold axis of the dimer (Wilce et al., 1995).

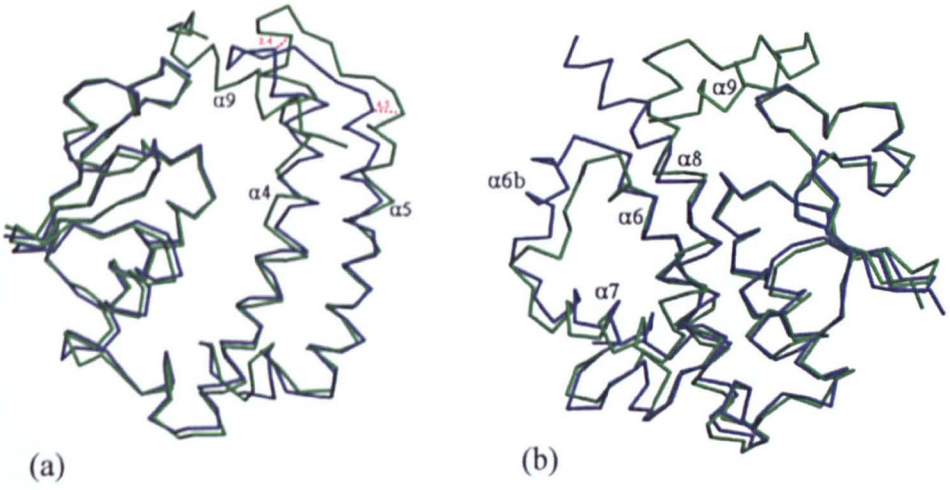


Figure 7-10: Superimposition of *At*GSTT1 and hGSTT2-2 monomer structures. Picture (a) views the monomer from the dimer interface and indicates movement of α4 and α5 together with conformational change in loop region, (helices α6, α7 and α8 were removed for clarity). Picture (b) is a view of the monomer from the rear of monomer and displays formation of helical element α6b and lack of helix α9 in the *At*GSTT1 structure (helices α4 and α5 were removed for clarity). *At*GSTT1 is displayed in blue and hGSTT2-2 is shown in green.

Original in colour

The subunit interface of the *At*GSTT1 structure is around 1193Å^2, which is consistent with that observed for other GSTs. It is predominantly mediated by hydrophobic and van der Waals contacts between helices. As observed for both the insect and the human Theta class GSTs there are very few polar interactions between side chains at the dimer interface, with only one hydrogen bond formed between Ser99 and the main chain of Phe66 in the *At*GSTT1 structure. As described in other GST structures an intersubunit 'lock and key' hydrophobic mechanism is observed for *At*GSTT1 as discussed previously for the Theta class enzymes (Rossjohn et al., 1998). However, this key is found in the C-terminal domain not in the N-terminal domain in contrast to the Alpha, Mu, Pi and Phi class structures. In the human enzyme hGSTT2-2 and the insect GST structure the 'lock and key' is formed by Phe150 and Phe140 respectively, whilst in the *At*GSTT1 structure Trp152 performs the role of the phenylalanine residue. An additional interaction across the dimer interface, not observed in either the insect or the Phi class structures, is formed by a pair of symmetry-related histidine residues, that form Van der Waal stacking interactions at the base of the interface.

7.4.1 The active site

Despite the *At*GSTT2-2 structure lacking the long C-terminal extension (α9) found in the hGSTT2-2 structure, the active site region of *At*GSTT1 is remarkably similar to that of hGSTT2-2, particularly within the glutathione binding site. The conservation of the stacking interaction between Phe46 and Tyr7 of *At*GSTT1, which is similar to that of Phe45 with Phe6 in the human enzyme, promotes the β2-β3 loop region to form a very similar structure in these enzymes (Phe53 stacks with Tyr4 in insect and Phe45 stacks with Phe6 in Phi class). Within the glutathione binding site many residues are also strongly conserved and superimpose closely between the two structures. In addition to the strongly conserved residues Ser12, *cis*-Pro56 and Glu67 previously shown to interact with glutathione the residues Asp9, Gln13, Pro14, Ser15 and Lys54 (figure 7-11) are also strongly conserved between the Theta class structures. It is not only within the G-site that strong conservation of the residues is observed,

but also within the H-site which typically shows variability of amino acids, even within GST enzymes of the same class (Prade et al., 1998).

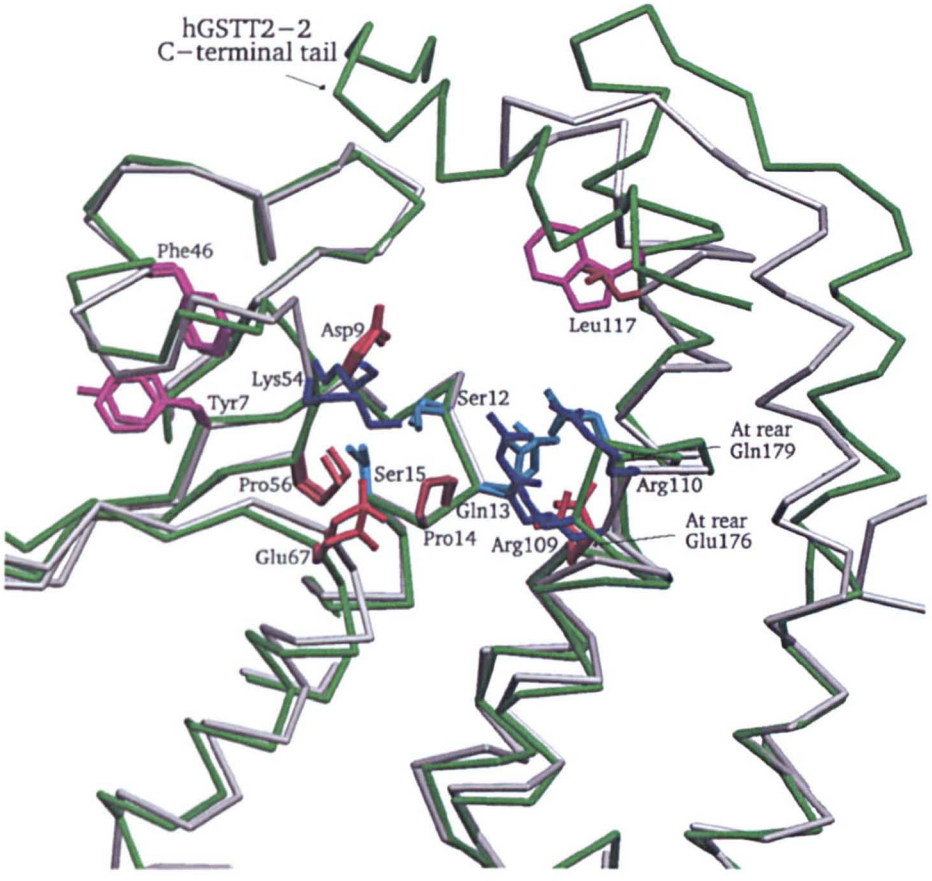


Figure 7-11: Superimposed active site for the Theta class GSTs, hGSTT1 and *At*GSTT1. Several residues in the active site (Asp9, Ser12, Gln13, Pro14, Ser15 Lys54, Val55, Pro56 and Arg109) are strictly conserved maintaining the same overall shape and charge distribution for both enzymes suggesting that they might both catalyse the same reaction.

Calculation of the molecular surface (figure 7-12) and electrostatic potential of *At*GSTT1 indicates that the top region of α4 that forms the H-site is hydrophobic, while the lower part forms a hydrophilic pocket. The hydrophilic pocket is reasonably well conserved between *At*GSTT1 and the human Theta class structure, sharing residues Arg109, Glu176 and Gln179. However several differences are observed such as Arg110 in *At*GSTT1 which is substituted by a glycine residue in hGSTT2-2. The substitution of Trp115 to Leu117 and the lack of the C-terminal helix α9 in *At*GSTT1 mean that the active site of *At*GSTT1 is slightly larger than that of the human enzyme. The similarity in the residues around the active site suggests similar types of compounds might bind to both the human and *Arabidopsis thaliana* enzyme.

Original in colour

Figure 7-12: Cut away molecular surface representation of *At*GSTT1. Position of glutathione moiety (modelled from 1LJR structure (Rossjohn et al., 1998)) is shown in red and position of thiol group of GSH cysteine is shown in light blue.

Deletion of the C-terminal extension in hGSTT2-2 has been shown to lead to a reduction in catalytic activity of Menaphthyl sulphate, cumene hydroperoxide and ethacrynic acid (Rossjohn et al., 1998). As *At*GSTT1 lacks this C-terminal element it would be interesting to determine if *At*GSTT1 is able to conjugate glutathione to these molecules and if so to determine the rate of catalysis. It was also suggested that the active site to the human enzyme was occluded by this C-terminal extension and it was postulated that the human enzyme had to alter its conformation 'breathe' to enable substrate access to the active site (Rossjohn et al., 1998). No studies have been performed to better understand the processes that occur on binding of substrate to the *At*GSTT1 enzyme.

7.4.2 Sulphate binding site

Structural determination of the hGSTT2-2 structure elucidated a sulphate binding site within the H-site, formed by the tetrahedral coordination of residues Gln12 (helix α1), Arg107 (helix α4), Trp115 (helix α4) and Arg239 (helix α9). It was proposed that the binding of sulphur was key factor which determined the

Original in colour

selectivity of substrate for this class of GST and the presence of this specific sulphate pocket was suggested to enable the sulphatase activity shown to occur with hGSTT2-2 (Rossjohn et al., 1998). Analysis of the *At*GSTT1 structure indicates that both the glutamine and arginine residues are conserved and that the residue Trp115 is substituted by a leucine. However Arg239 present in helix α9 of the human enzyme is not observed in the *A. thaliana* structure. In the *At*GSTT1 model these residues do not appear to coordinate a sulphate ion despite sulphate ions being present in the crystallisation mixture at a concentration of 0.2M. A sulphate ion was found to bind in all four monomers hydrogen bonded to Glu67 and Ser68, however since these residues are known to bind glutathione, the sulphate observed in this position is thought to be an artefact, which would be replaced if GSH was included in the crystallisation solution.

To better characterise the active site of *At*GSTT1 it would be valuable to bind a number of substrates or inhibitors in the active site. Unfortunately there has been relatively little characterisation of the enzymology of *At*GSTT1 to date making inhibitor and substrate selection difficult. The similarity in the active site topology between the *Arabidopsis thaliana* and human GSTs suggests that both enzymes might show very similar substrate specificity.

Chapter 8: Structure solution of Wheat J

Several crop species are protected from or given increased resistance to herbicide injury by safeners. Although the exact mode of action of herbicide safeners is unknown it is thought that they protect the plant by increasing one or more stages of herbicide metabolism.

Safener treatment of wheat seeds by fluxofenim or cloquintocet-mexyl has been shown to confer protection from injury from the chloracetamide herbicide, dimethenamid (Riechers et al, 1994; Riechers et al., 1996) via the glutathione conjugation system. More detailed study resolved that improved metabolism of dimethenamid is not due to increased levels of glutathione, but increased activity of specific GSTs. Characterisation of GST enzymes in unsafened and safened hexaploid (*Triticum aestivum L.*), tetraploid (*Triticum durum*) and diploid (*Triticum tauschii*) wheat determined that *T. aestivium* and *T. tauschii* contained safener induced GSTs with activity towards the herbicide dimethamid. As *T. aestivium* is hexaploid with genome ABD, *T. tauschii* is diploid with genome D, and *Triticum durum* has a tetraploid genome AB, it was proposed that the safener induced GSTs were encoded by genome D. Purification and characterisation of a safener induced GST from *T. tauschii* with activity toward dimethenamid revealed that it shared significant amino acid identity with safener inducible maize GSTs and auxin-inducible GSTs from tobacco. This sequence similarity placed the wheat GST in the Tau class of GSTs, a plant specific class most similar at that time to the mammalian Theta class GSTs (Riechers et al., 1997).

In Durham, Robert Edwards group had cloned the above enzyme from hexaploid wheat using probes from the Tauschii sequence. The significant herbicide conjugating ability of this enzyme together with the availability of this Tau class *T. tauschii* GST clone meant that it was of interest to include this wheat GST (wheat J) in the crystallisation programme at Glasgow.

As discussed in section 4.2.5, diffraction sized crystals for wheat J were successfully grown and data collected to a maximum resolution of 2.3Å. Initial attempts to determine the structure of the wheat J enzyme however, were hampered by difficulties in processing the data collected from crystals of this enzyme. Although the crystals of wheat J appeared to show hexagonal morphology and did not appear twinned, indexing of the first two data sets proved very difficult. The features described in section 8.1.1 meant it was difficult to ascertain if the correct spacegroup had been chosen in which to process the data and subsequently solve the phase problem. A third data set was found to process and scale well in the spacegroup $P6_122$ and this data was used to gain the molecular replacement solution.

8.1 Structure solution

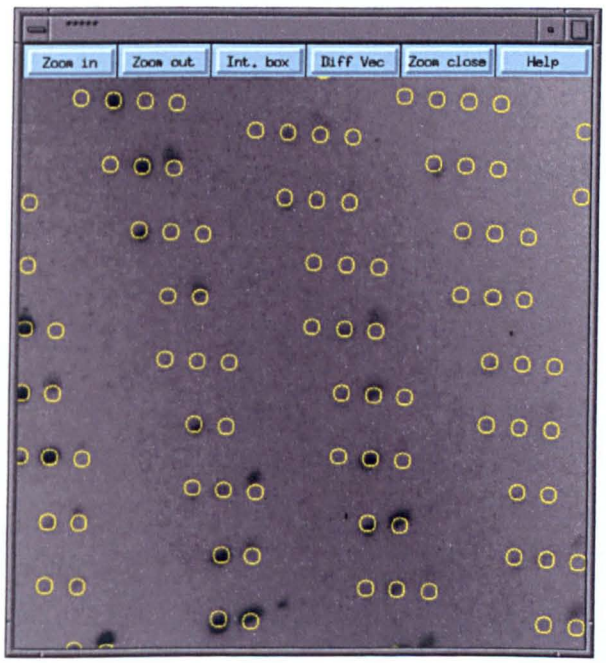
8.1.1 Data Processing problems

The initial two data sets for wheat J enzyme, which showed diffraction to 2.3Å and 3.0Å, were collected at the microfocus beamline ID13, ESRF Grenoble and station 7.2, SRS Daresbury respectively. Although the diffraction spots from these data collections were streaky and varied in shape significantly throughout the image as displayed in figure 8-1a (mosaicity of around 1.0), they were well resolved from each other. Auto indexing in DENZO (Otwinowski & Minor, 1997) suggested that the spots might be indexed in the hexagonal lattice system (Table 8-1). However visual examination of the diffraction images showed that although many of the spots were correctly predicted using the hexagonal lattice a region of the detector was not well predicted with certain partials significantly shifted from the centre of the predicted spot. (figure 8-1b)

Lattice	Metric tensor distortion index	Best cell (without symmetry restrains) Best cell (symmetrized)					
primitive cubic	21.96%	89.44 109.76	88.84 109.76	150.99 109.76	90.19 90.00	91.00 90.00	119.40 90.00
I centred cubic	25.56%	89.95 146.35	174.94 146.35	174.15 214.62	52.33 90.00	75.88 90.00	76.35 90.00
F centred cubic	24.80%	174.94 188.33	175.44 188.33	146.35 188.33	53.80 90.00	53.50 90.00	60.94 90.00
P rhombohedral	7.57%	174.15 166.43 89.14	174.16 166.43 89.14	150.99 166.43 475.81	31.09 30.51 90.00	30.90 30.51 90.00	29.55 30.51 120.00
P hexagonal	0.62%	89.95 89.39	88.84 89.39	150.99 150.99	89.81 90.00	91.18 90.00	119.97 120.00
P tetragonal	13.30%	88.84 89.14	89.44 89.14	150.99 150.99	91.00 90.00	89.81 90.00	60.60 90.00
I tetragonal	15.49%	88.84 89.14	89.44 89.14	313.32 313.32	82.56 90.00	98.06 90.00	60.60 90.00
P orthorhombic	13.30%	88.84 88.84	89.44 89.44	150.99 150.99	91.00 90.00	89.81 90.00	60.60 90.00
C orthorhombic	0.54%	88.84 88.84	155.85 155.85	150.99 150.99	91.25 90.00	89.81 90.00	90.37 90.00
I orthorhombic	15.49%	88.84 88.84	89.44 89.44	313.32 313.32	82.56 90.00	98.06 90.00	60.60 90.00
F orthorhombic	6.54%	89.95 89.95	153.93 153.93	313.32 313.32	90.32 90.00	105.50 90.00	90.4 90.00
P monoclinic	0.52%	88.84 88.84	150.99 150.99	89.44 89.44	91.00 90.00	119.40 119.40	90.19 90.00
C monoclinic	0.17%	155.85 155.85	88.84 88.84	150.99 150.99	90.19 90.00	91.25 91.25	89.63 90.00
P triclinic	0.00%	88.84	89.44	150.99	91.00	90.19	119.40

Table 8-1: Suggested lattices from autoindexing in DENZO.

(a)



(b)

Figure 8-1: Image (a) shows typical indexing of data collected on the microfocus beamline when using hexagonal lattice system. Image (b) shows a close up of an area of the image overlayed with the predictions from DENZO (a). Increasing the mosaicity to predict all the spots on the image does not improve the positioning of predictions, resulting in poor scaling.

Original in colour

Attempts to index the images using lower symmetry lattice systems, C centered orthorhombic or Primitive monoclinic, gave similar results with stronger spots, present over a number of images, observed to lie to the left of the prediction in the first image, correctly predicted on the second image and to the right of the prediction on the third image. This incorrect positioning of some of the predictions prevented the intensity of a number of spots on each image from being correctly measured. This was most evident following scaling of the images and comparison of the symmetrically equivalent reflections. The symmetrically equivalent reflections should have had very similar intensity measurement in agreement with Freidels law, however this was not found to be true. In addition to trying to process the data in DENZO attempts to process the first two datasets were also made using MOSFLM (Leslie, 1992) and D*TREK (Molecular Structure Corporation (incorporates 3D profile fitting) with the program authors help, but no solutions to the processing problems were found.

8.1.2 Solution of phases for Wheat J

A third data set of a native crystals were grown in similar conditions to those previously and again these were cryocooled using mother liquor containing 20% glycerol. This enabled data to be collected to 2.9Å, and this time the data could be processed in DENZO (Otwinowski & Minor, 1997) with the spots being correctly predicted when the hexagonal lattice system was chosen (section 4.2.5). Scaling of the data gave high R_{merge} statistics, however comparison of statistics for spacegroups of P6, P3 and C orthorhombic suggested that the data might belong to $P6_522$ or $P6_122$, with the appropriate systematic absences being observed (l=6n) (Otwinowski & Minor, 1997).

Comparison of the wheat J sequence with other GST structures in the PDB indicates it shares closest identity (21%) with *Zm*GSTF3 (Neuefeind et al., 1997b). Attempts using the molecular replacement AMoRe (Navaza, 1994) and EPMR (Kissinger et al, 1999) were made to gain phase information for the wheat J data using models of *Zm*GSTF3, *Zm*GSTF1 and *At*GSTF2, (Neuefeind et al., 1997a; Neuefeind et al., 1997b; Reinemer et al., 1996) however no correct solution was found. During the search for phase information for wheat J the *At*GSTT1 structure was being refined (shares 18% identity with wheat J) and so a

polyalanine model of a monomer of *At*GSTT1 was also tried in $P6_122$ using EPMR. Using this polyalanine trace of the *At*GSTT1 model a solution was detected which had a significantly higher correlation coefficient than the background (Table 8-2) and packed as a dimer with no clashes between symmetry generated molecules. The electron density maps calculated from this solution were very poor, although they did appear to provide details of the wheat J structure. Following solvent flattening using DM (Cowtan, 1994) a number of attempts were made to build and refine the wheat J structure into the poor, heavily biased electron density, however this proved very difficult.

$P6_122$	Rotation			Translation			Correlation coefficient	R factor
Soln 1	71.45	71.45	155.13	-19.88	45.35	109.33	0.405	0.868
Soln 2	71.06	71.06	154.15	-19.83	45.56	109.33	0.405	0.870
Soln 3	99.71	99.71	341.29	-9.71	29.28	43.46	0.307	0.924
Soln 4	96.61	96.61	319.97	-2.66	70.29	124.49	0.316	0.905
Soln 5	71.10	71.10	94.24	38.73	5.75	133.83	0.405	0.868
Soln 6	99.56	99.56	16.18	-11.32	26.82	126.37	0.318	0.907
Soln 7	81.12	81.12	165.26	-11.67	24.17	117.47	0.307	0.905
Soln 8	107.78	107.78	31.60	24.76	30.57	91.88	0.329	0.892
Soln 9	86.57	86.57	202.18	-10.75	72.28	78.65	0.304	0.908
Soln 10	107.56	107.56	332.97	-13.76	36.96	43.13	0.328	0.894

Table 8-2: Solutions gained using the polyala monomer structure of *At*GSTT1, in spacegroup $P6_122$ with molecular replacement program EPMR (Kissinger et al, 1999). Three solutions sharing a similar rotation function were found, indicated by grey shading which had a higher correlation coefficient and lower R factor than the background.

The structure of a human GST, defined as a new Omega class had been described by Board and coworkers at the GST 2000 conference. This Omega class GST shared 23% sequence identity with the wheat J enzyme (Figure 8-2), and it was thought this structure might be a better probe for molecular replacement than the *At*GSTT1 structure. The coordinates for the human Omega structure became available in the latter part of August 2000 allowing molecular replacement using the Omega structure to be attempted, and using these coordinates a clear solution was gained using the program EPMR (Table 8-2) (Board et al., 2000).

```
                  10         20         30         40         50
Wheat J                      MA GGDDLKLLGA WPSPFVTRVK LALALKGLSY
Hu omega  XMSGESARSL GKGSAPPGPV PEGSIRIYSM RFCPFAERTR LVLKAKGIRH
                  60         70         80         90        100
Wheat J   EDVEEDLYKK SELLLKSNPV HKKIPVLIHN GAPVCESMII LQYIDEVFAS
Hu omega  EVININLKNK PEWFFKKNPF GLVPVLENSQ GQLIYESAIT CEYLDEAYP-
                 110        120        130        140        150
Wheat J   TGPSLLPADP YERAIARFWV AYVDDKLVAP WRQWLRGKTE EEKSEGKKQA
Hu omega  -GKKLLPDDP YEKACQKM-IL ELFSKVPSL VGSFIRSQNK EDYAGLKEEF
                 160        170        180        190        200
Wheat J   FAAVGVLEGA LRECSKGGGF FGGDGVGLVD VALGGVLSWM KVTEALSGDK
Hu omega  RKEFTKLEEV LT--NKKTTF FGGNSISMID YL---IWPWF ERLEAMKLNE
                 210        220        230        240
Wheat J   IFD-AAKTPL LAAWVERFIE LDAAKAALPD VGRLLEFAKA REAAAAASK
Hu omega  CVDHTPKLKL WMAAMKEDPT VSALLTSEKD WQGFLELYLQ NSPEACDYGL
```

Figure 8-2: Sequence comparison of wheat J and the recently solved Omega class GST from human (PDB code 1EEM). Conserved residues are shown in red and similar residues in blue. Sequence alignment was performed using MULTIALIGN (Corpet, 1988).

The 2Fo-Fc electron density map calculated from the Omega probe molecular replacement solution was readily interpretable and enabled a series of features of the wheat J structure to be built using the program QUANTA (Molecular Simulations Inc.). This initial model was then refined using the simulated annealing option in CNS (Brunger et al, 1998). Following two rounds of building and refinement, this initial model (R_{cryst} = 33.7 and R_{free} = 41.2) built using the 2.5Å dataset, was used to molecular replace the higher resolution data gained at the microfocus beamline. Only data from the first 74 frames before translation of the crystal of this higher resolution dataset (2.3Å) were used, as they appeared to have suffered less from radiation damage. The data was scaled in $P6_122$ following processing, which gave a R_{merge} statistic of 12.9% and it was hoped that the higher resolution data would provide additional information which would allow further improvement to the model.

Original in colour

Solution	Rotation	Translation	CC	R factor
1	354.87 105.57 177.76	39.21 6.04 48.13	0.281	0.691
2	356.29 107.03 298.73	30.13 39.56 11.62	0.349	0.650
3	356.08 111.11 62.04	63.73 30.91 45.05	0.246	0.694
4	195.12 32.23 199.08	20.89 39.67 72.71	0.239	0.707
5	176.28 73.06 121.20	19.20 45.82 37.31	0.349	0.650
6	2.89 105.36 296.18	-6.53 70.85 65.25	0.247	0.696
7	356.52 107.13 59.03	63.24 30.37 36.09	0.349	0.649
8	176.85 76.69 242.90	-14.22 36.47 121.63	0.277	0.678
9	189.07 83.90 178.31	54.30 53.62 43.94	0.228	0.702
10	240.94 19.98 89.56	37.08 9.17 81.32	0.247	0.703

Table 8-3: Solutions gained for molecular replacement of wheat J X31 data (resolution 2.5Å with human Omega class GST (Board et al, 2000) using EPMR (Kissinger et al, 1999). (CC = correlation coefficient)

The molecular replacement solution gained using the higher resolution data gave significantly improved maps. These improved maps allowed corrections to the Cα backbone and the addition of further sidechains. The R_{free} set was extended to include the higher resolution shells and the program ARP (Lamzin & Wilson, 1993) was used to add waters to the structure during refinement using the maximum likelihood refinement program REFMAC (Murshudov et al., 1997).

As stated earlier the higher resolution data (2.3Å) from the microfocus beamline was difficult to process due to difficulties in predicting the spot profile. The difficulties in processing meant that the R_{merge} statistics were probably not very reliable and it was difficult to determine whether the spacegroup was $P6_122$ or $C222_1$ (Overall R_{merge} of $P6_122$, 12.9%; Overall R_{merge} of $C222_1$ 11.9%). Crystals of spacegroup $C222_1$ have similar cell dimensions as $P6_122$ with the non-crystallographic symmetry present in $C222_1$ corresponding to the crystallographic symmetry of the $P6_122$ crystals. The electron density maps calculated from the data processed in $C222_1$ was clearer than that calculated from data processed in $P6_122$, therefore the final rounds of refinement and building of the essentially complete model were performed in $C222_1$.

During refinement of the structure, electron density for the inhibitor molecule S-hexylglutathione became visible, however this was not included in the phasing until the protein chain was almost completely built and the inhibitor density was

unambiguous. After several cycles of refinement and rebuilding the R_{cryst} and R_{free} remained at 18.0% and 22.7%.

8.1.3 Quality of the model

The final model of wheat J in the spacegroup $C222_1$ contained three protein monomer units consisting of 221 amino acids in the asymmetric unit. Each of the monomers contained a molecule of S-hexylglutathione bound in the active site and a total of 716 water molecules. Additional density, resembling that as might be expected for a histidine was found at the side chain of residue six, which according to the sequence was an aspartate. A histidine was therefore included at this position, and the collaborators are in the process of checking the sequence of the wheat J clone. Although the mainchain could be traced within the loop region connecting α4 to α5, the side chains of residues Lys120, Thr121, Glu123 could not be traced in any of the three monomers. Further, the side chain for Arg118 could not be traced in monomer B or C and poor density prevented the tracing of the sidechain of Glu122 in monomer C. The α4-α5 loop region in the C-terminal domain of the structure exhibited higher B factors than any other region of the structure. This suggests that this loop might show some flexibility as suggested for the Phi class structures. The quality of the refined structure was assessed with PROCHECK (Laskowski et al., 1993) with the Ramachandran plot indicating that 94.5% of the residues lie within the most favoured regions. The model shows good stereochemistry with an estimated coordinate error (r.m.s.) of 0.2Å (Cruikshank, 1999) and a final r.m.s. deviation from ideal bond lengths and angles of 0.006Å and 1.7° respectively. The final $2F_o - F_c$ electron density map contoured at 1σ is of good quality and shows continuous, well-defined density. The parameters for the refined model are summarised in Table 8.4.

Resolution Range (Å)	30-2.3
No of unique reflections in resolution range	52618
R_{Work} (%)[a]	18.0
R_{Free} (%)	22.7
Model	
Total number of amino acids	661
No. of Solvent atoms (excl. H)	717
No. of Heteroatoms	78
Stereochemistry[b]	
Ramachandran quality, % in	
Most favoured regions	94.4
Allowed regions	5.1
R.M.S. deviation[c]	
Bond length (Å)	0.006
Bond angle (degrees)	1.7

[a] $R = \Sigma_{hkl}(|F_o - F_d|) / \Sigma|F_d|$.
[b] Calculated using the program PROCHECK (Laskowski et al., 1993)
[c] (Cruikshank, 1999)

Table 8-4: Refinement statistics for wheat J structure.

8.1.4 Overall structure

The N-terminal domain of wheat J is a βαβ element (β1 7-11, α1 15-29, β2 33-36) connected to a ββα element (β3 58-61, β4 64-67, α3 68-80) by an irregular solvent exposed region, which contains a short α-helical section (α2 43-50). A ten residue linker region (residues 81-90) joins the N-terminal domain to the larger C-terminal domain that folds in a similar fashion to that of the Phi class plant GST *At*GSTF2 (Reinemer et al., 1996) and the human Omega class GST (Board et al, 2000). As in the Phi and Omega structures, an aliphatic residue, Leu87 packs between the two domains, rather than an aromatic residue as observed in the Theta class structures. Comparison of the C-terminal domain of wheat J with other plant GSTs shows the orientation and length of α4 and α5 (α4 91-119 α5 121-147) are different. Helices α6, α7 and α8 (α6 159-180, α7 189-202 and α8 202-209) correspond closely to similar regions in most of the other GST classes, while α9 (211-224) folds back over the top of the N-terminal domain in a similar fashion as α9 in the human Omega class structure.

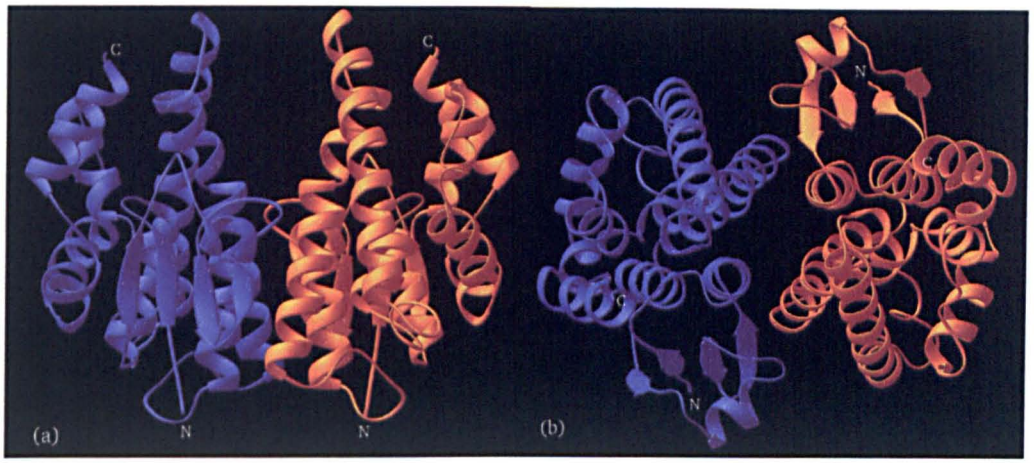


Figure 8-3: Ribbon diagram of the dimeric wheat J structure. The 2 fold axis relating to the dimer subunits normal to the plane in (a) and on the plane of the drawing in (b). Subunit A is coloured lilac and subunit B, gold. The N and C termini are indicated with N and C respectively. This figure was made with RIBBONS (Carson, 1997).

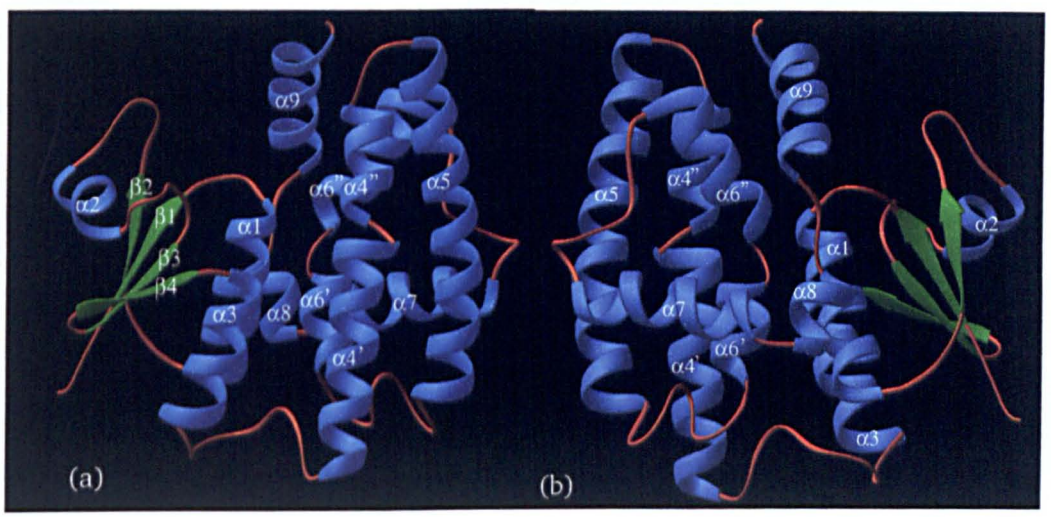


Figure 8-4: Ribbon representation of wheat J monomer structure annotated with secondary structure elements. Diagram (a) shows monomer viewed from dimer interface, while (b) shows rear of monomer structure. The α-helices are drawn in blue spirals and the β-strands are in green. This figure was made with RIBBONS (Carson, 1997).

The N and C-terminal domain interface is comprised of a series of van der Waal interactions between aliphatic residues and two polar interactions between Arg20-Asp106 and Gln74-Arg99. These structurally important residues are strongly conserved between Tau class GSTs, (figure 8.5) but not between Tau and Phi class GSTs. The dimer interface is formed predominantly by hydrophobic contacts, the most significant of which form the hydrophobic 'lock

Original in colour

and key' found in other GST classes (Rossjohn et al., 1998). In contrast to the Phi class GSTs that utilise a phenylalanine residue, a conserved valine (Val52) lies in a hydrophobic pocket formed by the conserved residues Phe100, Trp101 and Tyr104. A non conserved valine residue also forms part of this pocket in the wheat J structure. Additional hydrophobic interactions are formed by the strictly conserved Tyr73 that forms an edge to face packing arrangement with the conserved residue Tyr93 (observed to be a histidine residue in some Tau class members) of the opposite monomer.

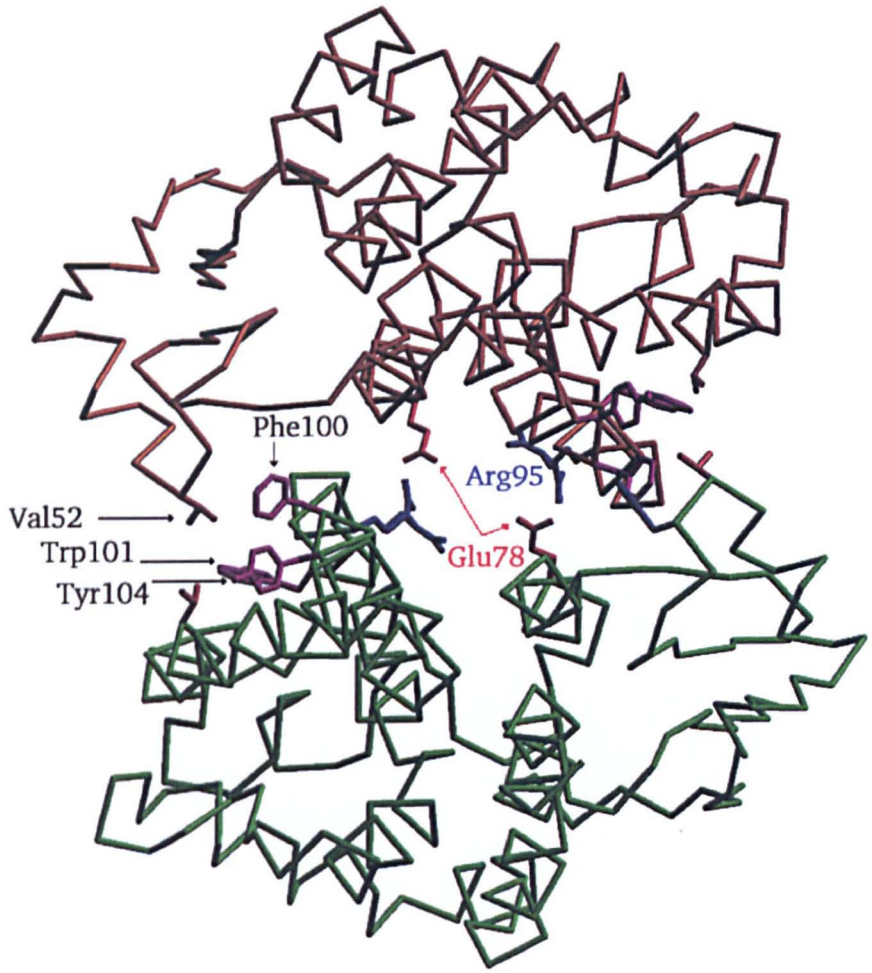


Figure 8-5: The hydrophobic 'lock and key' motif interacts at the edge of the dimer interface, (residues Val52, Phe100, Trp101 and Tyr104) while salt bridges are formed closer to the crystallographic dyad (residues Glu78 and Arg95). The non conserved valine, present on the same chain region as the aromatic groups, also forms a surface of the hydrophobic pocket and is shown in the above figure.

Original in colour

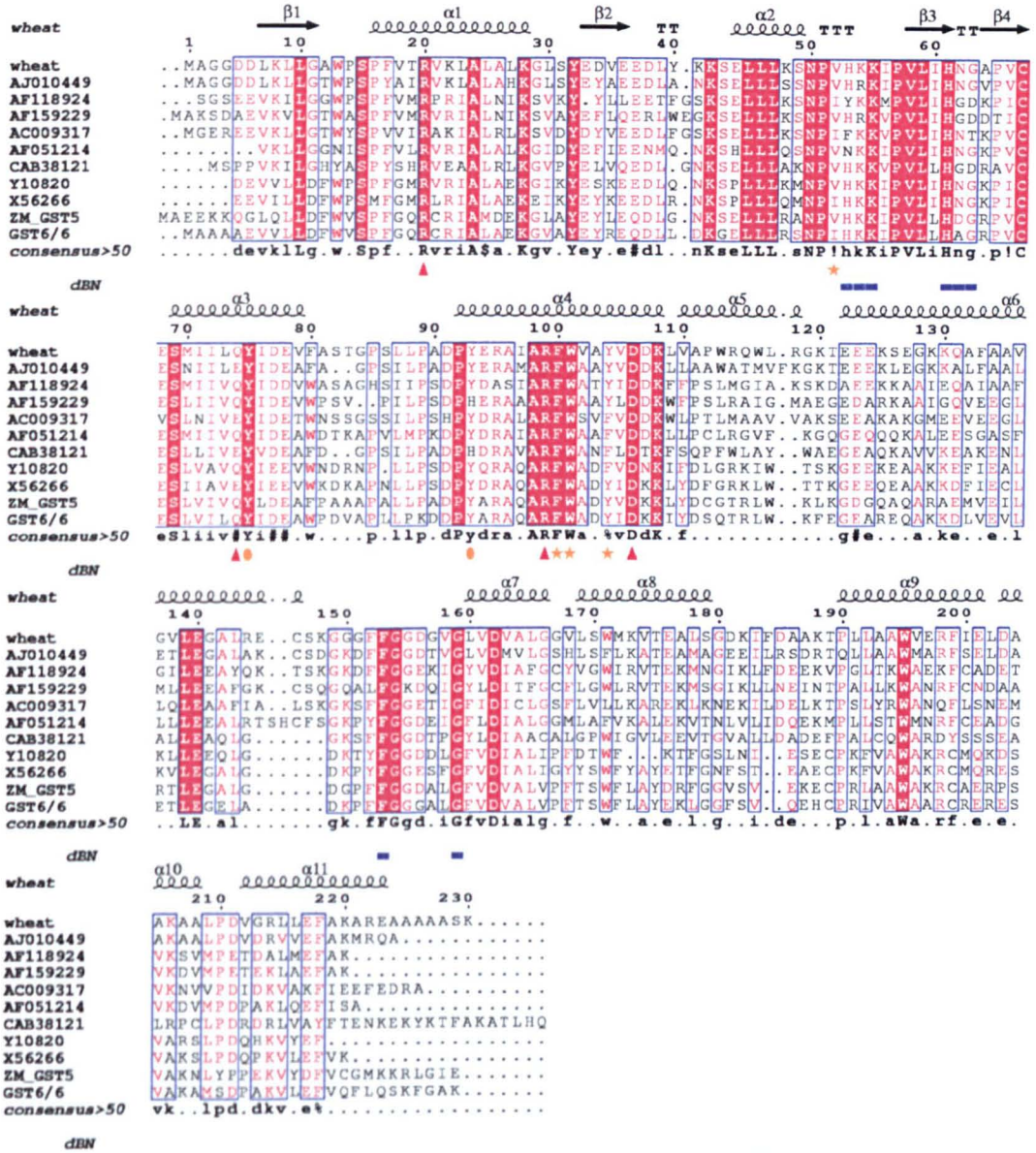


Figure 8-6: Sequence alignment of Tau Class sequences together with secondary structure elements of wheat J. Those residues which form polar interactions across the dimer interface are marked by ▲. Residues involved in the hydrophobic lock and key mechanism are marked by * and the hydrophobic interaction between Tyr75 and Tyr93 by ●. Conserved residues identified by Droog (Droog, 1997) are marked by _ (a blue line). This diagram was produced using ESPRIT (Gouet et al, 1999).

Original in colour

In addition to the hydrophobic interactions at the dimer interface two salt bridges are formed closer to the crystallographic dyad between residues Glu78-Arg95 and Glu78-Arg99. A hydrogen bond between the hydroxyl of Tyr93 and the mainchain carbonyl of Pro65 is also formed.

8.1.5 The GSH-binding site (G site)

In each monomer, one molecule of S-hexylglutathione is bound in the GSH site, (figure 8-7). The mainly polar binding of the glutathione moiety is consistent with that previously observed in other classes of GST (Garcia-saez et al, 1994; Oakley et al, 1997; Reinemer et al., 1996; Zeng et al, 1994). The γ-glutamyl moiety points downwards towards the core of the molecule and forms hydrogen bonds with Glu68 (Glu68 shows unfavourable Phi/Psi angles in this structure, as is typical for GSTs), the mainchain amide of Ser69 and the hydroxyl group of Ser69.

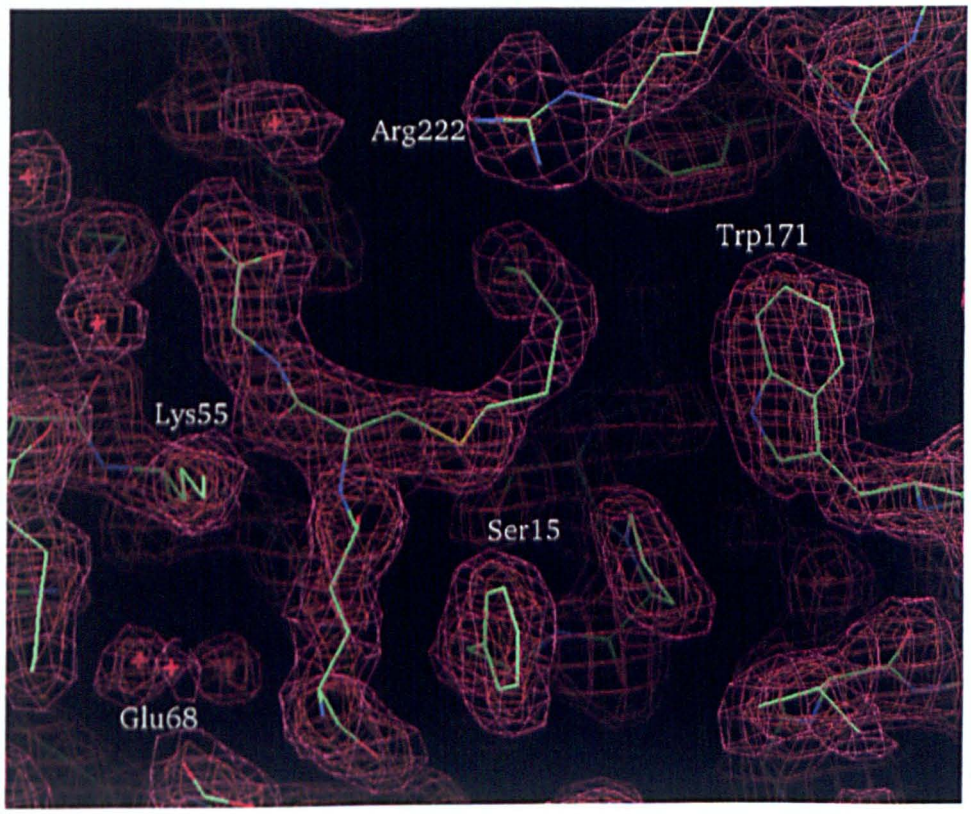


Figure 8-7: Density around S-hexylglutathione inhibitor molecule in the active site of wheat J. Residues of the active site; Ser15 (catalytic serine), Glu68 (binds to γ-glutamyl moiety, Arg222 of C-terminal tail, Lys55 of β2-β3 loop region and Trp171 are marked accordingly. This diagram was produced using QUANTA (Molecular Simulations Inc.).

Original in colour

A clearer representation of the interactions that binds the inhibitor into active site is shown in the LIGPLOT diagram, figure 8-8. In addition to the interactions discussed, the cysteinyl moiety of the S-hexyl glutathione molecule forms two hydrogen bonds. One of the hydrogen bonds is to the mainchain carbonyl of *cis*-Pro57, a conserved residue in all GST structures and crucial for recognition and binding to glutathione. The second hydrogen bond is formed between the cysteinyl moiety and the amide bond of Ile56 (Ji et al., 1995; Ji et al., 1992; Reinemer et al., 1992; Reinemer et al., 1996; Sinning et al., 1993; Wilce et al., 1995). A further hydrogen bond is formed between the glycyl part of the S-hexylglutathione and residue Lys42. As described for the Phi and Delta class structures from *A. thaliana, Zea mays,* (Neuefeind et al., 1997a; Reinemer et al., 1996) and *L. cuprina* (Wilce et al., 1995), the wheat J structure shows no interactions between the glutathione portion of the bound ligand and residues of the C-terminal domain. This contrasts the situation observed in the Alpha, Beta, Mu, Pi, Sigma and mammalian Theta classes where a salt bridge is typically found between the amino terminal nitrogen of GSH and an acidic residue on helix α4 of the C-terminal domain (Board et al, 2000).

8.1.6 The Hydrophobic binding site (H-site)

GSTs have been shown to bind a wide range of substrates and this diversity is a reflection of the low sequence identity and thus structural variation in the C-terminal domain, which contributes the bulk of the residues involved in binding the electrophiles to be conjugated to glutathione.

The H-site region within the C-terminal domain of GSTs is typically hydrophobic in nature (Ji et al., 1995; Ji et al., 1992; Reinemer et al., 1992; Reinemer et al., 1996; Sinning et al., 1993; Wilce et al., 1995), however in the case of the Zeta (section 6) and Omega classes of GST it also contains residues of hydrophilic character (Board et al, 2000). The H-site region of wheat J is formed by residues from helices α4, α6 and α9. These secondary structure elements contribute a number of residues however the aromatic groups Trp113, Trp171 and Phe218 form the bulk of the molecular surface in this region.

Interestingly the topmost region of helix α4 contains two arginine residues (Arg114 and Arg118), that together with Arg222 from helix α9 make the approach to the H-site relatively basic.

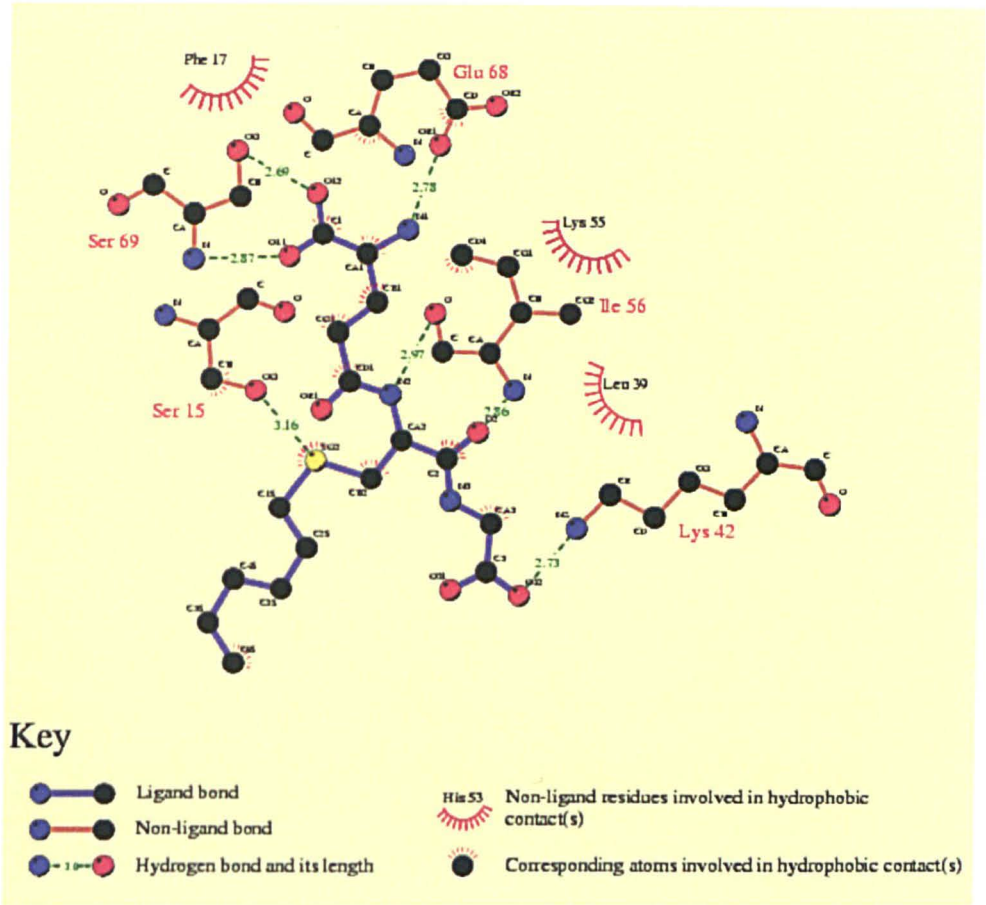


Figure 8-8: Schematic diagram showing the residues contacting the substrate, S-hexylglutathione. As typically observed for GSTs Glu68 and Ser79 hydrogen bond to the glutamyl moiety of glutathione, while the mainchain of Ile56 hydrogen bonds to the cystyl portion. The interaction of Lys42 with the glycyl portion of glutathione is particular to the Tau class GSTs. The figure was generated using the program LIGPLOT (Wallace et al., 1995).

None of these residues are strongly conserved among Phi class GSTs and the presence of these polar groups capping the H-site could have a significant effect on the substrate specificity of this isoform. Although helix α9 folds back over the active site in wheat J, this secondary structure element is relatively short when compared to the C-terminal helical element observed in the Alpha and mammalian Theta class GSTs. The H-site is therefore probably more accessible

Original in colour

in the wheat J structure than in the Alpha and mammalian Theta classes which have been described as having to 'breathe' to allow entry and egress of substrate.

8.2 Discussion

8.2.1 Analysis of primary sequence of wheat J

Since the elucidation of the first plant glutathione structure *At*GSTF2 (Reinemer et al., 1996) there has been a number of papers published comparing sequences of the Phi, Zeta and Tau class GSTs (Droog, 1997; Neuefeind et al., 1997c). Following sequence alignment of full length plant GSTs, Droog characterised Tau class GSTs as having several conserved sections of sequence most notably His-Lys-Lys at position 53-58, His-Asn-Gly at position 60-62, Phe156 and Gly159. (These residues are marked with -- in figure 8-6, note numbering in this figure is for wheat J and is not the same as used by Droog (Droog, 1997)). Based on the *At*GSTF2 structure, Droog proposed that the conserved lysine residues of His-Lys-Lys might interact with the glycyl moiety of glutathione, however this is not observed in the wheat J structure. In agreement with suggestions based on the *At*GSTF2 structure Glu68, Ser69 and Lys42 are found to interact with glutathione in the wheat J structure, however Lys41 and Lys55 do not. The conserved His-Asn-Gly motif present in the Tau class (Droog, 1997) is explained by the wheat J structure and clearly through the interaction of a number of hydrogen bonds plays a structural role in the orientation of helix α3 with respect to the β-sheet. These hydrogen bonds are formed between residues His61 (strand β4), Asp5 (strand β1) and the conserved Tyr75 (helix α3). An additional hydrogen bond to the mainchain amide of His6 by Asn62 also stabilises this loop region (figure 8.9).

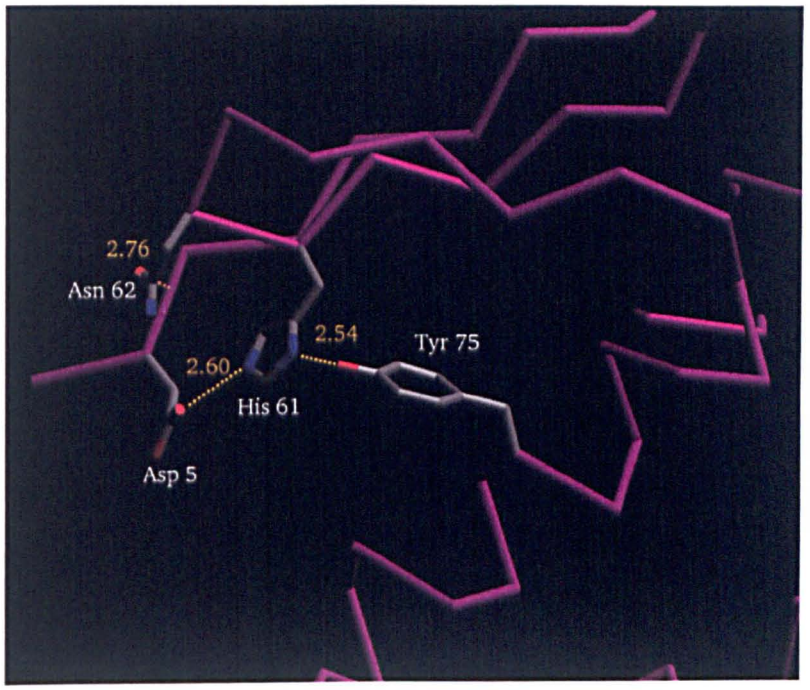


Figure 8-9: Diagram showing the structural role of the conserved motif His-Asn-Gly (residues 61-63). Distances of the respective hydrogen bonds are shown in Angstroms. This picture was made in SETOR (Evans, 1993).

8.2.2 Comparison of the wheat J enzyme with other GST structures

The tertiary structure of wheat J was compared with isoenzymes of the classes; Alpha (Sinning et al., 1993), Mu (Ji et al., 1992), Pi (Reinemer et al., 1991), mammalian and plant Theta (Rossjohn et al., 1998), Sigma (Ji et al., 1995), Beta (Nishida et al., 1998), Phi (Reinemer et al., 1996), Delta (Wilce et al., 1995), Zeta, Omega (Board et al, 2000) and GST from *Schistosoma japonicum* (McTigue et al., 1995) (Table 8-5 and Figure 8-10, note Delta class coordinates are not available).

Original in colour

	Alpha	Mu	Pi	Theta	Theta	Sigma	Beta	Phi	Delta	Zeta	Omega	*S. jap.*
Sequence identity (%)	18	20	20	21	19	17	19	18	21	23	23	18
r.m.s.d (Å)	1.69	1.55	1.72	1.58	1.55	1.64	1.50	1.36		1.37	1.34	1.59
Over no. of residues	140	147	154	172	182	153	163	167		153	179	150

Table 8-5; Root mean square deviations (r.m.s.d.) for each enzyme in comparison to wheat J together with sequence identity. Classes compared were; Alpha (Sinning et al, 1993), Mu (Ji et al., 1992), Pi (Reinemer et al, 1991), mammalian and plant Theta (Rossjohn et al, 1998), Sigma (Ji et al., 1995), Beta (Nishida et al., 1998), Phi (Reinemer et al, 1996), Delta (Wilce et al, 1995), Omega (Board et al., 2000), Zeta, and the GSTs from *Schistosoma japonicum* (McTigue et al, 1995) (table 8-5 and figure 8-8, note Delta class coordinates are not available).

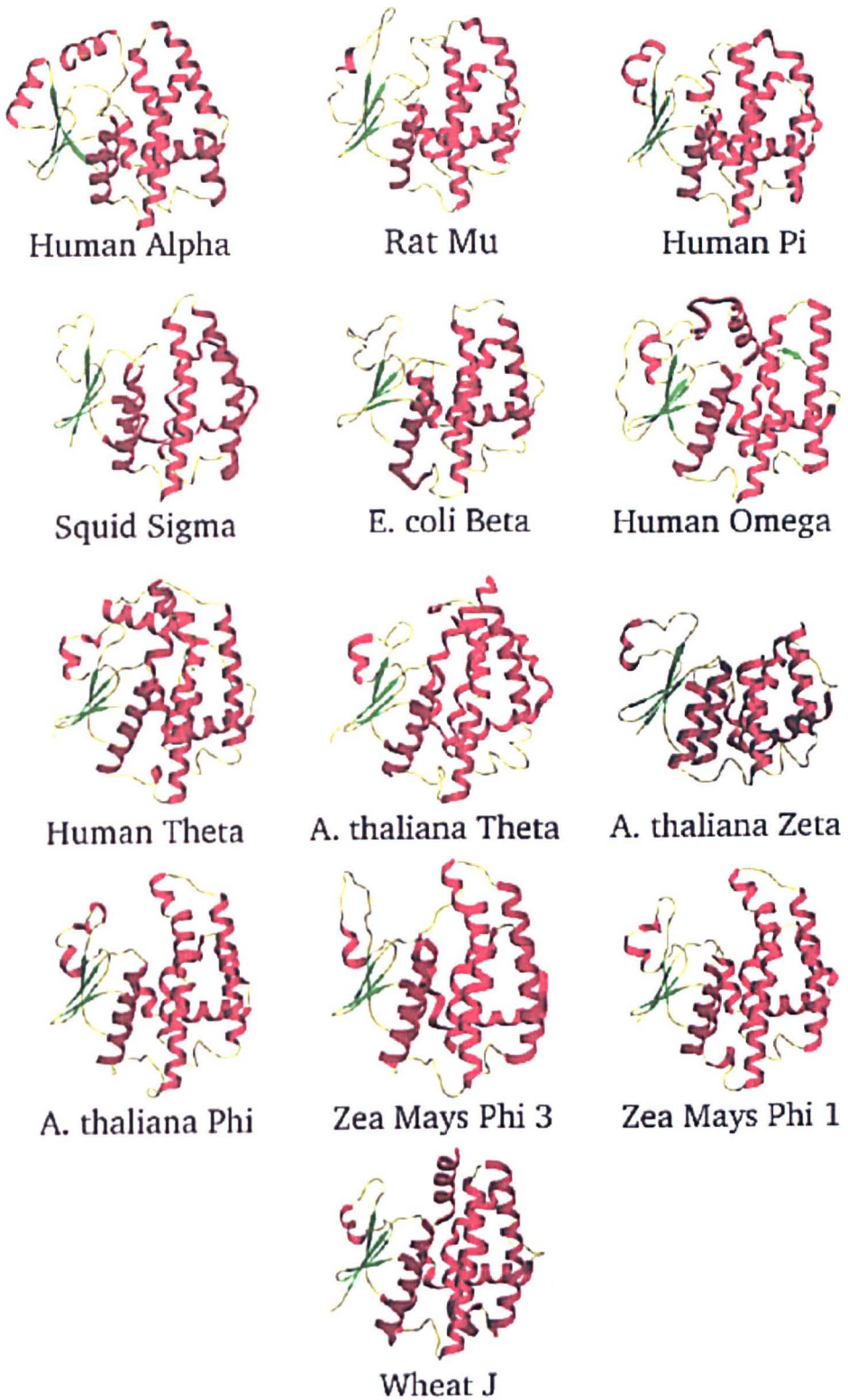


Figure 8-10: Superimposition of monomer structures of the GST classes solved to date on wheat J. Structures are 1GUH (human Alpha), 4GST (rat Mu), 1GSS (human Pi), 2GSQ (squid Sigma), 1A0F (*E. coli* Beta), 1EED (human Omega), 1IJR (human Theta), *A. thaliania* Type Theta (not yet deposited), 1E6B (*A. thaliania* Zeta), 1GNW (A. *thaliana* Phi), 1AW9 (*Zea mays* F3) and 1AXD (*Zea mays* F1). Images in this figure were drawn using SETOR (Evans, 1993).

Original in colour

When the structural core of wheat J is superimposed on GST structures from other classes, the degree of structural similarity decreases in the order, N-terminal domain > monomer > dimer. The monomer structure of the wheat J isoform superimposes most closely onto the human Omega class enzyme, although it lacks the unique proline rich segment observed at the N-terminus of the Omega class GST (Board et al, 2000). Whilst wheat J is generally similar to the Omega, Phi, Zeta and Theta N-terminal domains, the loop structure between β2-β3 of the wheat J enzyme adopts a different conformation. This alternative conformation does not contain the previously described aromatic stacking interactions observed in the structures of the other classes (Reinemer et al., 1996; Rossjohn et al., 1998; Wilce et al., 1995). Within this loop region a network of hydrogen bonds are formed between the conserved residues Glu37 to the hydroxy and mainchain amide of Ser43; Ser49 to the mainchain carbonyl of Leu45, and the amine group of Lys42 to the carbonyl of Lys54. Following the conserved *cis*-proline after the novel conformation of this loop region in the wheat J enzyme, the Omega, Phi, Zeta and Theta N-terminal domains again superimpose well on the wheat J structure as it forms the characteristic glutathione binding site.

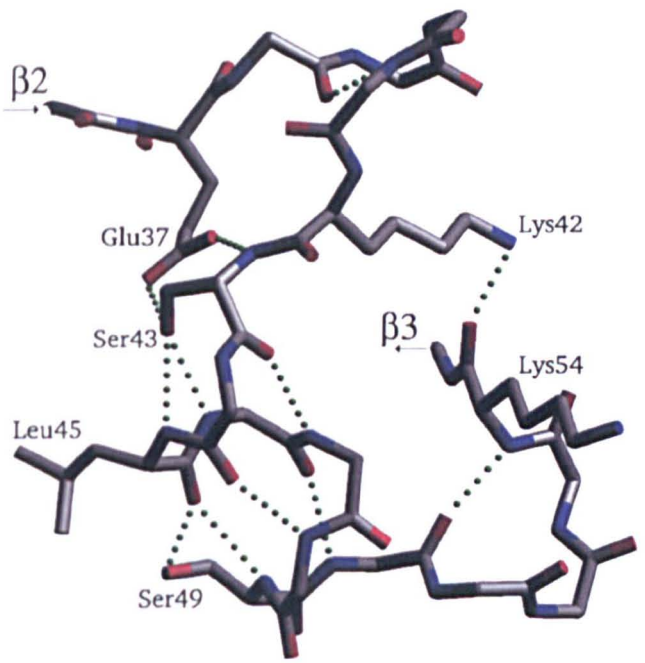


Figure 8-11: Hydrogen bonding within β2-β3 loop region. Residues mentioned in the text are marked appropriately. This figure was created using SETOR (Evans, 1993).

Original in colour

The C-terminal domain of wheat J displays some significant differences to the other structurally determined plant GSTs. Although the helical element α4' is of similar orientation and length to that previously observed (Neuefeind et al., 1997a; Neuefeind et al., 1997b; Reinemer et al., 1996), α4'' observed in wheat J is shorter and tilted at an angle of around 15° in comparison α4'' of *At*GSTF2 (Reinemer et al., 1996). This movement of the α4-α5 loop region with respect to Phi class structures means that wheat J adopts a conformation more similar to that of the Beta class structures (Nishida et al., 1998) within this part of the structure. This alternation in the positioning of the α4-α5 loop region promotes the Tau class structure to be significantly more open to substrates approaching from the dimer interface.

The other secondary structure elements within the C-terminal domain of the wheat J structure are generally similar to those of other plant GSTs, however the loop regions that join the secondary structure elements form larger loops in the wheat J structure. Aspartate (residue 162) is conserved in the Tau class GSTs and as in other GST structures it is buried in the hydrophobic core of the C-terminal domain. Buried in the core of the C-terminal domain it forms hydrogen bonds to the mainchain amide groups of Phe153, Gly154, Gly159 and the nitrogen of the indole group on Trp195.

Together with the different orientation of helices α4 and α5, the most significant feature of the C-terminal domain in wheat J is the long C-terminal tail (α9). Although this C-terminal helix lies at the rear of the C-terminal domain and is orientated toward the active site, it does not appear to occlude the active as is observed for C-terminal helix of the mammalian Alpha and Theta class GSTs. Helix α9 shares significant structural similarity with that described for the human Omega class GST and is similar in both length and orientation (figure 8-12). Of particular note is that the charged residue Arg222 of wheat J lies in a similar position as Tyr229 in the Omega Class structure (Board et al., 2000). In both cases the guanido group of the arginine and the hydroxyl moiety of the tyrosine are pointed into the active site providing a potential hydrogen bond. The presence of these residues in the respective enzymes causes the active site to be

considerably less hydrophobic than it would be otherwise. In addition, Arg118 of the wheat J enzyme shares a similar position as Arg132 in the Omega class structure and is situated such that it caps the H-site.

The exact nature of the substrate best suited to bind the large, mostly hydrophobic H-site of the Omega and Tau class structures, but also able to interact with these polar groups is unknown.

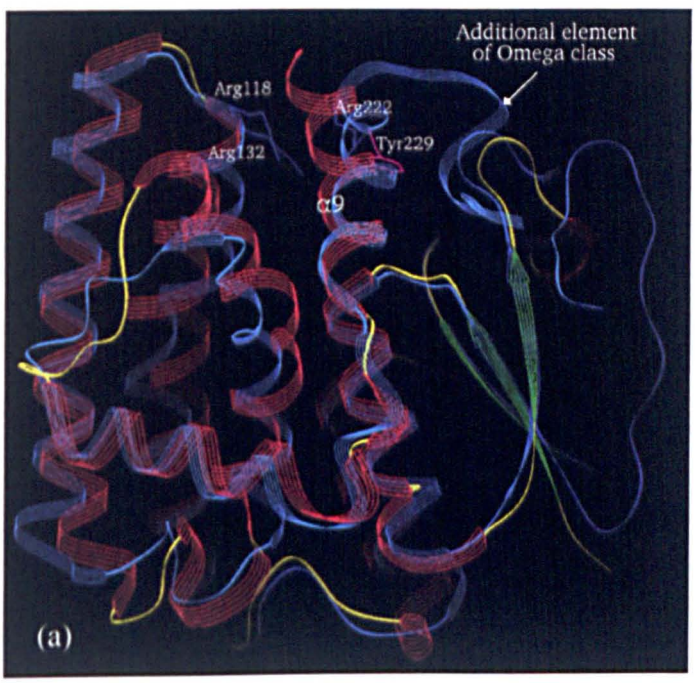


Figure 8-12: Comparison of wheat J and human Omega class GST opposite side of the monomer from the dimer interfacce. Helix α9 is of similar length and orientation for both classes. Further both enzymes contain charged residues in similar positions orientated into the otherwise hydrophobic H-site. In contrast to the Omega class structure the wheat GST has a shorter C-terminus and does not form an additional element after helix α9. Wheat J is shown in red, while the human Omega class GST (Board et al., 2000) is shown in blue. This picture was made using SETOR (Evans, 1993).

8.2.3 Oxidative stress.

Wheat J is known to bind a number of herbicides including Fenoxaprop and dimethenamid, however no potential endogenous substrates for this isoenzyme have been established. The elucidation that common features are present between the active site of wheat J and the human Omega class (GSTO 1-1) suggests that these GSTs might bind similar substrates.

Original in colour

It has been shown that a mouse ortholog of the human Omega class GST is overexpressed in radiation sensitive cells leading to the postulation that the Omega class GSTs may provide protection from oxidative stress (Board, 2000). If Omega class GSTs do indeed function to protect the cell from products of oxidative stress then the shared features between the active sites of GSTO 1-1 and wheat J suggests the Tau class GSTs might undertake a similar endogenous role. This hypothesis that Tau class GSTs might function to protect the cell during oxidative stress is supported by a recent publication which describes that a GST from tomato acts as an inhibitor against the apoptosis inducing protein BAX (Kampranis et al., 2000).

Apoptosis, or programmed cell death is a mechanism conserved through evolution having a role in organogenesis and removal of autoreactive clones in the immune systems of mammals, and also occurring in both development and at sites of pathogen invasion in plants. Bax is thought to generate its proapoptotic effect through alteration of organelle function by localization to the outer mitochondrial membrane and formation of an ion channel (Zha et al, 1996). It has been suggested that Bax toxicity might be related to oxidative control processes. In plants Bax expression causes localized tissue collapse, similar to as observed during the hypersensitive response (Lacomme & Cruz, 1999).

The above-mentioned publication describes the use of a yeast two-hybrid system designed to screen for plant proteins able to protect against Bax induced apoptosis. The yeast two hybrid system revealed that the best suppressor of Bax activity detected from a tomato library was a glutathione S-transferase. Although the tomato GST which conferred protection from apoptosis was classed as a member of the Theta class by Kampranis et al., sequence comparison clearly indicates that the tomato GST belongs to the Tau class. It is unknown the way in which the GST conferred protection from apoptosis however it was shown that the GST did not directly interact with Bax (Kampranis et al, 2000).

```
            1                                                  50
  AAF22647   MANDEVILL DFWPSMFGMR LRIALAEKEI KYEYRDEDL- --RNKSPLLL
wheat_J_GS  MAGGDDLKLL GAWPSPFVTR VKLALALKGL SYEDVEEDL- --YKKSELLL
Blow_Fly_D        XMDFY YLPGSAPCRS VLMTAKALGI ELNKKLLNLQ AGEHLKPEFL
Human_Thet      XMGLELF LDLVSQPSRA VYIFAKKNGI PLELRTVDLV KGQHKSKEFL
ZmGSTF1_Ph       APMKLY GAVMSWNLTR CATALEEAGS DYEIVPINFA TAEHKSPEHL
 Consensus  .......... ....S..... .......... .......... .........L
            51                                                 100
  AAF22647  QMNPIHKKIP VLIHNGKPIC ESIIGVEYID EVWKDKAP-F LPSDPYERAQ
wheat_J_GS  KSNPVHKKIP VLIHNGAPVC ESMIILQYID EVFASTGPSL LPADPYERAI
Blow_Fly_D  KINPQHT-IP TLVDGDFALW ESRAIMVYLV EKY-GKNDSL FPKCPKKRAV
Human_Thet  QINSLGK-LP TLKDGDFILT ESSAILIYLS CKY-QTPDHW YPSDLQARAR
ZmGSTF1_Ph  VRNPFGQ-VP ALQDGDLYLF ESRAICKYAA RK--NKPELL REGNLEEAAM
 Consensus  ..N......P .L........ ES.....Y.. .......... .........A.
            101                                                150
  AAF22647  ARFWADYIDK KLYDSGRKLW TTK------- GGEQET-AKK DFIECLKVLE
wheat_J_GS  ARFWVAYVDD KLVAPWRQWL RGK------- TEEEKSEGKK QAFAAVGVLE
Blow_Fly_D  INQRLYFDMG TLYKSFADYY YPQIFAK--- APADPELYKK MEAAFD----
Human_Thet  VHEYLGWHAD CIRGTFGIPL WVQVLGPLIG VQVPEEKVER NRTAMDQALQ
ZmGSTF1_Ph  VDVWIEVEAN QYTAALNPIL FQVLISPMLG GTTDQKVVDE NLEKLKKVLE
 Consensus  .......... .......... .......... .......... ..........
            151                                                200
  AAF22647  GALGEKP--- -YFGGGNFGC VDIALIGYYS WFYAYESYAN ISV--EAECP
wheat_J_GS  GALRECSKGG GFFGGDGVGL VDVALGGVLS WMKVTEALSG DKIFDAAKTP
Blow_Fly_D  -FLNTFLEGH QYVAGDSLTV ADLALLASVS TFEVAGF-D- FS-----KYA
Human_Thet  WLEDKFLGDR PFLAGQQVTL ADLMALEELM QPVALGY-EL FE-----GRP
ZmGSTF1_Ph  VYEAR-LTKC KYLAGDFLSL ADLNHVSVTL CLFATPYASV LD-----AYP
 Consensus  .......... ....G..... .D........ .......... ..........
            201                                                250
  AAF22647  KFVAWAKNCM LRDSVAKSLP DQHKVCEFVK VLRQKFGIE
wheat_J_GS  LLAAWVERFI ELDAAKAALP DVGRLLEFAK AREAAAAASK
Blow_Fly_D  NVAKWYANAK TV--APGFDE NWEGCLX
Human_Thet  RLAAWRGRVE AFLGAELCQE AHSIILSILE QAAKKTLPTP SPEAYQAMLL
ZmGSTF1_Ph  HVKAWWSGLM E-------RP SVQKVAALMK PSA
 Consensus  ....W..... .......... .......... .......... ..........
            251
  AAF22647
wheat_J_GS
Blow_Fly_D
Human_Thet  RIARIPX
ZmGSTF1_Ph
 Consensus  .......
```

Figure 8-13: The peptide sequence of the novel tomato GST/GPX detected by Kampranis *et al.* (AAF22647), compared to the sequences of wheat J (AF004358), a Delta class GST from *Lucilia cuprina* (L23126), a Theta class GST from Human (P30712) and a Phi class GST from Maize (P12653). Identical residues between the tomato GST and wheat J are highlighted yellow, characteristic sequence motifs of the Tau class are highlighted in green.

The conjugation of cytotoxic alkenals by glutathione S-transferases is well documented in animals (Hayes & Pulford, 1995) and both Phi GSTs from

sorghum and Tau class GSTs from wheat have been shown to detoxify 4-hydroxynonenal, a toxic alkenal released following oxidative damage in plants (Droog, 1997; Marrs, 1996). In addition numerous plant GSTs have been shown to be induced by agents including auxins, cytokinins, ABA, heavy metals, SA, H_2O_2, ethanol and conditions such as wounding, cold, flooding and pathogen invasion (see Marrs, 1996 & Droog, 1997 for review). The common step in all these factors is the generation of active oxygen species (AOS). It is thus conceivable that a range of GSTs, with particular substrate specificities towards compounds generated by the reactive oxygen species, might exist to protect cells from damage and that the tomato GST is acting to 'mop' up the compounds created by the action of Bax on the cell.

8.2.4 Mutagenesis of Maize GSTs

In conjunction with the crystallisation studies being performed in Glasgow, David Dixon (Dr Robert Edward's group, Crop Protection Unit, Durham University) was undertaking to modify the substrate specificities of previously cloned GSTs, to either increase activity against a chosen substrate or alter activity towards previously unrecognised substrates. This was performed using reassembly PCR (Stemmer, 1994) in which plasmids which express the Tau class GSTs *Zm*GSTU1 and *Zm*GSTU2 were individually digested using DNase I to produce DNA fragments, 200 base pairs in length. These fragments were then used as templates for up to 60 cycles of assembly PCR (no primers used) where at each cycle the partially overlapping fragments annealed and were extended to produce double stranded DNA. A screening protocol based on the detection of yellow coloured products following glutathione conjugation with the herbicide flurodifen, enabled detection GST mutants with increased activity towards the flurodifen substrate (personal communication with David Dixon).

Using the reassembly PCR technique together with the above screening method several mutants were discovered with activity towards flurodifen at least five times that of wild type enzyme (Table 8-6 and Figure 8-12).

Enzyme	Activity in crude bacterial lysate	
	CDNB (ΔOD_{340}/30s)	Fluorodifen (ΔOD_{400}/1hr)
pBluescript-*Zm*GSTU1	1.13	~0.6
pBluescript-*Zm*GSTU2	1.02	~0.6
56-5-D5	0.05	3.9
56-12-E1	0.37	4.3
56-30-B9	0.40	6.1
56-33-A12	1.68	14.1

Table 8-6: Adapted from table presented by David Dixon. Comparison of activities in crude bacterial lysate of wild type and mutated GSTs. Mutated maize GSTs, *Zm*GSTU1 and *Zm*GSTU2, were produced by random mutatagenisis followed by selection of mutants with increased activity toward flurodifen.

```
ZM_GST5
   GST6/6
  56-5-D5  MTMITPSSKL TLTKGNKSWS STAVAAALEL VDPPGCRNSA RA--NAGGE
 56-33-A12 MTMITPSSKL TLTKGNKSWS STAVAAALEL VDPPGCRNSA RA--NAGGE
  56-30-B9 MTMITPSSKL TLTKGNKSWS STAVAAALEL VDPPGCRNSA RA--NAGGE
  56-12-E1 MTMITPSSKL TLTKGNKSWS STAVAAALEL VDPPGCRNSA RALIHFGIS

  ZM_GST5        MAEE KKQGLQ---- --------LL DFWVSPFGQR CRIAMDEKG
   GST6/6                       MAAAAEVVLL DFWVSPFGQR CRIALAEKG
  56-5-D5  NKKKLDMAEE KKQGLQ---- --------LL DFWVSPFGQR CRIAMDEKG
 56-33-A12 NKKKLDMAEE KKQGLQ---- --------LL DFWVSPFGQR CRIAMDEKG
  56-30-B9 NKKKLDMAEE KKQGLQ---- --------LL DFWVSPFGQR CRIAMDEKG
  56-12-E1 NAFALPATSL RGRSTEQLVA MAAAAEVVLL DFWVSPFGQR CRIALAEKG

  ZM_GST5  LAYEYLEQDL GNKSELLLRA NPVHKKIPVL LHDGRPVCES LVIVQYLDE
   GST6/6  VAYEYREQDL LDKGELLLRS NPIHKKIPVL LHAGRPVCES LVILQYIDE
  56-5-D5  LAYEYLEQDL GNKSELLLRS NPIHKKIPVL LHAGRPVCES LVILQYIDE
 56-33-A12 LAYEYLEQDL GNKSELLLRA NPVHKKIPVL LHAGRPVCES LVILQYIDE
  56-30-B9 LAYEYLEQDL GNKSELLLRS NPIHKKIPVL LHAGRPVCES LVILQYIDE
  56-12-E1 VAYEYREQDL LDKGELLLRS NPIHKKIPVL LHAGRPVCES LVILQYIDE

  ZM_GST5  AFPAAAPALL PADPYARAQA RFWADYVDKK LYDCGTRLWK LKGDGQAQA
   GST6/6  AWPDVAPLLP KDDPYARAQA RFWADYIDKK IYDSQTRLWK FEGEAREQA
  56-5-D5  AWPDVAPLLP KDDPYARAQA RFWADYIDKK IYDSLTRLWK FEGEAREQA
 56-33-A12 AWPDVAPLLP KDDPYARAQA RFWADYIDKK LYDCGTRLWK FEGEAREQA
  56-30-B9 AWPDVAPLLP KDDPYARAQA RFWADYVDKK LYDCGTRLWK LKGDGQAQA
  56-12-E1 AWPDVA-LLP -ADPYARAQA RFWADYVDKK LYDCGTRLWK FEGEAREQA

  ZM_GST5  RAEMVEILRT LEGALGDGPF FGGDALGFVD VALVPFTSWF LAYDRFGGV
   GST6/6  KKDLVEVLET LEGELADKPF FGGGALGFVD VALVPFTSWF LAYEKLGGF
  56-5-D5  KKDLVEVLET LEGELADKPF FGGGALGFVD VALVPFTSWF LAYEKLGGF
 56-33-A12 KKDLVEVLET LEGELADKPF FGGDALGFVD VALVPFTSWF LAYEKLGGF
  56-30-B9 KKDLVEVLET LEGELADKPF FGGGALGFVD MALVPFTSWF LAYEKLGGF
  56-12-E1 KKDLVEVLET LEGELADKPF FGGGALGFVD VALVPFTSWF LAYEKLGGF

  ZM_GST5  SVEKECPRLA AWAKRCAERP SVAKNLYPPE KVYDFVCGMK KRLGIE
   GST6/6  SVQEHCPRIV AWAARCRERE SVAKAMSDPA KVLEFVQFLQ SKFGAK
  56-5-D5  SVQEHCPRIV AWAARCRERE SVAKAMSDPA KVLEFVQFLQ SKFGAK
 56-33-A12 SVQEHCPRIV AWAARCRERE SVAKAMSDPA KVLEFVQFLQ SKFGAK
  56-30-b9 SVQEHCPRIV AWAARCRERE SVAKAMSDPA KVLEFVQFLQ SKFGAK
  56-12-E1 SVQEHCPRIV AWAARCRERE SVAKAMSDPA KVLEFVQFLQ SKFGAK
```

Figure 8-14: Multiple alignment of flurodifen high activity mutants with wild type *Zm*GSTU1 and *Zm*GSTU2 presented by David Dixon. Regions of *Zm*GSTU1 are shown in blue while regions of *Zm*GSTU2 are shown in red, conserved regions between the two enzymes are shown in black. The amino acid substitutions in 56-5-D5 (Gln to Leu) and 56-39-B9 (Val to Met) are highlighted yellow, together with deletions present in 56-12-E1.

Biochemical characterisation of these mutants by David Dixon, determined that the mutant 56-5-D5 was expressed at significantly lower levels than the others were. Taking this level of expression into account means that 56-5-D5 is around 100 times more active towards flurodifen than the wild type GSTs, *Zm*GSTU1 and *Zm*GSTU2. The mutant 56-5-D5 was also shown to have increased activity

Original in colour

towards the diphenylether herbicides fomesafen and acifluoren (personal communication with David Dixon).

The chimeric GST 56-5-D5 displayed a point mutation at residue 115 in comparison to either the wild type enzymes *Zm*GSTU1 and *Zm*GSTU2. During the course of these mutation studies no structural information for the Tau class GSTs was available and thus it was difficult to suggest the role this substituted amino acid might have in substrate binding. To further characterise this mutant biochemically, David Dixon created several constructs substituting the leucine present in 56-5-D5 with a series of amino acids with different characteristics. Site directed mutagenesis of the substituted leucine residue of the mutant 56-5-D5 to Glu, Asp, Phe, Ala was performed to determine which properties of residues in this position resulted in increased flurodifen activity. The asparagine (shorter than glutamine, but same charge characteristics) and glutamic acid (similar structure to glutamine, but different charge) mutants were found to express almost entirely as insoluble protein. In comparison the phenylalanine and alanine constructs were found to express soluble protein. CDNB:flurodifen activity ratios revealed that the alanine substitution showed increased activity toward flurodifen in comparison to the phenylalanine substitution and that overall the substitution of the wild type glutamine residue to uncharged residues resulted in increased flurodifen activity.

Attempts to understand these biochemical results using the models of the Phi class GSTs *Zm*GSTF1, *Zm*GSTF3 and *At*GSTF2 (share around 25%, 28% and 23% identity with *Zm*GSTU1 and *Zm*GSTU2 respectively) was unsuccessful. The Phi class structures suggested that the substituted residue did not lie in the active site, but was involved in packing within the C-terminal domain.

Structural determination of the wheat J enzyme (shares 41% and 42% sequence identity with the *Zm*GSTU1 and *Zm*GSTU2 sequences respectively) enabled a better comparison to be performed. Comparison of maize Tau class sequences *Zm*GSTU1 and *Zm*GSTU2 with the wheat J structure indicated that due to the

considerable movement of the α4 -α5 loop region the leucine residue of the mutant 56-5-D5 would form part of the hydrophobic binding site.

The sequence similarity between wheat J, *Zm*GSTU1 and *Zm*GSTU2 allowed a theoretical model to be generated in Swiss Model based on the wheat J structure (Guex & Peitsch, 1997; Peitsch, 1995; Peitsch, 1996).

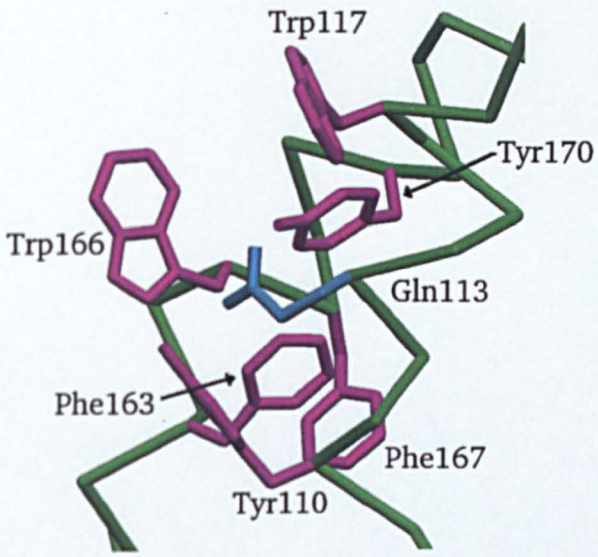


Figure 8-15: Analysis of the structural environment around Gln113 in the theoretical models of the maize Tau class GSTs (substituted to a leucine residue in the 56-5-D5 mutant) suggests that this residue is situated in a hydrophobic pocket.

Examination of the environment surrounding Gln113 in the theoretical model of *Zm*GSTU2 agrees with the biochemical evidence obtained by David that the substitution of a charged residue in this region would lead to unfavourable interactions. It would be expected that mutants containing a charged residue in this position would have difficulty folding leading to their insoluble expression. The increased activity of the *Zm*GSTU2 mutants toward flurodifen may be due to increased flurodifen binding affinity or an increase in the turnover rate. Flurodifen is a hydrophobic molecule that shows limited solubility. It can be envisaged that the active site of phenylalanine mutant might bind flurodifen molecules with higher affinity and thus reduce the turnover rate in comparison to the alanine mutant. Alternatively the phenylalanine mutant might have less

Original in colour

activity than the alanine mutant due to steric hindrance of the substrate on approach to the active site or efflux of the product from the active site.

A lack of knowledge of the binding of the flurodifen substrate in the Tau class GSTs means that the enhanced catalytic activity of the alanine mutant cannot be fully explained. Further structural and biochemical activity / binding studies will be required to determine what key features for substrate binding and catalysis are present in Tau Class GSTs.

```
                    10         20         30         40
wheat_J_GS    MAGGDDLK LLGAWPSPFV TRVKLALALK GLSYEDVEED LYKKSELLL
   ZM_GST5  MAEEKKQGLQ LLDFWVSPFG QRCRIAMDEK GLAYEYLEQD LGNKSELLL
    GST6/6    MAAAAEVV LLDFWVSPFG QRCRIALAEK GVAYEYREQD LLDKGELLL
 Consensus  .......... LL..W.SPF. .R...A...K G..YE..E.D L..K.ELLL
            50         60         70         80         90
wheat_J_GS  KSNPVHKKIP VLIHNGAPVC ESMIILQYID EVFASTGPSL LPADPYERA
   ZM_GST5  RANPVHKKIP VLLHDGRPVC ESLVIVQYLDE AFPAAAPAL LPADPYARA
    GST6/6  RSNPIHKKIP VLLHAGRPVC ESLVILQYIDE AWPDVAPLL PKDDPYARA
 Consensus  ..NP.HKKIP V..........  ..........  .........  .........
            100        110        120        130        140
wheat_J_GS  IARFWVAYVD DKLVAPWRQW LRGKTEEEKSE GKKQAFAAVG VLEGALRE
   ZM_GST5  QARFWADYVD KKLYDCGTRL WKLKGDG-QAQ ARAEMVEILR TLEGALGD
    GST6/6  QARFWADYID KKIYDSQTRL WKFEGEA-REQ AKKDLVEVLE TLEGELAD
 Consensus  ..........  ..........  ..........  .........  .LEG.L..
            150        160        170        180        190
wheat_J_GS  CSKGGGFFG GDGVGLVDVAL GGVLSWMKVTEA LSGDKIFDAAK TPLL
   ZM_GST5  ----GPFFG GDALGFVDVAL VPFTSWFLAYDR FGGVSV--EKE CPRL
    GST6/6  ----KPFFG GGALGFVDVAL VPFTSWFLAYEK LGGFSV--QEH CPRI
 Consensus  ......FFG G...G.VDVAL ....SW......  ..G......    .P.. .
                   200        210        220        230
wheat_J_GS  AAWVER FIELDAAKAA LPDVGRLLEF AKAREAAAAA SK
   ZM_GST5  AAWAKR CAERPSVAKN LYPPEKVYDF VCGMKKRLGI E
    GST6/6  VAWAAR CRERESVAKA MSDPAKVLEF VQFLQSKFGA K
 Consensus  .AW..R ..E.......  .........F ..........  ..
```

Figure 8-16: Alignment of Wheat J with *Zm*GSTU1 and *Zm*GSTU2. The glutamine residue that is altered to leucine in the site directed mutation present in the 56-5-D5 and the corresponding tryptophan referred to in the above text is highlighted in yellow. The threonine residue that H-bonds to the tryptophan in wheat J together with the corresponding tyrosine residue, are highlighted in green.

Original in colour

The limited knowledge concerning the factors which determine substrate specificity means that it is difficult to speculate why the two residue deletion within mutant 56-12-E1 which occurs in the linker region between domains promotes an increase in activity. In addition the single site mutation observed for the mutant 65-30-B9 also appears to be associated to interdomain packing and it is difficult to determine what changes are causing the increase in activity.

The mutations produced by Dr David Dixon have enabled some insight into which residues are important in the binding of substrate molecules in the Tau class enzymes, however further structural characterisation of these mutations will be required before it is clear why these mutants show increased activity.

Chapter 9: General Discussion and Conclusions

9.1 Discussion of Ligandin Function

The major focus of research into GSTs not only in plants, but also in animals and insects has been their ability to metabolise a range of toxic exogenous compounds such as herbicides, drugs, antibiotics or insecticides. As a billion dollar market exists for maize herbicides alone, the commercial reasoning behind this research is clear. Unfortunately as a consequence relatively little research into the variety of additional functions GSTs possess has been carried out. These additional functions range from non enzymatic binding, transport (ligandin function) and cellular protection from oxidative stress to isomerase activity and these alternative functions may represent the true endogenous roles that GSTs have within the cell. Despite the major focus of this work being directed toward achieving a structural insight into the differential metabolism of herbicides by GSTs between species, considerable insight into the additional functions of GSTs has also been gained. In particular the structural determination of the Zeta class (maleylacetoisomerase) GST discussed in chapter six has provided the structure of a GST with a specialised role within the cell rather than a generalised role in detoxification. The structures of wheat J, *At*GSTT1 and *Zm*GSTF1 all provide an insight into the variety of plant GSTs and have enabled some of the key features of these enzymes to be better understood.

One of the first *in vivo* functions to be associated with plant GSTs was the last genetically defined step in anthocyanin synthesis, the movement of the endogenous substrate cyandidin-3-glucoside from the cytoplasm to the vacuole (Marrs et al., 1995). This activity, associated with the GSTs *Zm*GSTU4 from maize (Bronze-2 gene) and AN9 from petunia (Alfenito et al., 1998), was originally thought to involve conjugation of the anthocyanin molecule, cyandinin-3-glucoside, with glutathione. This process of glutathione conjugation was thought to target anthocyanin tagged molecules for removal from the cytosol to the vacuole via the GS-X pump (Marrs et al., 1995). However there have been

no reports of naturally occurring glutathione conjugates or cysteine conjugates (a common breakdown product of herbicide GSH conjugates) of anthocyanin. In addition, efforts to produce flavonoid-glutathione conjugates *in-vitro* by incubating various flavonoids with [35S]GSH and recombinant AN9 failed to produce radio-labelled flavonoids (Mueller et al., 2000).

In view of the above results together with a series of equilibration dialysis experiments recent proposals suggest that certain glutathione S-transferases may act as flavanoid carrier proteins (Mueller et al., 2000). Interestingly although AN9 and *Zm*GSTF4 can functionally complement each other, they share only 12% amino acid identity (figure 9-1) and do not appear to contain a common sequence motif suggestive of a ligandin binding site.

```
          1                                        50
BRONZE-2    MRVLGGEV SPFTARARLA LDLRGVAYEL LDEPLGP--K KSDRLLAANP VYGKIPVLLL
PETUNIA   MVVKVHGSAM AACPQRVMVC LIELGVDFEL IHVDLDSLEQ KKPEFLVLQP -FGQVPVIED
                                                100
BRONZE-2  PDGRAICESA VIVQYIEDVA RESGGAEAGS LLLPDDPYER AMHRFWTAFI DDKFWPALDA
PETUNIA   GDFR-LFESR AIIRYYAAKY EVKGSKLTGT TLEEKALVDQ WLEVESNNYN DLVYNMVLQL
                                      150
BRONZE-2  VSLAPTPGAR AQAAEDTRAA LSLLEEAFKD RSNGRAFFSG GDAAPGLLDL ALGCFLPALR
PETUNIA   L-VFPKMGQT SDLTLVTKCA -NKLENVFDI YE--QRLSKS KYLAGEFFSL ADLSHLPSLR
                              200                                       240
BRONZE-2  ACERLHGLSL IDASATPLLD GWSQRFAAHP AAKRVLPDTE KVVQFTRFLQ VQAQFRVHVS
PETUNIA   FLMNEGGFSH L-VTKRKCLH EWYLDISSRD SWKKVLDLMM KKISEIEAVS IPAKEEAKV
```

Figure 9-1: Sequence comparison of Bronze-2 (*Zm*GSTF4) and Petunia (AN9) amino acid sequences. Conserved residues are coloured red.

Flavonoid treatment of *Zm*GSTF3, *Zm*GSTF1 and AN9 was found to inhibit CDNB activity to the same extent for *Zm*GSTF3 and AN9 and to a lesser extent for *Zm*GSTF1. This was proposed to suggest that a common flavonoid binding site may be present on GSTs in the H site region of the active site (Mueller et al., 2000).

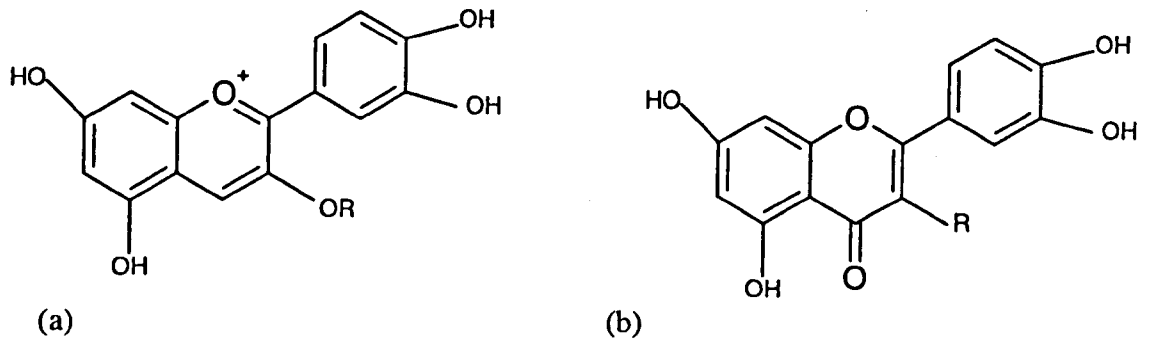


Figure 9-2: Structure of the flavonoids tested by Muller and coworkers (Mueller, 2000). (a) Structure of cyanidin (R=H) and cyanidin 3-glucoside (R=Glc). (b) Structure of quercetin (R=OH), luteolin (R=H), and isoquercitrin (R=glucosyl)

Protein fluorescence measurements of AN9 have shown that fluorescence is quenched very strongly following addition of the flavonoids luteolin and cyanidn, but not following the addition of cyanidin-3-glucoside and isoquercitin. This suggests that the binding of different flavonoid molecules may occur at different locations in the protein (Mueller et al., 2000).

In view of the results reported by Mueller together with the structures of several ligandin bound human Pi class GST enzymes (Ji et al, 1996; McTigue et al, 1995; Oakley et al, 1999) and plant GSTs (Neuefeind et al., 1997a; Neuefeind et al., 1997b; Reinemer et al., 1996), it was deemed worthwhile to try and determine if there was particular regions that flavonoid molecules might bind.

Several structures of a human Pi class GST complexed to number of large ligands including, sulfasalazine, Cibacron blue and bromosulfophthalein (figure 9-3) have been determined. These structures show these large ligands accommodated in the H-site of the enzyme, but not conjugated to glutathione. In addition, enzyme kinetic studies show that the binding of these molecules non-competitively inhibits 1-chloro-2,4-dinitrobenzene (CDNB) (Ketley et al, 1975; Mannervik & Danielson, 1988; Oakley et al, 1999).

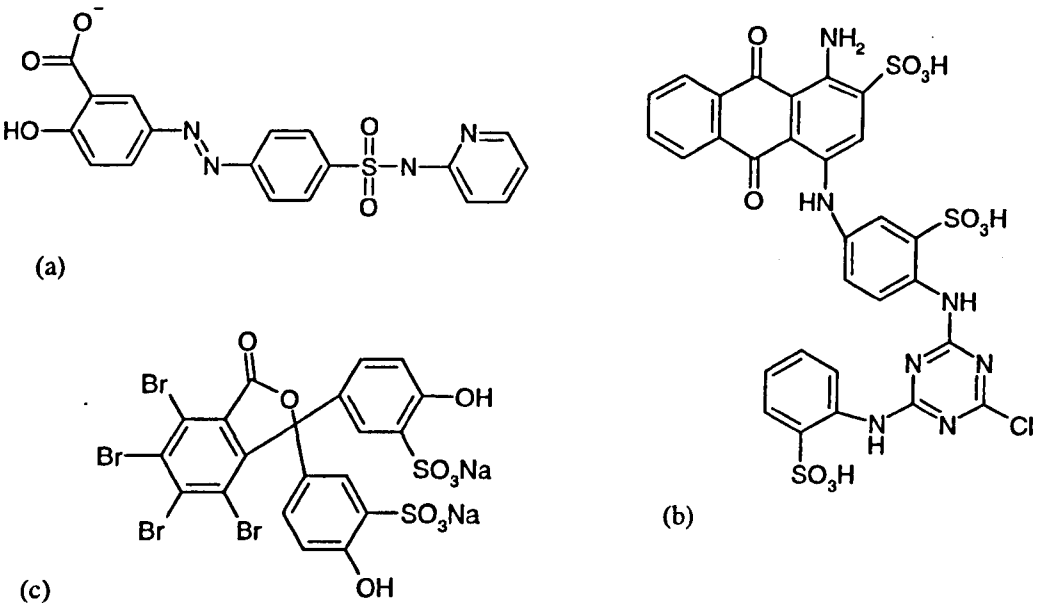


Figure 9-3: Schematic representation of the ligandin molecules (a) Sulfasalazine, (b) Cibacron blue and (c) bromosulfophthalein structurally characterised in complex with human Pi class GST (Oakley et al, 1999).

AN9 and *Zm*GSTU4 have yet to be structurally characterised and therefore cannot be directly compared to the human Pi class structures. However, as AN9 shares around 38% sequence identity with the structurally characterised GSTs *Zm*GSTF1 and *Zm*GSTF3, and *Zm*GSTU4 shares 40% sequence identity with wheat J, theoretical modelling strategies utilising structural templates were employed to determine the residues that might surround the H-site region in these two enzymes.

Three programs were used to attempt to model the two structures of AN9 and *Zm*GSTU4. It was hoped that by employing more that one program an estimate of how accurate the generated structures might be would be gained. The programs used were SWISS-MODEL (Guex & Peitsch, 1997; Peitsch, 1995; Peitsch, 1996), CPHmodels (Lund et al, 1997) and 3D-PSSM (Kelley et al, 2000).

These programs were chosen for the following reasons;

- They could be accessed from a web interface; http://www.expasy.ch/swissmod/SWISS-MODEL.html, http://www.cbs.dtu.dk/services/CPHmodels/ and http://www.bmm.icnet.uk/~3dpssm/)
- They were based on homology modelling or threading
- They allowed minimal user input
- The modelled proteins were provided as PDB files

The results obtained from these programs with respect to the two proteins AN9 and *Zm*GSTU4 were as follows.

9.1.1 Generation of theoretical models for AN9

SWISS-MODEL – Swiss model was run using the first approach mode for AN9 using both automatic template selection and providing the structures of *Zm*GSTF1, *Zm*GSTF3 and *At*GSTF2 as templates. In each case the automatic model building component of the program failed, suggesting that the alignment method was not sufficient to allow model generation.

CPHmodels – This program only allowed the user to enter a primary sequence of the protein to be modelled. It was able to generate a model of AN9 based on alignment with *At*GSTT2, (PDB entry 1GNW (Reinemer et al., 1996)) but many of the sidechains of the C-terminal domain were not modelled.

3D-PSSM – This program also only allowed the user to enter a primary sequence of the protein to be modelled. Several models are generated by this program based on a selection of template structures. However the model with the highest score was generated based on the structure of *Zm*GSTF3 (PDB entry 1AW9 (Neuefeind et al., 1997b), unfortunately only the Cα positions were detailed in the PDB file produced.

The models generated by CPHmodels and 3D-PSSM were calculated to have a Cα r.m.s.d. of 2.2Å using the program LSQKAB (CCP4, 1994). Visual inspection of the superimposed molecules revealed that the major differences

occurred in the loop region between helices α1 and α2, the linker region between the two domains and at the C-terminal helix.

9.1.2 Generation of theoretical models for ZmGSTU4

SWISS-MODEL – Swiss-model was run using the first approach mode as before. Using the auto template selection (trial 1), the model generated was based on the Human Omega structure (Board et al., 2000). As the Omega structure shares less sequence identity with *Zm*GSTU4 than the wheat J structure, trials were also run using the wheat J structure as a template alone (trial 2) and using both the wheat J and the human Omega structure as templates (trial 3). The models generated in trials 2 and 3 were very similar and as wheat J has a higher sequence similarity to *Zm*GSTU4 the model of *Zm*GSTU4 from trial 3 was adopted from this program and subjected to further analysis.

CPHmodels – The model generated by this program was based on the *At*GSTF2 structure (1GNW) (Reinemer et al., 1996) and this is probably because only a selection of structures from the PDB are included in the structural library of this program. The structure generated by this program contained all the Cα positions of the residues, but only around 10% of the sidechains were modelled.

3D-PSSM – The model of *Zm*GSTU4, with the highest score, generated by this program was based on the structure of *Zm*GSTF1 (PDB entry 1AXD (Neuefeind et al., 1997a)). The theoretical model produced was incomplete within the C-terminal domain, with several residues missing from the first two helices forming this domain. In addition, only the Cα positions were determined for the rest of the residues in the model.

Comparison of the three models of *Zm*GSTU4 generated revealed that they had a Cα r.m.s.d. of between 4.5-6Å with each other. As with the AN9 models and as expected the most significant differences between the structures existed in the linker region between the two domains, the C-terminus and the loop regions between the secondary structure elements. Despite the significant rms deviations obtained between the models generated by the different programs, visual

inspection of the superimposed structures revealed that the general orientation and length of the secondary structural elements were similar.

9.1.3 Discussion of the theoretical models and possible ligandin binding sites

Comparison of the human Pi class GST with the theoretical models of AN9 and *Zm*GSTU4, indicated large structural differences between the enzymes (figures 9-4 and 9-5). The different orientation of helix α4 of the C-terminal domain observed between the Pi class structure and the plant models provides the plant enzymes with a more open approach to the active site that the Pi class enzyme. Larger molecules would therefore not be sterically hindered from binding the plant GSTs in similar position to the ligandin binding site in the Pi class GST. However, there is virtually no sequence conservation or similar distinct structural features in this region between the Pi class GST and the theoretical models of AN9 and ZmGSTU4 to indicate a similar ligandin binding motif exists in this area of the plant structures.

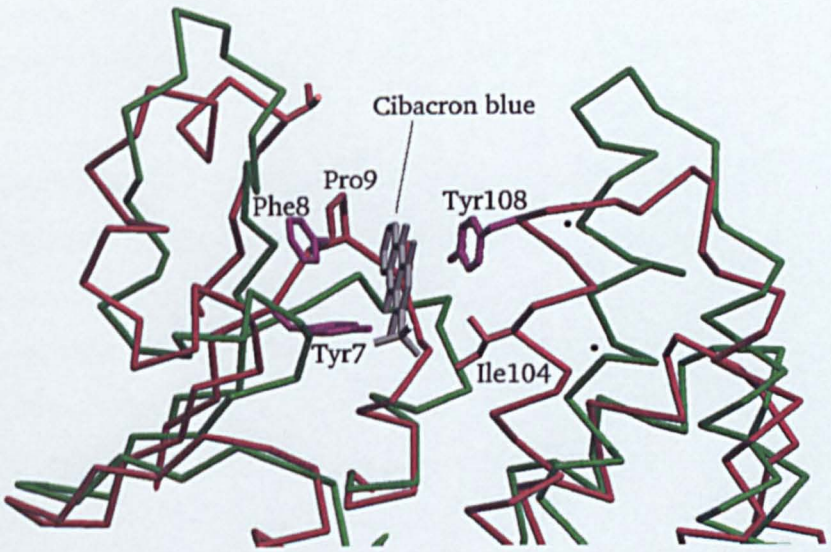


Figure 9-4: Comparison of human Cibracon blue ligandin bound Pi class structure (red) with the model of AN9 generated from 3D-PSSM (model from CPHmodels not shown to improve clarity). The residues discussed by Oakley et al., as binding the Cibracon blue ligandin are shown and the positions of corresponding residues in the H site of the AN9 molecule are marked• (Oakley et al, 1999). These marked residues are Asn113 and Val120 corresponding to Ile104 and Tyr108 of the Pi class molecule. In the N-terminal domain the residues Tyr7, Phe8, Pro9 and Val10 of the Pi class GST correspond Gly7, Ser8, Ala9 and Met10 of AN9.

Original in colour

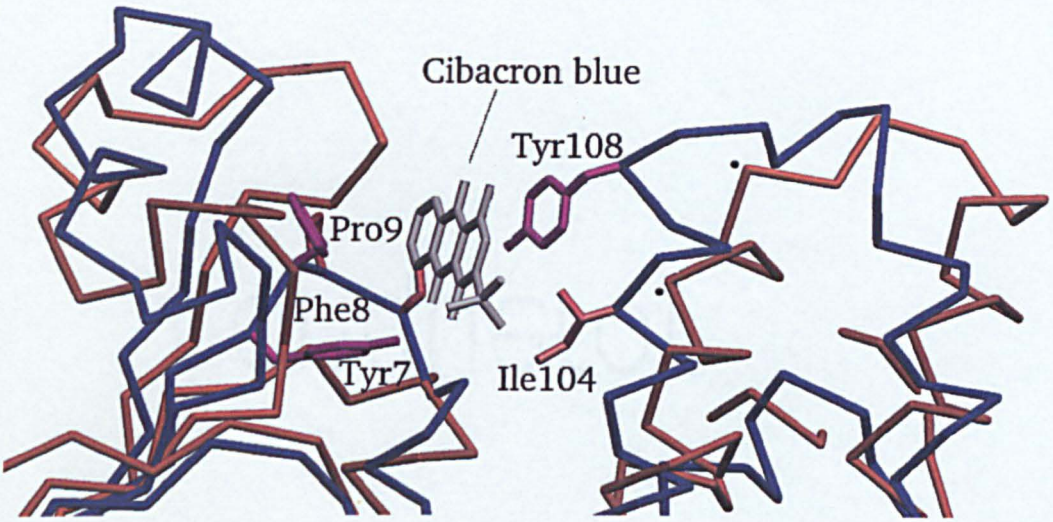


Figure 9-5: Comparison of human Cibracon blue ligandin bound Pi class structure (blue) with the model of *Zm*GSTU4 generated by SWISS-MODEL (This model was chosen as it was based on structures with closest sequence homology). The residues discussed by Oakley et al., as binding the Cibracon blue ligandin are shown and the positions of corresponding residues in the H site of the *Zm*GSTU4 molecule are marked •(Oakley et al, 1999). The marked residues are Leu115 and Asp119, which correspond to Ile104 and Tyr108 of the Pi class structure. The Pi class residues Tyr7, Phe8, Pro9 and Val10 of the N-terminal domain correspond to Gly5, Gly6, Glu7 and Val8. of *Zm*GSTU4 structure

In addition to the ligandin binding site detected in the human Pi class structure, the complexed structures of the antischistosomal drug Praziquantel with a *Schistosoma japonicum* GST (McTigue et al., 1995) and S-(3-iodobenxyl) glutathione with a squid (Sigma class) GST (Ji et al, 1996) have been described. Both of these structures indicate that the ligandin molecules bind at the dimer interface. Although the cavity in which these ligandins is present at the dimer interfaces in both structures, it does not occupy the same position. The Praziquantel molecule being found to bind more deeply between the two monomer subunits than the S-(3-iodobenxyl) glutathione. As shown in figure 9-6, structural alignment of both the *S. japonicum* GST and the squid GST indicates that with the exception of the Arg108 of *S. japonicum* which aligns with Lys103 of the squid GST, the residues which bind to the ligandin molecules lie in different positions in each structure.

Original in colour

Figure 9-6; Superimposition of the complex of the *Schistosoma japonicum* GST bound to Praziquatel (McTigue et al., 1995) (shown in yellow) with the ligandin bound S-(3-iodobenxyl) glutathione Sigma class GST (Ji et al, 1996), (shown in blue). Praziquatel is bound significantly more deeply between the dimer interface. Residue Arg108 of *S. japonicum* GST and Lys103 of Squid GST are also shown.

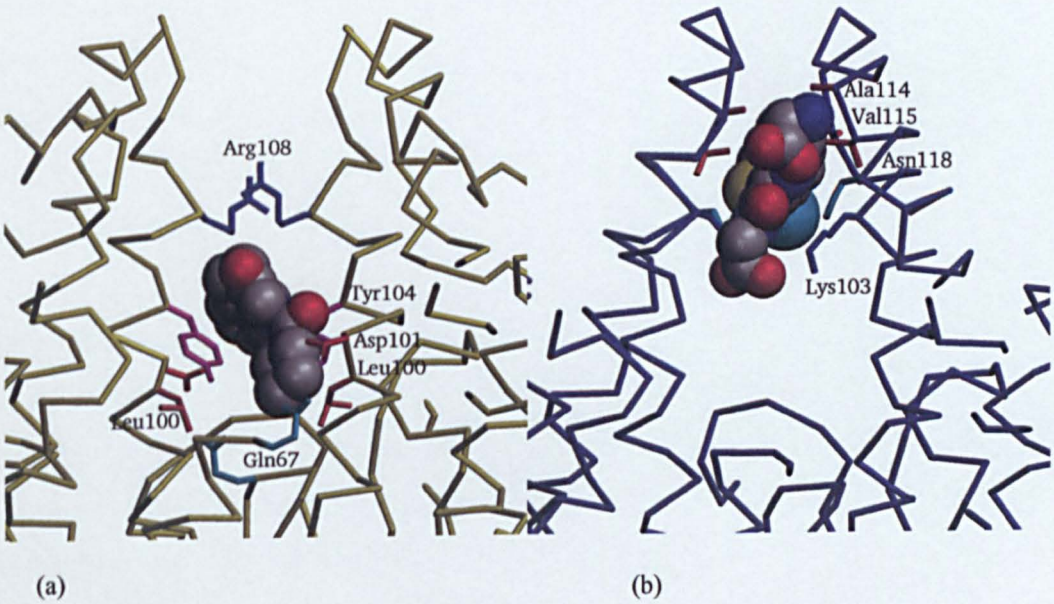


Figure 9-7: (a) Binding of ligandin molecule, Praziquatel, in *Schistosoma japonicum* GST (McTigue et al., 1995), (b) binding of S-(3-iodobenxyl)glutathione in the ligandin position in the Sigma class GST (Ji et al, 1996).

Original in colour

The structural differences between the theoretical plant models and the structures of the *S. japonicum* and squid GSTs means that like the Pi class models the ligandin binding site found within these GSTs does not appear to exist in the theoretical plant models. In particular, this is due to the relative change in orientation of helices α4-α5 of the C-terminal domain between the molecules. The change in orientation of these elements leads to the upper regions of helices 4 and 5 lying further away from their dimer opposite. This prevents the formation of the cleft described for the *S. japonicum* and squid GSTs.

It is clear from the plant structures discussed in this text together with the GST structures deposited in the PDB from various organisms, that the glutathione S-transferase backbone is a stable fold that can show considerable variation. This variation can be particularly observed at the N and C terminal tails and within specific loop regions between the secondary structure elements.

Although the mammalian Alpha, Mu and Pi class GSTs have been well structurally characterised over the last decade, it is now clear from the characterisation of the Theta, Omega and Zeta classes that much remains structurally unknown for this ubiquitous enzyme. Structural and functional characterisation of a wider range of substrates for *Zm*GSTF1, *At*GSTF2 and *Zm*GSTF3 would be preferable to gain a deeper understanding of the structural basis of substrate recognition.

Although some insight into the potential substrates of a particular GST can be derived from the primary sequences and tertiary structures determined to date, Further structural characterisation together with detailed biochemical characterisation of a broad range of enzymes will be required to accurately map the determinants of substrate specificity within plant GSTs.

Additional methods will also be required to reveal the full spectrum of endogenous substrates that GSTs might affect either though GSH conjugation or by a ligandin function.

Gene disruption enabled the identification of the role of the GSTs, AN9 and *Zm*GSTU4 within the anthocyanin pathway (Alfenito et al., 1998) and further use of this technique in the future is likely.

Recently DNA microarrays have been used to examine the levels of expression of particular GST enzymes during different stages of development, under different stress conditions, and in response to different herbicide and safener treatments.

The results of these experiments together with those described above and those gained from parallel experiments with the mammalian GST classes will provide a deeper understanding of the function of the many GSTs, which exist in plants. This will aid determination of the actual endogenous substrates, which bind GSTs and enable design of GST enzymes with altered substrate specificity.

References:

Abelson, P. H. and Hines, P. J. (1999). The plant revolution. *Science* **285**, 367-368.

Adams, C. A., Blee, E. and Casida, J. E. (1983). Dichloroacetamide herbicide antidotes enhance sulfate metabolism in corn roots. *Pesticide Biochemistry and Physiology* **19**, 350-360.

Alfenito, M. R., Souer, E., Goodman, C. D., Buell, R., Mol, J., Koes, R. and Walbot, V. (1998). Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *Plant Cell* **10**, 1135-1149.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

Anandarajah, K., Kiefer, P. M., Donohoe, B. S. and Copley, S. D. (2000). Recruitment of a double bond isomerase to serve as a reductive dehalogenase during biodegradation of pentachlorophenol. *Biochemistry* **39**, 5303-5311.

Armstrong, R. N. (1993). Glutathione S-transferases - Structure and mechanism of an archetypical detoxication enzyme. *Advances in Enzymology and Related Areas of Molecular Biology* **69**, 1-44.

Barycki, J. J. and Colman, R. F. (1997). Identification of the nonsubstrate steroid binding site of rat liver glutathione S-transferase, isozyme 1-1, by the steroid affinity label, 3 beta-(iodoacetoxy)dehydroisoandrosterone. *Archives of Biochemistry and Biophysics* **345**, 16-31.

Bayley, C., Trolinder, N., Ray, C., Morgan, M., Quisenberry, J. E. and Ow, D. W. (1992). Engineering 2,4-D resistance into cotton. *Theoretical and Applied Genetics* **83**, 645-649.

Benfey, P. N., Ren, L. and Chua, N. H. (1989). The Camv S-35 enhancer contains at least 2 domains which can confer different developmental and tissue-specific expression patterns. *EMBO Journal* **8**, 2195-2202.

Berfors, T. (1999). *Protein crystallisation techniques - Stratagies and Tips*, IUL Biotechnology Series, La Jolla.

Bernal, J. D. and Crowfoot, D. (1934). X-ray photographs of crystalline pepsin. *Nature* **794**, 133-134.

Bilang, J. and Sturm, A. (1995). Cloning and characterization of a glutathione s-transferase that can be photolabeled with 5-azido-indole-3-acetic acid. *Plant Physiology* **109**, 253-260.

Blundell, T. L. and Johnson, L. H. (1976). *Protein Crystallography*, Academic Press, New York.

Bricogne, G. (1986). *Proceedings of the Cooperative Workshop on Position-Sensitive Detector Software (Phase III), Paris, LURE.*

Board, P. G., Coggan, M., Wilce, M. C. J. and Parker, M. W. (1995). Evidence for an essential serine residue in the active-site of the Theta-class glutathione transferases. *Biochemical Journal* **311**, 247-250.

Board, P. G., Coggan, M., Chelvanayagam, G., Easteal, S., Jermin, L. S., Schulte, G. K., Danley, D. E., Hoth, L. R., Griffor, M. C., Kamath, A. V., Rosner, M. H., Chrunyk, B. A., Perregaux, D. E., Gabel, C. A., Geohegan, K. F. and Pandit, J. (2000). Indentification, characterization and crystal structure of Omega class glutathione transferases. *Journal of Biological Chemistry* **275**, 24798-24806.

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

Boot, K. J. M., Vanderzaal, B. J., Velterop, J., Quint, A., Mennes, A. M., Hooykaas, P. J. J. and Libbenga, K. R. (1993). Further characterization of expression of auxin-induced genes in tobacco (Nicotiana-Tabacum) cell-suspension cultures. *Plant Physiology* **102**, 513-520.

Booth, J., Boyland, E. and Sims, P. (1961). An enzyme from rat liver catalyziing conjugation with glutathione. *Biochemical Journal* **79**, 516-424.

Brown, H. M. and Neighbours, S. M. (1987). Soybean metabolism of chlorimuron-ethyl: physiological basis for soybean selectivity. *Pesticide Biochemistry and Physiology* **29**, 112-120.

Brunger, A. T. (1990). Extension of molecular replacement - a new search strategy based on Patterson correlation refinement. *Acta Crystallographica Section A* **46**, 46-57.

Brunger, A. T. (1992). X-PLOR (Version 3.1): A system for X-ray crystallography and NMR. Yale University.

Brunger, A. T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu,N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G. (1998). Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Crystallographica* **D54**, 905-921.

Buchananwollaston, V., Snape, A. and Cannon, F. (1992). A plant selectable marker gene based on the detoxification of the herbicide Dalapon. *Plant Cell Reports* **11**, 627-631.

Buetler, T. M. and. Eaton, D. L. (1992). Glutathione s-transferases - Amino-acid-sequence comparison, classification and phylogenetic relationship. *Environmental Carcinogenesis & Ecotoxicology Reviews-Part C of Journal of Environmental Science and Health* **10**, 181-203.

Cameron, A. D., Sinning, I., Lhermite, G., Olin, B., Board, P. G., Mannervik, B. and Jones, T. A. (1995). Structural-analysis of human alpha-class glutathione transferase A1-1 in the apo-form and in complexes with ethacrynic-acid and its glutathione conjugate. *Structure* **3**, 717-727.

Campbell, J. W. (1997). *CCP4 Newsletter.* **33**, 5-16.

Carson, M. (1997). RIBBONS. *Methods in Enzymology* **277**, 493-505.

CCP4. (1994). The CCP4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* **50**, 760-763.

Chang, G. and Lewis, M. (1997). Molecular replacement using genetic algorithims. *Acta Crystallographica Section D-Biological Crystallography* **53**, 279-289.

Chelvanayagam, G., Parker, M. W. and Board, P. G. (2000). Fly fishing for GSTs: a unified nomenclature for mammalian and insect glutathione transferases. *GST 2000 International Conference of Glutathione S-Transferases, Uppsala, Sweden.*

Chen, W., Graminski, G. F. and Armstrong, R. N. (1988). Dissection of the catalytic mechanism of isoenzyme 4-4 of glutathione s-transferases with alternative substrates. *Biochemistry* **27**, 647-654.

Chen, W. Q., Chao, G. and Singh, K. B. (1996). The promoter of a H_2O_2-inducible, *Arabidopsis* glutathione s- transferase gene contains closely linked OBF- and OBP1-binding sites. *Plant Journal* **10**, 955-966.

Chen, W. Q. and Singh, K. B. (1999). The auxin, hydrogen peroxide and salicylic acid induced expression of the *Arabidopsis* GST6 promoter is mediated in part by an ocs element. *Plant Journal* **19**, 667-677.

Cobb, A. (1992). *Herbicides and Plant Physiology*, Chapman and Hall, London.

Coggan, M., Liu, D., Chelvanayagam, G., Anderson, W. G., Anders, M. W. and Board, P. G. (2000). Evaluation of possible active site residues in GSTZ 1-1. *Proceedings of GST 2000 International Conference on Glutathione Transferases, Uppsala, Sweden.*

Cohen, J. D. and Bandurski, R. S. (1982). Chemistry and physiology of the bound auxins. *Annual Review of Plant Physiology and Plant Molecular Biology* **33**, 403-430.

Cole, D. J. (1994). Detoxification and activation of agrochemicals in plants. *Pesticide Science* **42**, 209-222.

Cole, D. J., Cummins, I., Hatton, P. J., Dixon, D. and Edwards, R. (1997). Glutathione transferases in crops and major weeds. *Regulation of Enzymatic Systems Detoxifying Xenobiotics in Plants* **37**, 139-154.

Cole, D. J. Edwards, R. (2000). Secondary metabolism of agrochemicals in plants. In *Metabolism of Agrochemicals in Plants* (Roberts, T., ed.), pp. 107-154. John Wiley and Sons Ltd.

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

Comai, L., Facciotti, D., Hiatt, W. R., Thompson, G., Rose, R. E. and Stalker, D. M. (1985). Expression in plants of a mutant Aroa gene from *Salmonella- typhimurium* confers tolerance to glyphosate. *Nature* **317**, 741-744.

Combes, B. and Stakellum, G. S. (1961). A liver enzyme that conjugates sulfobromophtalein sodium with glutathione. *Journal Clin. Invest.* 40, 981-988.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881-10890.

Cottingham, C. K. and Hatzios, K. K. (1991). Influence of the safener Benoxacor on the metabolism of metolachlor in corn. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences* **46**, 846-849.

Cowtan, K. (1994). DM: An automated procedure for phase improvement by density modification. In *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, Vol. 31, pp. 34-38.

Crowther, R. A. (1972). *The molecular replacement method* (Rossmann, M. G., Ed.), Gordon and Breach, New York.

Cruikshank, D. W. J. (1999). Remarks about protein structure precision. *Acta Crystallographica Section D-Biological Crystallography* **55**, 583-601.

Cummins, I., Cole, D. J. & Edwards, R. (1997). Purification of multiple glutathione transferases involved in herbicide detoxification from wheat (*Triticum aestivum L.*) treated with the safener fenchlorazole-ethyl. *Pesticide Biochemistry and Physiology* **59**, 35-49.

Daniel, V. (1993). Glutathione s-transferases - Gene structure and regulation of expression. *Critical Reviews in Biochemistry and Molecular Biology* **28**, 173-207.

Dauter, Z., Dauter, M., de La Fortelle, E. Bricogne, G. and Sheldrick, G. M. (1999). Can anomalous signal sulphur become a tool for solving protein crystal structures? *Journal of Molecular Biology* **289**, 83-92.

Davies, J. and Caseley, J. C. (1999). Herbicide safeners: a review. *Pesticide Science* **55**, 1043-1058.

Dean, J. V., Devarenne, T. P., Lee, I. S., and Orlofsky, L. E. (1995). Properties of a maize glutathione s-transferase that conjugates coumaric acid and other phenylpropanoids. *Plant Physiol.* **108**, 985-994.

Dean, J. V. and Devarenne, T. P., (1997). Peroxidase-mediated conjugation of glutathione to unsaturated phenylpropanoids. Evidence against glutathione s-transferase involvement. *Physiol Plant.* **99**, 271-278.

Deblock, M., Botterman, J., Vandewiele, M., Dockx, J., Thoen, C., Gossele, V., Movva, N. R., Thompson, C., Vanmontagu, M. and Leemans, J. (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *EMBO Journal* **6**, 2513-2518.

Devine, M. D. (1993). Physiology of Herbicide action, Prentice Hall, Englewook Cliffs. NJ. USA.

Dirr, H., Reinemer, P. and Huber, R. (1994). X-Ray Crystal-Structures of cytosolic glutathione s-transferases - Implications for protein architecture, substrate recognition and catalytic function. *European Journal of Biochemistry* **220**, 645-661.

Dixon, D. (1998). Glutathione transferases in maize (*Zea mays*), Thesis, University of Durham.

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

Dixon, D. P., Cole, D. J. and Edwards, R. (1997b). Characterization and regulation of multiple glutathione transferases in corn (Zea mays). *Plant Physiology* **114**, 746.

Dixon, D. P., Cole, D. J. and Edwards, R. E. (2000). Characterisation of *Arabidopsis thaliana* glutathione transferase with a putative role in tyrosine catabolism. *Archives in Biochemistry and Biophysics.* **384**, 407-412.

Dixon, D. P., Cummins, I., Cole, D. J. and Edwards, R. (1998). Glutathione-mediated detoxification systems in plants. *Current Opinion in Plant Biology* **1**, 258-266.

Dodge, A. D. (1982). Oxygen radicals and herbicide action. *Biochemical Society Transactions* **10**, 73-75.

Drenth, J. (1994). *Principles of Protein X-ray Crystallography.* Springer Advanced Texts in Chemistry (Cantor, C. R., Ed.), Springer-Verlag, New York

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

Droog, F. N. J., Hooykaas, P. J. J. and Vanderzaal, B. J. (1995). 2,4-Dichlorophenoxyacetic acid and related chlorinated compounds inhibit 2 auxin-regulated type-III tobacco glutathione s-transferases. *Plant Physiology* **107**, 1139-1146.

Ducruix, A. and Giege, R. (1992). *Crystallisation of Nucleic Acids an Proteins - A practical Approach.* The Practical Approach Series (Rickwood, D. Hames, B. D., Ed.), Oxford University Press, New York.

Ebert, E. and Ramsteiner, K. (1984). Influence of metolachlor and the metolachlor protectant CGA 43089 on the biosynthesis of epicuticular waxes on the primary leaves of sorghum-(bicolor moench). *Weed Research* **24**, 383-389.

Edwards, R., Dixon, D. P. and Walbot, V. (2000). Plant glutathione s-transferases: enzymes with multiple functions in sickness and in health. *Trends in Plant Science* **5**, 193-198.

Ekler, Z., Dutka, F. and Stephenson, G. R. (1993). Safener effects on acetochlor toxicity, uptake, metabolism and glutathione s-transferase activity in maize. *Weed Research* **33**, 311-318.

Ellis, J. G., Tokuhisa, J. G., Llewellyn, D. J., Bouchez, D., Singh, K., Dennis, E. S. and Peacock, W. J. (1993). Does the ocs-element occur as a functional component of the promoters of plant genes. *Plant Journal* **4**, 433-443.

Eshdat, Y., Holland, D., Faltin, Z. and BenHayyim, G. (1997). Plant glutathione peroxidases. *Physiologia Plantarum* **100**, 234-240.

Esterbauer, H. and Cheeseman, K. H. (1987). Lipid-peroxidation - pathological implications - introduction. *Chemistry and Physics of Lipids* **45**, 103.

Evans, S. V. (1993). SETOR - Hardware-lighted 3-dimensional solid model representations of macromolecules. *Journal of Molecular Graphics* **11**, 134.

Ezra, G. and Stepherson, G. R. (1985). Comparative metabolism of atrazine and EPTC in Proso Millet (Panicum miliaceum L.) and corn. *Pesticide Biochemistry and Physiology* **24**, 207-212.

Fahey, R. C. and Sundquist, A. R. (1991). Evolution of glutathione metabolism. *Advances in Enzymology and Related Areas of Molecular Biology* **64**, 1-53.

Farago, S. and Brunold, C. (1990). Regulation of assimilatory sulfate reduction by herbicide safeners in *Zea mays* L. *Plant Physiology* **94**, 1808-1812.

Feher, G. and Kam, Z. (1985). Nucleation and growth of protein crystals: general principles and assays. *Methods in Enzymology* **114**, 77-111.

Fernández-Cañón, J. M. and Penalva, M. A. (1998). Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. *Journal of Biological Chemistry* **273**, 329-337.

Fournier, J. M., Bride, M., Poire, J. B., Berge, F. W. and Plapp, Jr. (1992). Insect glutathione s-transferases. *Journal of Biological Chemistry* **267**, 1840-1845.

Frear, D. S. and Swanson, H. R. (1970). Biosynthesis of S-(4-ethylamino-6-isopropylamino-2-s-triazino)glutathione: partial purification and properties of glutathione s-transferase from corn. *Phytochemistry* **9**, 2123-2132.

Frear, D. S. and Swanson, H. R. (1973). Metabolism of substituted diphenylether herbicides in plants. I. Enzymatic cleavage of flurodifen in peas (Pisum sativum L.). *Pesticide Biochemistry and Physiology* **3**, 473-482.

Frear, D. S. S., Swanson, H. R. and Thalacker, F. W. (1991). Induced microsomal oxidation of diclofop, triasulfuron, chlorsulfuron, and linuron in wheat. *Pesticide Biochemistry and Physiology* **41**, 274-287.

Friling, R. S., Bergelson, S. and Daniel, V. (1992). 2 Adjacent Ap-1-like binding-sites form the electrophile-responsive element of the murine glutathione s-transferase Ya subunit gene. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 668-672.

Fromm, H., Katagiri, F. and Chua, N. H. (1989). An Octopine synthase enhancer element directs tissue-specific expression and binds Asf-1, a factor from tobacco nuclear extracts. *Plant Cell* **1**, 977-984.

Fuerst, E. P. and Gronwald, J. W. (1986). Induction of rapid metabolism of metolachlor in sorghum (Sorghum- Bicolor) shoots by CGA-92194 and other antidotes. *Weed Science* **34**, 354-361.

Garcia-saez, I., Parraga, A., Phillips, M.F., Mantle, T. J. and Coll, M. (1994). Molecular structure at 1.8 Angstroms of mouse liver class Pi glutathione s-transferase complexed with S-(P-Nitrobenzyl) glutathione and other inhibitors. *Journal of Molecular Biology* **237**, 298-314.

Gorog, K., Muschinek, G., Mustardy, L. A. and Faludidaniel, A. (1982). Comparative studies of safeners for the prevention of EPTC injury in maize. *Weed Research* **22**, 27-33.

Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.

Graminski, G. F., Kubo, Y. and Armstrong, R. N. (1989a). Spectroscopic and kinetic evidence for the thiolate anion of glutathione at the active site of glutathione s-transferase. *Biochemistry* **28**, 3562-3568.

Graminski, G. F., Zhang, P., Sesay, M. A. Ammon, H. L. and Armstrong, R. N. (1989b). Formation of the 1-(S-glutathionyl)-2,4,6-trinitrocyclohexadienate anion at the active site of glutathione s-transferase. *Biochemistry* **28**, 6252-6258.

Gressel, J. (1985). Plant-cell cultures for herbicide action, screening and obtaining resistant crops. *Abstracts of Papers of the American Chemical Society* **190**, 13-AGO.

Gronwald, J. W., Andersen, R. N. and Yee, C. (1989). Atrazine resistance in velvetleaf (Abutilon-Theophrasti) due to enhanced atrazine detoxification. *Pesticide Biochemistry and Physiology* **34**, 149-163.

Gronwald, J. W., Fuerst, E. P., Eberlein, C. V. and Egli, M. A. (1987). Effect of herbicide antidotes on glutathione content and glutathione s-transferase activity of sorghum shoots. *Pesticide Biochemistry and Physiology* **29**, 66-76.

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

Habig, W. H. (1983). Glutathione s-transferases: versatile enzymes of detoxification. In *Radioprotectors and Anticarcinogens,* (Nygaard, O. F. and Simic, M. G., ed.), pp. 169-190. Academic Press, San Diego.

Habig, W. H. Jakoby, W. B. (1981). Assays for differentiation of glutathione s-transferases. *Methods in Enzymology* **77**, 398-405.

Harper, J. L. (1956). The evolution of weeds in relation to resistance in herbicides. *Proceedings. British. Weed. Control. Conf.* **3**, 179.

Harris, J. M., Meyer, D. J., Coles, B. and Ketterer, B. (1991). A novel glutathione transferase (13-13) isolated from the matrix of rat-liver mitochondria having structural similarity to class Theta-enzymes. *Biochemical Journal* **278**, 137-141.

Hayes, J. D. and Chalmers, J. (1983). Bile acid inhibition of basic and neutral glutathione s-transferases in rat liver. *Biochemical Journal* **215**, 581-588.

Hayes, J. D. and Mantle, T. J. (1986). Inhibition of hepatic and extrahepatic glutathione s-transferases by primary and secondary bile acids. *Biochemical Journal* **233**, 407-415.

Hayes, J. D. and McLellan, L. I. (1999). Glutathione and glutathione-dependent enzymes represent a Co ordinately regulated defence against oxidative stress. *Free Radical Research* **31**, 273-300.

Hayes, J. D. and Pulford, D. J. (1995). The glutathione s-transferase supergene family: Regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Critical Reviews in Biochemistry and Molecular Biology* **30**, 445-600.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.

Hendrickson, W. A. and Ogata., C. M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymology* 276, 494-523.

Hershey, H. P. and Stoner., T. D. (1991). Isolation and characterisation of cDNA clones for RNA species induced by substituted benzenesulfonamides in corn. *Plant Molecular Biology* **17**, 679-690.

Hiratsuka, A., Yokoi, A., Sebata, N., Watabe, T., Satoh, K., Hatayama, I. & Sato, K. (1989). Glutathione conjugation of the flurophotometric epoxide substrate7-glycidoxycoumarin (GOC), by rat liver glutathione transferase isoenzymes. *Biochemical Pharmacology* **38**, 2609-2613.

Hoagland, R. E. (1989). Biochemical responses of plants to pathogen attack. *Abstracts of Papers of the American Chemical Society* **197**, 120-AGRO.

Hoesch, R. M. and Boyer, T. D. (1989). Localization of a portion of the active-site of 2 rat-liver glutathione s-transferases using a photoaffinity label. *Journal of Biological Chemistry* **264**, 17712-17717.

Hoffman, O. L. (1953). Inhibition of auxin effects by 2,3,6 trichlorophenoxyacetic acid. *Plant Physiology* **28**, 622-628.

Hoffman, O. L. (1969). Chemical antidotes for EPTC on corn. *Abstracts of Weed Science Society America* **9**, 12.

Holt, D. C., Lay, V. J., Clarke, E. D., Dinsmore, A., Jepson, I., Bright, S. W. J. and Greenland, A. J. (1995). Characterization of the safener-induced glutathione s-transferase isoform-II from maize. *Planta* **196**, 295-302.

Huckle, K. R., Tait, G. H., Millburn, P. and Hutson, D. H. (1981). Species variations in the renal and hepatic conjugation of 3- phenoxybenzoic acid with glycine. *Xenobiotica* **11**, 635-644.

Huskey, S. E. W., Huskey, W. P. and Lu, A. Y. H. (1991). Contributions of thiolate "desolvation " to catalysis by glutathione s-transferase isoenzymes1-1 and 2-2: evidence from kinetic solvent isotropic effects. *Journal Am. Chem. Soc.* **113**, 2283-2290.

Hutchison, J. M., Shapiro, R. and Sweetser, P. B. (1984). Metabolism of chlorsulfuron by tolerant broadleaves. *Pesticide Biochemistry and Physiology* **22**, 243-247.

Irzyk, G. P. and Fuerst, E. P. (1993). Purification and characterization of a glutathione s-transferase from benoxacor-treated maize (*Zea Mays*). *Plant Physiology* **102**, 803-810.

Ishigaki, S., Abramovitz, M. and Listowsky, I. (1989). Glutathione s-transferases are major cytosolic thyroid binding proteins. *Archives of Biochemistry and Biophysics* **273**, 265-272.

Jakoby, W. B. (1978). The glutathione s-transferases: a group of multifunctional detoxification proteins. *Advances in Enzymology and Related Areas of Molecular Biology* **46**, 383-414.

Jakoby, W. B., Ketterer, B. and Mannervik, B. (1984). Glutathione transferases - nomenclature. *Biochemical Pharmacology* **33**, 2539-2540.

Jancarik, J. and Kim, S. H. (1991). Sparse-matrix sampling - a screening method for crystallization of proteins. *Journal of Applied Crystallography* **24**, 409-411.

Ji, X., von Rosenvinge, E. C., Johnson, W. W. Armstrong, R. N. and Gilliland, G. L. (1996). Location of a potential transport binding site in a Sigma class glutathione transferase by X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8208-8213.

Ji, X. H., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N. and Gilliland, G. L. (1995). 3-dimensional structure, catalytic properties, and evolution of a Sigma-class glutathione transferase from squid, a progenitor of the lens s-crystallins of Cephalopods. *Biochemistry* **34**, 5317-5328.

Ji, X. H., Zhang, P. H., Armstrong, R. N. and Gilliland, G. L. (1992). The 3-Dimensional structure of a glutathione s-transferase from the Mu-gene class - structural-analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-Angstrom resolution. *Biochemistry* **31**, 10169-10184.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **26**, 795-800.

Kampranis, S. C., Damianova, R., Atallah, M., Toby, G., Kondi, G., Tsichlis, P. N. and Makris, A. M. (2000). A novel plant glutathione s-transferase/peroxidase suppresses Bax lethality in yeast. *Journal of Biological Chemistry* **275**, 29207-29216.

Kelley, L. A., MacCallum, R. M. and Sternberg M. J. E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *Journal of Molecular Biology* **299**, 501-522.

Ketley, J. N., Habig, W. H. and Jakoby, W. B. (1975). Binding of non-substrate ligands to the glutathione s-transferases. *Journal of Biological Chemistry* **250**, 8670-8673.

Ketterer, B., Meyer, D. J. and Clark, A. G. (1988). *Soluble glutathione transferase isoenzymes*. Glutathione Conjugation (Sies, H. and Ketterer, B., Ed.), Academic Press, London.

Ketterer, B., Ross-Mansell, P. and Whitehead, J. K. (1967). The isolation of carcinogen-binding protein form livers of rats given 4-methylaminoazobenzene. *Biochemical Journal* **103**, 316-324.

Kirsch, R., Fleischner, G., Kamisaka, K. and Arias, I. M. (1975). Structural and functional studies of ligandin, a major renal organic anion-binding protein. *Journal Clin. Invest.* **55**, 1009-1019.

Kissinger, C. R., Gehlhaar, D. K. Fogel, D. B. (1999). Rapid automated molecular replacement by evolutionary search. *Acta Crystallographica Section D-Biological Crystallography* **D55**, 484-491.

Kosower, E. M. (1976). *Chemical properties of glutathione.* Glutathione: metabolism and function (Jakoby, W. B. and Irwin. M. A. Ed.), Raven Press, New York.

Knox, W. E. (1955). Enzymes involved in conversion of tyrosine to acetoacetate. *Methods in Enzymology* **2**, 287-300.

Knox, W. E. and Edwards, S. (1956). Homogentisate metabolism: The isomerisation of maleylacetoacetate by an enzyme which requires glutathione. *Journal of Biological Chemistry* **220**, 79-91.

Kreuz, K., Tommasini, R. and Martinoia, E. (1996). Old enzymes for a new job - herbicide detoxification in plants. *Plant Physiology* **111**, 349-353.

Lacomme, C. and Cruz, S. S. (1999). Bax-induced cell death in tobacco is similar to the hypersensitive response. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 7956-7961.

Lamb, C. D., R. A. (1997). The oxidative burst in plant disease resistance. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**, 251-275.

Lamzin, V. S. and Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallographica Section D-Biological Crystallography* **49**, 129-147.

Lamoureux, G. L. and Rusness, D. G. (1986). Tridiphane [2-(3,5-dichlorophenyl)-2-(2,2,2-trichloroethyl)oxirane] an atrazine synergist - Enzymatic conversion to a potent glutathione s-transferase inhibitor. *Pesticide Biochemistry and Physiology* **26**, 323-342.

Lamoureux, G. L. and Rusness, D. G. (1992). The mechanism of action of BAS-145-138 as a safener for chlorimuron ethyl in corn - Effect on hydroxylation, glutathione conjugation, glucoside conjugation, and acetolactate synthase. *Pesticide Biochemistry and Physiology* **42**, 128-139.

Lamoureux, G. L., Rusness, D. G., Schroder, P. and Rennenberg, H. (1991). Diphenyl ether herbicide metabolism in a spruce cell-suspension culture - the identification of 2 novel metabolites derived from a glutathione conjugate. *Pesticide Biochemistry and Physiology* **39**, 291-301.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. and Thornton, J. M. (1993). Procheck - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291.

Lay, M. M. and Casida, J. E. (1976). Dichloroacetamide antidotes enhance thiocarbamate sulfoxide detoxification by elevating corn root glutathione content and glutathione s-transferase activity. *Pesticide Biochemistry and Physiology* **6**, 442-456.

LeBarron, H. M. and Gressel, J. (1982). *Herbicide resistance in Plants*, John Wiley and sons, New York.

Lee, K. Y., Townsend, J., Tepperman, J., Black, M., Chui, C. F., Mazur, B., Dunsmuir, P. and Bedbrook, J. (1988). The molecular-basis of sulfonylurea herbicide resistance in tobacco. *EMBO Journal* **7**, 1241-1248.

Lee, Y. T., Chang, A. K. and Duggleby, R. G. (1999). Effect of mutagenesis at serine 653 of Arabidopsis thaliana acetohydroxyacid synthase on the sensitivity to imidazolinone and sulfonylurea herbicides. *FEBS Letters* **452**, 341-345.

Leslie A. G. W. (1992). Mosfilm, *Joint CCP4 and ESRF-EACMD newsletter on protein crystallography*, Daresbury Laboratory, UK

Lim, K., Ho, J. X., Keeling, K., Gilliland, G. L., Ji, X. H., Ruker, F. and Carter, D. C. (1994). 3-Dimensional structure of *Schistosoma-japonicum* glutathione s- transferase fused with a 6-amino acid conserved neutralizing epitope of Gp41 from HIV. *Protein Science* **3**, 2233-2244.

Listowski, I., Abramovitz, M. Homma, H. and Niitsu, Y. (1988). Intracellular binding and transport of hormones and xenobiotics by glutathione s-transferases. *Drug Metabolism Review* **19**, 305-318.

Lo Bello, M. P., A., Petruzelli, R., Parker, M., Wilce, M., Federici, G. and Ricci, G. (1993). Conformational states of human placental glutathione transferase as probed by limited proteolysis. *Biochemical and Biophysical Research Communications* **194**, 804-810.

Lu, Y. P., Li, Z. S., Drozdowicz, Y. M., Hortensteiner, S., Martinoia, E. and Rea, P. A. (1998). AtMRP2, an *Arabidopsis* ATP binding cassette transporter able to transport glutathione s-conjugates and chlorophyll catabolites: Functional comparisons with AtMRP1. *Plant Cell* **10**, 267-282.

Lu, Y. P., Li, Z. S. and Rea, P. A. (1997). ABC transporter (AtMRP1) gene of *Arabidopsis* encodes a glutathione conjugate pump. *Plant Physiology* **114**, 49.

Lund, O., Frimand, K., Gorodkin, J., Bohr, H., Bohr, J., HansenJ. and Brunak.S. (1997). Protein distance constraints predicted by neural networks and probability density functions. *Protein Engineering* **10**, 1241-1248.

Malthus, T. (1798). An essay on the principle of population. In *An Essay on the principle of population, its affects the future improvement of society with remards on the speculation of Mr. Godwin, M. Condorcet, and other writers.* Printed for J. Johnson, in St. Paul's Church-yard, 1798, London. http://marxists.org/reference/subject/economics/malthus/

Mann, C. C. (1999). Crop scientists seek a new revolution. *Science* **283**, 310-316.

Mannervik, B. and Danielson, U. H. (1988). Glutathione transferases - structure and catalytic activity. *Critical Reviews in Biochemistry and Molecular Biology* **23**, 283-337.

Mannervik, B., Alin, P., Guthenberg, C., Jensson, H., Tahir, M. K., Warholm, M. and Jornvall, H. (1985). Identification of 3 classes of cytosolic glutathione transferase common to several mammalian-species - correlation between structural data and enzymatic-properties. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 7202-7206.

Mannervik, B. Gluthenberg, C. (1981). Glutathione Transferase (human placenta). In *Methods in Enzymology* (Jakoby, W. B., ed.), 77, 231-235. Academic Press Inc., London.

Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology* **33**, 491-497.

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

Marrs, K. A. (1996). The functions and regulation of glutathione s-transferases in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**, 127-158.

Martinola, E., Grill, E., Tommasini, R., Kreuz, K. and Amrhein, N. (1993). ATP-dependent glutathione s-conjugate transport 'export' pump in the vacuolar membrane of plants. *Nature* **346**, 247-249.

McCarthy, D. L., Louie, D. F. and Copley, S. D. (1997). Identification of a covalent intermediate between glutathione and cysteine13 formed during catalysis by tetrachlorohydroquinone dehalogenase. *Journal of the American Chemical Society* **119**, 11337-11338.

McCarthy, D. L., Navarrete, S., Willett, W. S., Babbitt, P. C. and Copley, S. D. (1996). Exploration of the relationship between tetrachlorohydroquinone dehalogenase and the glutathione s-transferase superfamily. *Biochemistry* **35**, 14634-14642.

McRee, D. E. (1999). *Practical Protein Crystallography*, Academic Press, London.

McPherson, A. (1990). Current approaches to macromolecular crystallisation. *Eur J. Biochem.* **189**, 1-23.

McTigue, M. A., Williams, D. R. and Tainer, J. A. (1995). Crystal-structures of a schistosomal drug and vaccine target - glutathione s-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug Praziquantel. *Journal of Molecular Biology* **246**, 21-27.

Meyer, D. J., Coles, B., Pemble, S. E., Gilmore, K. S., Fraser, G. M. and Ketterer, B. (1991a). Theta, a new class of glutathione transferases purified from rat and man. *Biochemical Journal* **274**, 409-414.

Meyer, R. C., Goldsbrough, P. B. and Woodson, W. R. (1991b). An ethylene-responsive flower senescence-related gene from carnation encodes a protein homologous to glutathione s-transferases. *Plant Molecular Biology* **17**, 277-281.

Meyer, D. J. (1993). Significance of an unusually low K_m for glutathione in glutathione transferases of the Alpha, Mu and Pi classes. *Xenobiotica* **23**, 823-834.

Milcamps, A. and deBruijn, F. J. (1999). Identification of a novel nutrient-deprivation-induced *Sinorhizobium meliloti* gene (hmgA) involved in the degradation of tyrosine. *Microbiology-UK*, **145**, 935-947.

Mine, A., Miyakado, M. and Matsunaka, S. (1975). The mechanism of bentazon selectivity. *Pesticide Biochemistry and Physiology* **5**, 566-574.

Minor, W., Tomchick, D. R. and Otwinowski, Z. (2000). Strategies for macromolecular synchrotron crystallography. *Structure with Folding and Design* **8**, R105-R110.

Moreland, D. E., Corbin, F. T. and McFarland, J. E. (1993). Effects of safeners on the oxidation of multiple substrates by grain- sorghum microsomes. *Pesticide Biochemistry and Physiology* **45**, 43-53.

Morgenstern, R. and Depierre, J. W. (1983). Microsomal glutathione transferase - purification in unactivated form and further characterisation of the activation process, substrate - specificity and amino-acid composition. *European Journal of Biochemistry* **134**, 591-597.

Mozer, J. J., Tiemeier, D. C. and Jaworski, E. G. (1983). Purification and characterisation of corn glutathione s-transferases. *Biochemistry* **22**, 1068-1072.

Mueller, L. A., Goodman, C. D., Silady, R. A. Walbot, V. (2000). AN9, a Petunia glutathione s-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiology* **123**, 1561-1570.

Murata, T., Hatayama, I., Satoh, K., Tsuchida, S. and Sato, K. (1990). Activation of rat glutathione transferases in class mu by active oxygen species,. *Biochemical and Biophysical Research Communications* **171**, 845-851.

Murshudov, G. N., Dodson, E. J. and Vagin, A. A. (1996). Application of maximum likelihood methods for macromolecular refinement. In *Macromolecular Refinement* (Dodson, E., Moore, M. and Bailey, S., ed.), pp. 93-104. Daresbury Laboratory, Warrington, UK.

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

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. and Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallographica Section D-Biological Crystallography* **55**, 247-255.

Navaza, J. (1994). AMoRe - an automated package for molecular replacement. *Acta Crystallographica Section A* **50**, 157-163.

Nebert, D. W., Nelson, D. R., Coon, M. J., Estabrook, R. W., Feyereisen, R., Fujiikuriyama, Y., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Loper, J. C., Sato, R., Waterman, M. R. and Waxman, D. J. (1991). The P450 Superfamily - Update on new sequences, gene-mapping, and recommended nomenclature. *DNA and Cell Biology* **10**, 1-14.

Neuefeind, T., Huber, R., Dasenbrock, H., Prade, L. and Bieseler, B. (1997a). Crystal structure of herbicide-detoxifying maize glutathione s-transferase-I in complex with lactoylglutathione: evidence for an induced-fit mechanism. *Journal of Molecular Biology* **274**, 446-453.

Neuefeind, T., Huber, R., Reinemer, P., Knablein, J., Prade, L., Mann, K. and Bieseler, B. (1997b). 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. *Journal of Molecular Biology* **274**, 577-587.

Neuefeind, T., Reinemer, P. and Bieseler, B. (1997c). Plant glutathione s-transferases and herbicide detoxification. *Biological Chemistry* **378**, 199-205.

Nicholson, D. W., Ali, A. Vaillancourt. J. P. Calaycay, J. R., Mumford, R. A., Zamboni, R. J. and Ford-Hutchinson, A. W. (1993). Purification to homogeneity and the N-terminal sequence of human leukotriene C_4 synthase: A homodimeric glutathione s-transferase composed of 18-kDa subunits. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2015-2019.

Nishida, M., Harada, S., Noguchi, S., Satow, Y., Inoue, H. and Takahashi, K. (1998). Three-dimensional structure of *Escherichia coli.* glutathione s- transferase complexed with glutathione sulfonate: Catalytic roles of Cys10 and His106. *Journal of Molecular Biology* **281**, 135-147.

Oakley, A. J., Bello, M. L., Battistoni, A., Ricci, G., Rossjohn, J. Villar, H. O. and Parker, M. W. (1997). The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution. *Journal of Molecular Biology* **274**, 84-100.

Oakley, A. J., Bello, M. L., Ricci, G., Federici, G. and Parker, M. W. (1998). Evidence for induced-fit mechanism operating in Pi class glutathione transferases. *Biochemistry* **37**, 9912-9917.

Oakley, A. J., Bello, M. L., Nuccetelli, M., Mazzetti, A. P. and Parker, M. W. (1999). The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site). *Journal of Molecular Biology* **291**, 913-926.

Okuda, A., Sakai, M. & Muramatsu, M. (1987). The structure of the rat glutathione s-transferase Pi gene and related pseudogenes. *Journal of Biological Chemistry* **262**, 3858-3863.

Orser, C. S., Dutton, J., Lange, C., Jablonski, P., Xun, L. Y. and Hargis, M. (1993). Characterization of a flavobacterium glutathione s-transferase gene involved in reductive dechlorination. *Journal of Bacteriol.* **175**, 2640-2644.

Otwinowski, Z. (1991). *Maximum likelihood refinement of heavy atom parameters.* In data collection and processing (Sawyer, L., Isaacs, N. and Bailey, S, Ed.), SERC, Daresbury Laboratory, Warrington, UK.

Otwinowski, Z. and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography, Pt A*, Vol. 276, pp. 307-326.

Oxtoby, E. and Hughes, M. A. (1990). Engineering herbicide tolerance into crops. *Trends in Biotechnology* **8**, 61-65.

Parry, K. P. (1989). *Herbicide use and invention.* Herbicides and plant metabolism (Dodge, A. D., Ed.), Cambridge University Press, Cambridge.

Pemble, S. E. and Taylor, J. B. (1992). An evolutionary perspective on glutathione transferases inferred from class-Theta glutathione transferase CDNA sequences. *Biochemical Journal* **287**, 957-963.

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

Peitsch, M. C. (1995). Protein modeling by E-mail. *Bio/Technology* **13**, 658-660.

Peitsch, M. C. (1996). ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochemical Society Transactions* **24**, 274-279.

Pickett, C. B. and Lu, A. Y. H. (1989). Glutathione s-transferases - gene structure, regulation, and biological function. *Annual Review of Biochemistry* **58**, 743-764.

Pinstrup-Andersen (1994). World food trends and future food security. *IFPRI publications: Food Policy Statement* **18**, http://www.cgiar.org/ifpri/pubs/fps/fps18.htm.

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

Reinemer, P., Dirr, H. W., Ladenstein, R., Huber, R., Lobello, M., Federici, G. and Parker, M. W. (1992). 3-Dimensional structure of class-Pi glutathione s-transferase from human placenta in complex with s-hexyl glutathione at 2.8 Angstrom resolution. *Journal of Molecular Biology* **227**, 214-226.

Reinemer, P., Dirr, H. W., Ladenstein, R., Schaffer, J., Gallay, O. and Huber, R. (1991). The 3-dimensional structure of class-Pi glutathione s-transferase in complex with glutathione sulfonate at 2.3 Å Resolution. *EMBO Journal* **10**, 1997-2005.

Reinemer, P., Prade, L., Hof, P., Neuefeind, T., Huber, R., Zettl, R., Palme, K., Schell, J., Koelln, I., Bartunik, H. D. and Bieseler, B. (1996). Three-dimensional structure of glutathione s-transferase from *Arabidopsis thaliana* at 2.2 Angstrom resolution: Structural characterization of herbicide-conjugating plant glutathione s-transferases and a novel active site architecture. *Journal of Molecular Biology* **255**, 289-309.

Riechers, D. E., Fuerst, E. P. and Boerboom C. M. (1994). Optimization and mechanism of action of seed-applied herbicide safeners in winter wheat. *Abstracts of Weed Science Society America* **34**, 62.

Riechers, D. E., Irzyk, G. P., Jones, S. S. and Fuerst, E. P. (1997). Partial characterization of glutathione S-transferases from wheat (*Triticum spp.*) and purification of a safener-induced glutathione s-transferase from *Triticum tauschii*. *Plant Physiology* **114**, 1461-1470.

Riechers, D. E., Yang, K., Irzyk, G. P., Jones, S. S. and Fuerst, E. P. (1996). Variability of glutathione s-transferase levels and dimethenamid tolerance in safener-treated wheat and wheat relatives. *Pesticide Biochemistry and Physiology* **56**, 88-101.

Rosegrant, M. W., Agcaoili-Sombilla, M. and Perez (1995). Global food projections to 2020: Implications for Investment. *International Good Policy Research Institute, N. P. Food, Agriculture, and the Environment.* Discussion paper 5.

Rosegrant, M. W., Agcaoili-Sombilla, M. A., Gerpacio, R. V. and Ringler, C. (1997). Meeting the demand for food in the twenty-first century: challenges and opportunities for Illinois agriculture. *Illinois World Food and Sustainable Agriculture Program Conference.*

Rossini, L., Jepson, I., Greenland, A. J. and Gorla, M. S. (1996). Characterization of glutathione s-transferase isoforms in three maize inbred lines exhibiting differential sensitivity to alachlor. *Plant Physiology* **112**, 1595-1600.

Rossjohn, J., Board, P. G., Parker, M. W. and Wilce, M. C. J. (1996). A structurally derived consensus pattern for Theta class glutathione transferases. *Protein Engineering* **9**, 327-332.

Rossjohn, J., McKinstry, W. J., Oakley, A. J., Verger, D., Flanagan, J., Chelvanayagam, G., Tan, K. L., Board, P. G. and Parker, M. W. (1998). Human Theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site. *Structure* **6**, 309-322.

Rossmann, M. G. and vanBleek, C. G. (1999). Data Processing. *Acta Crystallographica Section D-Biological Crystallography* **55**, 1631-1641.

Rossmann, M. G. and Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* **15**, 24-31.

Rost, B., Sander, C. and Schneider, R. (1994). PHD-an automatic server for protein secondary structure prediction. *CABIOS* **10**, 53-60.

Roxas, V. P., Smith, R. K., Allen, E. R. and Allen, R. D. (1997a). Overexpression of glutathione s-transferase glutathione peroxidase enhances the growth of transgenic tobacco seedlings during stress. *Nature Biotechnology* **15**, 988-991.

Roxas, V. P., Wang, J., Lodhi, S. and Allen, R. D. (1997b). Engineering stress tolerance in transgenic plants. *Acta Physiologiae Plantarum* **19**, 591-594.

Ryan, G. F. (1970). Resistance of common groudsel to simizine and atrazine. *Weed Science* **18**, 614.

Sabeh, F., Wright, T. and Norton, S. J. (1993). Purification and characterization of a glutathione-peroxidase from the Aloe Vera plant. *Enzyme and Protein* **47**, 92-98.

Sambrook, J., Fritsch, E. F. and Maniatis. T. (1989). Polyacrylamide gel electrophoresis. In *Molecular Cloning a Laboratory Manual* (Nolan, C., ed.), Vol. 3, pp. 18.47-18.59. Cold Spring Harbor Laboratory Press.

SanchezFernandez, R., ArdilesDiaz, W., VanMontagu, M., Inze, D. and May, M. J. (1998). Cloning and expression analyses of AtMRP4 a novel MRP-like gene from *Arabidopsis thaliana. Molecular and General Genetics* **258**, 655-662.

Sandermann, H. (1992a). Plant-metabolism of xenobiotics. *Trends in Biochemical Sciences* **17**, 82-84.

Sandermann, H. (1992b). Plant-metabolism of xenobiotics - the green liver concept. *Abstracts of Papers of the American Chemical Society* **204**, 62-AGRO.

Sandermann, H. (1994). Higher-plant metabolism of xenobiotics - the green liver concept. *Pharmacogenetics* **4**, 225-241.

Sato, K., Fukuba, Y., Mitsuda, T., Hirai, K. and Moriya, K. (1992). Observation of lattice defects in orthorhombic hen-egg white lysozyme crystals with laser scattering tomography. *Journal of Crystal Growth* **122**, 87-94.

Schmitt, R. and Sandermann, H. (1982). Specific localisation of β-D- glucoside conjugates of 2,4 dichlorophenoxyacetic acid in soybean vacuoles. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences* **37c**, 772-777.

Schmidt-Krey, I., Murata, K., Hiarai, T., Mitsuoka, K., Cheng, Y., Morgenstern, R., Fujiyoshi, Y. and Hebert, H. (1999). The projection structure of the membrane protein microsomal glutathione transferase at 3 Angstrom resolution as determined from two-dimensional hexagonal crystals. *Journal of Molecular Biology* **288**, 243-253.

Sheldrick, G. M. (1990). Phase annealing in Shelx-90 - Direct methods for larger structures. *Acta Crystallographica Section A* **46**, 467-473.

Sheriff, S., Klei, H. E. and Davis, M. E. (1999). Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement. *Journal of Applied Crystallography* **32**, 98-101.

Shimabukuro, R. H., Frear, D. S. and Swanson, J. R. and Walsh, W. C. (1971). Glutathione conjugation: an enzymatic basis for atrazine resistance in corn. *Plant Physiology* **47**, 10-14.

Shimabukuro, R. H., Swanson, J. R. and Walsh, W. C. (1970). Glutathione conjugation: atrazine detoxification mechanisms in corn. *Plant Physiology* **46**, 103.

Singh, A. P. and Brutlag, D. L. (1997). Hierarchical protein superimposition using both secondary structure and atomic representations. *Proceedings of Fifth International conference on Intelligent Systems for Molecular Biology.*

Singh, B. R. and Shaw, R. W. (1988). Selective-inhibition of oat glutathione s-transferase activity by tetrapyrroles. *FEBS Letters* **234**, 379-382.

Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X. H., Board, P. G., Olin, B., Mannervik, B. and Jones, T. A. (1993). Structure determination and refinement of human-Alpha class glutathione transferase-A1-1, and a comparison with the Mu-class and Pi-class enzymes. *Journal of Molecular Biology* **232**, 192-212.

Snyder, M. J. and Maddison, D. R. (1997). Molecular phylogeny of glutathione s-transferases. *DNA and Cell Biology* **16**, 1373-1384.

Stalker, D. M., McBride, K. E. and Malyj, L. D. (1988). Herbicide resistance in transgenic plants expressing a bacterial detoxification gene. *Science* **242**, 419-423.

Stenberg, G., Abdallah, A and Mannervik, B. (2000). The dimeric state of glutathione s-transferases; role of a key residue at the subunit interface in human GSTP1-1. *GST 2000, Uppsala, Sweden.*

Stemmer, W. P. C. (1994). DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 10747-10751.

Sturrock, S. S., Dryden, D.T.F. (1997). A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Research* **25**, 3408-3414.

Subramaniam, K., Ye, Z., Buechley, G., Shaner, G., Solomos, T. and Ueng, P. P. (1999). Isolation of a Zeta class wheat glutathione s-transferase gene. *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1447**, 348-356.

Sussman, J. L., Lin, D. W., Jiang, J. S., Manning, N. O., Prilusky, J., Ritter, O. and Abola, E. E. (1998). Protein Data Bank (PDB): Database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica Section D-Biological Crystallography* **54**, 1078-1084.

Sweetser, P. B., Schow, G. S. and Hutchison, J. M. (1982). Metabolism of chlorsulfuron by plants - biological basis for selectivity of a new herbicide for cereals. *Pesticide Biochemistry and Physiology* **17**, 18-23.

Takahashi, Y., Sakai, T., Ishida, S. and Nagata, T. (1995). Identification of auxin-responsive elements of Parb and their expression in apices of shoot and root. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6359-6363.

Talalay, P., DeLong, M. J. and Prochaska, H. J. (1988). Identification of a common chemical signal regulating the induction of enzymes that protect against chemical carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 8261-8265.

Tecle, B., Dacunha, A. and Shaner, D. L. (1993). Differential routes of metabolism of imidazolinones - basis for soybean (Glycine-Max) selectivity. *Pesticide Biochemistry and Physiology* **46**, 120-130.

Timmerman, K. P. (1989). Molecular characterization of corn glutathione s-transferase isoenzymes involved in herbicide detoxification. *Physiologia Plantarum* **77**, 465-471.

Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R. and Jansoniu, J. N. (1981). Repeated seeding technique for growing large single crystals of proteins. *Journal of Molecular Biology* **147**, 465-469.

Tomarev, S. I. and Zinovieva, R. D. (1988). Squid major lens polypeptides are homologous to glutathione s-transferases subunits. *Nature* **336**, 86-88.

Tomarev, S. I., Zinovieva, R. D., Chang, S. & Piatigorsky, J. (1993). S-crystallins and glutathione s-transferase of Cephalopods - evolution of lens specificity. *Investigative Ophthalmology & Visual Science* **34**, 759.

Tong, Z., Board, P. G. and Anders, M. W. (1998a). Glutathione transferase Zeta catalyses the oxygenation of the carcinogen dichloroacetic acid to glyoxylic acid. *Biochemical Journal* **331**, 371-374.

Tong, Z., Board, P. G. and Anders, M. W. (1998b). Glutathione transferase Zeta-catalyzed biotransformation of dichloroacetic acid and other alpha-haloacids. *Chemical Research in Toxicology* **11**, 1332-1338.

Ulmasov, T., Hagen, G. and Guilfoyle, T. (1994). The Ocs element in the soybean Gh2/4 promoter is activated by both active and inactive auxin and salicylic-acid analogs. *Plant Molecular Biology* **26**, 1055-1064.

Ulmasov, T., Ohmiya, A., Hagen, G. and Guilfoyle, T. (1995). The soybean Gh2/4g gene that encodes a glutathione s-transferase has a promoter that is activated by a wide-range of chemical-agents. *Plant Physiology* **108**, 919-927.

United Nations. (1999). World population prospects: the 1998 Revision. http://www.un.org/esa/population/wpp2000h.pdf

VanderZaal, B. J., Droog, F. N. J., Pieterse, F. J. and Hooykaas, P. J. J. (1996). Auxin-sensitive elements from promoters of tobacco GST genes and a consensus as-1-like element differ only in relative strength. *Plant Physiology* **110**, 79-88.

Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995). Ligplot - a program to generate schematic diagrams of protein ligand interactions. *Protein Engineering* **8**, 127-134.

Walton, J. D. and Casida, J. E. (1995). Specific binding of a dichloroacetamide herbicide safener in maize at a site that also binds thiocarbamate and chloroacetanilide herbicides. *Plant Physiology* **109**, 213-219.

Wiegand, R. C., Shah, D. M., Mozer, T. J., Harding, E. I., Diaz-Collier, J., Saunders, C., Jaworski, E. G. and Tiemeier, D. C. (1986). Messenger RNA encoding a glutathione s-transferase responsible for herbicide tolerance in maize is induced in response to herbicide treatment. *Plant Molecular Biology* **7**, 235-243.

Wilce, M. C. J., Board, P. G., Feil, S. C. and Parker, M. W. (1995). Crystal-structure of a Theta-class glutathione transferase. *EMBO Journal* **14**, 2133-2143.

Wilce, M. C. J. and Parker, M. W. (1994). Structure and function of glutathione s-transferases. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1205**, 1-18.

Wilson, K. S., Davies, G., Ashton, A. W. and Bailey, S. (1997). Recent advances in phasing. *Proceedings of the CCP4 study weekend, Daresbury.*

Young, B. G. and Hart, S. E. (1997). Control of volunteer sethoxdim-resistant corn (*Zea mays*) in soybean (Glycine max). *Weed Technology* **11**, 649-655.

Yu, S. J. (1984). Interactions of allelochemicals with detoxifcation enzymes of insecticide-susceptible and resistant Armyworms. *Pesticide Biochemistry and Physiology* **22**, 60-68.

Zeng, K., Rose, J. P., Chen, H. C., Strickland, C. L., Tu, C. P., Wang, B. C. (1994). A surface mutant (G82R) of a human Alpha-glutathione s-transferase shows decreased thermal stability and a new mode of molecular association in the crystal. *Proteins* **20**, 259-263.

Zha, H. B., Fisk, H. A., Yaffe, M. P., Mahajan, N., Herman, B. and Reed, J. C. (1996). Structure-function comparisons of the proapoptotic protein Bax in yeast and mammalian cells. *Molecular and Cellular Biology* **16**, 6494-6508.

Zhang, B. and Singh, K. B. (1994). Ocs element promoter sequences are activated by auxin and salicylic- acid in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2507-2511.

Zimmerlin, A. and Durst, F. (1992) Aryl hydroxylation of the herbicide diclofop by a wheat cytochrome-P-450 monooxygenase – substrate-specificity and physiological-acivity, *Plant Physiology.* **100**, 874-881.