# Structural Studies of Herbicide Detoxifying Enzymes

by

## Alastair Graham McEwen



# UNIVERSITY
## *of*
# GLASGOW

A thesis submitted for the degree of Doctor of Philosophy in
the Faculty of Physical Sciences, University of Glasgow

ProQuest Number: 10753985

ProQuest 10753985

# 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.

© Alastair Graham McEwen, 18th April, 2006

Signed                              Date

**Alastair Graham McEwen        18/04/06**

# Abbreviations

| | |
|---|---|
| 2,4-D | 2,4-Dichlorophenoxyacetic acid |
| 2,4,5-T | 2,4,5-Trichlorophenoxyacetic acid |
| ABC | ATP-Binding Cassette |
| AOS | Active Oxygen Species |
| ATP | Adenosine Triphosphate |
| AU | Asymmetric Unit |
| CD | Circular Dichroism |
| CDNB | Chloro-2,4-dinitrobenzene |
| CLIC | Chloride Intracellular Channels |
| DCA | Dichloroacetic acid |
| DCM | Dichloromethane |
| DEPT | Distortionless Enhancement by Polarization Transfer |
| DMSO | Dimethyl Sulfoxide |
| DNP-GSH | $S$-(2,4-dinitrophenyl)-glutathione |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetate |
| EPTC | $S$-ethyl dipropylthiocarbamate |
| EST | Expressed Sequence Tag |
| FAA | Fumarylacetoacetate |
| GPOX | Glutathione Peroxidase |
| GS⁻ | Thiolate anion of glutathione |
| GSH | Reduced glutathione |
| G-site | Glutathione-binding site |
| GSSG | Oxidised glutathione |
| GST | Glutathione $S$-transferase |
| GS-X | Glutathione-Conjugate |
| HEPES | $N$-(2-Hydroxyethyl)piperazine-$N'$-(2-ethanesulfonic acid) |
| HRMS | High Resolution Mass Spectroscopy |
| H-site | Hydrophobic-binding site |
| IAA | Indole acetic acids |
| IPTG | Isopropyl-$\beta$-D-thiogalactoside |

| IR | Infrared |
|---|---|
| LB | Miller's Luria Broth |
| MAA | Maleylacetoacetate |
| MAAI | Maleylacetoacetate Isomerase |
| MAD | Multi-Wavelength Anomalous Dispersion |
| MAPEG | Membrane-Associated Proteins in Eicosanoid and Glutathione Metabolism |
| MCPA | 4-Chloro-2-methylphenoxyacetic acid |
| MDR | Multidrug Resistance |
| MES | 2-Morpholinoethanesulfonic acid |
| MME | Mono-Methyl Ether |
| MOPS | 3-Morpholinopropanesulfonic acid |
| MP | Melting Point |
| mRNA | Messenger Ribonucleic Acid |
| MRP | Multidrug Resistance-Associated Protein |
| MS | Mass Spectroscopy |
| NMR | Nuclear Magnetic Resonance |
| NTFP-GSH | $S$-[2-Nitro-4-(trifluoromethyl)-phenyl]-glutathione |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PEG | Polyethylene Glycol |
| Pgen | Protoporphyrinogen |
| PPIX | Protoporphyrin IX |
| ROS | Reactive Oxygen Species |
| SDS | Sodium Dodecyl Sulphate |
| SRS | Synchrotron Radiation Source |
| TB | Terrific Broth |
| THF | Tetrahydrofuran |
| TLC | Thin Layer Chromatography |
| Tris | Tris(hydroxymethyl)aminomethane |
| UV | Ultraviolet |

# Summary

Glutathione S-transferases (GSTs) [EC 2.5.1.18] are a ubiquitous family of multifunctional enzymes that are widely distributed in nature. GSTs have been identified from bacteria, fungi, insects, cephalopods, fish, amphibians, reptiles, avians, mammals and plants. The major role of GSTs is the conjugation of the tripeptide glutathione (GSH; γ-L-glutamyl-L-cysteinyl-glycine) to a range of electrophilic substances.

Ten GSTs from rice, wheat, petunia and *Arabidopsis thaliana* were overexpressed in *Escherichia coli* and purified using an Orange A agarose column, S-hexyl glutathione sepharose column, or a nickel chelation column. Using a variety of crystallization screens, crystals were grown for six of the proteins. The X-ray diffraction data collected for two of these proteins enabled their structure to be solved.

The first structure to be solved was that of *Os*GSTU1-1 in complex with GSH, which is a different subgroup of Tau class GSTs from that of the previously reported *Ta*GSTU4-4. *Os*GSTU1 is similar in sequence to the well characterized *Zm*GSTU1 and *Zm*GSTU2, safener induced enzymes from maize highly active against the diphenyl ether herbicide fluorodifen.

In order to determine the structural basis of *Os*GSTU1 substrate specificity a number of GSH conjugates were prepared. A conjugate of CDNB with GSH was prepared by enzymatic routes using the enzyme *Zm*GSTF1. The reaction product of the herbicide fluorodifen with GSH was prepared by a two-step synthetic route by preparing a sulfonamide of the phenyl ring and reacting this with GSH. The structure of *Os*GSTU1-1 was solved with both of these conjugates bound as well as with that of a conjugate of the chloracetanilide herbicide metolachlor with GSH, which was a gift from Prof. R. Edwards, University of Durham, UK.

The structure with metolachlor-GSH bound was useful in structurally characterizing the mode of binding of chloracetanilide herbicides. The structures with the CDNB conjugate with GSH and the reaction product of fluorodifen with GSH were less useful in determining the mode of binding of diphenyl ether herbicides. Both ligands were similar to each other and were found to be sitting between the two lobes of the active site with the nitro group facing inwards. This suggests that the orientation of the molecule within the active site predicted by Dixon *et al.*, 2003 may be incorrect.

The structure of *Os*GSTU4-4 has also been solved in complex with GSH. This enzyme is similar to *Ta*GSTU4-4 and is highly expressed in rice under a variety of stress conditions. A comparison of the three Tau class GST structures now available, along with enzyme assays against a range of xenobiotics substrates has helped to partially understand the activity of Tau class GSTs in herbicide detoxification.

The H-site of Tau class GSTs has two lobes. The previously reported structure of *Ta*GSTU4-4 in complex with *S*-hexylglutathione has the hexyl chain sitting in lobe 'A', but in the structure with metolachlor-GSH the ligand is sitting in lobe 'B' with the gatekeeper Tyr 116 swung away from the active site and disordered.

Conformational changes in plant GSTs were also investigated. A comparison of the structures of apo-form *Zm*GSTF1-1 with the structures of the enzyme in complex with herbicides showed that a conformation change seemed to occur upon ligand binding. Using UV Difference spectroscopy and circular dichroism a change in the conformation of *Zm*GSTF1-1 was seen upon binding of GSH. However, there were problems with the reproducibility of these results. In the case of *At*GSTT1-1, which shows a large difference in conformations between molecules in the apo-form of the crystal structure, both UV difference spectroscopy and circular dichroism showed no spectroscopic changes upon GSH binding.

# Acknowledgements

# Table of Contents

# List of Figures

# List of Tables

# List of Equations

# Chapter 1      Introduction

## 1.1    Importance of Modern Agriculture

In his essay on population of 1798 Thomas Malthus stated that, "Population, when unchecked, increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will shew the immensity of the first power in comparison of the second." This theory postulated that if allowed to grow unchecked, the population of the world would increase at such a rate that the means of food production would fall below the level where the population could be sustained (Malthus, 1798). Although population has doubled since 1960, cereal grain harvests have more than doubled in this period (FAOSTAT data, 2004; Mann, 1999; Figure 1.1). This is due to several factors including plant breeding, agrochemical discovery and farm management techniques known as the 'green revolution.' So far, this has prevented Malthus' prediction from becoming a reality.
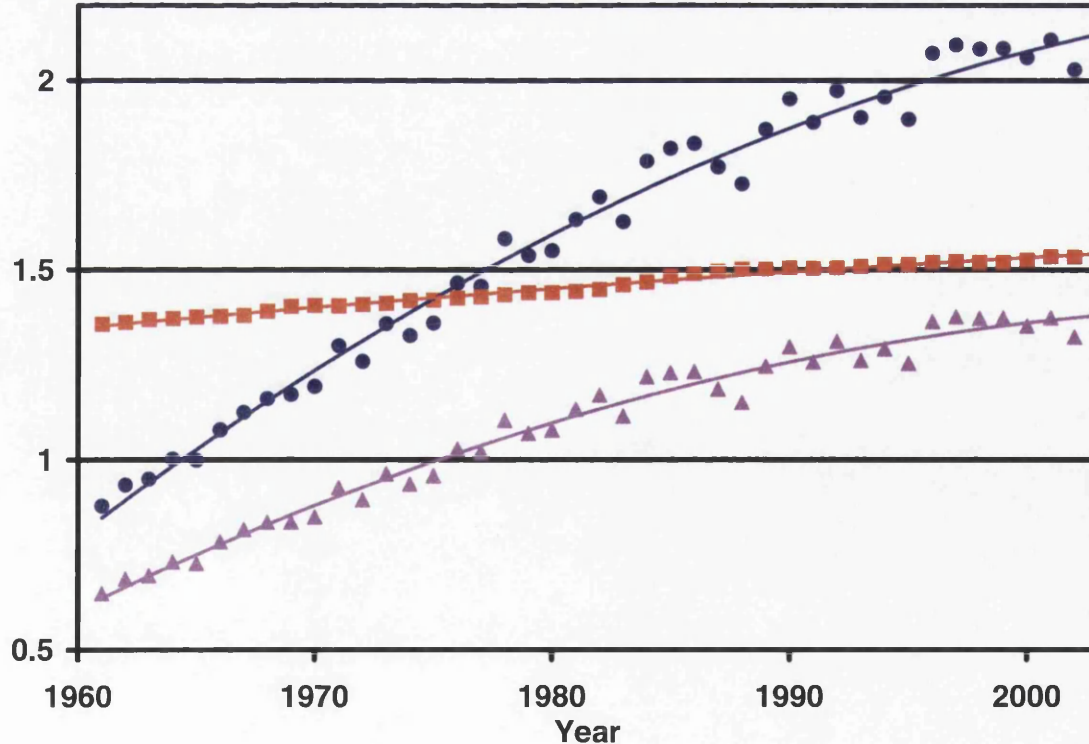


**Figure 1.1.** Growth of grain production since 1960. ● Total grain production (billion tons). ■ Total arable area (billion hectares). ▲ Total grain production per hectare (tons) (FAOSTAT data, 2004).

Projections suggest that global population will rise by seventy-four million per year (FAOSTAT data, 2004; Rosegrant *et al.*, 2001) from now until 2020. This will cause an increase in the global demand for rice (*Oryza sativa*), wheat (*Triticum aestivum*) and maize (*Zea mays*) of 1.35% (Rosegrant *et al.*, 2001). Despite the fact that crop yields have increased over the last forty years, the increase in total crop production per hectare is dropping, from 2.2% per year during 1967 – 1982 to 1.5% per year during 1982 –1994 (figure 1.1; Abelson & Hines, 1999). Combined with the increase in population the total amount of grain produced per person per year has been dropping since the late 1980s (figure 1.2; FAOSTAT data, 2004).
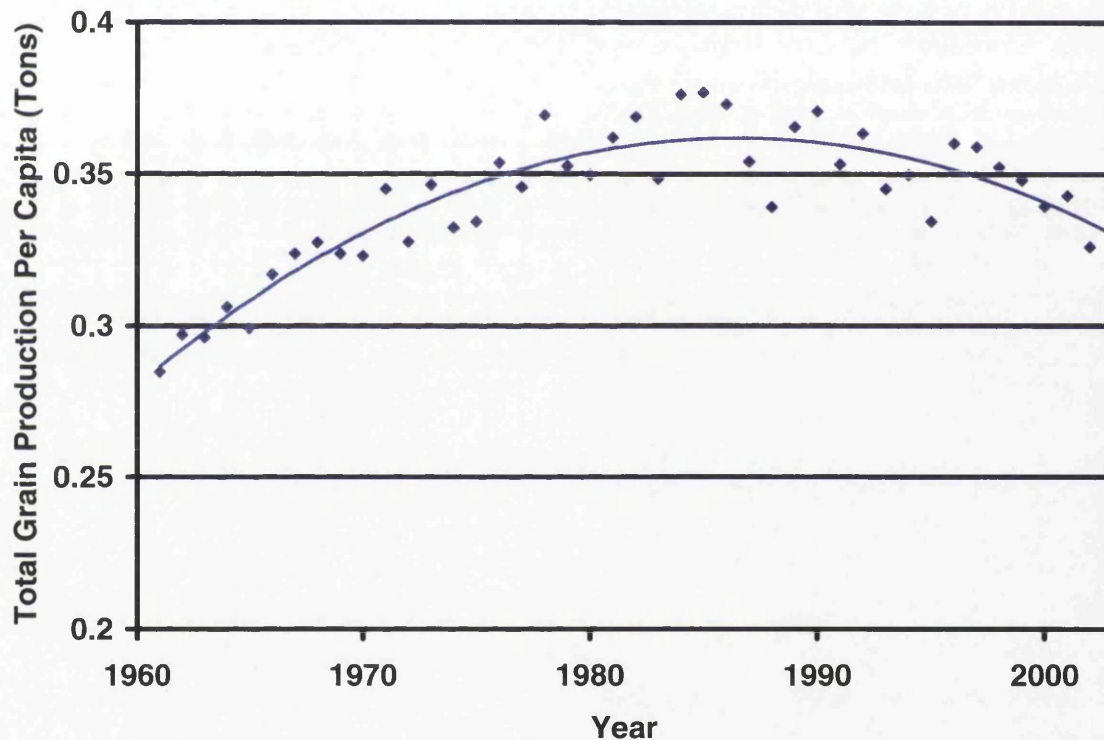


**Figure 1.2.** Total grain production per capita per year (FAOSTAT data, 2004).

Most of the gains in cereal production over the last forty years have been due to improvements in crop yields as most of the world's arable land is already under cultivation (figure 1.3). The exception is Africa where until recently the increase in crop production has been due to cultivation of new land, but gains from now on must be made by increasing the crop yield per hectare or cultivating environmentally sensitive areas (Pinstrup-Andersen, 1994).
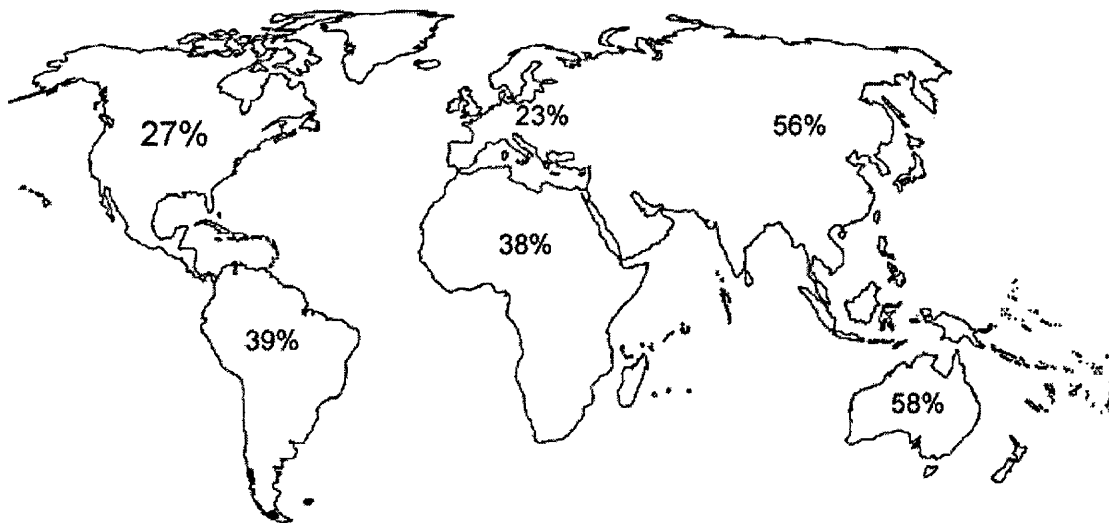
**Figure 1.3.** Domestic land usage. Approximately 39% of the earth's landmass has been converted to agricultural area and urban uses. This figure shows the domesticated land area in Africa (38%), Asia (56%), Europe including Russia (23%), Latin America and the Caribbean (39%), North America (27%) and Australasia & Oceania (58%). (Harrison & Pearce, 2000).

There also seems to be a ceiling on crop yields. Although the average crop yield of maize in the United States of America has risen from five metric tons per hectare (t/ha) to eight t/ha over the thirty years from 1967 to 1997, the maximum yields in experimental plots has remained fairly constant at around twenty t/ha (Mann, 1999). Ceilings on the yields of rice and wheat have also been seen and it seems that farmers are reaching the physical limits on cereal harvests (Cassman, 1999; Conway & Toenniessen, 1999; Mann, 1999).

Other factors contributing to difficulties in feeding the global population are increasing urbanization and a burgeoning demand for 'Western-style' diets. The increase in demand for these western style diets, i.e. more meat and dairy products, is greatest in developing countries with improved standards of living. Therefore these countries have an increased need of cereals to raise domestic livestock that is not needed to support their more traditional diets.

## 1.2 Herbicides

Regardless of the farming system used, the development and use of a range of herbicides is one of the most significant advances that has allowed an increase in crop yields. There is a large range of herbicides (over five hundred) available on the market that can control weeds by a variety of methods. These agrochemicals interfere with vital plant specific processes such as; photosynthetic electron transfer and the biosynthesis of amino acids, porphyrins, caroteinoids, fatty acids or cellulose (table 1.1; Cole, 1994; Oxtoby & Hughes, 1990; Shaner, 2003). A diagram of representative classes of herbicides is shown in figure 1.4.
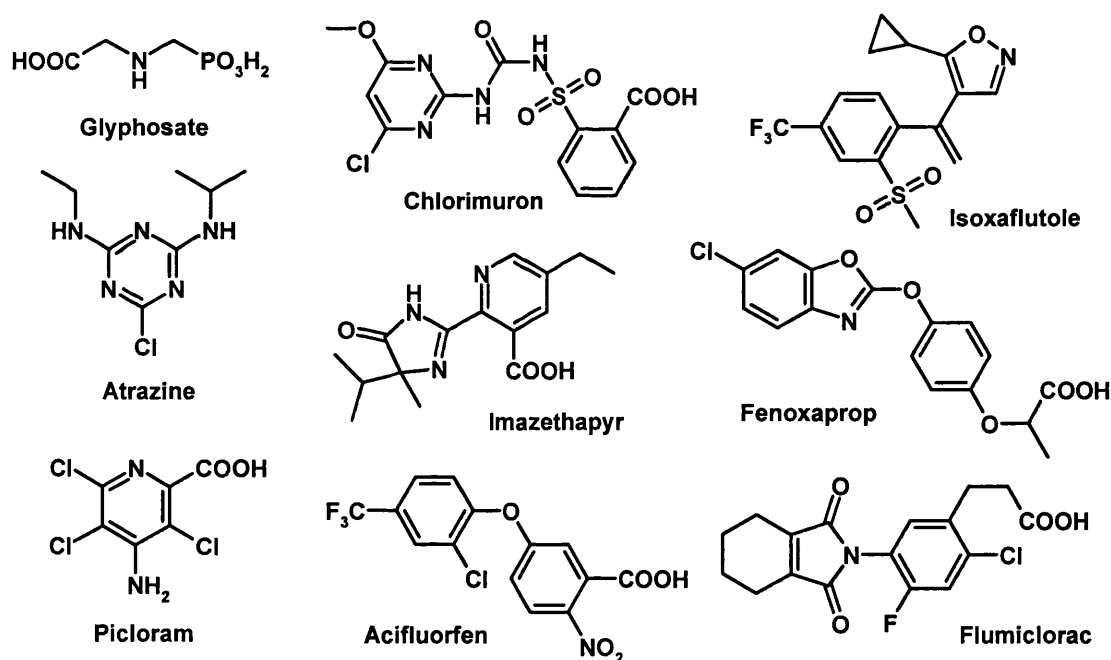
**Figure 1.4.** Representatives of common herbicide classes. The herbicides (and their classes) in this diagram are glyphosate (organophosphorous), chlorimuron (sulfonylurea), isoxaflutole (cyclopropylisoxazole), atrazine (triazine), imazethapyr (imidazoleinone), fenoxaprop (aryloxyphenoxy propionate), picloram (picolinic acid), acifluorfen (diphenyl ether) and flumiclorac (dicarboximide). Image generated with ISIS/Draw (MDL Information Systems, Inc).

| Mechanism of Action | Target | Herbicide Family | Examples |
|---|---|---|---|
| Inhibition of aromatic amino acid biosynthesis | 5-enolpyruvylshikimate-3-phosphate synthase | Organophosphorous | Glyphosate |
| Inhibition of branched-chain amino acid biosynthesis | Acetolactate synthase | Sulfonylurea | Chlorimuron, Nicosulfuron, Thifensulfuron, Halosulfuron, Primisulfuron, Prosulfuron, Rimsulfuron. |
| | | Imidazoleinone | Imazethapyr, Imazaquin |
| Glutamine biosynthesis | Glutamine synthetase | Organophosphorous | Glufosinate |
| Inhibition of photosynthetic electron transport | Photosystem II | Triazine | Atrazine, Cyanazine, Simazine |
| | | Triazinone | Metribuzin |
| | | Nitrile | Bromoxynil |
| | | Quaternary ammonium | Paraquat |
| Auxin mimics | | Phenoxyacetic acid | 2,4-D, 2,4,5-T, MCPA |
| | | Benzoic acid | Dicamba |
| | | Picolinic acid | Picloram |
| | | Pyridine | Triclopyr, Clopyralid |
| Inhibition of fatty acid biosynthesis | Acetyl-CoA carboxylase | Cyclohexanedione | Mesotrione |
| | | Cyclohexene oxime | Sethoxydim, Clethodim |
| | | Aryloxyphenoxy propionate | Fluazifop, Quizalofop, Fenoxaprop, Diclofop |
| Inhibition of carotenoid biosynthesis | 4-Hydroxyphenylpyruvate dioxygenase | Cyclopropylisoxazole | Isoxaflutole |
| Inhibition of chlorophyll biosynthesis | Protoporphyrinogen IX oxidase | Triazolone | Carfentrazone |
| | | Diphenyl ether | Acifluorfen, Lactofen, Fomesafen |
| | | Dicarboximide | Flumiclorac |
| Inhibition of very long chain fatty acid biosynthesis | Very long chain fatty acid elongase | Acetanilide | Alachlor, Metolachlor, Acetochlor |
| | | Anilide | Flufenacet |
| | | Thiocarbamate | EPTC, Butylate |
| Generation of reactive oxygen species | Photosystem I | Quaternary ammonium | Paraquat, Diquat |

Table 1.1. Mechanisms of herbicide action.

Weeds are considerable pests that can: -

- Reduce crop yields significantly, depending on the crop system and geography, by competing for nutrients, water and sunlight.

- Decrease crop quality by contamination with weed material (may be toxic or unpalatable).

- Act as pest and disease reservoirs increasing likelihood that crop will be attacked.

- Hinder use of mechanical harvesting equipment.

Weeds can be controlled by several methods including mechanically (tilling and hoeing), culturally (crop rotation) or by application of herbicides. The use of agrochemicals is effective and its economic benefits mean that mechanical and cultural methods are rarely used alone to control weeds in modem agriculture (Cobb, 1992; Parry, 1989). The major requirement of herbicides used in agriculture is that they can kill the weed species with little or no damage to the crop plants. Herbicides can be divided into two broad types, selective and non-selective.

### 1.2.1  Non-Selective Herbicides

Non-selective herbicides are generally used to destroy all plant material in an area, for example, railroad tracks, roadsides and industrial sites. Until recently non-selective herbicides could only be used in agriculture when no crops were growing (i.e., before sowing or after harvest). This could lead to weed control problems during the growing season. This problem has been partially remedied by using genetically enhanced crops (table 1.2). Crops have been genetically enhanced to provide resistance to previously phytotoxic non-selective herbicides either by increasing the sensitivity of target enzymes for the agrochemical (LeBarron & Gressel, 1982) or by introducing resistance genes from other sources (Gressel, 1985). Monsanto's resistant crop Roundup Ready Soybean is an example of a genetically modified crop that is resistant to the 'environmentally favourable' herbicide glyphosate. The use of these genetically enhanced crop varieties allows use of non-selective herbicides during the growing season to control a wide range of weed species.

They can also offer several potential advantages over the use of selective herbicides, for example, reduced cost, fewer applications and the use of only one herbicide. However, there are some potential disadvantages of herbicide resistant crops, such as the evolution of resistant or tolerant weeds, a shift in spectrum in the range of troublesome weeds or an increase in the quantity of soil borne plant pathogens (Termorshuizen & Lotz, 2002).

| Resistant Crop | Herbicide/Active Ingredient | Herbicide Class | Resistance Gene Product | Reference |
|---|---|---|---|---|
| Imidazolinone tolerant Maize & Tobacco hybrids | Pursuit, Contour, Resolve, Lightning (imazethapyr) | Imidazolinone | Mutant acetolactate synthase (ALS) genes | (Lee *et al.*, 1999) |
| Sethoxydim resistant Maize hybrids | Poast, Poast Plus, Headline (sethoxydim) | Cyclohexendione | Mutant acetyl-CoA carboxylase (ACCase) | (Young & Hart, 1997) |
| Liberty Link or Glufosinate Maize, Wheat, Rice, Cotton & Soybean hybrids | Liberty (glufosinate) | Organo phosphorous | Introduced bacterial *N*-Acetyl transferase | (Deblock *et al.*, 1987) |
| Roundup Ready Soybean & Maize, Tobacco & Tomato hybrids | Roundup Ultra (glyphosate) | Organo phosphorous | EPSPS gene amplification, introduced bacterial AroA gene | (Comai *et al.*, 1985) |
| Soybean, Tobacco | Synchrony (Chlorimuron Ethyl, Thifensulfuron Methyl), Reliance | Sulfonylurea | Mutant Acetolactate synthase (ALS) genes | (Lee *et al.*, 1988) |
| Oilseed Rape, Cotton, Potato | Bromoxynil | Nitrile | Introduced Nitrilase | (Stalker *et al.*, 1988) |
| Cotton | 2,4-D | Phenoxyacetic Acid | Introduced Monooxygenase from gene *tfd*A | (Bayley *et al.*, 1992) |
| Tobacco | Dalapon | Halogenated Aliphatic Acids | Introduced Dehalogenase | (Buchanan Wollaston *et al.*, 1992) |
| Tobacco | Glufosinate | Organo phosphorous | Introduced bacterial Glutamine synthase gene amplification bar gene | (Deblock *et al.*, 1987) |

**Table 1.2.** Genetically enhanced crops available on the market.

In 1956, John Harper first proposed that weeds that were previously susceptible to herbicide control might develop resistance (Harper, 1956). The first cases of herbicide resistance were discovered in the late 1960's in a pine nursery where triazine herbicides had been used repeatedly (Ryan, 1970). It has since been shown that a weed can become resistant to an herbicide when it has been used repeatedly for four to ten years (Heap, 1994). Resistance to a particular class of herbicide in weeds can occur due to mutations to the herbicide target enzyme. Resistance to chemically unrelated herbicides however is the result of the plant developing up-regulating mechanisms to metabolize herbicides before they can act.

The emergence of herbicide resistant weeds has been recognized as a problem in modern farming methods and techniques such as crop rotation and the use of herbicide mixtures are now being used to minimize the risk resistance arising (Cole, 1994).

### 1.2.2 Selective Herbicides

Unlike non-selective herbicides, selective herbicides are able to exploit the difference between species of plants such as herbicide uptake, movement, and metabolism, which allows them to be used during the growing season with little or no crop damage.

To minimize the occurrence of resistant weeds, a range of selective herbicides that undergo different metabolic detoxification pathways are applied to the crop (figure 1.5). Although differences in uptake and transport of the herbicide can be factors in its selectivity (Devine *et al.*, 1993), the plants ability to detoxify the herbicide rapidly is the most important factor in herbicide resistance.

The difference in plant structure can be utilized by selective herbicides, for instance broad-leafed weeds are more susceptible to leaf bleaching agents than upright, narrow-leaved crop plants (Cobb, 1992). However, when the crop and weed species are similar, the problem is more challenging. To

design herbicides able to discern between similar species, knowledge of the subtle differences in their biochemistry is required.



**Figure 1.5.** Selective grass control in *Z. mays*. Plants treated with isoxaflutole and metolachlor are shown in the centre two rows with untreated plants on either side. Photo courtesy of David J. Cole.

## 1.3   Xenobiotic Detoxification Systems in Plants

Research over the past few decades has revealed that higher plants have an extensive ability to metabolize xenobiotics. In order to detect differences in metabolism between species, many agrochemical companies have focused their research to understand how plants metabolize foreign compounds. These differences may be effectively exploited in the design of new agrochemicals. A brief discussion of plant herbicide metabolism has been included in this text describing the various roles played by enzymes in these processes, and the points at which the glutathione *S*-transferase family act. Details of herbicide metabolism and plant physiology have been covered in detail in a number of texts (Cobb, 1992; LeBarron & Gressel, 1982) and articles (Cole & Edwards, 2000; Dodge, 1982; Sanderman, 1994; Van Eerd *et al.*, 2003)

The chemical reactions that both plants and animals utilize to metabolize toxic compounds can be classified into one of three phases:

- Phase I – Transformation
- Phase II – Conjugation
- Phase III – Compartmentalization

### 1.3.1  Phase I: Transformation

The most common first stage of detoxification of xenobiotics is oxygenation, reduction or hydrolysis by, for example cytochrome P450s, peroxidases and polyphenol oxidases (Van Eerd *et al.*, 2003). The most important family of phase I enzymes are the cytochrome P450s, and these are the most comprehensively studied (Barrett, 2000). Cytochrome P450s most commonly catalyse the hydroxylation of the agrochemical (figure 1.6) but can also catalyse dehydration, dimerization, deamination, dehydrogenation, heteroatom dealkylation, epoxidation, reduction and C-C or C=N cleavage (Van Eerd *et al.*, 2003).
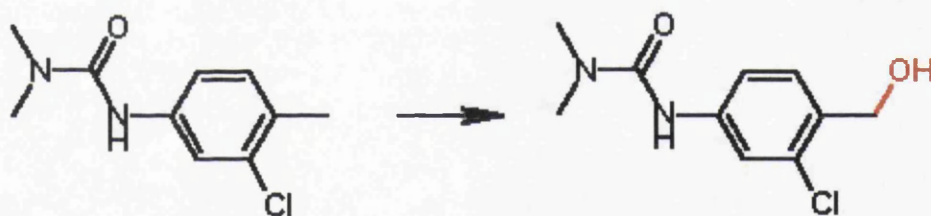


**Figure 1.6.** Hydroxylation of chlorotoluron by cytochrome P450 in *T. aestivum*. The reaction requires oxygen, NADPH and NADPH-Cyt-P450 reductase. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Many herbicides are converted by cytochrome P450s into a range of different metabolites. For example, the herbicide prosulfuron is converted into three different metabolites in wheat, by ring hydroxylation, alkyl hydroxylation and *O*-demethylation (figure 1.7; Frear, 1995). The fact that these herbicides can be metabolized in different ways by different species can be exploited in the mechanisms of selective herbicides. The phenylurea herbicide chlortoluron can be metabolized by either ring-methyl hydroxylation or by di-*N*-demethylation. In tolerant wheat the half-life of chlortoluron is less than

twenty-four hours and the main metabolite is ring-methyl hydroxylated. However, in susceptible weeds, the main metabolite is the mono-*N*-demethylated metabolite, which retains phytotoxicity (Mougin *et al.*, 2000).
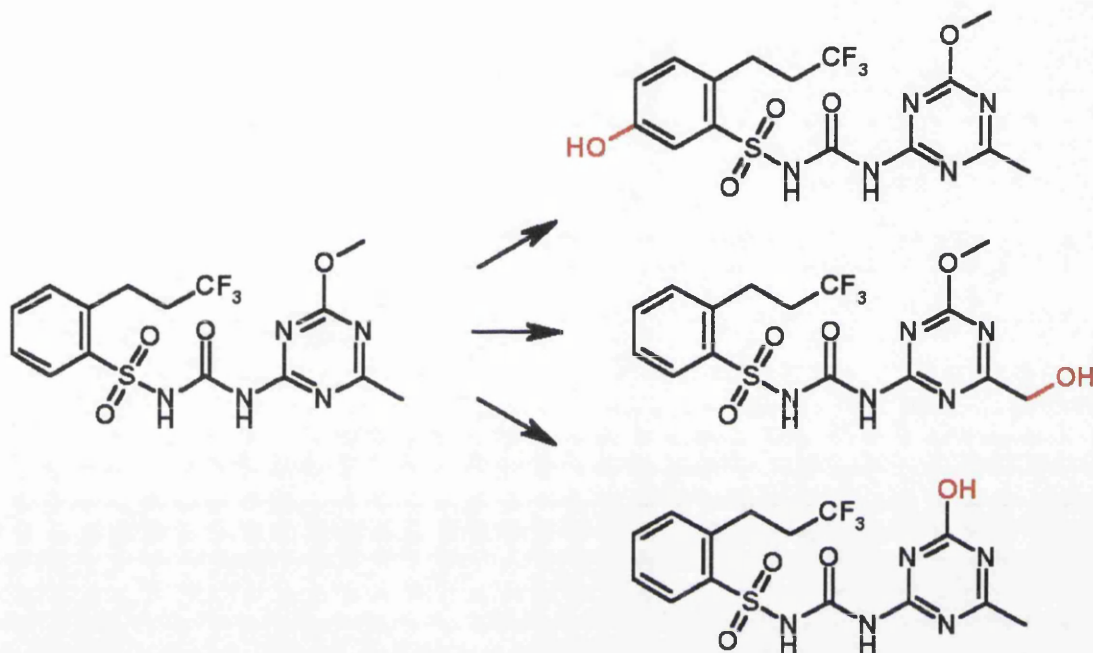
**Figure 1.7.** Oxidation of prosulfuron by cytochrome P450s in *T. aestivum*. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Although cytochrome P450 enzymes can detoxify agrochemicals they can also activate pro-herbicides through hydroxylation of inactive esters. These pro-herbicides may have more favourable physiochemical properties for uptake and transport by the plant. For instance, the pro-herbicide diclofop-methyl is inactive *per se*, but on entry to the plant it is rapidly converted to phytotoxic diclofop (Shimabukuro *et al.*, 1979).

### 1.3.2  Phase II: Conjugation

In phase II metabolism the functionalized toxin from phase I can undergo a number of conjugation reactions. These include glycosylation and conjugation with glutathione, malonic acid and amino acids. Conjugation increases the water solubility of the molecule and thereby decreases its reactivity, toxicity and mobility. This also allows the targeting of the

1994). In a few cases lipophilic conjugates of xenobiotics have been identified, for example esters formed with fatty acids (Lamoureux & Rusness, 1986).

These phase II metabolism reactions have been shown to be important in herbicide resistance among crop plants such as maize, rice, soybean (*Glycine max*) and wheat (Cole *et al.*, 1997; Schröder, 1997). The variance in both the types of enzymes involved in secondary metabolism and their differential expression between species allows for the production of selective herbicides or tolerant crop plants. Over-expression of particular detoxifying enzymes, either through up-regulation with safeners (chemicals that stimulate expression of detoxifying enzymes) or heterologous expression, or alteration of the substrate specificity of the enzymes are methods which may allow the development of new selective herbicides (Cole & Edwards, 2000).

### 1.3.2.1 Glycosylation

Where a hydroxyl, thiol, amino or carboxylic acid group is introduced to the agrochemical by phase I metabolic enzymes (or is present on the parent compound), the molecule can undergo rapid glycosylation, yielding a much less mobile conjugate. The imidazolinone herbicide imazethapyr is detoxified in soybean in a two-step process. Hydroxylation by phase I enzymes to HO-imazethapyr reduces its phytotoxicity by 2-3 times, but the herbicide is only completely inactivated by glycosylation (figure 1.8; Little & Shaner, 1991; Tecle *et al.*, 1993).

Although other sugars are conjugated to endogenous compounds in plants, glucose appears to be the major sugar involved in the detoxification of xenobiotics (Cole & Edwards, 2000). Conjugates of xenobiotics with glucuronic acid have been identified from several plant species (Bokern *et al.*, 1996; Kolb & Harms, 2000), and conjugates with galactose have tentatively been identified from several species (van der Krol *et al.*, 1995).
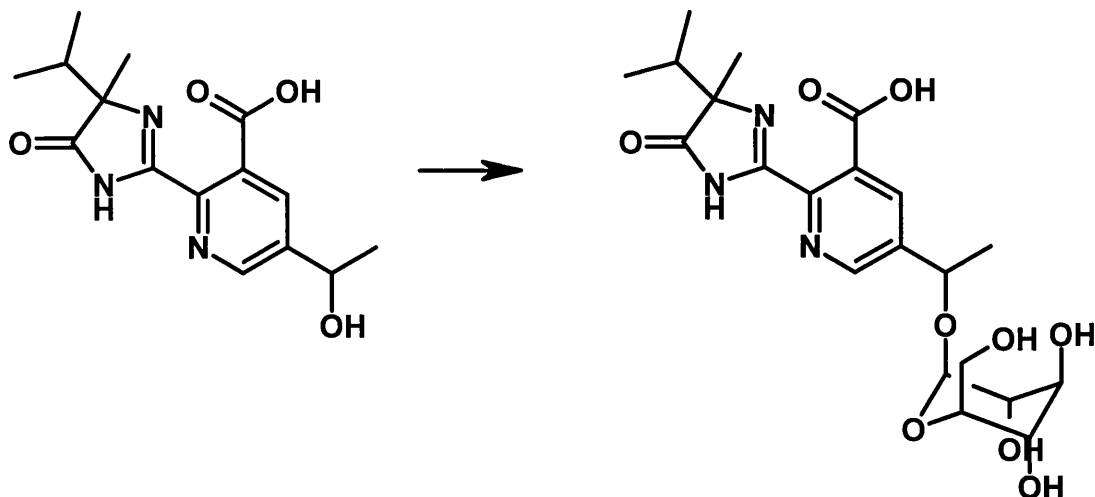
**Figure 1.8.** Glucosylation of HO-imazethapyr in *G. max*. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Although not common in primary glycosylation of xenobiotics, sugars like xylose and arabinose are often attached to a glucoside to form a disaccharide of the compound (More *et al.*, 1978; Wright *et al.*, 1980). A second molecule of glucose can also be added to different positions on the glucoside to form the disaccharides gentiobiose ($\beta$-(1→6) linked diglucoside; Hodgson *et al.*, 1974; Schmidt *et al.*, 1997), cellobiose ($\beta$-(1→4) linked; Mikami *et al.*, 1984) and sophorose ($\beta$-(1→2) linked; Mikami *et al.*, 1985).

One problem of glycosylation is that conjugates formed with carboxylic acid groups can undergo ester hydrolysis to re-yield the parent compound and, therefore, glycosylation cannot be regarded as an effective detoxification step for these compounds.

### 1.3.2.2 Amino Acid Conjugation

While amino acid conjugation has been well characterized in animals as a mechanism for xenobiotic detoxification, it has limited importance in the detoxification of agrochemicals, although studies have shown that dicotyledonous species are more likely to use amino acid conjugation than monocotyledonous species (Mumma & Davidonis, 1983). One of the few cases of amino acid conjugation to agrochemicals is the amide linkage of amino acids, mainly aspartate and glutamate, to the herbicide 2,4-

dichlorophenoxyacetic acid (2,4-D) (figure 1.9). This mechanism has been described in soybean and is proposed to be analogous to the metabolism of auxin molecules, which it mimics (Cohen & Bandurski, 1982). Amino acid conjugates have also been identified in soybean as metabolites of the pyridinyloxyacetic acid herbicide triclopyr (Lewer & Owen, 1989, 1990), and a conjugate of the carbanilate herbicide chlorpropham with cysteine has been identified from oat (*Avena sativa*; Still and Rusness, 1977).
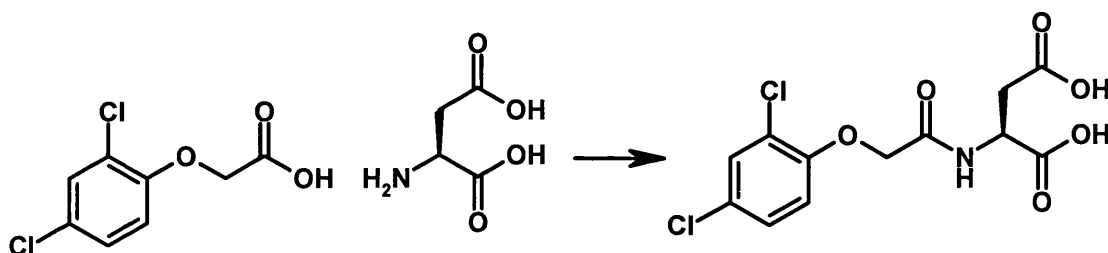
Figure 1.9. Conjugation of 2,4-D with aspartic acid in *G. max*. Image generated with ISIS/Draw (MDL Information Systems, Inc).

### 1.3.2.3 Glutathione Conjugation

One of the major methods with which plants detoxify xenobiotics is the conjugation of the compounds with the cellular thiol, reduced glutathione ($\gamma$-L-glutamyl-L-cysteinyl-glycine [GSH]). Several legume species use an analogue of glutathione known as homoglutathione ($\gamma$-L-glutamyl-L-cysteinyl–$\beta$-alanine; Macnicol, 1987) and hydroxymethylglutathione ($\gamma$-L-glutamyl-L-cysteinyl-L-serine; Mason *et al.*, 1987) has been identified as the major tripeptide in some grass species. Glutathione S-transferases (GSTs) stabilize the formation of the thiolate anion of glutathione (GS⁻). This anion can then attack an electrophilic centre on the agrochemical to form the conjugate with associated displacement of a nucleophile (e.g. a halogen, phenolate or alkyl sulfoxide). Since GSH has a high number of hydrophilic functional groups but a relatively low molecular weight it has a high water solubility (Kosower, 1976).

A great many xenobiotics are directly susceptible to GSH conjugation in plants, including several major herbicide classes such as the triazines, chloracetanilides, sulfonylureas, aryloxyphenoxy propionates, thiocarbamates

and diphenylethers (figure 1.10; Lamoureux and Russness, 1993; Cole *et al.*, 1997; Schröder, 1997), however electrophilic centres introduced by phase I enzymes allow the metabolism of many additional xenobiotics by GSTs. As well as reducing the activity of the agrochemical and increasing its polarity, the GSH also acts as a tag, signalling for the removal of the conjugated herbicide from the cell cytoplasm, either to the cell vacuole, apoplast or plant matrix (Sanderman, 1994).
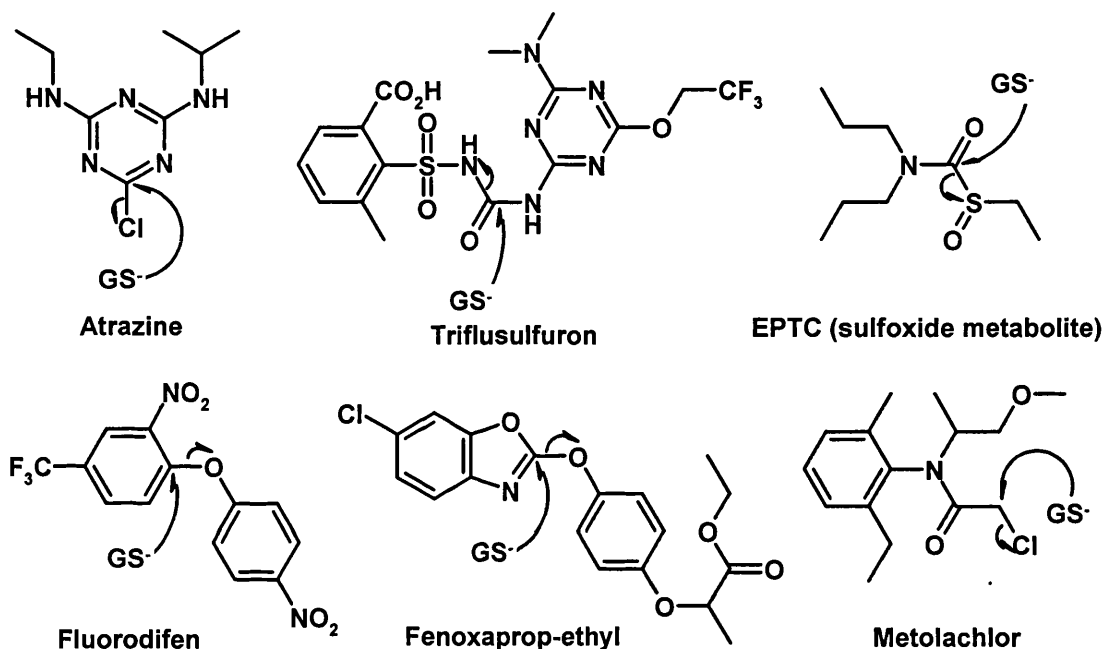


**Figure 1.10.** Conjugation of thiolate anion of glutathione (GS⁻) with atrazine (triazine), triflusulfuron (sulfonurea), the sulfoxide metabolite of EPTC (thiocarbamate), fluorodifen (diphenyl ether), fenoxaprop-ethyl (aryloxyphenoxy propionate) and metolachlor (chloracetanilide). Image generated with ISIS/Draw (MDL Information Systems, Inc).

*1.3.2.4 Malonyl Conjugation*

While malonic acid is not conjugated directly to xenobiotics it is often introduced to glucosides or products of the catabolism of glutathione-conjugates (Cole & Edwards, 2000) and is seen as a significant final step in the metabolism of xenobiotics. Also, malonylation seems to provide a signal for the removal of the conjugate, either to the vacuole (Mackenbrock *et al.*, 1992) or across the plasma membrane and out of the cell (Winkler & Sanderman, 1989; Höhl & Barz, 1995). An example of the malonylation of the glucoside of pentachlorophenol in soybean and wheat is shown in figure 1.11 (Schmitt *et al.*, 1985).
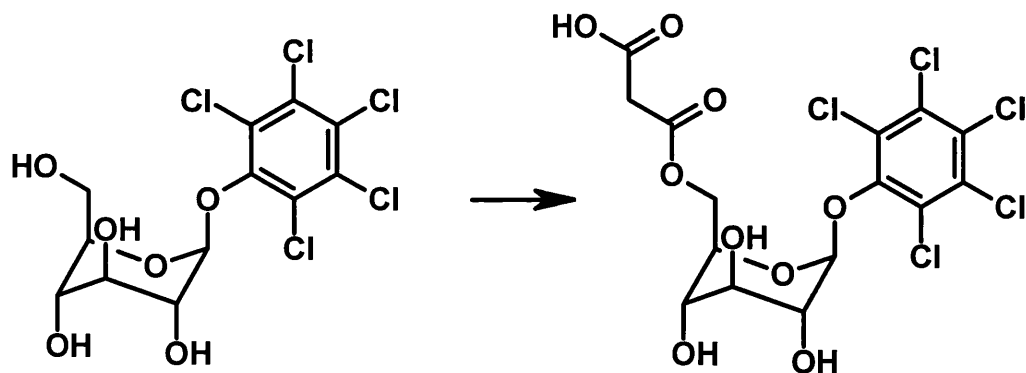
**Figure 1.11.** Conjugation of malonic acid with the O-β-D-glucoside of pentachlorophenol in *G. max* and *T. aestivum*. Image generated with ISIS/Draw (MDL Information Systems, Inc).

### 1.3.2.5 Lipophilic Conjugation

In a few cases lipophilic conjugates of agrochemicals have been identified from plants. Esters of triton detergents, a wetting agent used with many herbicides, with palmitic and linoleic acids have been identified from a crude enzyme preparation of corn shoots and, in vivo, small amounts of the conjugates have been identified from rice and barley (*Hordeum vulgare*) tissues (Frear *et al.*, 1977). Lipophilic conjugates of the herbicides picloram and 2,4-D have been isolated from radish (*Raphinus sativus*) and mustard (*Brassica kaber*) plants (Chkanikov *et al.*, 1984). These conjugates are formed with 4-hydroxystyryl-mustard, vinyl-mustard and allyl-mustard, and for 2,4-D, conjugates with mustard oils are the only conjugates identified (figure 1.12).
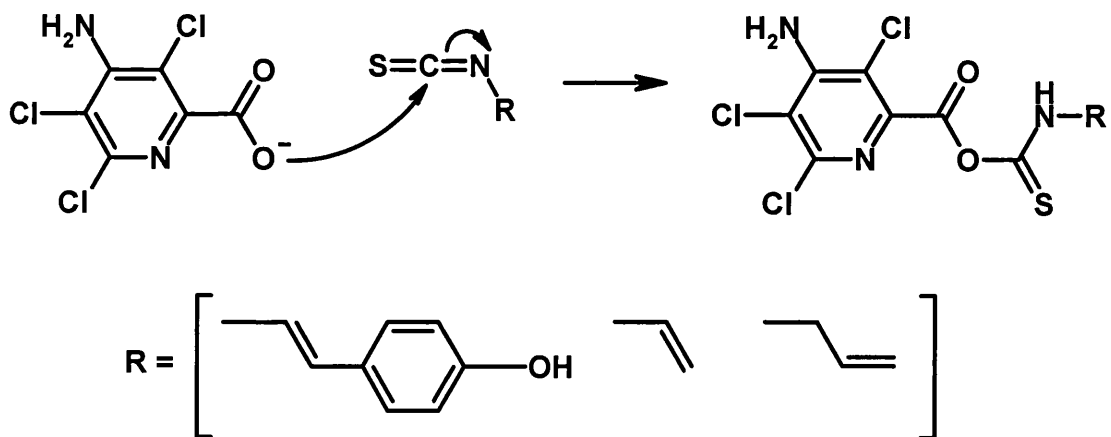


**Figure 1.12.** Lipophilic conjugation of picloram with mustard oils in *R. sativus* and *B. kaber*. Image generated with ISIS/Draw (MDL Information Systems, Inc).

### 1.3.3 Phase III: Compartmentalization

The final step of detoxification is removal of the xenobiotic from the cytosol – either to the vacuole, the extracellular space, or deposited in the cell wall as a 'bound-residue' (Sandermann, 1992). Since the conjugates from phase II have such a reduced partition coefficient they would be unable to diffuse across the membrane. The main route of removal of these conjugates is transport into the vacuole by a member of the ABC (ATP-Binding Cassette) superfamily of transporters (Martinoia et al., 1993). Proteins of the ABC superfamily all share four 'core' domains. There are two membrane-spanning domains that are each followed by an ATP-binding domain. The family is divided into two subclasses – the multidrug resistance proteins (MDR; Bellamy, 1996) and the multidrug resistance-associated proteins (MRP; Cole et al., 1992).

The transporter involved in the removal of glutathione-conjugates to the vacuole (GS-X pump) is a member of the second class of ABC transporters. The GS-X pump requires MgATP to function but cannot use free ATP (Li et al., 1995) and can transport a range of conjugates, including oxidized glutathione (GSSG) but will not transport free GSH or cysteine conjugates (Martinoia et al., 1993). Once transported to the vacuole the conjugates can undergo further processing by carboxypeptidases and dipeptidases that can remove the glycinyl and glutamyl moieties sequentially (Wolf et al., 1996; Coleman et al., 1997).

Another method of disposing of the xenobiotic is covalent binding of the molecule to the cell wall. Xenobiotics can be bound on to several different components of the cell wall, namely proteins, lignins, hemicelluloses and pectins (Skidmore et al., 1998; Schmidt, 1999; Sandermann et al., 2001). Recently, the different types of linkage to the four mentioned cell wall components have been proposed by Heinrich Sandermann Jr. (2004; figure 1.13).

A.                         B.                         C.                         D.
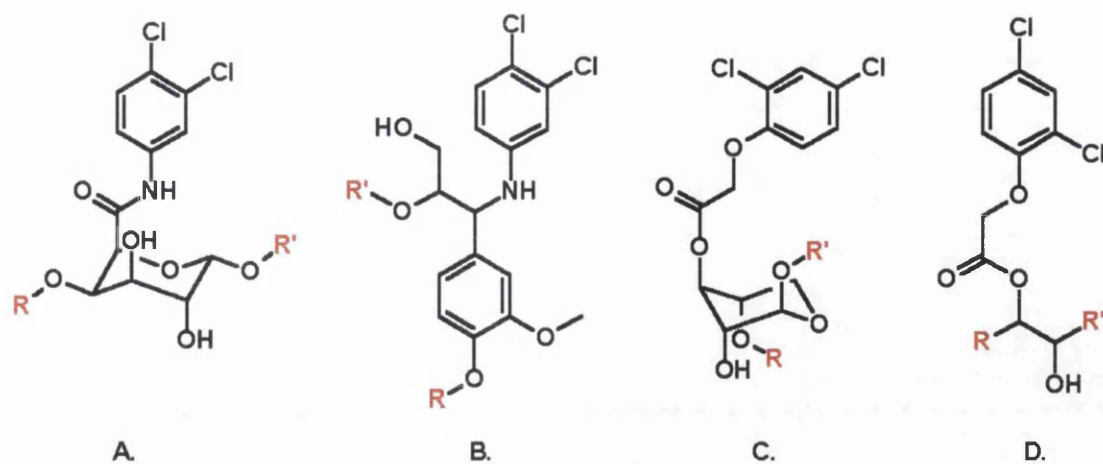
**Figure 1.13.** Binding of xenobiotics to the plant cell wall. R and R' represent points of attachment to the various polymer chains. **A.** Chloroaniline bound to pectin. **B.** Chloroaniline bound to lignin. **C.** Phenoxyacetic acid bound to hemicellulose. **D.** Penoxyacetic acid bound to cutin. Image generated with ISIS/Draw (MDL Information Systems, Inc).

## 1.4    Glutathione *S*-Transferase Family

A number of good review articles discuss the glutathione *S*-transferase (GST) family in detail (Vuilleumier *et al.*, 1997; Edwards *et al.*, 2000; Sheehan *et al.*, 2001; Dixon *et al.*, 2002b; Frova, 2003; Enayati *et al.*, 2005; Hayes *et al.*, 2005). This section provides a brief overview.

GSTs are a large family of catalytic and binding proteins that are distributed throughout highly diverse aerobic organisms. They have many and varied activities, including the detoxification of xenobiotics, mediation of oxidative stress, amino acid degradation, and the non-catalytic binding of hormones and other endogenous compounds. Within this family there are many distinct classes of GSTs based on sequence similarity, immunological cross-reactivity, and intron number and position. There have been eight classes of GSTs identified in mammals, seven of which are cytosolic and are classified as Alpha (Kamisaka *et al.*, 1975), Mu (Warholm *et al.*, 1983), Pi (Guthenberg & Mannervik, 1981), Theta (Meyer *et al.*, 1991), Sigma (Buetler & Eaton, 1992), Zeta (Board *et al.*, 1997), and Omega (Board *et al.*, 2000). There is also the Kappa class that is found in mitochondria (Ladner *et al.*, 2004) and peroxisomes (Morel *et al.*, 2004).

The intracellular chloride ion channel (CLIC) that is found in Chordata can exist in two forms, an integral membrane bound form and a soluble, monomeric form. The soluble form has been shown to be structurally similar to Omega class GSTs and contains an intact G-site (Harrop et al., 2001). Under oxidizing conditions the N-terminal domain can rearrange, the protein will dimerize and insert into the membrane to form the ion channel (Littler et al., 2004).

Five distinct classes have been identified in plants of which the major two classes, Phi and Tau, are plant specific. Theta and Zeta classes have also been identified in plants along with Lambda, and a further GST-like class named DHAR (proteins with GSH dependent dehydroascorbate reductase activity; Dixon et al., 2002). Several other classes of GSTs have been identified in other species including the bacterial Beta class (Rossjohn et al., 1998a), the insect Delta (Board et al., 1997), Epsilon (Ranson et al., 2001), and U classes (or unclassified; Ding et al., 2003), the fish specific Rho class (Konishi et al., 2005), the GST related S-crystallins from cephalopods (Tomarev & Zinovieva, 1988), and the microsomal MAPEG superfamily (Membrane Associated Proteins in Eicosanoid and Glutathione Metabolism) which has been identified in mammals, plants, fungi and bacteria (Morgenstern et al., 1981; Jakobsson et al., 1999).

### 1.4.1 Evolution of Glutathione S-Transferases

Proteins capable of detoxifying xenobiotics have been present in both prokaryotes and eukaryotes for well over 2.5 billion years, long before animals and plants are thought to have diverged (~1.2 billion years; Nebert, 1994; Nebert & Dieter, 2000). It appears that the GST superfamily evolved from an ancestral thioredoxin-like protein, possibly in response to the development of oxidative stress (Koonin et al., 1994; Martin, 1995). Several other GSH- and cysteine-binding proteins contain a thioredoxin-like fold (Martin, 1995) and it seems clear that GSTs share sequence and structural similarity with a number of stress-related proteins from a vast range of species (Rossjohn et al., 1996).

In 2001 David Sheehan proposed a pattern of evolution for the members of the GST superfamily that had been identified, based on sequence comparisons from several sources (figure 1.14; Board et al., 2000; Kanaoka et al., 1997; Snyder & Maddison, 1997). Based on their ubiquitous distribution throughout prokaryotes and eukaryotes, it has been suggested that the Theta and Zeta classes of GSTs developed from ancestral GSTs before the animal-plant divergence (Dixon et al., 1998, 2002b; Edwards et al., 2000). Following this event these classes have undergone only one or two duplication events yielding a maximum of three paralogs in all plant species that have been analysed (Wagner et al., 2002).
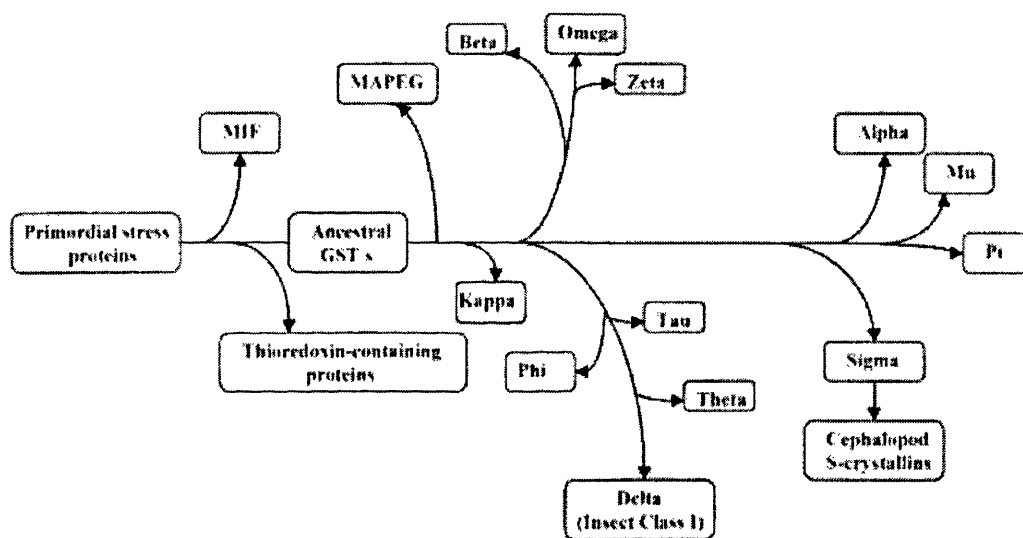


**Figure 1.14.** Possible pattern of evolutionary divergence of the GST superfamily. Image taken from Sheehan et al., 2001.

The plant specific Tau and Phi classes however are thought only to have duplicated after the divergence of monocotyledonous plants from dicotyledonous plants around two hundred million years ago (Wolfe et al., 1989). After this point extensive duplication events occurred in both monocots and dicots separately, yielding several paralogous proteins in each class. Evidence for this can be seen the clustering of both Tau and Phi class GSTS into several subclasses in both monocots and dicots (McGonigle et al., 2000; Dixon et al., 2002b), and in the extensive clustering of genes within chromosomes (figure 1.15).
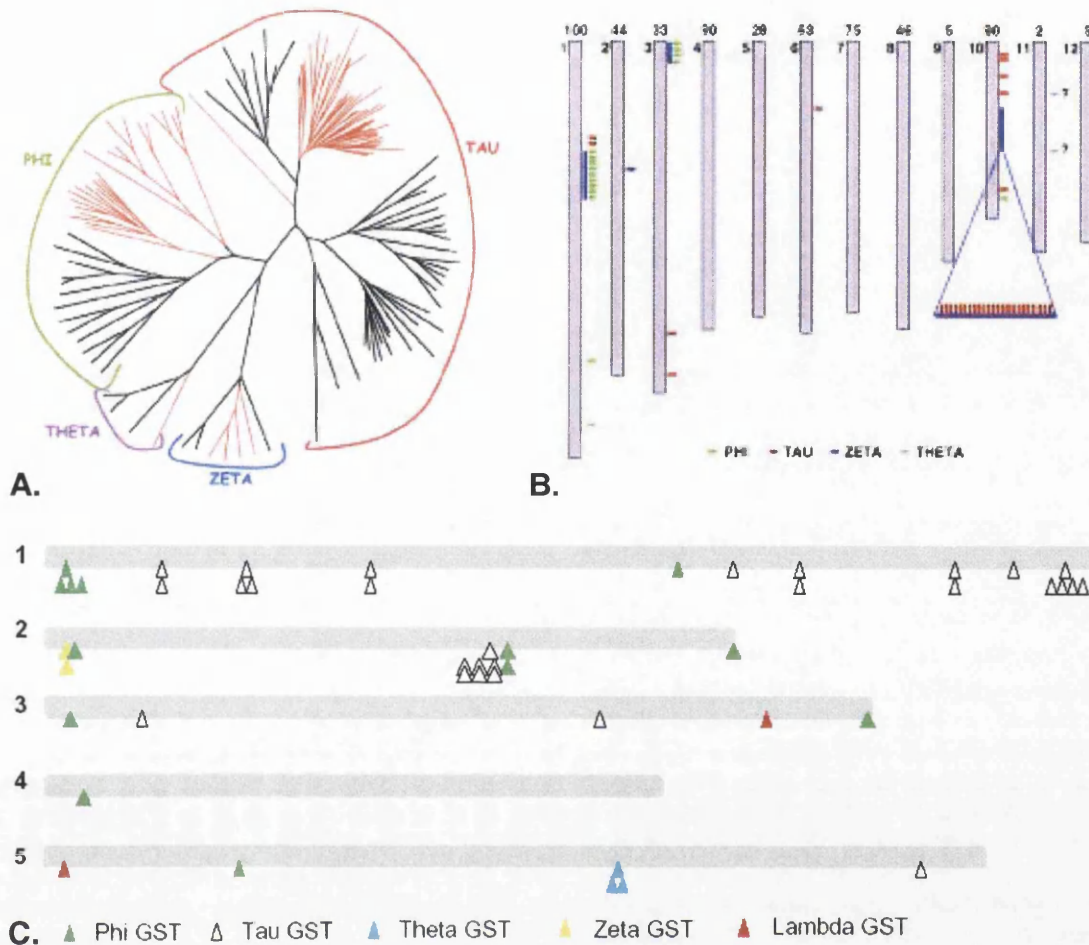
**Figure 1.15.** Divergence of GSTs within plants. **A.** Phylogenetic tree of cytosolic plant GSTs showing clustering of Phi and Tau class GSTs into several subclasses. Sequences from monocotyledonous rice and maize are shown in red and sequences from dicotyledonous soybean and *Arabidopsis* are shown in black. Image taken form Soranzo *et al.*, 2004. **B.** Distribution of GSTs within the rice genome showing clustering of proteins. Numbers at the top indicate percent of chromosome sequenced. Image taken from Frova, 2003. **C.** Distribution of genes within the *Arabidopsis* genome showing clustering of proteins. Image adapted from Dixon *et al.*, 2002b.

In rice, a cluster of twenty-three Tau class genes has been identified on a 239 kilo base (kb) long contig on chromosome ten (OSJNBa0034L04 & OSJNBb0005J14). This contig contains sixty-seven genes, which is an average of one gene every 3.6 kb. This is nearly double the average for rice chromosome ten, which has around one gene every 6.5 kb (The Rice Chromosome 10 Sequencing Project, 2003), and triple the average of one gene every 10.5 kb calculated for the complete rice genome (Goff *et al.*, 2002). This again suggests rapid duplication in this region of the genome.

The fact that there are few copies of Theta and Zeta class genes in plant genomes suggests that these proteins have been subjected to little selective pressure. Zeta class GSTs are involved in primary cell metabolism and still share high sequence similarity between phyla. Human Zeta GSTs share 49% sequence identity with *Caenorhabditis elegans* Zeta GSTs and 38% identity with carnation (*Dianthus caryophyllus*) Zeta GSTs (Board *et al.*, 1997). The Theta class of enzymes shares less sequence similarity in general than other classes of GSTs (Droog, 1997). Their activity is also less well understood. Although mammalian Theta class GSTs can detoxify some xenobiotics they also show very high GST-GPOX activity. In plants, some of the Theta class GSTs have GST-GPOX activity but the endogenous role of the family is not yet known. Phi and Tau class GSTs have evolved to specialize in the conjugation of GSH to a range of toxic xenobiotics. Due to plants sessile nature they are unable to avoid toxins from their environment, which may change quickly. Therefore plants must be able to adapt quickly to changing environmental stresses, and the selective pressure for gene duplication and diversification of substrate specificity must have been strong.

## 1.5   The Structure of Glutathione *S*-Transferases

Cytosolic GSTs are a dimer of two approximately 26 kDa monomers which have a fairly well conserved canonical fold comprised of two distinct domains (figure 1.16; Wilce & Parker, 1994; Sinning *et al.*, 1993). The N-terminal domain is composed of a thioredoxin like fold composed of a four-strand $\beta$-sheet flanked by three $\alpha$-helices, two beneath the sheet and packing against the C-terminal domain, and one above the sheet exposed to the solvent (Martin, 1995). This fold is composed of two separate motifs, which are composed of $\beta\alpha\beta$ and $\beta\beta\alpha$ elements. These two motifs are linked by a loop containing another $\alpha$-helix. This loop contains a *cis*-proline residue that is highly conserved throughout all GSTs. While this residue does not play any part in binding of GSH it is necessary to maintain the protein in a catalytically competent structure (Allocati *et al.*, 1999).
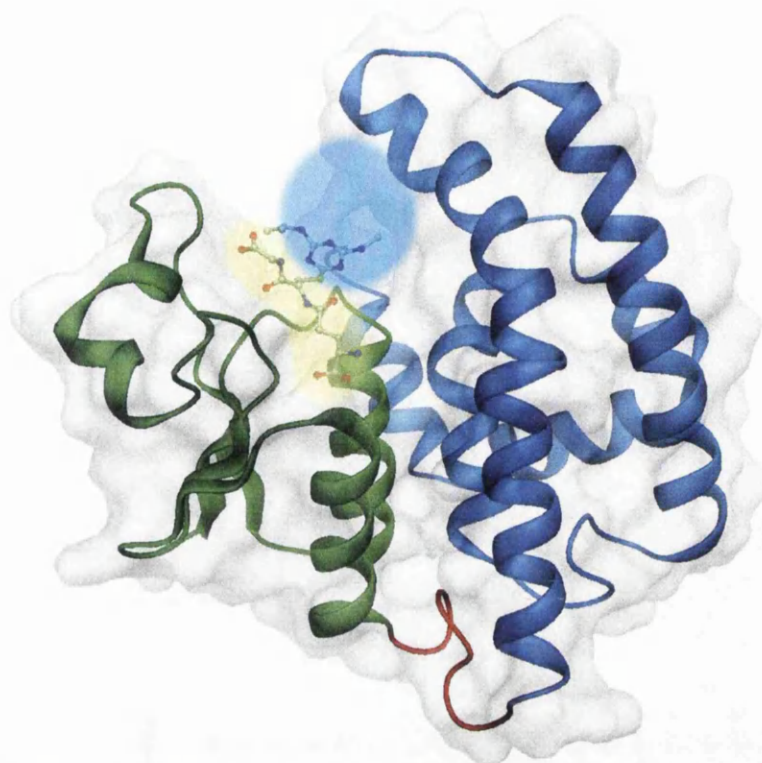
**Figure 1.16.** Ribbon diagram of the structure of *Zm*GSTF1 with an atrazine/GSH conjugate shown as ball and stick (PDB ID 1BYE; Prade *et al.*, 1998). The N-terminal domain is shown in green and the C-terminal domain in blue. The GSH-binding site (G-site) is highlighted in yellow and the hydrophobic binding site (H-site) in blue. Diagram taken from Dixon *et al.*, 2002b).

The N-terminal domain contains the G-site and helps stabilize the thiolate anion of glutathione, GS⁻. The GS⁻ anion is stabilized by a hydroxyl group of a tyrosine (Alpha, Mu, Pi) or serine (Theta, Zeta, Phi, Tau) residue, which are essential for activity of the protein (Dirr *et al.*, 1994; Thom *et al.*, 2001a). Omega, Beta, Lambda and DHAR class GSTs all contain an active site cysteine and form a mixed disulphide with GSH (Board *et al.*, 2000; Nishida *et al.*, 1998; Rossjohn *et al.*, 1998a; Dixon *et al.*, 2002a).

The C-terminal domain of GSTs is far less conserved than the N-terminal domain and contains the H-site that recognizes the varied substrates of GSTs. This domain is composed of α-helices, with a varying number of helices between the different classes. The first three helices form a three-helix bundle that is the core of the domain. There then follows between two and four helices that form the rest of the domain (Sheehan *et al.*, 2001). A diagram of the monomer structures of the different GST classes is shown in figure 1.17.
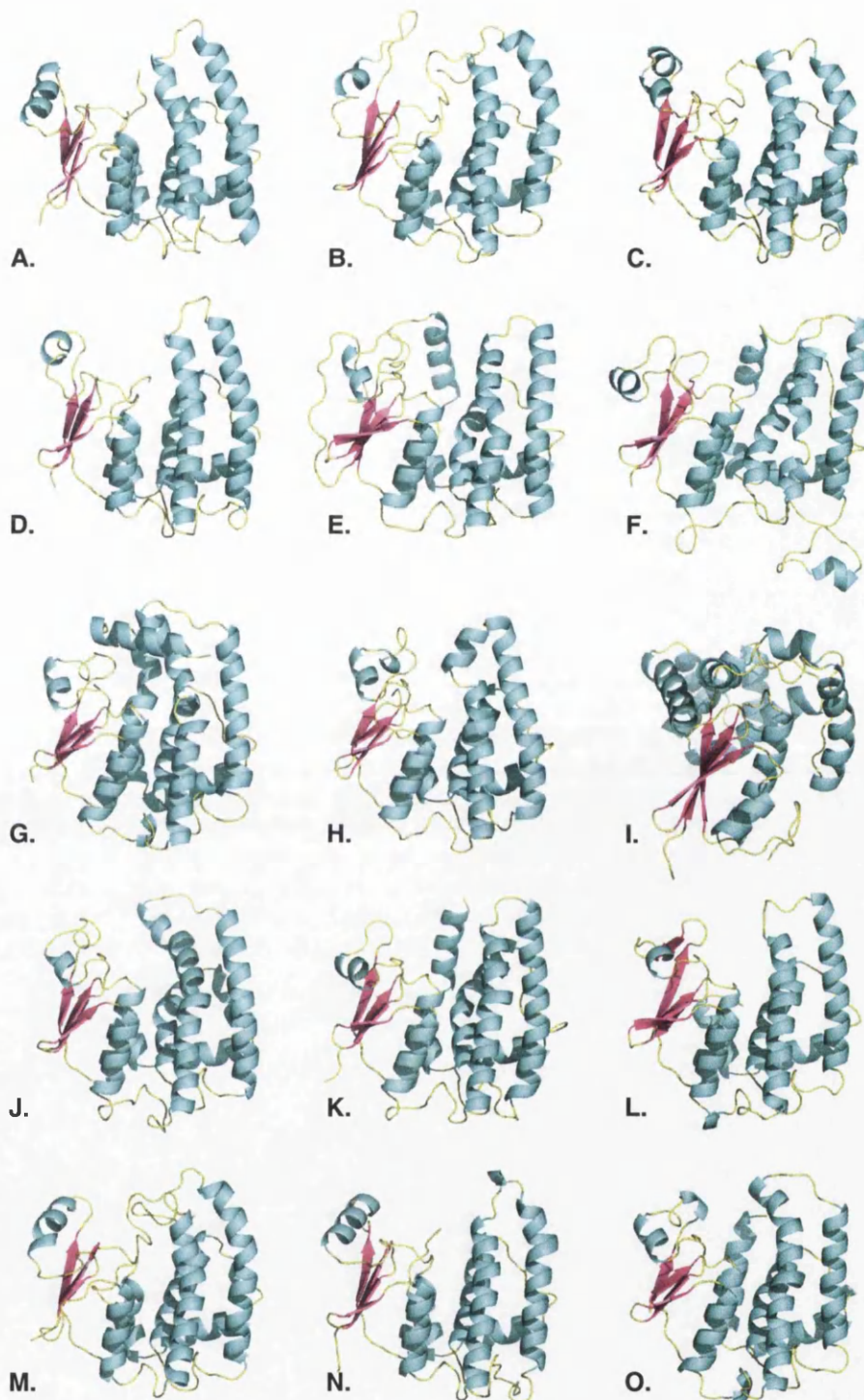
**Figure 1.17.** Comparison of the different classes of GSTs. Structures were superimposed onto the structure of the Tau class enzyme *Ta*GSTU4-4. The structures used are **A.** Alpha (*H. sapiens*; 1GSD; Cameron *et al.*, 1995) **B.** Mu (*R. rattus*; 4GST; Ji *et al.*, 1993) **C.** Pi (*H. sapiens*; 1GSS; Reinemer *et al.*, 1992) **D.** Sigma (*Ommastrephes sloani*; 2GSQ; Ji *et al.*, 1996) **E.** Omega (*H. sapiens*; 1EEM; Board *et al.*, 2000) **F.** CLIC (*H. sapiens*; 1K0M; Harrop *et al.*, 2001) **G.** Theta (*H. sapiens*; 1LJR; Rossjohn *et al.*, 1998b) **H.** Zeta (*H. sapiens*; 1FW1; Polekhina *et al.*, 2001) **I.** Kappa (*R. norvegicus*; 1R4W; Ladner *et al.*, 2004) **J.** Phi (*A. thaliana*; 1GNW; Reinemer *et al.*, 1996) **K.** Tau (*T. aestivum*; 1GWC; Thom *et al.*, 2002) **L.** Beta (*E. coli*; 1A0F; Nishida *et al.*, 1998) **M.** *S. Japonica* (*Schistosoma japonica*; 1GTA; McTigue *et al.*, 1995) **N.** *P. falciparum* (*Plasmodium falciparum*; 1OKT; Fritz-Wolf *et al.*, 2003) **O.** Delta (*Anopheles dirus*; 1JLV; Oakley *et al.*, 2001). Image generated with PyMOL (Delano, 2002).

## 1.6    Functions of Glutathione *S*-Transferases

### 1.6.1  Detoxification of Xenobiotics

The ability of plant GSTs to detoxify xenobiotics has been discussed in section 1.3.2 so this section will briefly look at the importance of human GSTs in detoxification. In humans several GSTs have been identified in detoxifying carcinogens from smoking, and thus protecting against lung cancer. Mu and Pi class GSTs can detoxify polycyclic aromatic hydrocarbons (Hayes & Pulford, 1995), and Theta class GSTs can detoxify smaller compounds such as ethylene oxide and monohalomethane (Landi, 2000). Polymorphisms in all these genes have been identified and null genotypes for Mu and Theta GSTs are present in 50% and 20% of Caucasian population, respectively (Garte *et al.*, 2001). Two mutations of the Pi class GST have also been identified. These are an isoleucine to valine mutation at position 105 and an alanine to valine mutation at position 114 (Ali-Osman *et al.*, 1997).

All three polymorphisms can result in an increased risk of lung cancers in both smokers and non-smokers. The Mu null-genotype alone caries only a small increased risk of lung cancer with an overall risk of 1.17 (Benhamou *et al.*, 2002). Mutations in the Pi class GST also did not significantly increase the risk of cancer. Someone with a Theta null-genotype has a relative risk of getting lung cancer of 1.60 but this increases to 5.02 for squamous cell carcinoma. Combinations of these polymorphisms will also increase the chance of lung cancer. A Theta and Mu null-genotype has a relative risk of developing cancer of 2.81, and a Theta null genotype with homozygous Pi mutations has a relative risk of 6.11 (Sørensen *et al.*, 2004).

However, the presence of these enzymes can also decrease the chance of survival from lung cancer. In patients with nonsmall cell lung carcinoma who express Pi class GSTs the response to chemotherapy with cisplatin is much lower (Bai *et al.*, 1996; Arai *et al.*, 2000). The chemotherapy agent taxol has also been shown to increase the level of expression of both Theta and Pi class GSTs (Mathieu *et al.*, 2004).

### 1.6.2 Glutathione Dependent Peroxidase Activity

Under oxidative stress a range of GSTs can be up-regulated in plant cells and can act as glutathione dependent peroxidases (GST-GPOX; Board *et al.*, 1997). During oxidative stress, oxygen radicals are produced and can generate cytosolic hydroperoxides. These can be detoxified by GST-GPOXs to less toxic monohydroxy alcohols (Edwards *et al.*, 2000). This occurs in two steps, firstly the attack by glutathione on the hydroperoxide to produce the monohydroxy alcohol, and then the spontaneous reaction of the sulfenic acid of GSH with another molecule of GSH to produce oxidized glutathione (GSSG; figure 1.18; Habig, 1983).
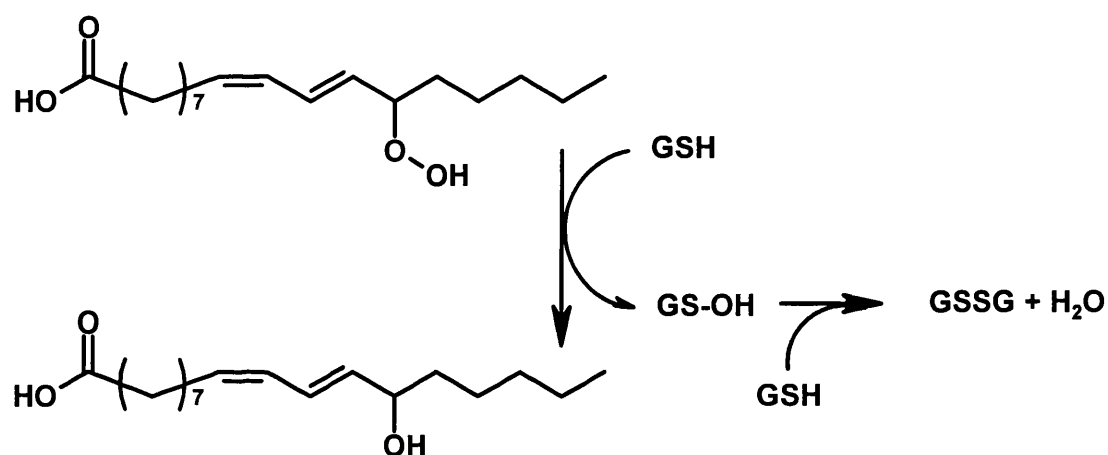
**Figure 1.18.** Reduction of 13-hydroperoxylinoleic acid to 13-hydroxylinoleic acid by GST-GPOX. The resultant sulfenic acid spontaneously reacts with a molecule of GSH to produce GSSG. Image generated with ISIS/Draw (MDL Information Systems, Inc).

### 1.6.3 Aromatic Amino Acid Degradation

The various steps involved in the catabolism of phenylalanine and tyrosine were identified in the 1950s (figure 1.19; Knox, 1955) but it was not until 1998 that the enzyme responsible for the *cis-trans* isomerization of maleylacetoacetate (MAA) to fumarylacatoacetate (FAA) was identified and shown to be a Zeta class GST (Fernández-Cañón & Peñalva, 1998). This penultimate step in aromatic amino acid degradation involves the addition of GSH to the *cis* double bond of MAA. This bond can then rotate to the *trans* conformation before GSH is eliminated to produce FAA (Edwards & Knox, 1956). Zeta class GSTs from both mammals (Fernández-Cañón & Peñalva,

1998) and plants (Dixon *et al.*, 2000) have also been shown to detoxify dichloroacetic acid (DCA), a toxic contaminant in drinking water.
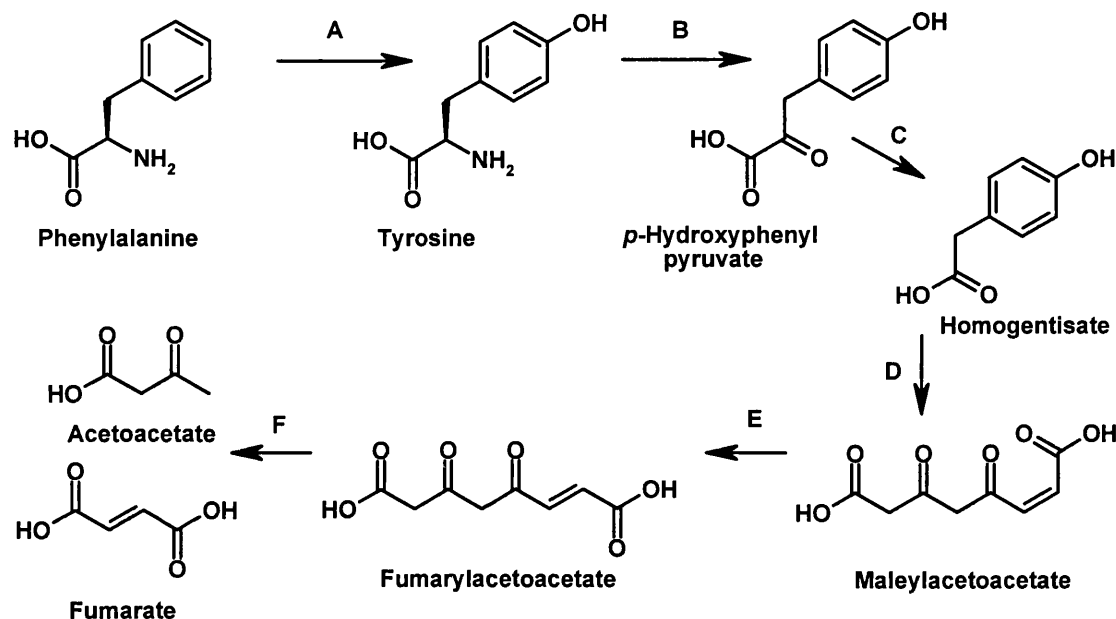
Figure 1.19. Aromatic amino acid degradation pathway. Enzymes involved are **A.** Phenylalanine Hydroxylase (PAH) **B.** Tyrosine Aminotransferase (TAT) **C.** *para*-Hydroxyphenylpyruvate Dioxygenase (HPPD) **D.** Homogentisate 1,2-Dioxygenase (HGO) **E.** Maleylacetoacetate Isomerase (MAAI) **F.** Fumarylacetoacetate Hydrolase (FAH). Image generated with ISIS/Draw (MDL Information Systems, Inc).

## 1.6.4 Anthocyanin Sequestration

The anthocyanin biosynthesis pathway generates flavanoid pigments that provide bright colours to most fruits and flowers. Synthesis of the pigment occurs in the cytoplasm but if not transported to the vacuole they will oxidize and polymerize and can cause cell death (figure 1.20; Marrs, 1996). Proteins from maize, petunia (*Petunia hybrida*) and carnation, necessary for localization of the anthocyanins into the cell vacuole have been identified as GSTs (Marrs *et al.*, 1995; Alfenito *et al.*, 1998; Larsen *et al.*, 2003).

It was initially thought that these enzymes were forming GSH conjugates with the anthocyanins (Marrs *et al.*, 1995; Alfenito *et al.*, 1998) but it has since been shown that no such conjugates are formed *in vivo* (Mueller *et al.*, 2000). The function of these GSTs seems to be binding and escort of the anthocyanins to the vacuole. Since flavanoids are cytotoxic and genotoxic it

appears that they require a ligandin protein to protect the cell during their synthesis (Ahmed *et al.*, 1994). It has been suggested that these enzymes present the anthocyanin to the GS-X pump where it is co-transported with GSH (Dixon *et al.*, 1998; Edwards *et al.*, 2000).
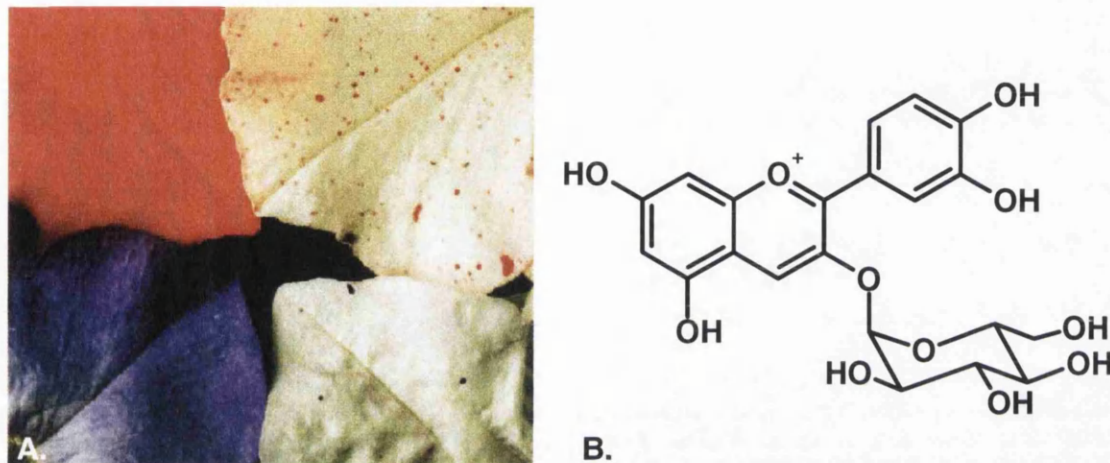


**Figure 1.20.** Anthocyanin sequestration in the vacuole. **A.** Two petunia cultivars (left) and their respective anthocyanin negative mutants (right). Image taken from Alfenito *et al.*, 1998. **B.** The structure of cyanidin-3-*O*-glucoside, a major anthocyanin in petunia flowers. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Despite that fact that the enzymes from maize (*Zm*GSTU4-4) and petunia (An9) perform the same role *in vivo* they are different classes of GST and share only 7% sequence identity. This suggests that the plants have developed this ability independently and that different GSTs can evolve to perform the same job. In fact *Zm*GSTU4-4 and An9 can functionally complement each other (Alfenito *et al.*, 1998) and in anthocyanin negative carnation tissue (Larsen *et al.*, 2003).

### 1.6.5 Porphyrin Binding and Transport

The biosynthesis of porphyrins occurs in several different locations in the plant cell and the various molecules along the pathway must be transported to the different organelles (figure 1.21; Rodermel, 2001). The tetrapyrrole precursor protoporphyrinogen (Pgen) is synthesized in the chloroplasts (Beale, 1999) and then exported to the mitochondrion and endoplasmic reticulum. This is easily oxidized to protoporphyrin IX

produce reactive oxygen species (ROS) causing oxidative damage. In fact this is the mode of action of peroxidizing herbicides (Wakabayashi & Böger, 1999). Since plant GSTs have long been recognized as bearing resemblance to mammalian porphyrin-binding proteins (Mozer *et al.*, 1983) and have the ability to prevent oxidative damage (Marrs, 1996), this indicates that they may be involved in the process.
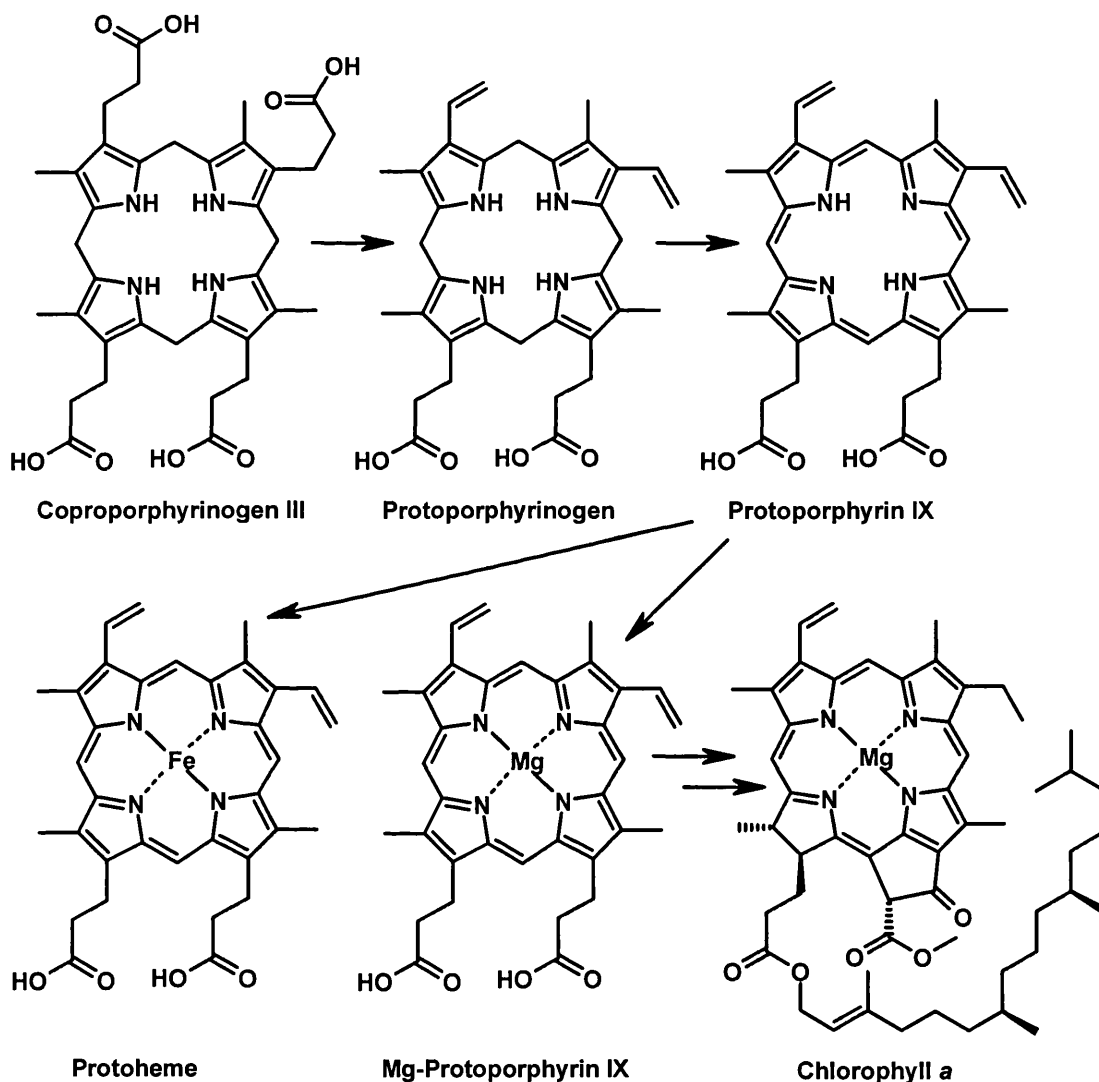
**Figure 1.21.** Biosynthesis of porphyrins. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Several GSTs from maize have been identified as being able to bind porphyrins. *Zm*GSTF1-1, *Zm*GSTF2-2 and *Zm*GSTF3-3 can all bind Pgen, PPIX, mesoporphyrin, coproporphyrin, uroporphyrin and Mg-protoporphyrin

(Mg-PPIX), but only ZmGSTF3-3 shows substantial ability to protect Pgen from autoxidation (Lederer & Böger, 2003).

In the chloroplast and mitochondrion PPIX is converted to heme (Fe-PPIX), which is required as a cofactor by many enzymes, and is steadily released into the cytosol (Thomas & Weinstein, 1990). In the cytosol hemin is easily degraded in the presence of GSH and peroxide, but when ZmGSTF2-2 is present the porphyrin is highly protected against damage. It has been suggested that the endogenous role of GSTs may in fact be as binding proteins rather than their ability to conjugate GSH to substrates (Walbot et al., 2000).

Another incidence of porphyrin binding by a GST is seen in the malarial parasite Plasmodium falciparum. PfGST can represent up to 10% of all cellular protein in the parasite and has been suggested as an alternative hemin detoxification pathway (Liebau et al., 2005). The main route of hemin detoxification is incorporation into the insoluble, dimeric haemozoin in the food vacuole of the parasite (Egan et al., 2002). However, it has been shown that Fe-PPIX can leak into the cytoplasm where it is toxic to the organism (Ginsburg et al., 1998). PfGST can bind Fe-PPIX and is thought to sequester the porphyrin and prevent it from damaging the parasite (Harwaldt et al., 2002). Chloroquine is an important antimalarial drug that inhibits heme polymerization, leading to free Fe-PPIX in the cytosol (Famin et al., 1999). In chloroquine resistant P. falciparum the levels of PfGST are significantly increased and it has been suggested that inhibitors of PfGST could work synergistically with chloroquine making it a good potential drug target (Srivastava et al., 1999).

### 1.6.6 Auxins

Auxins are a group of plant hormones that do not necessarily have any structural similarity but all possess the ability to induce shoot elongation (Tukey et al., 1954). A range of GSTs in plants can be induced by auxins

(Droog et al., 1993; 1995; Takahashi et al., 1995) and several auxin-binding proteins have been identified as GSTs (Jones, 1994). It has been shown that different auxins bind to different sites on the GST (Bilang & Sturm, 1995). The auxins Indole-3-acetic acid (IAA) and α-naphthoxyacetic acid (α-NAA) are seen to bind in a non-catalytic site, and the auxin mimics 2,4-D and 2,4,5-T bind to a catalytic site (figure 1.22).
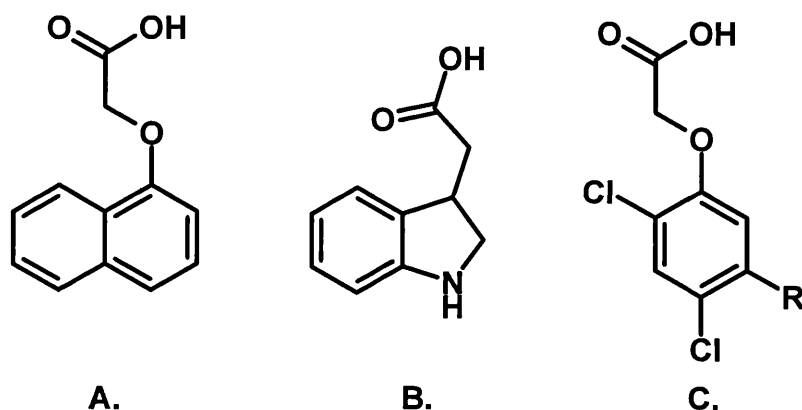


Figure 1.22. Plant auxins and mimics. A. α-naphthoxyacetic acid. B. Indole-3-acetic acid. C. Auxin mimics 2,4-D (R = H) and 2,4,5-T (R = Cl). Image generated with ISIS/Draw (MDL Information Systems, Inc).

It is thought that GSTs bind auxins without conjugating them to GSH. This is possibly for temporary storage of IAA and regulation of hormonal levels in the cell, or possibly for intracellular transport (Bilang & Sturm, 1995). Another possibility is that GSTs protect the auxins from oxidative damage by peroxides.

## 1.7   Project Aims

The aims of this study were to characterize a number of GSTs from various plant species and see how they interact with different substrates. The main points I wished to investigate were –

- The structure of Tau class GSTs and their mode of binding of herbicides.
- The structure of anthocyanin sequestering GSTs and their binding of anthocyanins and flavanoids.
- The structure of *Zm*GSTU4-4 and its mechanism of binding of porphyrins.
- Investigate induced conformational change of *Zm*GSTF1 upon ligand binding.
- Structural characterization of plant Lambda class GSTs.

# Chapter 2    Materials and Methods

## 2.1    Clones

The constructs for ZmGSTF1, ZmGSTU1 and AtGSTL2 were provided by Dr David Dixon (Dr Robert Edwards' group, University of Durham, UK) in vector pET-24 (Novagen, UK), while the constructs for OsGSTU1, OsGSTU4, OsGSTU18 and OsGSTF5 were provided by Dr David Cole (Aventis CropScience, UK) in vector pET-11 (Novagen, UK). Lukas Müeller (Dr Virginia Walbot's group, Stanford University, USA) provided the constructs for Petunia hybrida An9 (pAQ1) and ZmGSTU4 (pQBz2c). Details of the production and purification of ZmGSTF1, AN9 and AtGSTL2 have been previously published (Dixon et al., 1997a, Alfenito et al., 1998, Dixon et al., 2002a). The construct for TaGSTL1 was provided by Dr. Frederica Theodoulou (Crop Performance and Improvement Division, Rothamsted Research, UK) in pET16b (Novagen, UK).

## 2.2    Preparation of Competent *Escherichia coli* BL21 (DE3)

A 5ml culture of E. coli strain BL21 (DE3) was grown overnight at 37°C with shaking. This culture was added to 1L of Miller's Luria Broth (LB, Sigma, UK) and incubated at 37°C with shaking until the optical density at 600nm ($OD_{600}$) reached between 0.6 and 1.0. The culture was then centrifuged at 5,000g for thirty minutes at 4°C and the cell pellet was resuspended in 400ml ice-cold competence solution (100mM $CaCl_2$, 70mM $MnCl_2$, 40mM sodium acetate buffer pH 5.5). The cells were then centrifuged at 5,000g for thirty minutes at 4°C and the cell pellet was resuspended in 5ml ice-cold 85% competence solution, 15% glycerol. The competent cells were divided into 100μl aliquots in 1.5ml eppendorf tubes and snap frozen by dropping into dry-ice/ethanol.

## 2.3   Expression

To permit large quantities of protein to be expressed the clones for the GST enzymes had to be transfected into suitable over-expression strains of *E. coli*. The clones for An9 and *Zm*GSTU4 had previously been transfected into strains M15pREP4 and JM105 respectively.

### 2.3.1  Transformation Protocol

An aliquot of competent *E. coli* BL21 (DE3) was defrosted on ice for thirty minutes. 1μl of plasmid DNA was added to the cells with gentle mixing. The cells were left on ice for a further thirty minutes before heat shocking at 42°C for thirty seconds. 1ml of LB was added and the cells were incubated at 37°C with shaking for two hours. The cells were then plated out onto LB agar plates containing 100μg/ml ampicillin (50μg/ml kanamycin for *At*GSTL2, 100μg/ml carbenicillin for *Zm*GSTU4 and 50μg/ml kanamycin and 100μg/ml carbenicillin for AN9, All antibiotics from Melford, UK) and incubated overnight at 37°C. One colony from the plate was added to 100ml LB containing the appropriate antibiotics and grown overnight at 37°C with shaking. 0.75ml was removed and mixed with 0.75ml glycerol in a 2ml cryogenic vial (Sigma) and stored at -80°C

### 2.3.2  Over-Expression of Glutathione *S*-Transferase Enzymes - *Zm*GSTF1-1 and *Zm*GSTU1-1.

Cells from a glycerol stock of the transformed *E. coli* were streaked out onto an agar plate containing 100μg/ml ampicillin and grown overnight at 37°C. One colony was removed and added to 100ml LB containing 100μg/ml ampicillin and incubated overnight at 37°C with shaking. This was used to inoculate two litres of LB containing 100μg/ml ampicillin. These were incubated at 37°C with shaking until the $OD_{600}$ had reached 0.6. Isopropyl-$\beta$-D-thiogalactopyranoside (IPTG, Melford, UK) was added to a final concentration of 0.3mM. This allowed over-expression of the protein to occur as mRNA transcription of the clone is under the control of the *lac* operon, and IPTG is a non-hydrolysable analogue of lactose. The cultures were further

incubated at 37°C for three hours. The cells were then centrifuged at 3,000g for forty-five minutes at 4°C. The cell pellet was resuspended in 50ml 10mM potassium phosphate buffer pH7.0 and stored at -80°C.

### 2.3.3 Over-Expression of OsGSTU1-1, OsGSTU4-4, OsGSTU18-18 and OsGSTF5-5.

Cells from a glycerol stock of the transformed E. coli were streaked out onto an agar plate containing 100μg/ml ampicillin and grown overnight at 37°C. One colony was removed and added to 100ml LB containing 100μg/ml ampicillin and incubated overnight at 37°C with shaking. This was used to inoculate two litres of LB containing 100μg/ml ampicillin. These were incubated at 37°C with shaking until the $OD_{600}$ had reached 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.3mM, the temperature was reduced to 20°C and the cultures were incubated overnight. The cells were then centrifuged at 3,000g for forty-five minutes at 4°C. The cell pellet was resuspended in 50ml 20mM Tris HCl pH 7.8 and stored at -80°C.

### 2.3.4 Over-Expression of AtGSTL2 and An9.

Cells from a glycerol stock of the transformed E. coli were streaked out onto an agar plate containing 50μg/ml kanamycin (plus 100μg/ml carbenicillin for An9) and grown overnight at 37°C. One colony was removed and added to 100ml LB containing the appropriate antibiotics and incubated overnight at 37°C with shaking. This was used to inoculate two litres of LB, which were incubated at 37°C with shaking until the $OD_{600}$ had reached 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.3mM, the temperature was reduced to 30°C and the cultures were incubated for four hours. The cells were then centrifuged at 3,000g for forty-five minutes at 4°C. The cell pellet was resuspended in 50ml 20mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol and stored at -80°C.

### 2.3.5  Over-Expression of *Zm*GSTU4-4 and *Ta*GSTL1.

Cells from a glycerol stock of the transformed *E. coli* were streaked out onto an agar plate containing 100µg/ml carbenicillin and grown overnight at 37°C. One colony was removed and added to 100ml Terrific Broth (TB; Tartof & Hobbs, 1987) containing the appropriate 100µg/ml carbenicillin and incubated overnight at 37°C with shaking. This was used to inoculate two litres of TB, which were incubated at 37°C with shaking until the $OD_{600}$ had reached 0.6. Protein expression was induced by adding IPTG to a final concentration of 0.3mM, the temperature was reduced to 20°C and the cultures were incubated overnight. The cells were then centrifuged at 3,000g for forty-five minutes at 4°C. The cell pellet was resuspended in 100ml 10mM imidazole, 20mM Tris HCl pH 7.5, 300mM NaCl, 10% glycerol and stored at -80°C.

## 2.4  Purification

### 2.4.1  Sonication and Centrifugation

Before purification the cells were defrosted and lysed using a Philip Harris, US 200 sonicator at 50% power (ten runs of thirty seconds with one minute rest between runs to prevent the lysate overheating). The lysate was centrifuged at 20,000g for fifteen minutes at 4°C and the supernatant was decanted and used in later purification steps. For the proteins that were over expressed in TB, 1mg ml$^{-1}$ hen egg white lysozyme (Sigma, UK) was added to the cells one hour before Sonication to help lyse the cells as sonication alone left many cells unbroken.

### 2.4.2  Purification of *Zm*GSTF1-1

The purification protocol for *Zm*GSTF1-1 had been previously determined (Dixon et al., 1997a) using an Amicon® Orange A DyeMatrex gel column (Millipore, UK). A 5ml Orange A DyeMatrex column was equilibrated in five column volumes 10mM potassium phosphate buffer pH 6.0, before the crude protein solution was loaded. The column was washed with five column volumes 10mM potassium phosphate buffer pH 6.0 followed by five column volumes 50mM potassium phosphate buffer pH 7.0 to remove any non-

specifically bound proteins. *Zm*GSTF1-1 was eluted from the column using five column volumes 50mM potassium phosphate buffer pH 7.0, 5mM glutathione (Sigma, UK) before being dialysed overnight into 10mM potassium phosphate buffer pH 6.0.

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

### 2.4.3 Purification of *Zm*GSTU1-1, *Os*GSTU1-1, *Os*GSTU4-4, *Os*GSTU18-18 and *Os*GSTF5-5

These enzymes were purified by affinity chromatography using a *S*-hexyl-glutathione Sepharose (Amersham, UK) column using a previously determined protocol (Dixon et al., 1997b). A 6ml *S*-hexyl glutathione Sepharose column was equilibrated in five column volumes 20mM Tris HCl pH 7.8 before loading the crude protein suspension. The column was washed in five column volumes 20mM Tris HCl pH 7.8 followed by five column volumes 20mM Tris HCl pH 7.8, 50mM NaCl to remove any non-specifically bound protein. The enzymes were then eluted using 20mM Tris HCl pH 7.8, 5mM glutathione. The column was regenerated using 20mM Tris HCl pH 7.8, 3M NaCl and stored in 20mM Tris HCl pH 7.8, 0.02% sodium azide.

### 2.4.4 Purification of An9 and *At*GSTL2

An9 and *At*GSTL2 were cloned into vectors that incorporated a hexa-histidine tag onto the protein, as they were known to bind poorly to glutathione-based columns. The histidine tag binds strongly to nickel ions thus a nickel charged metal chelate column can be used as an affinity purification step. A 10ml Ni-NTA Superflow column (Qiagen, UK) was equilibrated in 20mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol before loading the crude protein solution. The column was washed with five column volumes 20mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol before being eluted in 250mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol. The column was stored in 20% ethanol to prevent bacterial growth. The buffer was exchanged by

concentrating the protein in an Amicon® pressure cell fitted with a 10kD cut-off ultrafiltration membrane (Millipore, UK) and diluting it with 50ml 20mM Tris HCl pH 7.8, 1mM EDTA, 0.3mM DTT. This was repeated before the protein was concentrated.

### 2.4.5 Purification of *Zm*GSTU4-4 and *Ta*GSTL1

*Zm*GSTU4-4 and *Ta*GSTL1 were also purified using the Ni-NTA Superflow column. The column was equilibrated in 10mM imidazole, 20mM Tris HCl, 300mM NaCl, 10% glycerol before the crude protein solution was loaded. The column was then washed in five column volumes 10mM imidazole, 20mM Tris HCL, 300mM NaCl, 10% glycerol to remove any unbound protein. Purification of the proteins was carried out with a step and gradient on an AKTA FPLC (Amersham, UK). Non-specifically bound proteins were removed using a three column volume gradient from 10mM to 40mM imidazole and holding this concentration for three column volumes. The protein was then eluted using a five column volume gradient to 100mM imidazole and holding this concentration for four column volumes. Any remaining protein was stripped from the column using a three column volume gradient to 250mM imidazole. The protein was dialysed overnight into 20mM Tris HCl pH 7.5.

### 2.4.6 Protein concentration Using Absorbance (280nm)

The concentration of a protein solution can be determined by measuring the absorbance at 280nm. Proteins absorb in this region due to their tyrosine, tryptophan and cysteine content. The specific absorption values for each protein can be determined using the following formula (Gill & von Hippel, 1989).

$$\varepsilon = 1280nY + 5960nW + 120nC$$

**Equation 2.1.** Calculation of the molar extinction coefficient of a protein. The extinction coefficient $\varepsilon$ ($M^{-1}$ $cm^{-1}$) can be calculated for a protein by inserting the Y, W and C with the number of tyrosine, tryptophan and cysteine residues present in the sequence

Using the calculated molar extinction coefficient ($\varepsilon$) and the absorbance value (A) at 280nm, the protein concentration can be calculated using equation 2.2.

$$A = \varepsilon c l$$

**Equation 2.2.** The Beer-Lambert Law. Rearranging the equation allows determination of the protein concentration (c) in Moles where (A) is the absorbance measured at 280nm and (l) is the path length of the spectrophotometer cell in cm.

The calculated extinction coefficient for the proteins used in this project are shown in table 2.1. All absorbance measurements were carried out using a Perkin Elmer (USA) Lambda 40 UV/Visible spectrophotometer and using 400µl quartz cuvettes (Hellma, UK).

| Protein | Extinction Coefficient ($\varepsilon$) | Abs 0.1% |
|---------|---------------------------|----------|
| *Zm*GSTF1-1 | 34520 | 1.451 |
| *Zm*GSTU1-1 | 40330 | 1.606 |
| *Zm*GSTU4-4 | 22310 | 0.833 |
| *Os*GSTU1-1 | 40210 | 1.554 |
| *Os*GSTU4-4 | 34970 | 1.368 |
| *Os*GSTU18-18 | 38690 | 1.460 |
| *Os*GSTF5-5 | 46940 | 1.855 |
| *Ta*GSTL1 | 28140 | 0.989 |
| *At*GSTL2 | 38240 | 1.129 |
| *Ph*An9 | 27550 | 1.007 |

**Table 2.1.** Extinction coefficients ($M^{-1} cm^{-1}$) and absorbance of a 0.1 % ($1mg\ ml^{-1}$) solution of various GSTs. The molar extinction coefficients and Abs 0.1% were calculated using the program ProtParam (Gasteiger *et al.*, 2005).

## 2.5   Electrophoresis

### 2.5.1  SDS-PAGE

SDS-PAGE is a method in which proteins can be separated by their size. This allows easy determination of the presence of the target protein and indicates if the sample is contaminated. Electrophoresis was carried out using an ECell SureLock Electrophoresis Cell with NuPAGE 4-12% Bis-Tris polyacrylamide gels (Invitrogen, UK) and using Low Molecular Weight Markers (Amersham, UK) as a reference.

### 2.5.2 Staining of SDS-PAGE Gels

Gels were stained using a solution of 0.1% Coomassie® Brilliant Blue R-250 (Fluka, Switzerland), 40% ethanol, 10% acetic acid. Staining was carried out by heating the gel in the solution for one minute and shaking for fifteen minutes. De-staining was carried out by four cycles of heating in deionised water for one minute and shaking for fifteen minutes.

## 2.6 Enzyme Assays

In order to understand the specificities of the different enzymes, OsGSTU1-1, OsGSTU4-4, OsGSTF5-5, ZmGSTU4-4 and An9 were assayed against a number of substrates.

### 2.6.1 Enzyme Assays with Chloro-2,4-dinitrobenzene

A cuvette containing 900µl of 0.1M potassium phosphate buffer pH 6.5 and 25µl of 40mM chloro-2,4-dinitrobenzene (CDNB, Sigma, UK) in ethanol was incubated in a water bath at 30°C for five minutes. After this 25µl of the enzyme sample (table 2.2) and 50µl of 100mM GSH NaOH pH7.0 was added and the cuvette transferred to a Unicam Helios alpha spectrophotometer and the change in absorbance at 340nm was measured for thirty seconds. The experiment was repeated with buffer added instead of enzyme to calculate the chemical rate of reaction.

| Protein | Concentration |
|---------|---------------|
| OsGSTU1-1 | 0.052mg ml$^{-1}$ |
| OsGSTU4-4 | 0.036mg ml$^{-1}$ |
| OsGSTF5-5 | 3.167mg ml$^{-1}$ |
| ZmGSTU4-4 | 1.184mg ml$^{-1}$ |
| PhAn9 | 3.168mg ml$^{-1}$ |

**Table 2.2.** Concentration of enzymes used in CDNB assays.

### 2.6.2 Enzyme Assays with Herbicides and Safeners

20µl of 10mM herbicide solution was added to an eppendorf tube containing 20µl of the protein solution diluted in 140µl of either 100mM potassium

phosphate buffer pH 6.8 (for fenclorim [Riedel-de Haën, Germany], atrazine, metolachlor, alachlor, and chlorimuron ethyl [Chem Service, Inc, USA]) or 100mM Tris HCl pH 9.5 (for fenclorim, fluorodifen, fenoxaprop-P, acifluorfen & fomesafen [Riedel-de Haën, Germany]). These were incubated at 37°C for five minutes before the reaction was started by adding 20µl of 100mM GSH NaOH pH 7.0. This was further incubated at 37°C for one hour. After this time the reaction was stopped by addition of 10µl of 3.6M HCl. The samples were then chilled at -20°C for ten minutes before being centrifuged at 13,000g for ten minutes.

The contents of the reaction mixture were analysed using an HPLC system consisting of a Beckman System Gold 125P Solvent Module fitted with a Phenomex Phenosphere ODS2 5µm reverse phase column, connected to a Beckman System Gold 166 Detector. The system was run at a rate of 1ml min$^{-1}$ with an increasing gradient of acetonitrile in methanol.

## 2.7   Crystallization

The sitting-drop vapour diffusion method of crystallization is a technique by which the supersaturation of a protein solution can be increased by removing water from the drop (figure 2.1). In this technique a reservoir is filled with a precipitant solution, 1µl of this solution is mixed with 1µl of protein solution in the central well, and the well is sealed with a cover slip. The drop in the central well is now at a lower concentration than the reservoir solution and therefore water will evaporate from it at a higher rate than the precipitant solution and bring the protein solution to a state of supersaturation from which crystal nucleation can occur.
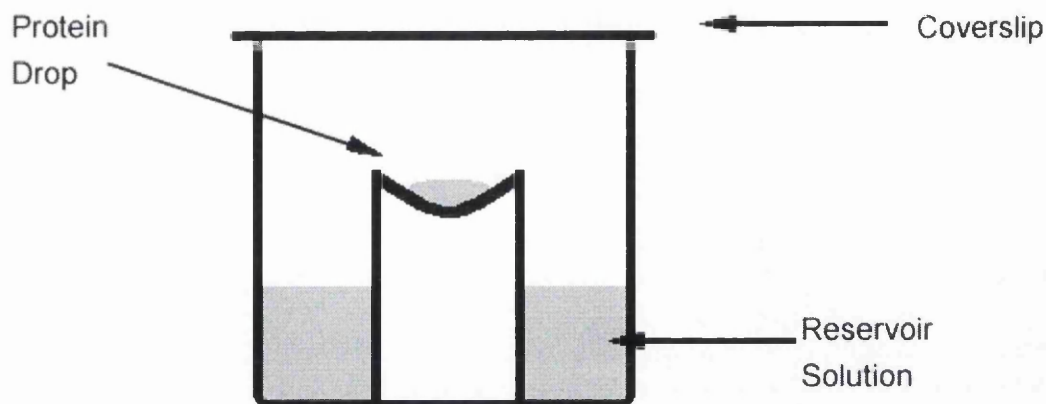
**Figure 2.1.** Sitting-drop vapour diffusion crystallization set-up.

### 2.7.1  Crystallization of Enzymes with Glutathione

GSH was added to An9, ZmGSTU4-4, AtGSTL2 and TaGSTL1 to a final concentration of 5mM. These enzymes, and the four rice enzymes were concentrated using an Amicon® pressure cell fitted with a 10kD cut-off ultrafiltration membrane (Millipore, UK; Table 2.3).

| Protein | Concentration |
|---------|---------------|
| OsGSTU1-1 | 12mg ml$^{-1}$ |
| OsGSTU4-4 | 9mg ml$^{-1}$ |
| OsGSTU18-18 | 8mg ml$^{-1}$ |
| OsGSTF5-5 | 16.5mg ml$^{-1}$ |
| An9 | 18mg ml$^{-1}$ |
| ZmGSTU4-4 | 3mg ml$^{-1}$ |
| AtGSTL2 | 6mg ml$^{-1}$ |
| TaGSTL1 | 5mg ml$^{-1}$ |

**Table 2.3.** Concentration of proteins used for crystallization experiments.

All crystallization experiments were carried out using twenty-four well Cryschem sitting drop trays (Hampton Research, USA). Sparse matrix crystallization trials were carried out using in-house screens – the M-screen, N-screen, I-screen (Westwater, 2002) and CAS screen (developed during this project, appendix A) – Hampton Research (USA) screens (Crystal screens I & II and PEG/Ion screen) and Emerald Biostructures (USA) screens (Wizard I & II and Cryo I & II).

### 2.7.2  Co-Crystallization of An9 and *Zm*GSTU4-4 with Anthocyans and Flavanoids

To a 500μl sample of An9 at 18mg ml$^{-1}$ (678μM) Cyanidin-3-O-glucoside chloride (Kuromanin, Extrasynthese, France) was added to a final concentration of 3.4mM (0.8mg). To 1ml of *Zm*GSTU4-4 at 3mg ml$^{-1}$ (117μM) Quercetin-3-O-glucopyranoside (Extrasynthese, France) was added to a final concentration of 584 μM (0.3mg). These solutions were used to set up sparse matrix crystallization trials using the screens listed in section 2.7.1.

### 2.7.3  Co-Crystallization of *Zm*GSTU1-1 with Hemin

*Zm*GSTU1-1 was concentrated to 5mg ml$^{-1}$ using an Amicon® pressure cell fitted with a 10kD cut-off ultrafiltration membrane (Millipore, UK). A 10mM solution of pig blood hemin was produced by dissolving 65.2mg hemin (Acros Organics, Belgium) in 10ml of a 1% ammonium hydroxide solution. Hemin solution was added to the *Zm*GSTU1-1 solution to a final concentration of 1mM. A PD-10 desalting column (Amersham, UK) was equilibrated in 25ml of 20mM Tris HCl pH 7.8 before loading 2.5ml of the protein. This was eluted in 3.5ml of 20mM Tris HCl pH 7.8 and then concentrated to 12mg ml$^{-1}$ using a 4ml, 10Kda cut-off Amicon® Ultra-4 centrifugal filter (Millipore, UK). This was used to set up sparse matrix crystallization trials using the screens listed in section 2.7.1.

# Chapter 3      Purification,      Crystallization      and Preliminary Characterization

## 3.1    *Os*GSTU1-1

*Os*GSTU1-1 is a Tau class GST that has been shown to be important in the detoxification of herbicides in rice. This enzyme is expressed at the highest level of the sixty-one GST genes in rice, making up 15% of all ESTs, and is the dominant protein expressed in callus tissue. The gene for *Os*GSTU1 is also unusual for being one of the few Tau GST genes in rice not located in the large cluster of genes on chromosome ten (Soranzo *et al.*, 2004). The enzyme was purified by *S*-hexyl GSH affinity as described in section 2.4.3 (figure 3.1). *Os*GSTU1-1 with GSH was screened using sparse matrix crystallization conditions. Crystals of different morphologies were grown in several conditions, including Wizard II condition three (20% PEG 8K, 0.2M MgCl$_2$, 0.1M Tris HCl pH 8.5), I screen condition six (30% PEG 4K, 0.2M CaCl$_2$, 0.1M HEPES NaOH pH 7.5), M screen conditions 110 (30% PEG 4K, 0.1M Tris HCl pH 8.5) and 114 (20% PEG 8K, 0.2M CaCl$_2$, 0.1M MOPS NaOH pH 6.5). After optimization, diffraction quality crystals were grown in 21.25% PEG 4K, 0.2M MgCl$_2$, 0.1M HEPES NaOH pH 7.74.
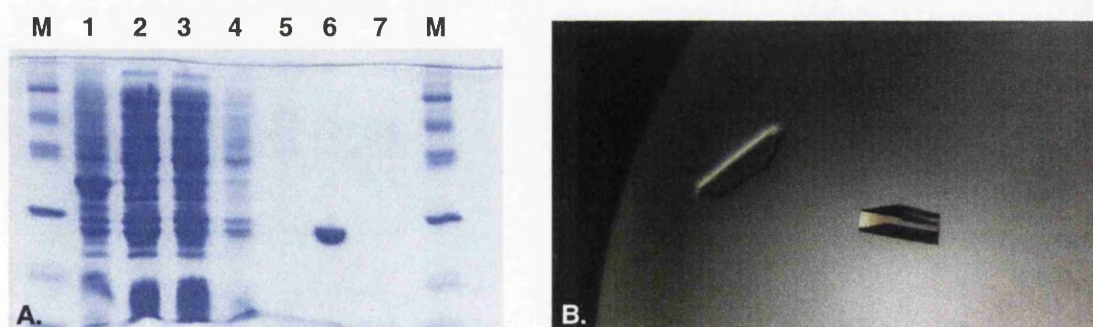


**Figure 3.1.** Purification and crystallization of *Os*GSTU1-1. **A.** Purification. **(M)** Low Molecular Weight Markers (Amersham, UK). **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Though. **(4)** Wash in 20mM Tris HCl pH 7.8. **(5)** Wash in 20mM Tris HCl pH 7.8, 50mM NaCl. **(6)** Elution in 20mM Tris HCl pH7.8, 5mM GSH. **(7)** Wash in 20mM Tris HCl 7.8, 3M NaCl. **B.** Crystals of *Os*GSTU1-1 with GSH.

These crystals were taken to Daresbury Synchrotron Radiation Source (SRS) where they were shown to diffract to 1.95Å. Autoindexing of the diffraction pattern with Denzo (Otwinowski & Minor, 1997) showed that the protein crystals belonged to a primitive orthorhombic space group with unit cell dimensions a = 76.829Å, b = 91.129Å, c = 165.034Å, α = β = γ = 90°. An example of diffraction is shown in figure 3.2.
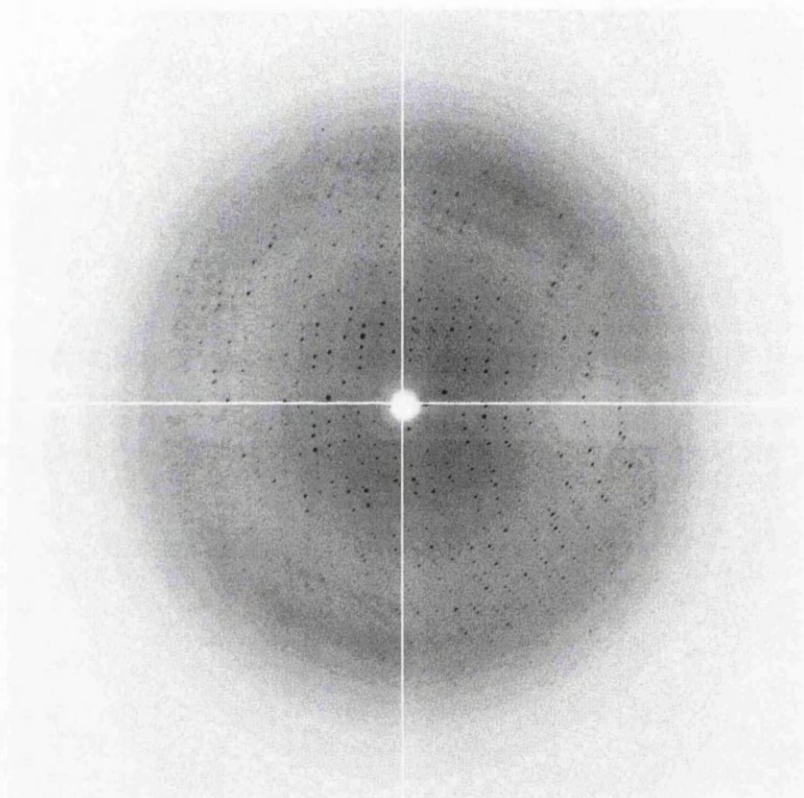


**Figure 3.2.** A representative image of the diffraction of *Os*GSTU1-1 with GSH crystals.

## 3.2    *Os*GSTU4-4

*Os*GSTU4-4 has been found to be highly expressed in rice, totalling 4.7% of all ESTs (Soranzo *et al.*, 2004). Under stress conditions with PEG, salicylic acid, NaCl, hydrogen peroxide, or heavy metals OsGSTU4-4 is highly expressed in the roots of rice within a very short time (Moons, 2003). The enzyme was purified by *S*-hexyl GSH affinity as described in section 2.4.3 (figure 3.3). Small, hexagonal crystals of *Os*GSTU4-4 with GSH were grown in several conditions including I screen condition forty-one (2.4M $K_2HPO_4$, 0.1M

NaCl, 0.1M MES NaOH pH 5.0), Crystal screen I condition forty-six (18% PEG 8K, 0.2M $Ca(OAc)_2$, 0.1M sodium cacodylate pH 6.5), and Wizard I condition forty-four (30% PEG 400, 0.2M $Ca(OAc)_2$, 0.1M NaOAc pH 4.5). After optimization, diffraction quality crystals were grown in 30% PEG 400, 0.2M $CaCl_2$, 0.1M HEPES NaOH pH 6.87.
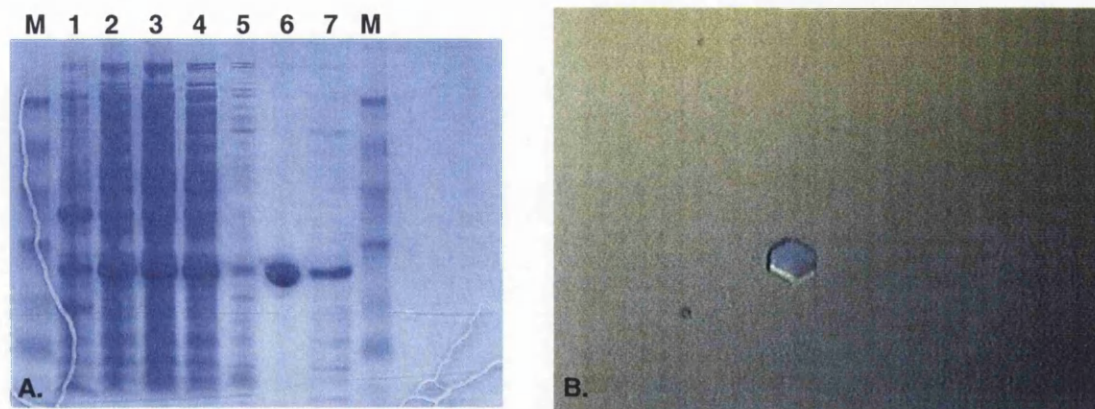


**Figure 3.3.** Purification and crystallization of *Os*GSTU4-4. **A.** Purification. **(M)** Low Molecular Weight Markers (Amersham, UK). **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Though. **(4)** Wash in 20mM Tris HCl pH 7.8. **(5)** Wash in 20mM Tris HCl pH 7.8, 50mM NaCl. **(6)** Elution in 20mM Tris HCl pH7.8, 5mM GSH. **(7)** Wash in 20mM Tris HCl 7.8, 3M NaCl. **B.** Crystal of *Os*GSTU4-4 with GSH.

These crystals were taken to Daresbury SRS where they were shown to diffract to 2.3Å. Autoindexing of the diffraction pattern using Denzo (Otwinowski & Minor, 1997) showed that the protein crystals belonged to a primitive tetragonal space group with unit cell dimensions $a = b = 56.093$Å, $c = 165.034$Å, $\alpha = \beta = \gamma = 90°$. An example of the diffraction is shown in figure 3.4.
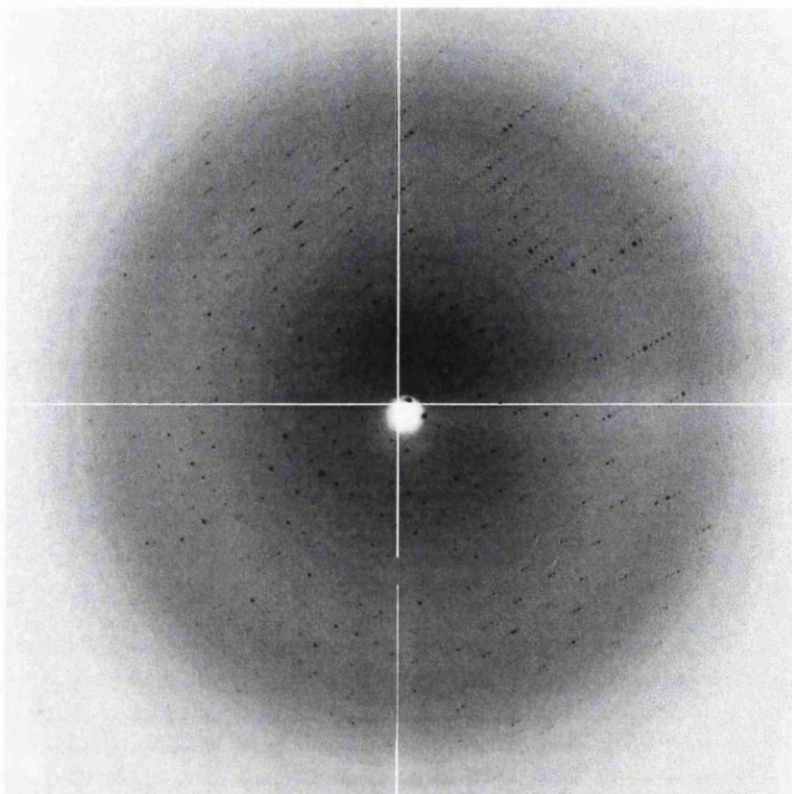
**Figure 3.4.** A representative diffraction image from crystals of *Os*GSTU4-4 with GSH.

## 3.3    *Os*GSTU18-18

*Os*GSTU18-18 is also a GST that is highly expressed in rice, totalling 5% of all ESTs and has been found to be expressed in response to infection (Soranzo *et al.*, 2004). The enzyme was purified by *S*-hexyl GSH affinity as described in section 2.4.3 (figure 3.5). Despite extensive screening of the protein, with and without GSH, no crystals of *Os*GSTU18-18 were grown during this project.
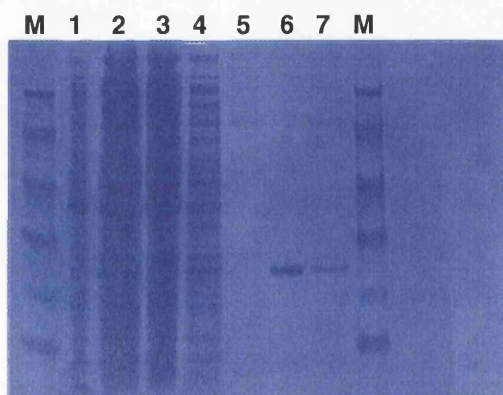


**Figure 3.5.** Purification of *Os*GSTU18-18. **(M)** Low Molecular Weight Markers (Amersham, UK). **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Though. **(4)** Wash in 20mM Tris HCl pH 7.8. **(5)** Wash in 20mM Tris HCl pH 7.8, 50mM NaCl. **(6)** Elution in 20mM Tris HCl pH7.8, 5mM GSH. **(7)** Wash in 20mM Tris HCl 7.8, 3M NaCl.

47

## 3.4    *Os*GSTF5-5

*Os*GSTF5-5, a Phi class GST active in herbicide detoxification, is the second highest expressed Phi class GST in rice and is expressed extensively throughout vegetative tissues (Soranzo *et al.*, 2004). The enzyme was purified by *S*-hexyl GSH affinity as described in section 2.4.3 (figure 3.6). Thin, plate-like crystals of *Os*GSTF5-5 with GSH were grown in 0.9M sodium tartrate, 0.1M NaOAc pH 4.6 (figure 3.6). No diffraction was ever seen from these crystals, either in-house or at Daresbury SRS and extensive screening failed to yield crystals in any different conditions.
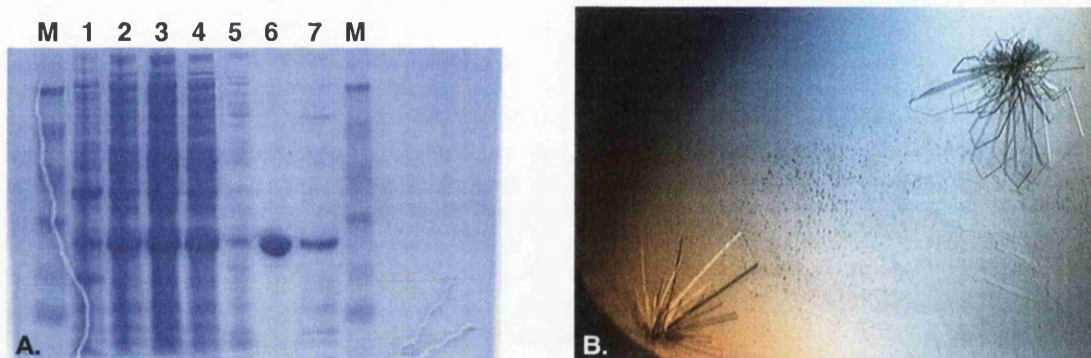


**Figure 3.6.** Purification and crystallization of *Os*GSTF5-5. **A.** Purification. **(M)** Low Molecular Weight Markers (Amersham, UK) **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Though. **(4)** Wash in 20mM Tris HCl pH 7.8. **(5)** Wash in 20mM Tris HCl pH 7.8, 50mM NaCl. **(6)** Elution in 20mM Tris HCl pH7.8, 5mM GSH. **(7)** Wash in 20mM Tris HCl 7.8, 3M NaCl. **B.** Crystals of *Os*GSTF5-5 with GSH.

## 3.5    *Petunia hybrida* An9

*Petunia hybrida* An9 is a Phi class GST involved in anthocyanin biosynthesis. It is a chaperonin protein that binds the anthocyanin in the cytoplasm and presents it to a GS-X pump in the vacuolar membrane where it is co-transported across with GSH (Dixon *et al.*, 1998; Edwards *et al.*, 2000). An9 was purified to homogeneity by Ni-NTA affinity as described in section 2.4.4 (figure 3.7 A). Hexagonal rod shaped crystals of An9 with GSH were grown in over fifty different conditions including 30% PEG 2K MME, 0.1M HEPES

NaOH pH 7.5. Similar crystals were also grown with cyanidin-3-*O*-glucoside (kuromanin). Examples of these crystals are shown in figure 3.7 B & C.
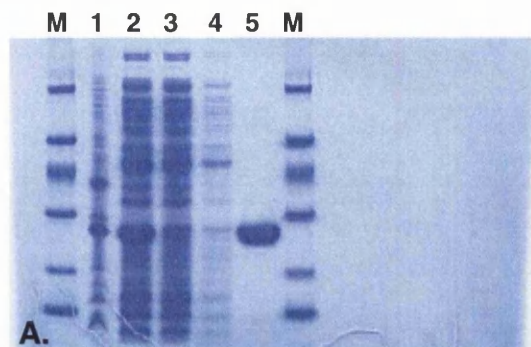


**Figure 3.7.** Purification and crystallization of An9. **A.** Purification of An9 **(M)** Low Molecular Weight Markers (Amersham, UK). **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Through. **(4)** Wash in 20mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol. **(5)** Elution in 250mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol. **B.** Crystals of An9. **C.** Crystals of An9 with Kuromanin.
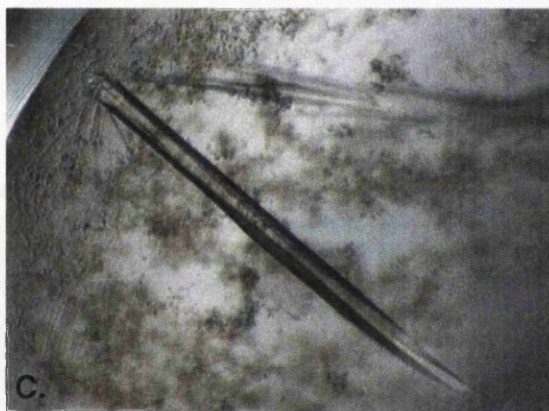
Many crystals from different conditions were tested for diffraction at Daresbury SRS. Diffraction was seen from these crystals to around 2.5Å along one axis but the pattern was characteristic of fibre diffraction (figure 3.8). Various different cryoprotectants were tried to see if these could improve diffraction but to no effect. A crystal was mounted in a capillary to try room temperature collection to see if cryo-cooling was damaging the crystal but the diffraction remained unchanged. This indicated that the crystals were only ordered in one dimension or that they were bundles of fine needles.
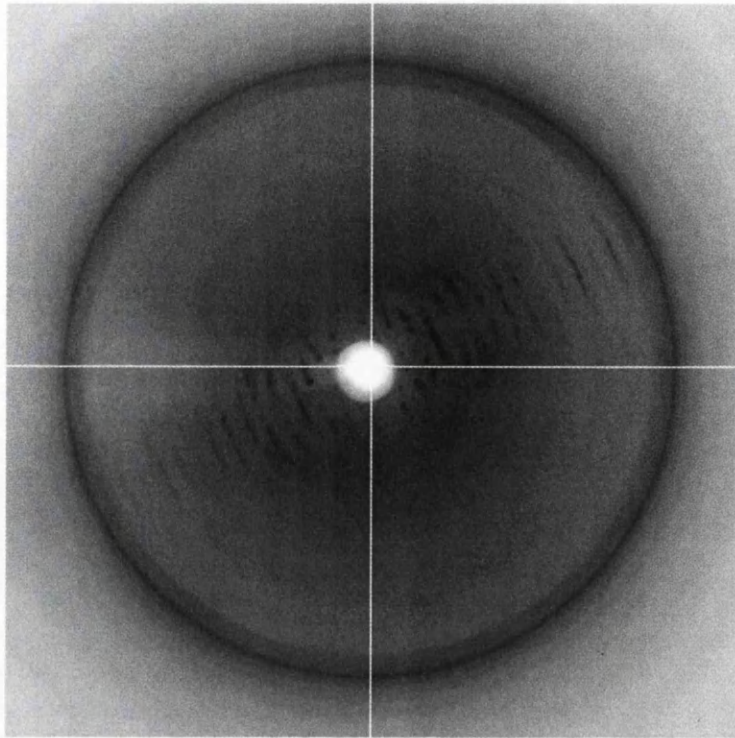
**Figure 3.8.** Fibre diffraction seen from a crystal of An9 with GSH.

## 3.6    *Zm*GSTU4-4

*Zm*GSTU4-4 is a Tau class GST from maize that also is involved in anthocyanin sequestration. Despite the fact that *Zm*GSTU4-4 and An9 are different classes of GST and have different structure they can functionally complement each other in *bz2* and *an9* mutant tissues (Alfenito *et al.*, 1998). Previous work in protein over-expression has only yielded inclusion bodies of this protein and any attempts to refold the protein result in no activity (private correspondence, Lukas Müeller, Stanford University, USA). During this project by altering growing conditions and media, soluble protein was obtained. This was purified by Ni-NTA affinity as described in section 2.4.5 (figure 3.9). This protein was very unstable and hard to keep in solution but small crystals were obtained in a solution of 5% PEG 8K, 0.05M $CaCl_2$, 0.1M Tris HCl pH 8.5 (figure 3.9). These crystals did not diffract and attempts to optimize the conditions produced no results.
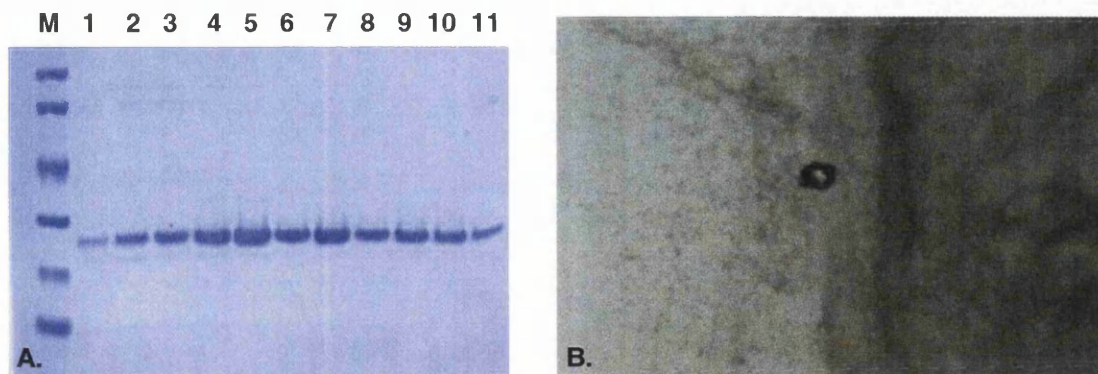
**Figure 3.9.** Purification and crystallization of *Zm*GSTU4-4. **A.** Purification **(M)** Low Molecular Weight Markers (Amersham, UK). **(1-11)** Samples from the elution peak of *Zm*GSTU4-4. **B.** Crystal of *Zm*GSTU4-4 with GSH.

## 3.7  *Zm*GSTU1-1

*Zm*GSTU1-1 is a Tau class GST from maize active in herbicide detoxification. Previous work on this protein had failed to yield diffraction quality crystals (Thom, 2001b). Recently it has been seen that *E. coli* overexpressing *Zm*GSTU1-1 became pigmented and were found to hyperaccumulate several porphyrins, one of which, harderoporphyrin, was shown to be glutathionylated (personal communication, Robert Edwards, University of Durham, UK). *Zm*GSTU1-1 was purified as described in section 2.4.3 by *S*-hexyl GSH affinity (figure 3.10). Small clusters of crystals of *Zm*GSTU1-1 with GSH and pig blood hemin were grown in a solution of 40% PEG 400, 0.1M NaOAc pH 4.5. These clusters showed diffraction to better than 3Å in-house but further optimization did not yield single crystals of this protein.
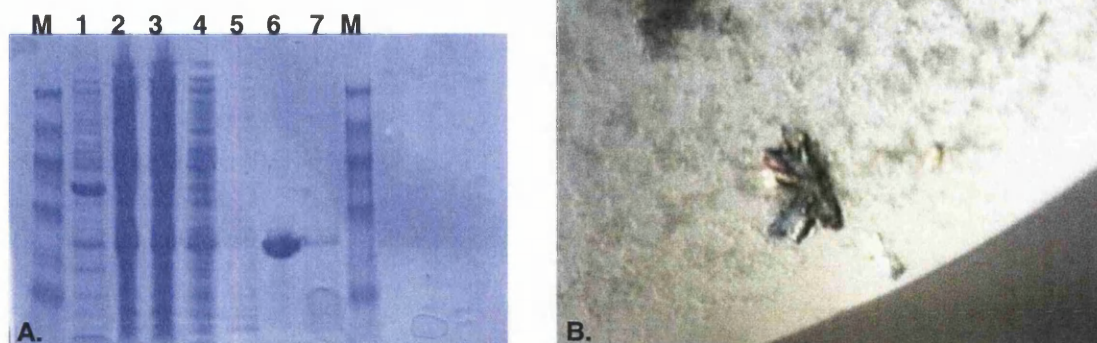
**Figure 3.10.** Purification and crystallization of *Zm*GSTU1-1. **A. (M)** Low Molecular Weight Markers (Amersham, UK). **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Though. **(4)** Wash in 20mM Tris HCl pH 7.8. **(5)** Wash in 20mM Tris HCl pH 7.8, 50mM NaCl. **(6)** Elution in 20mM Tris HCl pH7.8, 5mM GSH. **(7)** Wash in 20mM Tris HCl 7.8, 3M NaCl. **B.** Crystals of *Zm*GSTU1-1 with GSH and Hemin

## 3.8   *Zm*GSTF1-1

*Zm*GSTF1-1 was purified using an Orange A DyeMatrex gel column as described in section 2.4.2 (figure 3.11). This enzyme was not used for structural studies in this project but for preparation of GSH derivatives as discussed in the next chapter.
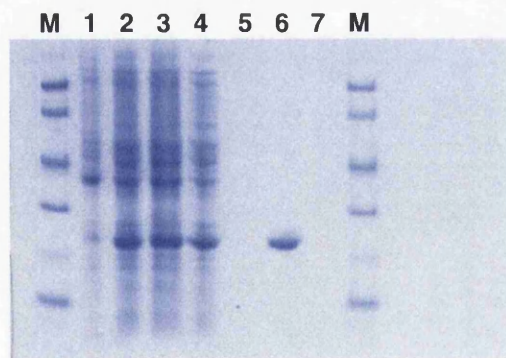


**Figure 3.11.** Purification of *Zm*GSTF1-1. **(M)** Low Molecular Weight Markers (Amersham, UK) **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Though. **(4)** Wash in 10mM potassium phosphate buffer pH 6.0. **(5)** Wash in 50mM potassium phosphate buffer pH 7.0. **(6)** Elution in 50mM potassium phosphate buffer pH7.0, 5mM GSH. **(7)** Wash in 1M NaCl.

## 3.9   *Ta*GSTL1

*Ta*GSTL1 is a Lambda class GST involved in redox homeostasis in wheat and is induced by xenobiotic treatment (Theodoulou *et al.*, 2003). This enzyme was purified by Ni-NTA chelation column as described in section 2.4.5 (figure 3.12). Unfortunately no crystals of this enzyme were grown during this project.
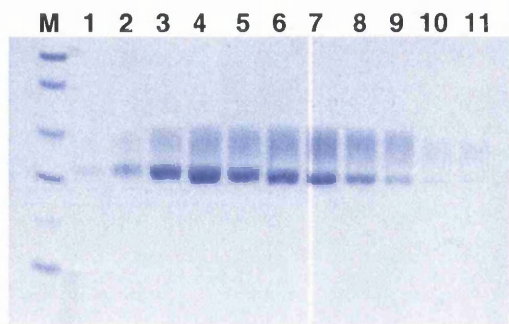
**Figure 3.12.** purification of *Ta*GSTL1. **(M)** Low Molecular Weight Markers (Amersham, UK). **(1-11)** Samples from the elution peak of *Ta*GSTL1.

## 3.10 *At*GSTL2

*At*GSTL2 is a lambda class GST from *A. thaliana* involved in redox homeostasis, and is located in the chloroplast. This enzyme is mildly induced by the xenobiotics fluorodifen and *t*-butyl hydroperoxide, and has demonstrated thiol transferase activity (Dixon *et al.*, 2002a). *At*GSTL2 was purified as described in section 2.4.4 (figure 3.13). Unfortunately no crystals of this enzyme were obtained during this project.
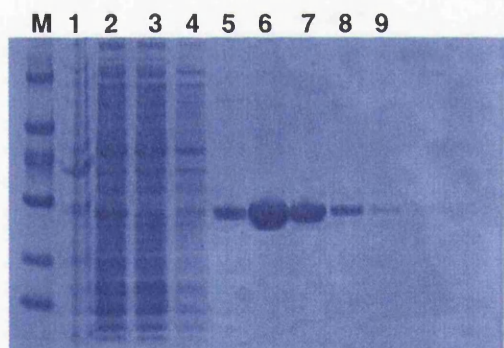


**Figure 3.13.** Purification of *At*GSTL2. **(M)** Low Molecular Weight Markers (Amersham, UK). **(1)** Cell Debris Pellet. **(2)** Crude Protein Solution. **(3)** Flow Through. **(4)** Wash in 20mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol. **(5-9)** Elution in 250mM imidazole, 50mM sodium phosphate buffer pH 8.0, 300mM NaCl, 10% glycerol.

# Chapter 4        Glutathione Derivatives

One of the aims of this thesis was to obtain the structure of a herbicide-glutathione conjugates bound to a Tau class GST, namely *Os*GSTU1-1. Several methods were tried to generate herbicide-glutathione conjugates to give structural details of the specificity of the GST

## 4.1    Soaking of Crystals with Herbicides

The first method tried was to introduce the herbicides to the protein by soaking them into crystals of *Os*GSTU1-1 with GSH. It was hoped that although the compounds have low solubility in aqueous solutions, they would slowly dissolve and diffuse into the active site. This method has been used successfully to produce GSH conjugates in other GSTs (Personal communication, Dr. M. Parker, St. Vincent's Institute of Medical Research, Australia).

Crystals of *Os*GSTU1-1 were grown using protein at 12mg ml$^{-1}$ with 20-25% PEG 4K, 0.2M MgCl$_2$, 0.1M HEPES NaOH pH 7.74 used as a precipitant. After two days the trays were opened and small crystals of CDNB, alachlor, acifluorfen, fluorodifen and fenclorim were added to the drops. Another tray of crystals was opened and 1µl of a pre-prepared metolochlor-GSH conjugate (*S*-[*N*-(2-ethyl-6-methylphenyl)-*N*-(2-methoxy-1-methylethyl)-acetamide]-glutathione), gifted to us by Robert Edwards, was added to the drops. These trays were resealed and left for one month to equilibrate. These crystals were similar in appearance to those shown in figure 3.8 and were found to diffract from 2.2-2.5Å at Daresbury SRS. After processing the data and solving the structures it was found that GSH was present in the active site but only in the crystals soaked with the metolachlor-GSH conjugate was any density seen in the H-site (see chapter 7).

## 4.2    Preparation of Glutathione Derivatives

The failure to soak herbicides into OsGSTU1-1-GSH crystals meant it was necessary to prepare GSH conjugates by chemical and enzymatic routes.

### 4.2.1   Reaction of Fluorodifen with Glutathione

The reaction product of fluorodifen with GSH (S-[2-nitro-4-(trifluoromethyl)-phenyl]-glutathione; NTFP-GSH) should be possible to synthesize chemically by reaction in basic conditions. 1ml samples of 10mM GSH with 10mM fluorodifen in a variety of 100mM basic buffers were prepared and shaken at room temperature for twenty-four hours. The reaction could be followed due to the production of 4-nitrophenol, which appears yellow in solution due to its absorbance at 405nm (figure 4.1). It was found that a yellow colour appeared in all solutions used but when the experiments were repeated without GSH the yellow colour was seen to appear indicating that the herbicide was undergoing base catalysed hydrolysis to form 4-nitrophenol and 2-nitro-4-(trifluoromethyl)-phenol. When the experiments were repeated in neutral solutions no yellow colouring appeared indicating no reaction was occurring.
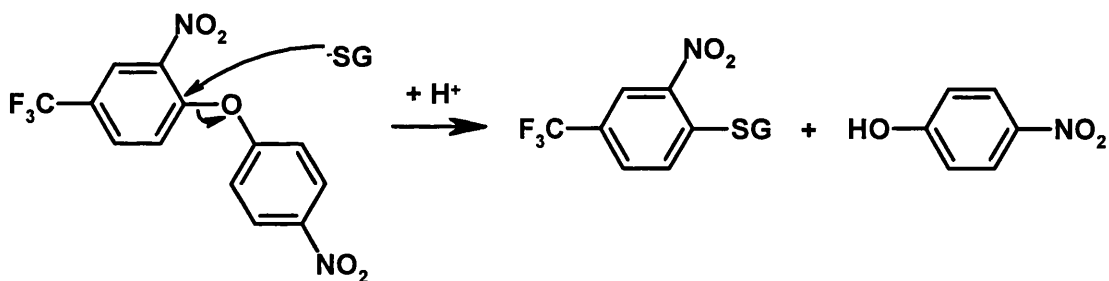


**Figure 4.1.** Reaction of fluorodifen with GSH. Image generated with ISIS/Draw (MDL Information Systems, Inc).

It was then attempted to perform the reactions in a variety of organic solvents. A 10mM fluorodifen, 10mM GSH solution in these solvents was prepared and shaken at room temperature for forty-eight hours. The only solution to develop a yellow colouring was the reaction taking place in dimethyl sulfoxide (DMSO). When the reaction was repeated in the absence of GSH no colouration

appeared indicating that in the initial reaction the NTFP-GSH conjugate was being produced (figure 4.2)

Due to its low volatility it is very difficult to evaporate off DMSO so a small quantity of this solution was added to drops containing crystals of *Os*GSTU1-1 with GSH. After one month these crystals were tested for diffraction at Daresbury SRS but unfortunately the DMSO appeared to have damaged the crystals as the diffraction appeared blurred due to an increase in mosaicity to over 4°.
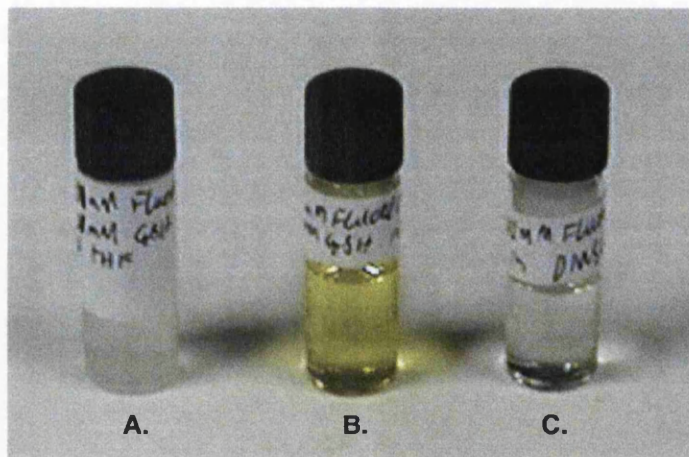


**Figure 4.2.** Attempts to produce NTFP-GSH conjugates. **A.** Fluorodifen with GSH in tetrahydrofuran (THF). **B.** Fluorodifen with GSH in DMSO. **C.** Fluorodifen in DMSO.

## 4.2.2  Preparation of Conjugate of CDNB with Glutathione

Since the reaction product of chloro-2,4-dinitrobenzene (CDNB) with GSH is homœomorphous with that of fluorodifen with GSH it was attempted to produce this conjugate (figure 4.3). To prepare the derivative of CDNB with GSH (*S*-[2,4-dinitrophenyl]-glutathione; DNP-GSH) 307mg of reduced L-glutathione (GSH) was dissolved in 90ml of 10mM potassium phosphate buffer pH 6.0. 101mg of chloro-2,4-dinitrobenzene (CDNB) was dissolved in 1.5ml of ethanol and added to the GSH solution and made up to 100ml with 10mM potassium phosphate buffer pH 6.0.

**Figure 4.3.** Reaction of CDNB with GSH. Image generated with ISIS/Draw (MDL Information Systems, Inc).

*Zm*GSTF1-1 in 10mM potassium phosphate buffer pH 6.0 was concentrated in an Amicon® pressure cell fitted with a 10kD cut-off ultrafiltration membrane (Millipore, UK) until it reached a concentration of 5mg ml$^{-1}$. 3ml of the *Zm*GSTF1-1 was placed into a 3ml 10kDa cut-off Slide-A-Lyzer® Dialysis Cassette (Pierce, USA) and placed in the 5mM CDNB, 10mM GSH solution at 4°C and stirred overnight.

1ml of this resultant solution was added to 12ml of 1mg ml$^{-1}$ *Os*GSTU1-1 which was then concentrated to 12mg ml$^{-1}$ using an Amicon® pressure cell fitted with a 10kD cut-off ultrafiltration membrane (Millipore, UK). This was used to set up crystallization trials as described in section 2.7. Crystals of *Os*GSTU1-1 with DNP-GSH were grown in a solution of 30% PEG 4K, 0.2M CaCl$_2$, 0.1M Tris HCl pH7.3 (figure 4.4).



**Figure 4.4.** Crystals of *Os*GSTU1-1 with the DNP-GSH conjugate.

These were taken to Daresbury SRS where they were seen to diffract to 3.1Å. Autoindexing showed that the protein had crystallized in a different space

group from the native enzyme, a primitive hexagonal space group with unit cell a = b = 76.613Å, c = 146.072Å, α = β = 90°, γ = 120°.

### 4.2.3 Preparation of Reaction Product of Fluorodifen with Glutathione

As OsGSTU1-1 has high activity towards fluorodifen it was decided that to obtain sufficient NTFP-GSH for crystallography it would be necessary to synthesize this compound. The synthesis followed a procedure outlined by Fukuyama et al. (1995) and was prepared in a two-step process with the help of Laurie Parker and Kathryn Nicoll.

*4.2.3.1 Preparation of 2-Nitro-4-(trifluoromethyl)-benzenesulfonamide*

1.80g of 2-nitro-4-(trifluoromethyl)-benzenesulfonyl chloride (6.21mM, Sigma, UK) was dissolved in 20ml dry dichloromethane (DCM, Fisher Scientific, UK). 50ml of a concentrated solution of ammonium hydroxide (25% $NH_3$, Riedel-de Haën, Germany) was heated at 50°C with stirring and the evolved ammonia was bubbled through the DCM solution for three hours under a nitrogen atmosphere with stirring (figure 4.5). The reaction mixture was filtered and the solvent was evaporated to yield the product. The product from this reaction was analysed by Thin Layer Chromatography (TLC), Proton and Carbon Nuclear Magnetic Resonance (NMR), Infrared Spectroscopy (IR), Mass Spectroscopy (MS), and High Resolution Mass Spectroscopy (HRMS).
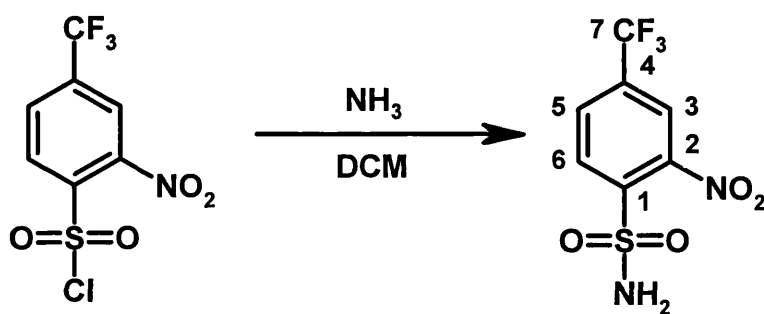


Figure 4.5. Reaction of 2-Nitro-4-(trifluoromethyl)-benzenesulfonyl chloride and ammonia to produce 2-Nitro-4-(trifluoromethyl)-benzenesulfonamide. The carbon atoms in the product are numbered for interpretation of NMR spectra. Image generated with ISIS/Draw (MDL Information Systems, Inc).

The TLC was run in a mixture of 75% hexane, 25% ethyl acetate. The product had an $R_f$ value of 0.11 and no trace of the starting material ($R_f$ = 0.64) was present so no further purification steps were required.

$^1$H NMR, $^{13}$C NMR and $^{13}$C DEPT45 (Distortionless Enhancement by Polarization Transfer) spectra were run on a sample of the product in $D_6$-DMSO (Aldrich, USA; figure 4.6). The peaks seen in the $^1$H NMR spectrum are at chemical shifts of 8.6 ppm (1H, s, 3-H), 8.4 ppm (1H, d, j=8.40, 5-H), 8.35 ppm (1H, d, j=8.40, 1-H), and 8.2 ppm (2H, s, $NH_2$). The $^{13}$C NMR spectrum shows seven peaks and the $^{13}$C DEPT spectrum shows only three peaks indicating that there are only three CH groups in the product. The peaks at a chemical shift of 40 ppm are from the $D_6$-DMSO.

In the infrared spectrum (figure 4.7), peaks were seen at 3361 cm$^{-1}$, 3255 cm$^{-1}$ (NH), 1537 cm$^{-1}$, 1326 cm$^{-1}$ ($NO_2$), 1348 cm$^{-1}$, 1164 cm$^{-1}$ (-$SO_2$-), and 1136 cm$^{-1}$ (CF).

In the electron impact mass spectrum (figure 4.7) the main peaks were seen at 270 (M$^+$, 12.1%), 254 (100%), 84 (57.8%), and 66 (59.9%). For $C_7H_5O_4N_2F_3S$ the calculated mass is 269.9922 and HRMS found 269.9921.
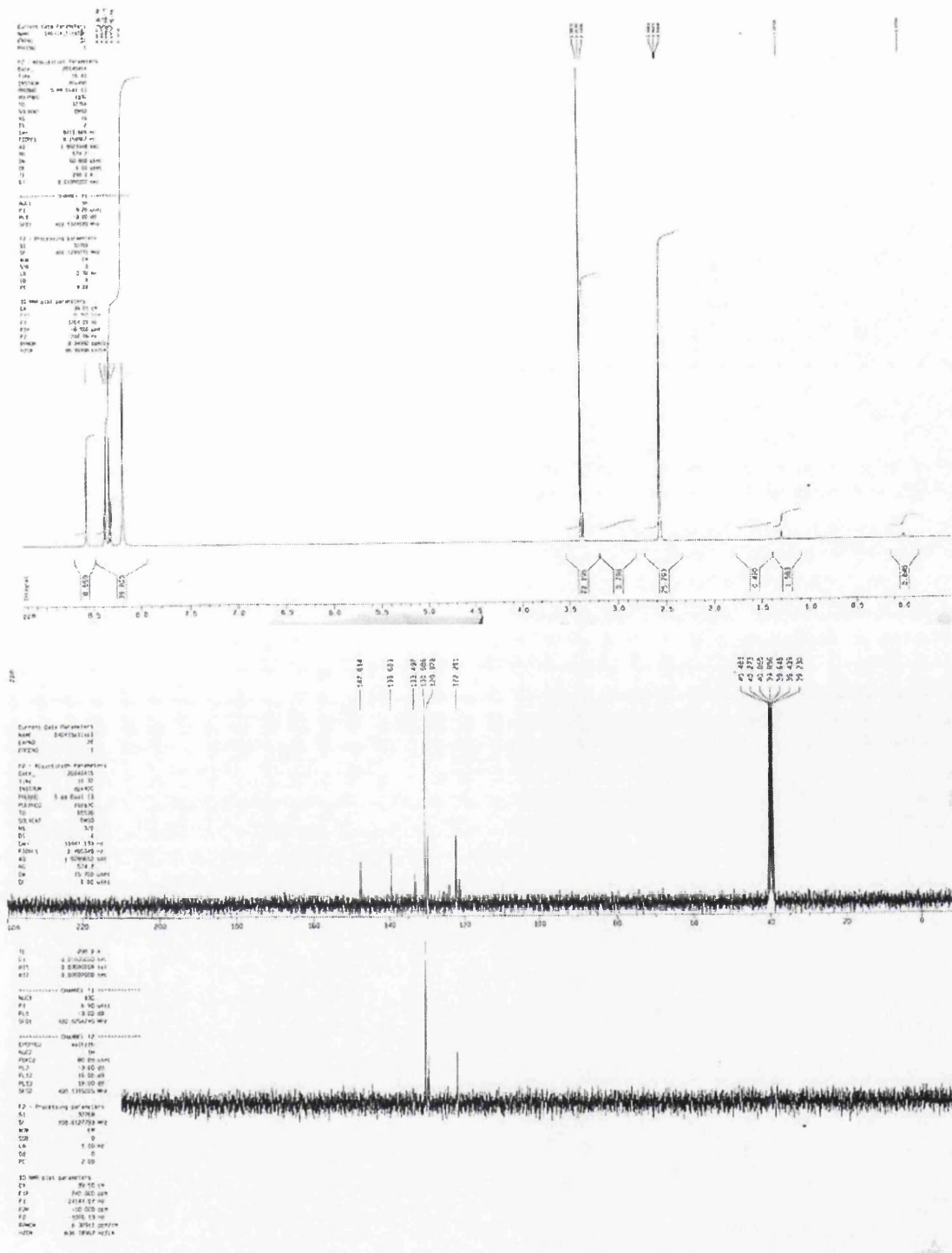
**Figure 4.6.** NMR spectra of 2-Nitro-4-(trifluoromethyl)-benzenesulfonamide in D$_6$-DMSO. The top spectrum is from the $^1$H NMR and the bottom spectra are the $^{13}$C NMR and $^{13}$C DEPT45 spectra.
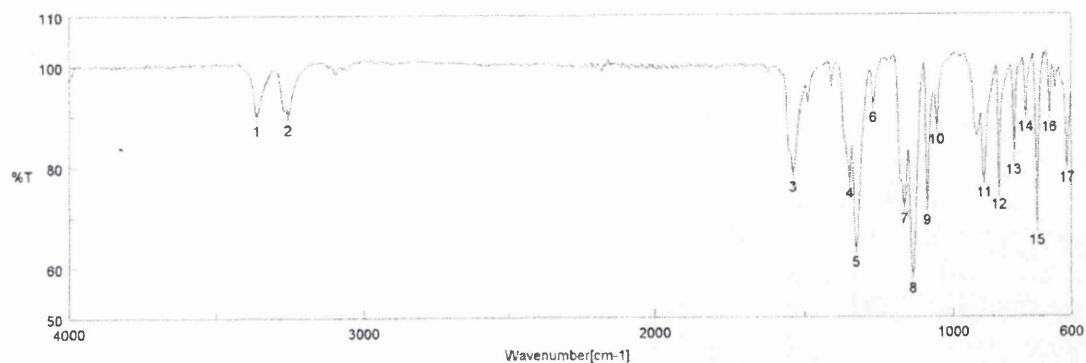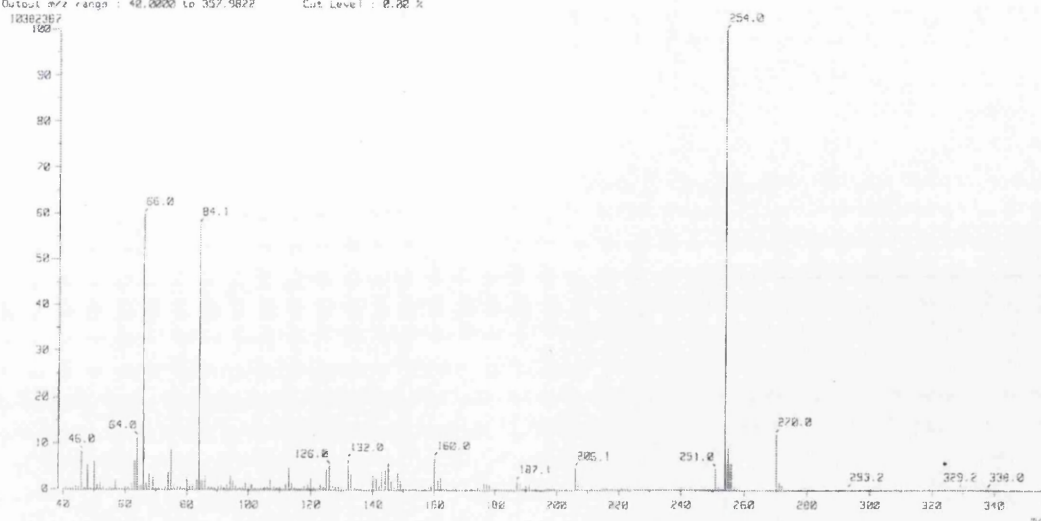
**Figure 4.7.** Infrared and mass spectra of 2-Nitro-4-(trifluoromethyl)-benzenesulfonamide.

## 4.2.3.2 Preparation of S-[2-Nitro-4-(trifluoromethyl)-phenyl]-glutathione

The procedure outlined in Fukayama *et al.* (1995) had to be altered for this stage due to the strength of the base used. The base $K_2CO_3$ used in the paper caused damage to the GSH moiety, and several other bases tried also caused the same problem. Pyridine was chosen as a solvent as it also has weakly basic properties.

0.41g of 2-nitro-4-(trifluoromethyl)-benzenesulfonamide (1.52mM) and reduced L-Glutathione (0.52g, 1.68mM) were dissolved in 60ml dry pyridine (Lancaster Synthesis, UK) and stirred at 80°C under a nitrogen atmosphere

**61**

for five days (figure 4.8). The solvent was evaporated off and the resulting solid was dissolved in 20ml methanol. 1g of silica was added and the methanol was evaporated off. This was loaded onto a 25g silica column and the column run in 1:3 ethyl acetate:methanol. The solvent was evaporated off to yield the product. The product from this reaction was analysed by TLC, Proton and Carbon NMR, IR, MS and by MP.
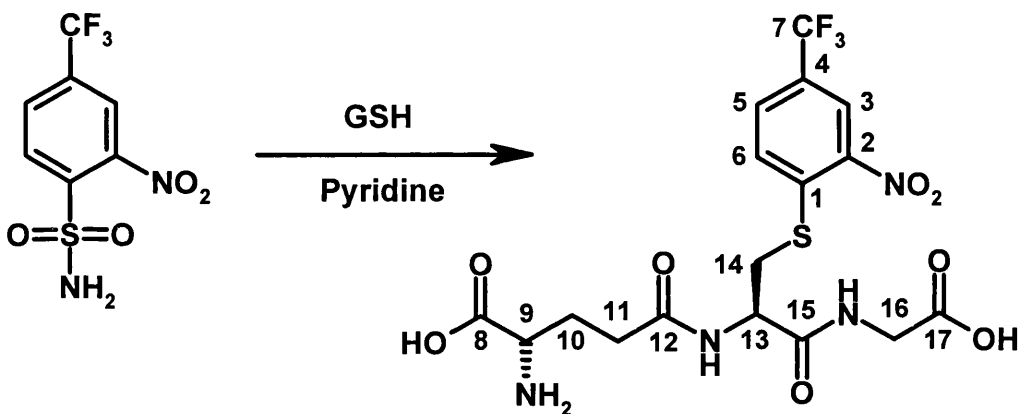


**Figure 4.8.** Reaction of 2-nitro-4-(trifluoromethyl)-benzenesulfonamide with GSH to produce S-[2-nitro-4-(trifluoromethyl)-phenyl]-glutathione. The carbon atoms in the product are numbered for interpretation of NMR spectra. Image generated with ISIS/Draw (MDL Information Systems, Inc).

After the reaction was completed the product was run on a TLC plate in 25% ethyl acetate, 75% methanol. This indicated that the reaction had not gone to completion as some starting material ($R_f$ = 0.94) was present with the product ($R_f$ = 0.38). To attempt to separate these from each other a silica column was run in 25% ethyl acetate, 75% methanol but TLC after the column indicated that some starting material was still present. All further analysis was performed on this mixture.

The $^1H$ NMR spectrum of this mixture was very complex and made interpretation very difficult. The $^{13}C$ DEPT135 spectrum was interpreted more easily. The peaks found in this spectrum were at chemical shifts of 24.59 ppm (14-C), 29.05 ppm (10-C), 34.51 ppm (11-C), 41.06 ppm (16-C), 54.62 ppm (13-C), 54.80 ppm (9-C), 116.00 ppm (6-C), 129.73 ppm (5-C), 130.70 ppm (1-C), and 131.93 ppm (3-C).
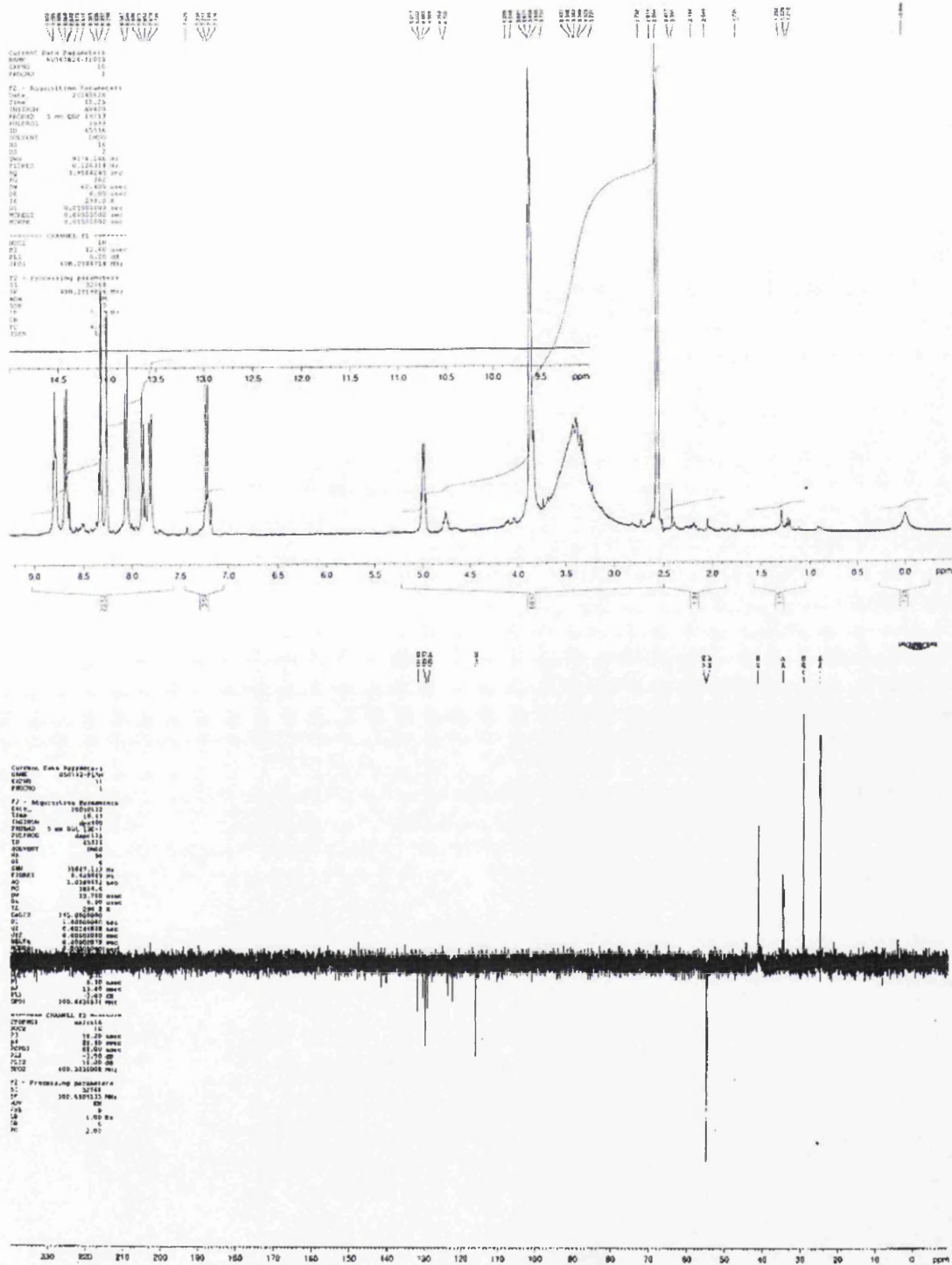
**Figure 4.9.** NMR spectra *S*-[2-Nitro-4-(trifluoromethyl)-phenyl]-glutathione in D₆-DMSO. The top spectrum is from the ¹H NMR and the bottom spectrum if from the ¹³C DEPT135.

A MP of 65-67°C was recorded for the product but due to the fact that it was a mixture no useable IR or MS spectra were recorded. It was decided to continue on with crystallization trials, as the main product of the reaction

would have a higher affinity for the enzyme active site than the sulfonamide starting material.

To a 1ml sample of OsGSTU1-1 at 12mg ml$^{-1}$, S-[2-nitro-4-(trifluoromethyl)-phenyl]-glutathione (NTFP-GSH) was added to a final concentration of 5mM (2.5mg). This was used to set up crystallization trials as described in section 2.7. Crystals of OsGSTU1-1 with NTFP-GSH were grown in a solution of 30% PEG 4K, 0.2M CaCl$_2$, 0.1M Tris HCl pH7.3 (figure 4.10). These were taken to Daresbury SRS where they were seen to diffract to 2.4Å. Autoindexing showed that the protein had crystallized in a primitive hexagonal space group with unit cell a = b = 77.129Å, c = 147.873Å, α = β = 90°, γ = 120°.
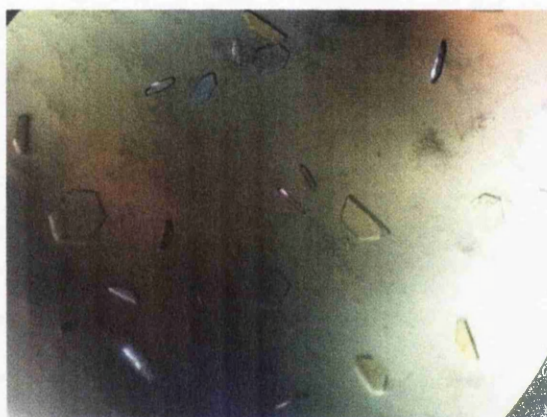


**Figure 4.10.** Crystals of OsGSTU1-1 with the NTFP-GSH conjugate.

# Chapter 5      X-Ray Diffraction and Data Processing

## 5.1   Introduction to Protein Crystallography

Ever since John Bernal and Dorothy Hodgkin first observed X-ray diffraction images from a crystal of pepsin in 1934 (Bernal & Crowfoot, 1934) the number of structures being solved has been steadily increasing. Although the technology of crystallography has improved dramatically the fundamental theory of structural prediction has changed very little since the early days. A variety of books have been written on crystallographic theory (Blundell & Johnson, 1976; Drenth, 1999; Rhodes, 1999; Ladd & Palmer, 2003) so only a brief overview has been included in this text.

### 5.1.1  Relationship of Diffraction to the Crystal and the Model

When a beam of electro-magnetic radiation like X-rays interacts with atoms within a molecule, electrons within the atoms, due to their small size, will oscillate at the same frequency as the incident wave. This is called Thomson, or coherent scattering, and the emitted radiation is at the same frequency as the incident radiation but differing in phase by 180°. The X-ray scattered from different atoms in the molecule will interfere with each other (reinforcing when they are in phase and cancelling out when out of phase) to produce an interference pattern called a 'molecular transform.' This gives rise to a characteristic pattern of intensities depending on the position of the atoms within the crystal (figure 5.1).   Diffraction spots are seen where there is constructive interference from each asymmetric unit (AU), therefore giving massive amplification of the signal.

The observed diffraction pattern for a protein crystal is the combination (convolution) of the diffraction pattern of the crystal lattice (reciprocal lattice) and the diffraction pattern of the molecule (molecular transform; Blundell & Johnson, 1976). From Bragg's law (equation 5.1) it can be seen that sin $\theta$ is inversely proportional to $d$, the spacing of the crystal lattice. Therefore, large

unit cell spacings will give rise to small separations in the diffraction pattern, and small unit cell spacings will give rise to large separations.
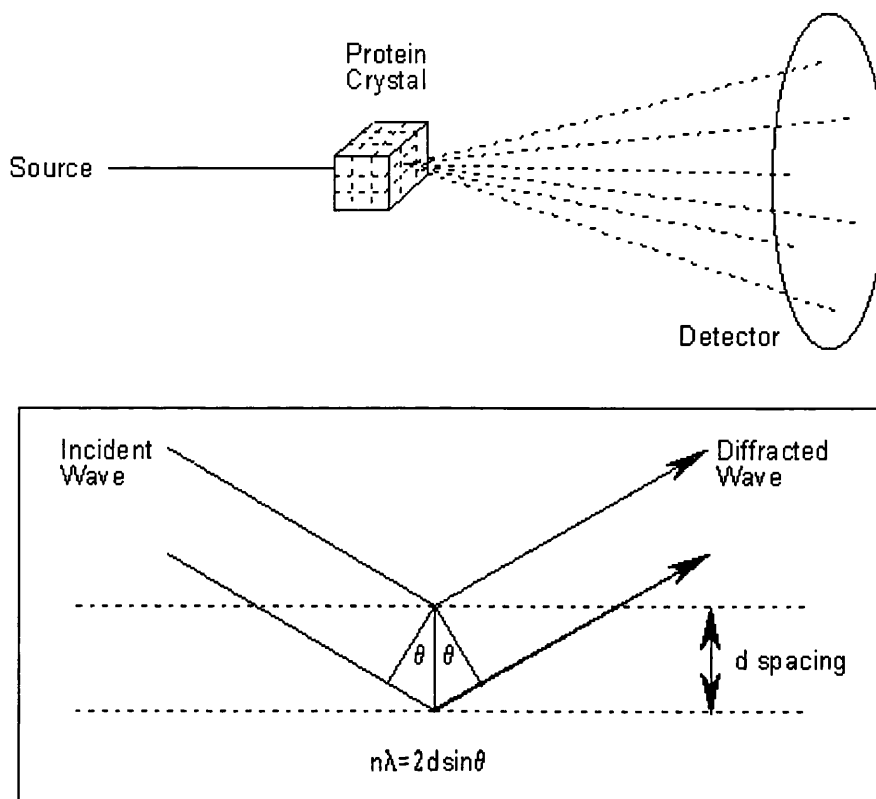


Figure 5.1. Generation of diffraction pattern from a protein crystal. Spots observed on the diffraction image relate to where scattered X-rays come into contact with the plate in phase with each other.

$$\sin\theta = \frac{n\lambda}{2d}$$

Equation 5.1. Bragg's Law

To be able to obtain a representation of the electron density for a structure from its diffraction pattern we must know the structure factors, $F_{(hkl)}$. The structure factors are complex numbers composed of an amplitude, $|F|_{(hkl)}$, and a phase, $\alpha_{(hkl)}$, for a given diffracted wave with Miller indices $hkl$, where $|F|_{(hkl)}$ is the square root of the intensity, $I_{(hkl)}$, of a reflection (equation 5.2). Since the diffraction pattern is the Fourier transform of the structure and, conversely, the structure is the Fourier transform of the diffraction pattern we

can use structure factors to calculate the electron density at any given point in the unit cell (equation 5.3). Although the amplitude component of the structure factor can be measured from the diffraction pattern the phase component cannot. This 'phase problem' is fundamental to crystallography and must be overcome by a number of direct methods, such as Isomorphous Replacement, Multi-Wavelength Anomalous Scattering (MAD) or Molecular Replacement.

$$F_{(hkl)} = \left| F \right|_{(hkl)} e^{i\alpha_{(hkl)}}$$

**Equation 5.2.** Calculation of structure factors, $F_{(hkl)}$, where $\left| F \right|_{(hkl)}$ is the amplitude and $\alpha_{(hkl)}$ is the phase of the diffracted wave.

$$\rho_{(xyz)} = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{(hkl)} e^{-2\pi i(hx+ky+lz)}$$

**Equation 5.3.** Calculation of electron density, $\rho_{(xyz)}$, where V is the volume of the unit cell and $F_{(hkl)}$ is the structure factors.

Data processing allows us to convert a series of diffraction images to a list of Miller indices with properly measured and corrected intensities assigned to them. From this we can then calculate the structure factor amplitudes $\left| F \right|_{(hkl)}$ for each reflection which forms the basis of the experimental X-ray data.

### 5.1.2 Molecular Replacement

One method of solving the phase problem is by molecular replacement with a protein of homologous structure. This technique was first developed in 1962 by Michael Rossmann and David Blow to utilise the presence of non-crystallographically related molecules in the AU to obtain phase information. By finding the transformation of a known structure to that of an unknown structure, its phases can be taken and used to estimate those for the unknown data. This can be done by using the Patterson function (equation 5.4) for which no phase information is required.

Two sets of vectors are present in a Patterson map. The first set is intramolecular vectors, which occur between atoms of a molecule and rely

only on the orientation of the molecule in the unit cell, not its position. The second set is intermolecular vectors, which occur between molecules and depend both on the orientation and position of the molecules in the unit cell.

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

**Equation 5.4.** The Patterson function. $P_{(u,v,w)}$ is a contour map with peaks representing vectors between atoms present in the structure.

The fact that these two sets of vectors can be separated allows the search for the correct orientation and position of a molecule in the unit cell to be divided into two three-dimensional searches. To find both the orientation and position in one step would require a six-dimensional search, which is a far larger problem.

### 5.1.2.1 Rotation Function

The intramolecular vectors in a Patterson map depend only on the orientation of the molecules within the unit cell, and even though the individual vectors are unresolved for a structure as large as a protein they can be isolated from the intermolecular vectors by defining a maximum Patterson radius. The rotation function is carried out by computing the product of the Patterson map of the probe structure and the Patterson map of the unknown structure. When the probe is correctly oriented the peaks from both Patterson maps will superimpose and the product will be large.

### 5.1.2.2 Translation Function

The translation function is used to discover the correct position of the correctly oriented probe within the unit cell. This is done in a similar way to the rotation function but utilises the intermolecular vectors rather than the intramolecular.

## 5.2    Introduction to Data Processing

With the increase in power from Synchrotron Radiation sources and the increase in speed of detector readout, macromolecular data processing software has had to be improved to enable optimal data collection. The software must be able to readily identify a crystals orientation, space group, unit cell dimensions and an estimation of its mosaicity. In addition, techniques like Multiwavelength Anomalous Dispersion (MAD) require the processing of large amounts of data. Recently a number of data processing and scaling programs have been developed, for example, XDS, Mosflm and the HKL Suite (Kabsch, 1993; Leslie, 1992; Otwinowski & Minor, 1997). During the course of this research the HKL Suite (Denzo, XdisplayF and Scalepack) have been used to process and scale all data. The above-mentioned data processing programs generally all perform the following functions;

- Account for detector characteristics.
- Determination and refinement of crystal orientation, space group and unit cell dimension.
- Predict position of reflections.
- Reflection integration and profile fitting.
- Scaling and refinement.

Although most data processing programs deal with the above steps in a similar manner, the points have been expanded on in relation to the HKL suite.

### 5.2.1   Detector Characteristics

Generally these days, crystallographic experiments are being carried out using the oscillation technique, and images are collected on an image plate system (IP, for example a Mar 345 Image Plate detector) or a charge coupled device (CCD, for example an ADSC Quantum 4 detector). The data processing software must be able to account for the specific characteristics of

different detectors to allow for correct indexing and integration of the diffraction images.

### 5.2.2 Determination of Crystal Parameters

A critical step in the analysis of macromolecular diffraction data is the ability to autoindex the diffraction patterns, i.e. the determination of the unit cell dimensions and the orientation of the crystal to the X-ray beam and the detector. Denzo uses a novel Fast Fourier Transform (FFT) algorithm for autoindexing, which allows for the independent identification of $h$, $k$ and $l$. It was suggested in 1986 by Gerard Bricogne (Bricogne, 1986) that a FFT could be a useful indexing tool but for large unit cells excessive amounts of memory and time would be required for the procedure (Campbell, 1997). After identifying these real-space vectors Denzo finds the three vectors that will give a minimum basic description of the unit cell. The program will then try to find the best fit for this primitive unit cell to the fourteen Bravais lattices and calculate the distortion necessary to fit this basic cell to a higher symmetry unit cell.

### 5.2.3 Prediction of Reflection Positions

Once a lattice has been picked and the unit cell defined the program uses the chosen dimensions to predict the positions of the reflections. The crystal and detector orientation parameters (for example, the X-beam & Y-beam position, unit cell dimensions, crossfire, and the cassette rotx & roty) are continually refined throughout processing to account for any small changes of orientation during the data collection.

### 5.2.4 Reflection Integration and Profile Fitting

Reflection integration is an important step in data processing as it generates the raw measurement of Intensity ($I_{(hkl)}$). To calculate the intensity of a reflection the detector background first has to be calculated. This background occurs from Compton scattering of the sample or from the Dark Current of the detector, and has to be calculated and removed from the reflection profile. To do this, Denzo calculates the average background in the vicinity of each

reflection and applies a series of statistical tests to remove pixels that deviate by more than $3\sigma$ from the best fit of the background, or are in the spot area of an adjacent reflection.

To improve the accurate measurement of all peaks Denzo uses a two step profile fitting process. Firstly, Denzo generates a profile of strong, well-measured reflections in the neighbourhood of the spot, and then applies the predicted profile to the actual profile of the weaker reflections. Profile fitting is most useful for weak reflections where the spot peak is not clear and is affected by noise. Profile fitting can also be useful in the case where you have overloaded spots. By matching the profile to the non-overloaded part of the spot an estimate of the intensity can be obtained. Profile fitting can only be a useful tool when the predicted spots match accurately with the actual spots. If there is systematic error in the predicted positions then noise will be added to the spots by broadening or displacing the profile.

The intensity of a diffracted spot is dependent on a variety of factors (equation 5.5). Firstly, the intensity of the incident radiation is important, as is the size of the crystal, with larger crystals generally producing greater diffraction. The volume of the unit cell is also important, with crystals with larger unit cells producing weaker diffraction than similar sized crystals with smaller unit cells. The scattering factors of the atoms within the unit cell contribute as well, with light atoms like carbon, oxygen and nitrogen as found in proteins scattering relatively weakly.

$$I_{(hkl)} = I_o \times \frac{\lambda^3}{\omega} \times \frac{V_{CR} \times L \times P \times A}{V^2} \times \left| F_{(hkl)} \right|^2$$

Equation 5.5. Calculation of the intensity of a diffracted beam from a crystal rotated with angular velocity $\omega$ depends on the intensity of the incident beam ($I_o$), the scattering from the atoms it contains ($F_{(hkl)}$), the wavelength of radiation ($\lambda$), the volume of the unit cell ($V$) and the volume of the crystal ($V_{CR}$). Additionally, corrections are made to the raw diffraction intensities ($L$ = Lorentz, $P$ = Polarization and $A$ = Absorption factor).

In a perfect crystal, diffraction spots would appear as sharp points, however observed diffraction spots generally show a small angular spread of around $0.25 - 0.5°$. This is due to the mosaic spread of the crystal as shown in figure 5.2. The mosaic spread is small imperfections in the crystal lattice, and therefore as this increase in size affects the intensity of a reflection this is taken into account by including the volume of the crystal.
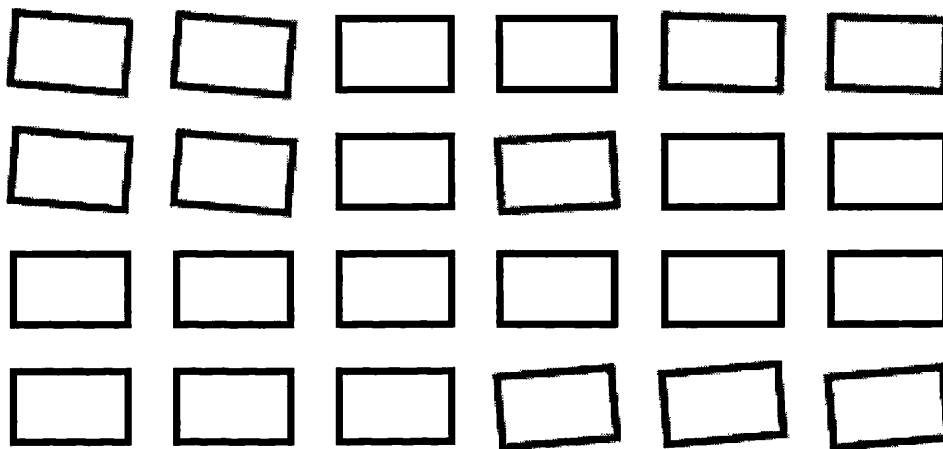
**Figure 5.2.** Mosaicity of a crystal.

The wavelength of the X-rays used for the diffraction experiment is generally determined by the available source, or by the need of a particular wavelength for an anomalous scattering experiment. However, the wavelength used affects the intensity of the scattered beam significantly. The longer the wavelength used the stronger the diffraction intensities, although longer wavelengths are absorbed more strongly by the crystal leading to increased radiation damage. The choice of copper $K\alpha$ (1.5418Å) for home sources is a good compromise between absorption and scattering.

During the data processing three corrections are made to the raw intensities. These are the Lorentz correction, the Polarization correction and the Absorption correction. The Lorentz correction is a measure of the velocity of a reflection as it passes through the Ewald sphere. The Polarization correction is applied when the incident beam is non-polarized, therefore this does not apply to synchrotron radiation or to radiation that has been collimated using a

monochromator. This correction is applied when non-polarized X-rays are used because when they reflect from the plane of a crystal they are split into distinct components, which will affect the intensity of the diffracted beam. A correction is applied for the absorption of X-rays by the crystal, and from absorption of X-rays by the air between the sample and the detector.

### 5.2.5 Scaling and Refinement

The corrected intensities from Denzo were then scaled using Scalepack. This program merges the intensities of symmetry related observations and allows for the rejection of statistically outlying reflections. Scaling also provides a method for the global refinement of the unit cell, crystal mosaicity and orientation, and provides statistics by which the data quality can be assessed. The merging R factor ($R_{merge}$) provides some idea of the data quality but can be influenced in a number of ways. The removal of weak data and decreasing the overall redundancy will both artificially decrease this value.

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

**Equation 5.6.** Calculation of $R_{merge}$. $R_{merge}$ should typically be less than 10% for most protein structures.

## 5.3    Data Collection and Processing

Individual data for the various crystals grown during this project were collected in-house, at Daresbury SRS, and at the ELETTRA Synchrotron Light Source in Trieste, Italy. Details of the different data sets collected during this project are detailed below.

### 5.3.1  *Os*GSTU1-1 with Glutathione

Crystals of *Os*GSTU1-1 with GSH were grown in a solution of 21.25% PEG 4K, 0.2M $MgCl_2$, 0.1M HEPES NaOH pH 7.74. The crystals were soaked in artificial mother liquor containing 20% glycerol as a cryoprotectant, and cooled

to 100K using an Oxford Instruments Cryojet. Data were collected to 1.95Å on station 14.2 at Daresbury Synchrotron Radiation Source (SRS) and processed using Denzo (Otwinowski & Minor, 1997). Autoindexing indicated that the protein had crystallized in a primitive orthorhombic space group with unit cell dimensions a = 76.829Å, b = 91.129Å, c = 165.034Å, α = β = γ = 90°. Scaling of the data was carried out using Scalepack (Otwinowski & Minor, 1997). From the scaling statistics and systematic absences (all odd reflections along h, k and l) it was clear that the space group was $P2_12_12_1$. This means that, in the favourable case, a minimum of 90° of data must be collected. 180° of data were collected at 1° oscillation per image to maximise the data redundancy. Data collection statistics are shown in table 5.1.

| Data collection | OsGSTU1-1 with GSH |
|---|---|
| Source | Daresbury SRS stn. 14.2 |
| Detector | ADSC Q4R CCD |
| Wavelength (Å) | 0.979 |
| Space Group | $P2_12_12_1$ |
| Resolution (Å) | 1.95 |
| Total number of reflections | 622923 |
| Total number of unique reflections | 85157 |
| Completeness (%) | 100 |
| Multiplicity[a] | 7.3 |
| Solvent content (%) | 56.02 |
| No. molecules in a.u. | 4 |
| $R_{merge}$ (%)[b] | 10.2 |
| Wilson B (Å$^2$) | 29.26 |

[a] Redundancy of data, defined as the ratio of the number of measured and the number of unique reflections.
[b] $R_{merge} = \sum_h \sum_i |I(h,i) - \langle I(h) \rangle| / \sum_h I(h,i)$, where $I(h,i)$ is the intensity value of the $i$th 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 5.1.** Data collection statistics for OsGSTU1-1 with GSH.

After Scalepack had merged the intensity data into a sca file they were converted into mtz format by the program scalepack2mtz (CCP4, 1994). The number of molecules in the unit cell and the solvent content was calculated (Matthews, 1968) and the intensities were converted to amplitudes using

Truncate (CCP4, 1994). The overall B factor was calculated using the Wilson statistics and the data was checked for twinning by examining the distribution of centric and acentric reflections. Solution to the phase problem for each structure is detailed in the following chapters.

### 5.3.2 Data Processing of OsGSTU1-1 with Glutathione and Herbicide Soaks

Crystals of OsGSTU1-1 with GSH were grown in a solution of 21.25% PEG 4K, 0.2M $MgCl_2$, 0.1M HEPES NaOH pH 7.74. To these crystals small crystals of CDNB, alachlor, acifluorfen, fluorodifen and fenclorim, or a small drop of a pre-prepared metolachlor-GSH conjugate was added to the drops and left to equilibrate for one month. The crystals were soaked in artificial mother liquor containing 20% glycerol as a cryoprotectant, and cooled to 100K using an Oxford Instruments Cryojet. Data were collected to 2.5Å for both the CDNB and acifluorfen soaks at Daresbury SRS station 14.1, to 2.4Å for alachlor and 2.2Å for fluorodifen and fenclorim at Daresbury SRS station 14.2 and to 2.4Å for metolachlor-GSH at Daresbury SRS station 9.6. All datasets were processed using Denzo and scaled with Scalepack and further processed with CCP4 software. Unit cell dimensions are shown in table 5.2 and data collection statistics are shown in table 5.3.

| Soak | Space Group | a (Å) | b (Å) | c (Å) |
|---|---|---|---|---|
| CDNB | $P2_12_12_1$ | 75.417 | 88.948 | 162.628 |
| Alachlor | $P2_12_12_1$ | 76.406 | 90.687 | 164.726 |
| Acifluorfen | $P2_12_12_1$ | 76.609 | 90.761 | 165.231 |
| Fluorodifen | $P2_12_12_1$ | 76.057 | 89.864 | 164.518 |
| Fenclorim | $P2_12_12_1$ | 76.387 | 90.638 | 164.531 |
| Metolachlor-GSH | $P2_12_12_1$ | 76.225 | 90.540 | 164.911 |

**Table 5.2.** Unit cell dimensions for herbicide soaks.

| Data Collection | CDNB | Alachlor | Acifluorfen | Fluorodifen | Fenclorim | Metolachlor-GSH |
|---|---|---|---|---|---|---|
| Source | SRS 14.1 | SRS 14.2 | SRS 14.1 | SRS 14.2 | SRS 14.2 | SRS 9.6 |
| Detector | ADSC Q4R CCD | ADSC Q4R CCD | ADSC Q4R CCD | ADSC Q4R CCD | ADSC Q4R CCD | ADSC Q4R CCD |
| Wavelength (Å) | 0.977 | 0.979 | 0.977 | 0.979 | 0.979 | 0.87 |
| Space Group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Resolution (Å) | 2.5 | 2.4 | 2.5 | 2.2 | 2.2 | 2.4 |
| Total number of reflections | 138992 | 199984 | 310135 | 577576 | 233507 | 238030 |
| Total number of unique reflections | 39187 | 45438 | 40785 | 58733 | 58593 | 45151 |
| Completeness (%) | 86.5 | 98.4 | 99.7 | 92.9 | 91.7 | 99.6 |
| Multiplicity[a] | 3.5 | 4.4 | 7.6 | 9.8 | 4.0 | 5.5 |
| Solvent content (%) | 55.72 | 55.63 | 55.77 | 54.81 | 55.39 | 55.12 |
| No. molecules in a.u. | 4 | 4 | 4 | 4 | 4 | 4 |
| $R_{merge}$ (%) | 8.0 | 9.3 | 10.0 | 7.8 | 11.3 | 10.2 |
| Wilson B (Å²) | 75.60 | 51.70 | 47.68 | 41.93 | 35.67 | 48.93 |

[a] Redundancy of data, defined as the ratio of the number of measured and the number of unique reflections.
[b] $R_{merge} = \sum_h\sum_i |I(h,i) - \langle I(h)\rangle| / \sum_h I(h,i)$, where $I(h,i)$ is the intensity value of the $i$th 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 5.3.** Data collection statistics for OsGSTU1-1 with GSH and herbicide soaks.

### 5.3.3 *Os*GSTU1-1 Co-Crystallization with DNP-GSH and NTFP-GSH Conjugates

Crystals of *Os*GSTU1-1 with both conjugates were grown in a solution of 30% PEG 4K, 0.2M $CaCl_2$, 0.1M Tris HCl pH 7.3. The crystals were soaked in artificial mother liquor containing 20% glycerol as a cryoprotectant, and cooled to 100K using an Oxford Instruments Cryojet. Data were collected to 3.1Å on station 14.2 at Daresbury SRS for the DNP-GSH crystals and to 2.4Å on station 14.1 for the NTFP-GSH crystals. The data sets were processed using Denzo and scaled using Scalepack (Otwinowski & Minor, 1997). This indicated that the protein had crystallized in the primitive hexagonal space group $P3_221$ for both datasets. The crystals with DNP-GSH had unit cell dimensions a = b = 76.613Å, c = 146.072Å, $\alpha = \beta = 90°$, $\gamma = 120°$ while the crystals with NTFP-GSH had unit cell dimensions a = b = 77.129Å, c = 147.873Å, $\alpha = \beta = 90°$, $\gamma = 120°$. Data collection statistics are shown in table 5.4.

| Data collection | DNP-GSH | NTFP-GSH |
|---|---|---|
| Source | Daresbury SRS stn. 14.2 | Daresbury SRS stn. 14.1 |
| Detector | ADSC Q4R CCD | ADSC Q4R CCD |
| Wavelength (Å) | 0.979 | 0.977 |
| Space Group | $P3_221$ | $P3_221$ |
| Resolution (Å) | 3.1 | 2.4 |
| Total number of reflections | 85722 | 173413 |
| Total number of unique reflections | 9971 | 23134 |
| Completeness (%) | 99.9 | 99.7 |
| Multiplicity[a] | 8.6 | 7.6 |
| Solvent content (%) | 48.67 | 49.48 |
| No. molecules in a.u. | 2 | 2 |
| $R_{merge}$ (%)[b] | 13.4 | 8.1 |
| Wilson B (Å²) | 55.17 | 49.55 |

[a] Redundancy of data, defined as the ratio of the number of measured and the number of unique reflections.
[b] $R_{merge} = \sum_h \sum_i |I(h,i) - \langle I(h) \rangle| / \sum_h I(h,i)$, where $I(h,i)$ is the intensity value of the $i$th 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 5.4.** Data collection statistics for *Os*GSTU1-1 with conjugates.

### 5.3.4   *Os*GSTU4-4 with Glutathione

Crystals of *Os*GSTU4-4 with GSH were grown in a solution of 30% PEG 400, 0.2M $CaCl_2$, 0.1M HEPES NaOH pH 6.87. The crystals were cooled to 100K using an Oxford Instruments Cryojet and data were collected to 2.3Å on station 14.1 at Daresbury SRS. The data were processed using Denzo and scaled using Scalepack (Otwinowski & Minor, 1997). This indicated that the protein had crystallized in a primitive tetragonal space group with unit cell dimensions a = b = 56.093Å, c = 165.034Å, $\alpha = \beta = \gamma = 90°$. From the scaling and the systematic absences (h = 4n+1, 4n+2, and 4n+3) it was clear that the point group was 422 and the space group either $P4_122$ or $P4_322$. Data collection statistics are shown in table 5.5.

| Data collection | OsGSTU4-4 with GSH |
|---|---|
| Source | Daresbury SRS stn. 14.1 |
| Detector | ADSC Q4R CCD |
| Wavelength (Å) | 0.977 |
| Space Group | $P4_122^c$ |
| Resolution (Å) | 2.30 |
| Total number of reflections | 112855 |
| Total number of unique reflections | 10588 |
| Completeness (%) | 99.7 |
| Multiplicity[a] | 10.7 |
| Solvent content (%) | 48.79 |
| No. molecules in a.u. | 1 |
| $R_{merge}$ (%)[b] | 9.8 |
| Wilson B (Å$^2$) | 44.08 |

[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 *i*th 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.
[c] Uniquely identified on structure solution.

**Table 5.5.** Data collection statistics for *Os*GSTU4-4 with GSH.

# Chapter 6      Structure of *Os*GSTU1-1

## 6.1   Structure Solution of *Os*GSTU1-1 with GSH

After collecting and processing data for *Os*GSTU1-1 with GSH to 1.95Å the sequence was compared with the sequence of the wheat protein *Ta*GSTU4 (1GWC, Thom *et al.*, 2002), the only previously solved structure of a Tau class GST. The two sequences share 42% identity and 55% similarity indicating that *Ta*GSTU4-4 would be a good search model for molecular replacement (figure 6.1). A dimer of the complete *Ta*GSTU4-4 structure (residues 4-224) was used as the original search model in the molecular replacement program AMoRe (Navaza, 1994) however this yielded no clear solution. It is likely that differences between the model and the X-ray structure account for the lack of success with molecular replacement. To minimize the difference between model and structure, the region 108-134 where there are large changes in sequence (for example Trp113 to Gly, Trp116 to Leu, and Leu117 to Trp) was removed.
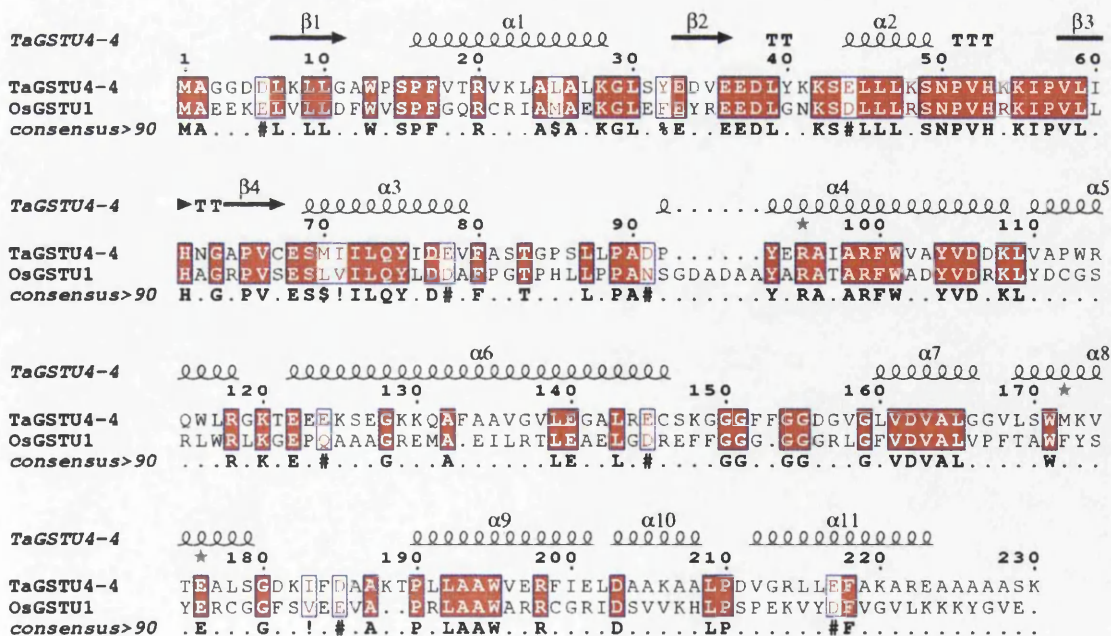


**Figure 6.1.** Sequence alignment between *Os*GSTU1 and *Ta*GSTU4 with identical residues outlined in red and similar residues indicated in red. The secondary structure elements of *Ta*GSTU4 are shown. Sequence alignment performed with MultAlin (Corpet, 1988) and figure generated with ESPript (Gouet *et al.*, 1999).

Using a dimer of this truncated structure as the search model for molecular replacement, a search for two dimers in the AU was performed with data from 10.0-4.0Å. A clear solution for two dimers was found using AMoRe giving a correlation coefficient of 45.1% and an R factor of 46.7% (table 6.1 and figure 6.2) compared to 41.3% and 48.1% for the next highest solution.

| | Correlation Of Intensities (%) | R-factor (%) |
|---|---|---|
| Rotation Function | 20.1 | 56.8 |
| Translation Function | 38.9 | 49.9 |
| Rigid Body Fitting | 45.1 | 46.7 |

Table 6.1. Solutions found by AMoRe for *Os*GSTU1-1.
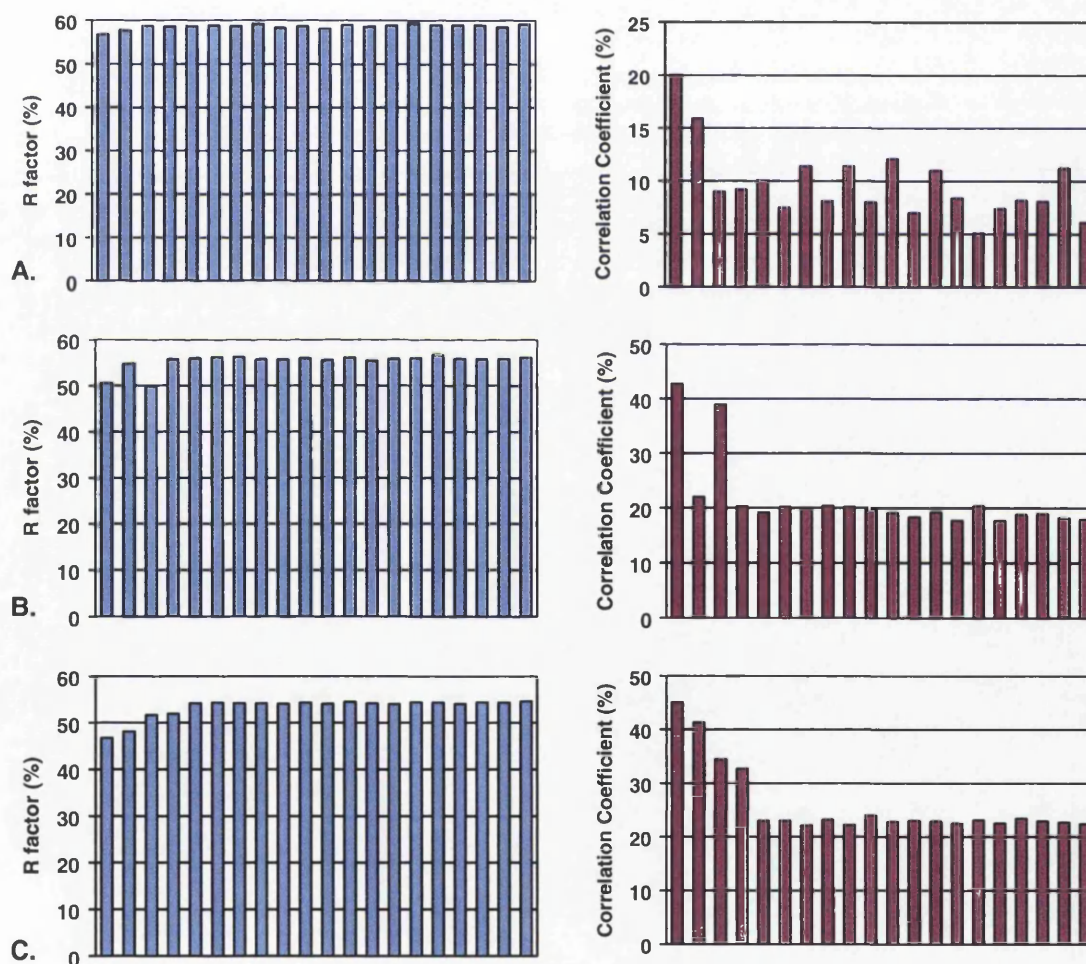


Figure 6.2. Statistics from each stage of AMoRe with Crystallographic R-factor shown in blue and the Correlation of Intensities shown in burgundy. From top to bottom they are statistics from **A.** the rotation function, **B.** two-body translation function and **C.** rigid-body refinement.

From this solution, two matrices were applied to transform the search model in to the unit cell using LSQKAB (CCP4, 1994). The orientation of these probes within the unit cell is shown in figure 6.3.

| First Matrix | | | |
|---|---|---|---|
| **Rotation Matrix** | 0.99998 | -0.00169 | -0.00542 |
| | 0.00166 | 0.99999 | -0.00508 |
| | 0.00543 | 0.00507 | 0.99997 |
| **Translation** | -0.07 | 0.19 | -0.57 |

| Second Matrix | | | |
|---|---|---|---|
| **Rotation Matrix** | 0.99996 | -0.00169 | -0.00927 |
| | 0.00155 | 0.99987 | 0.01579 |
| | -0.00930 | -0.01577 | 0.99983 |
| **Translation** | -0.97 | 45.97 | 38.49 |
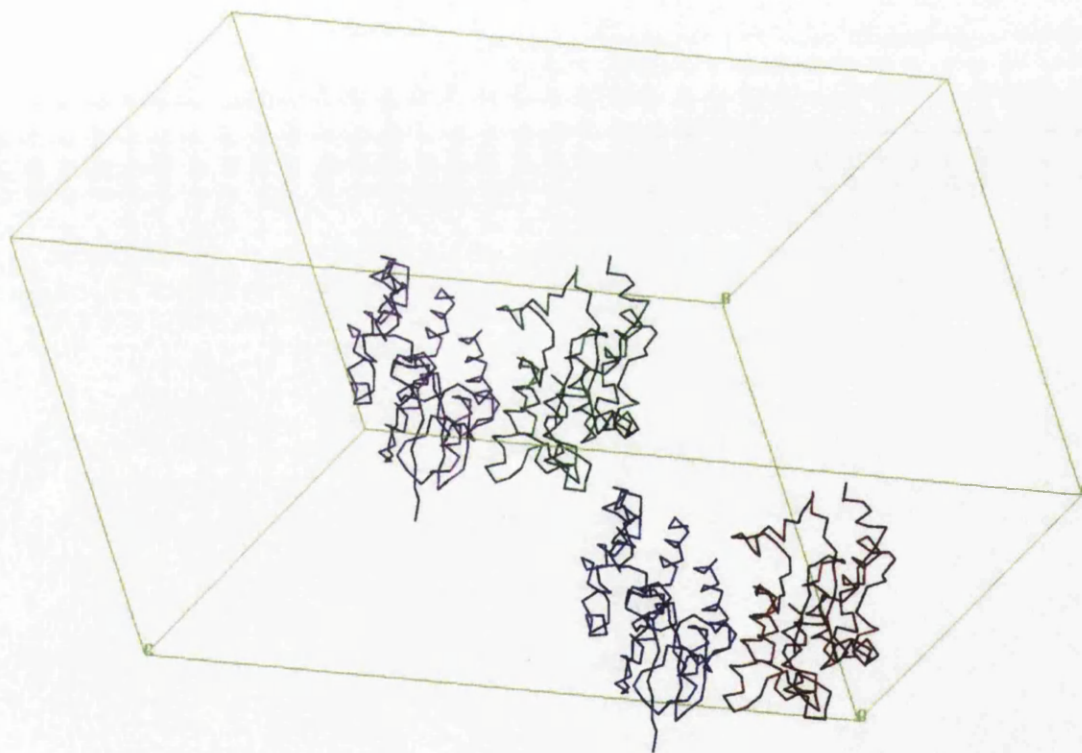


**Figure 6.3.** Orientation of probes within the Unit Cell. Image generated with SETOR (Evans, 1993).

The packing of the two dimers was scrutinized using the graphics program SETOR (Evans 1993). The absence of overlapping regions of structure and clear crystal contacts visually confirmed that the solution was correct. Due to the presence of four molecules in the AU, fourfold noncrystallographic symmetry (NCS) averaging could be used to improve the electron density

maps. Rotational and translational matrices were calculated using LSQKAB for the transformation of each of the three molecules to monomer A. NCSMASK (CCP4, 1994) was used to generate a mask around monomer A of the whole of the 1GWC structure (to average the missing regions of the structure). Both the mask and the NCS transformations were used in density modification of the initial electron density maps using the program DM (CCP4, 1994; Cowtan, 1994). A 2Fo-Fc map was generated using FFT (CCP4, 1994) and a fourfold averaged map of the protein was calculated using MAPROT (CCP4, 1994) over monomer A. Averaging is a powerful tool because it will amplify the signal but not any noise, resulting in more easily interpretable electron density maps.

After examining the average maps with the starting model in the graphics program QUANTA (Molecular Simulations Inc., 1986) a further twenty-eight residues were built into the structure and, where the sequence of the two proteins varied, the side chains of the residues were altered to the correct sequence, or where density was poor the residue was mutated to an alanine. The co-ordinates for the other three molecules were generated by PDBSET (CCP4, 1994) using the inverse of NCS matrices. The starting model was refined with the maximum likelihood refinement program REFMAC5 (Murshudov *et al.*, 1996) using all the X-ray data to 1.95Å and strict NCS restraints. After two more cycles of rebuilding and refinement with strict NCS restraints, the structure and density had improved to such a point that individual differences between the four chains were becoming obvious. From this point onwards the NCS restraints were not included, and each chain was built and refined separately. Eighteen more cycles of rebuilding and refinement were carried out, with the addition of 611 solvent molecules, seven GSH molecules, nine glycerol molecules, three magnesium ions, and twenty chloride ions. At this point the $R_{work}$ had been reduced to 17.8% and the $R_{free}$ to 22.7%. Refinement statistics are shown in table 6.2.

Geometric analysis of the structure showed bond and angle deviations from ideality of 0.023Å and 2.111° respectively. Analysis using the program

PROCHECK (Laskowski *et al.*, 1993) showed the geometry was better than average for a 1.95Å structure.

| | |
|---|---|
| Space Group | $P2_12_12_1$ |
| a (Å) | 76.829 |
| b (Å) | 91.129 |
| c (Å) | 165.034 |
| Resolution Range (Å) | 40.0 – 1.95 |
| No of unique reflections in resolution range | 68290 |
| $R_{Work}$ (%)[a] | 17.8 |
| No of reflections used in $R_{free}$ | 3635 (5.1%) |
| $R_{Free}$ (%) | 22.7 |
| Total number of protein atoms | 7126 |
| Total number of solvent atoms | 611 |
| Total number of hetero atoms | 217 |
| Average B factor ($Å^2$) | |
| Main-chain atoms | 30.1 |
| Side-chain atoms | 32.2 |
| Solvent atoms | 49.9 |
| Hetero atoms | 53.3 |
| Root mean square deviation | |
| Bonds (Å) | 0.023 |
| Bond angles (°) | 2.111 |
| Ramachandran quality, % in[b] | |
| Most favoured regions | 92.7 |
| Allowed regions | 7.0 |

[a] $R = \sum_{hkl}( |F_{obs} - F_{calc}|) / \sum|F_{obs}|$

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

**Table 6.2.** Refinement statistics for *Os*GSTU1-1 with GSH in $P2_12_12_1$.

## 6.2  Structure of *Os*GSTU1-1 with GSH

The *Os*GSTU1-1 structure is a homo-dimer, with each molecule composed of two domains. The N-terminal domain (residues 1-79) is a α/β domain and contains the GSH binding site. There is then a linker region (residues 80-94) that connects the N-terminal domain to the C-terminal domain (residues 95-

231), which is composed of 6 helices and contains the hydrophobic binding pocket. The structure of OsGSTU1-1 is shown in figure 6.4.



**A.**



**B.**
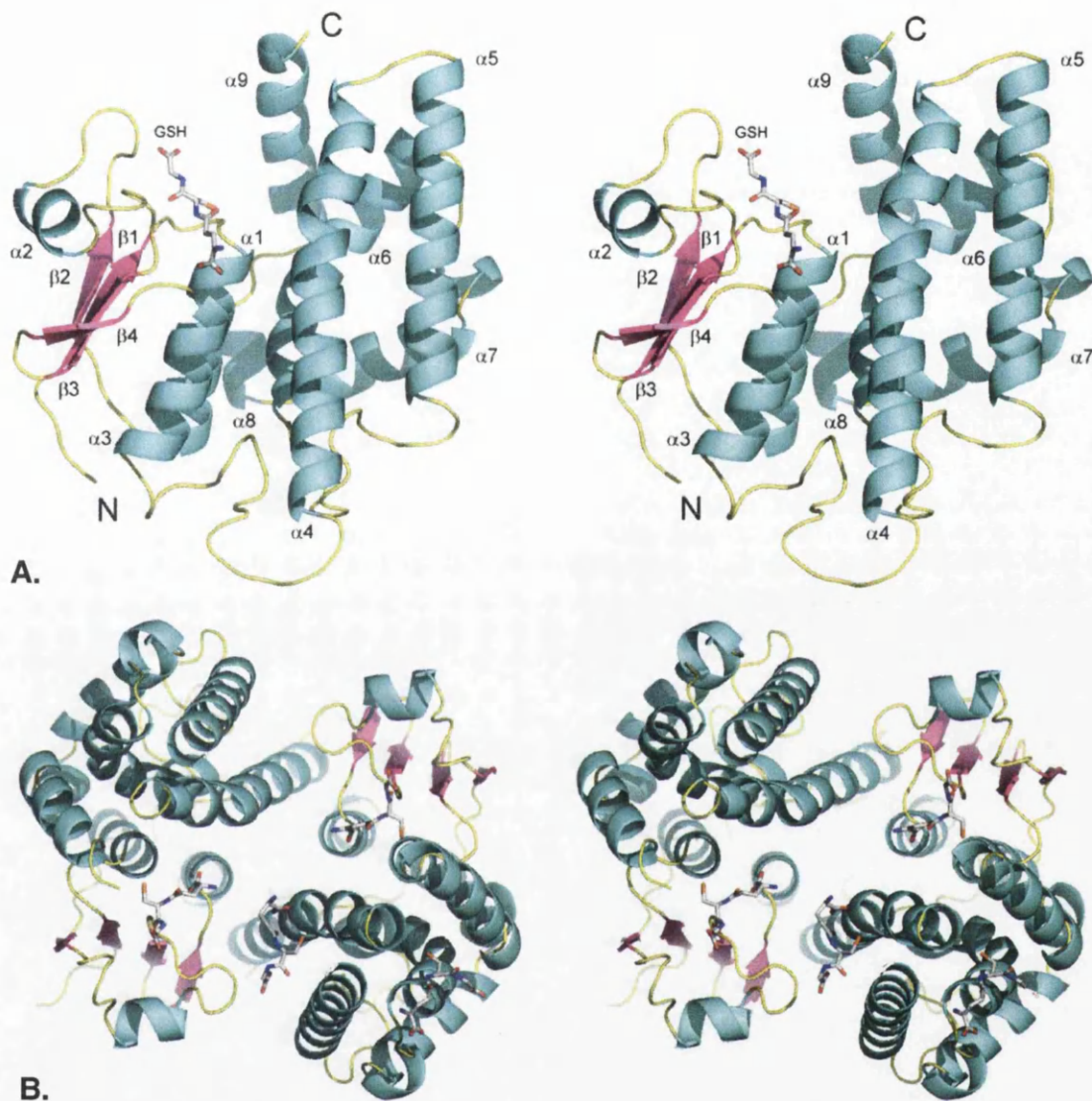
**Figure 6.4.** Stereo views of the OsGSTU1 monomer and dimer represented as a ribbon with GSH represented as ball and stick and coloured according to atom type. Images generated with PyMOL (Delano, 2002).

### 6.2.1  N-Terminal Domain

The N-terminal domain of OsGSTU1 consists of a four-strand β-sheet flanked by three α-helices (figure 6.4 A). This domain is composed of two motifs, a βαβ motif (β-1 7-11, α-1 16-29, β-2 33-36) and a ββα motif (β-3 58-61, β-4 65-67, α-3 69-79) joined by a loop containing a short helix (α-2 44-49) similar to the fold found in thioredoxin and T4 glutaredoxin. This first domain starts with

an N-terminal β-strand (β-1) followed by an α-helix (α-1) leading to a second β-strand (β-2) which runs parallel to the first strand. A loop containing a second helix (α-2) connects this motif to two sequential β-strands, which run antiparallel to each other and are followed by a third helix (α-3). The loop, which connects α-2 to β-3, contains a characteristic *cis*-proline residue that is highly conserved in GSTs. The four β-strands form a sheet in the centre of the domain with α-1 and α-3 behind the sheet and α-2 running perpendicular to the strands above the sheet and facing the solvent.

The sequence of the N-terminal domain is, to an extent, fairly well conserved between most GSTs, and the structure is remarkably well conserved, and contains the glutathione-binding site (G-site) in the cleft between the two motifs. The structure of *Os*GSTU1 contains a molecule of GSH in the G-site and it interacts with the protein in several ways (figure 6.5). The GSH molecule interacts with the side-chains of Lys 42 and Glu 68 and with the main-chain nitrogen and carbonyl oxygen of Ile 56. GSH also forms an interaction with the side-chain hydroxyl and main-chain nitrogen of Ser 69. Hydrophobic interactions are also seen between GSH and Phe 17, Leu 39 and Lys 55. The sulphur moiety of GSH is in close proximity to the hydroxyl of Ser 15, which is known to stabilise the formation of the thiolate anion GSH (GS⁻). Apart from these contacts, four water molecules are seen to form contacts with both GSH and the protein. The first water molecule forms contacts to O12 of GSH, the carbonyl oxygen of Pro 57 and to the main-chain nitrogen of Glu 68. The second water forms contacts with O11 of GSH, the main-chain nitrogen of Leu 70 and OD1 of Asp 112 in the C-terminal domain. The third forms contacts with N1 of GSH and again to OD1 of Asp 112. The fourth water forms contacts with N1 of GSH, NZ of Lys 55 and finally to NH2 of Arg 113 on the C-terminal domain of the second monomer.

There is a variable linker region (variable in sequence and length), which joins the N- and C-terminal domains. In *Os*GSTU1 the linker region is fifteen residues in length (80-94) and contains four proline residues.
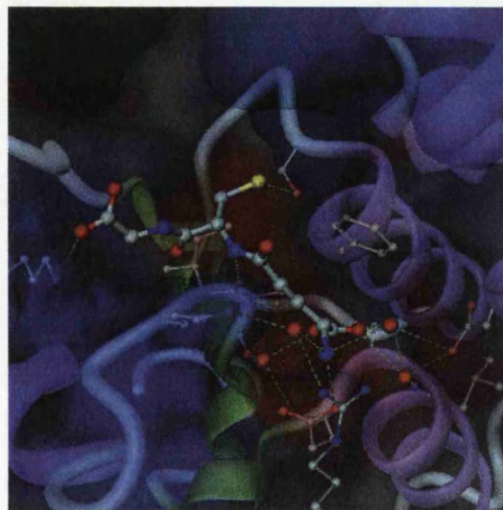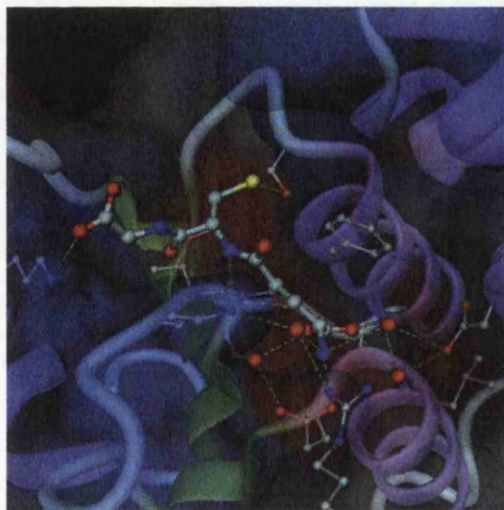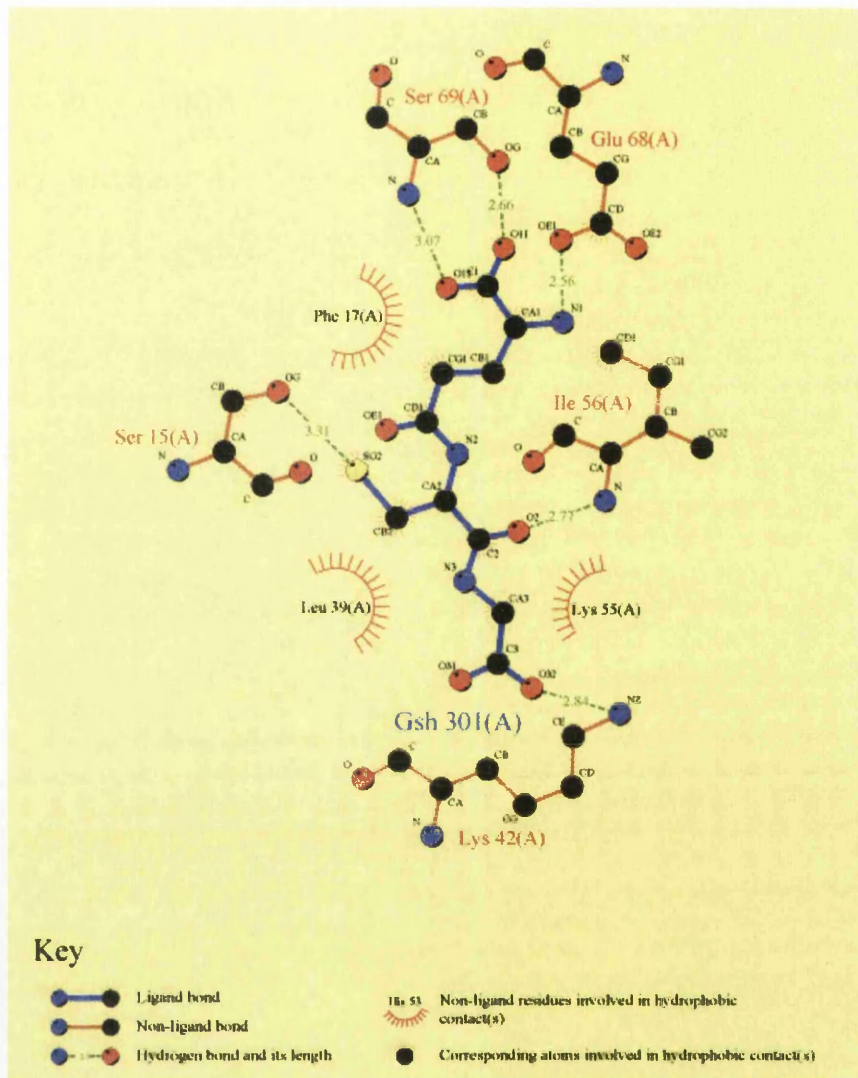
**Figure 6.5.** Diagram of interactions between GSH and the G-site of *Os*GSTU1. The top image is generated with LIGPLOT (Wallace *et al.*, 1995), and the bottom image is a stereo view of the active site generated with DINO (2003).

### 6.2.2  C-Terminal Domain

The C-terminal domain (95-231) of *Os*GSTU1 is composed of six, partially interrupted α-helices connected by a variety of loops. The first two helices (α-4 95-124, α-5 128-148) are relatively long and α-4 is broken at position 114 as is seen in many GSTs. Helix α-5 is connected to α-6 (164-183) by a fifteen residue long loop which contains a chain of six glycine residues. Helix α-6 is distorted by Pro 171, and with helices α-4 and α-5 forms a three-helix bundle that is the core of the C-terminal domain. Helix α-7 (192-203) runs almost perpendicular to α-5 and α-6, and is linked by one residue to the short helix α-8 (205-210) which runs at a 90° angle to α-7. Helix α-9 (214-228) crosses the top of α-5 and α-6 at around a 65° angle and it is this area that contains most of the hydrophobic-binding pocket (H-site). The proposed H-site of *Os*GSTU1 contains a deep pocket that is lined with the hydrophobic aromatic residues Trp 123 & 175, and Phe 172 & 220 as well as Leu 224. Although the H-site is mostly hydrophobic, there are also a few hydrogen-bonding partners available in Tyr 116 & 179, Ser 120, Arg 124 and Lys 227. A few residues from the N-terminal domain may also contribute to the H-site, namely Phe 12 & 17, Val 14 and Pro 16.

### 6.2.2.1 Glutathionylation Sites

The AU of *Os*GSTU1-1 contains two dimers and it was observed that each dimer had one monomer that was glutathionylated at one or two sites on the C-terminal domain. Monomer C is glutathionylated at Cys 118 & 182 (figure 6.6) while monomer A is glutathionylated only at Cys 118. In both cases these cysteine residues have a lower predicted pKa (PROPKA; Li *et al.*, 2005) compared to the pKa for free cysteine due to their proximity to arginine residues (121 & 181). The biological significance of these thiolation sites has not been identified but proteins with similarly acidic cysteine residues have been shown to scavenge free thiols and thiyl radicals (Starke *et al.*, 2003). Omega and lambda class GSTs contain an active site cysteine in place of serine or tyrosine and form mixed disulfides with glutathione in mediation of oxidative stress (Board *et al.*, 2000; Dixon *et al.*, 2002a).
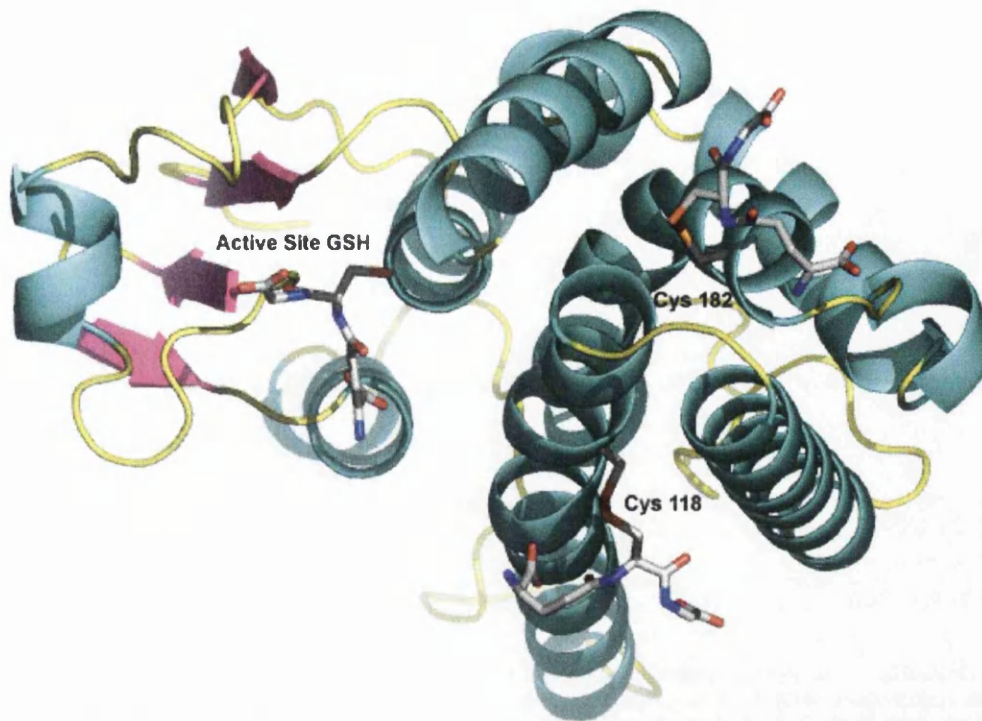
**Figure 6.6.** Monomer of *Os*GSTU1 showing the glutathionylation sites at Cys 118 & 182. Image generated with PyMOL (Delano, 2002).

It has been seen however that the glutathione molecule attached to Cys 118 of monomer A is essential for forming crystal contacts (figure 6.7). The contact is formed from an $Mg^{2+}$ ion coordinating to the oxygen for the glycine fragment of the glutathione and to the OD2 of Asp 205 of symmetry related monomer D. The glutathione orientation is stabilised by two hydrogen bonds to the NZ atom of Lys 114 on monomer A. These are from the OD1 atom of the glycine fragment (2.6Å) and the OE1 atom of the γ-glutamyl fragment. In conditions very similar to that in which *Os*GSTU1-1 was crystallised but without di-valent cations, very fine hexagonal plates were seen to grow, but no diffraction was seen from these crystals.
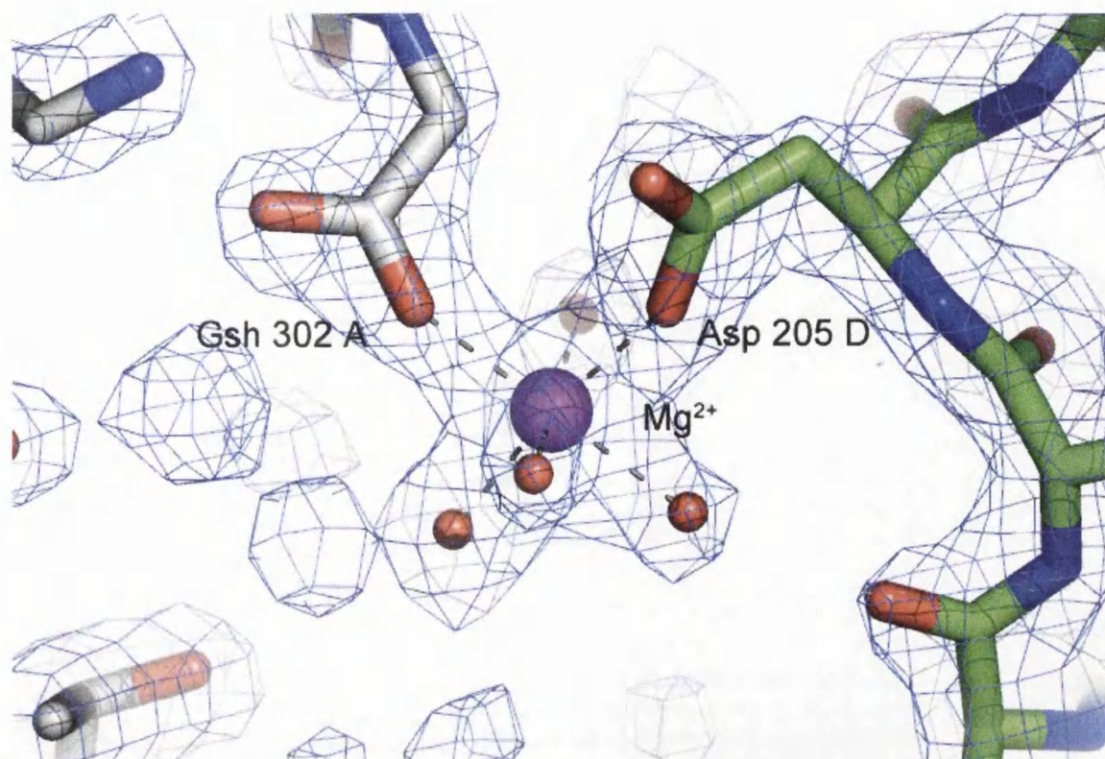
**Figure 6.7.** Crystal contact formed by an $Mg^{2+}$ Ion coordinating to the glutathione bound to Cys 118 of chain A, and OD2 atom of Asp 205 of chain D. Image generated with PyMOL (Delano, 2002).

### 6.2.3  Interactions Between Domains

The interface between the N-terminal and C-terminal domains has a buried surface area of $1803\text{Å}^2$ and is dominated by van der Waals interactions between hydrophobic and aromatic residues. The main residues involved in these interactions are Trp 13, Val 14, Pro 16, Phe 17, Ile 23 and Leu 70 & 73 in the N-terminal domain, and Ala 108 & 168, Tyr 116, Leu 162 & 211, Phe 164, 172 & 220, Val 167 & 170, Pro 171, Trp 175 and Lys 216 in the C-terminal domain. There are also seven hydrogen bonds between the two domains formed by three residues in the N-terminal domain to four residues in the C-terminal domain. These residues are Arg 20, Glu 27, Gln 74, Arg 105, Asp 112, Ala 168 and Ser 206. These interactions are detailed in table 6.3. In addition to these contacts, eight water molecules are seen to bridge this domain interface making eight contacts with the N-terminal domain and ten contacts with the C-terminal domain (table 6.4).

| Residue N-term domain | Residue C-term domain | Distance (Å) | Angle (°) |
|---|---|---|---|
| **Hydrogen Bonds** | | | |
| Arg 20 NH1 | Asp 112 OD2 | 2.6 | 141.0 |
| Arg 20 NH2 | Asp 112 OD2 | 2.9 | 161.4 |
| Arg 20 NH2 | Ala 168 O | 2.9 | 121.0 |
| Arg 20 NE | Ala 168 O | 2.9 | 113.3 |
| Glu 27 OE2 | Ser 206 OG | 2.7 | 124.9 |
| Gln 74 OE1 | Arg 105 NH1 | 3.1 | 170.7 |
| Gln 74 OE1 | Arg 105 NE | 2.8 | 141.6 |

**Table 6.3.** Details of the interactions between the N-terminal and C-terminal domains of OsGSTU1.

| Residue N-term | Distance (Å) | Water | Residue C-term | Distance (Å) |
|---|---|---|---|---|
| Arg 22 NH2 | 3.1 | 4 | His 210 ND1 | 2.7 |
| Leu 70 N | 3.0 | 11 | Asp 112 OD1 | 2.7 |
| Asp 78 OD2 | 2.7 | 66 | Arg 101 NH2 | 2.9 |
| Glu 27 OE1 | 2.9 | 92 | Arg 161 NH2 | 2.8 |
| | | | Leu 162 O | 2.9 |
| Gln 19 OE1 | 2.7 | 217 | Pro 171 O | 2.9 |
| | | | Trp 175 NE1 | 2.7 |
| Glu 27 OE1 | 2.8 | 219 | Leu 162 N | 2.7 |
| Asp 77 OD2 | 3.2 | 519 | Arg 101 NH1 | 3.1 |
| Glu 27 OE2 | 3.2 | 729 | Ser 206 N | 2.7 |

**Table 6.4.** Details of bridging water molecules between the domains of OsGSTU1.

## 6.2.4 Interactions at the Dimer Interface

The functional biological unit of OsGSTU1-1 is a homodimer with a buried surface accessible area of $2703\text{Å}^2$. This is at the lower end of the values seen in other classes of GST $(2700\text{-}3400\text{Å}^2)$ but is significantly larger than the buried surface area of TaGSTU4-4 $(2096\text{Å}^2$, Thom et al., 2002). The dimer interface of OsGSTU1-1 has several hydrophilic and hydrophobic interactions. There are two salt bridges formed between the monomers comprised of Arg 64 at the bottom of β-4 in the N-terminal domain and Asp 96' at the bottom of helix α-4 in the C-terminal domain. There are also six other hydrogen bonds formed between Pro 65 on β-4 to Tyr 99' on helix α-4, and between Asp 78 on helix α-3 to Arg 101' & 105' on helix α-4.

There are thirteen residues on each monomer involved in hydrophobic contacts and form three clusters. The first cluster is formed by Val 52 and His 53 on the loop between helix α-2 and β-3 and Glu 68 at the bottom of helix α-3 on one monomer, and Phe 106', Trp 107' and Tyr 110' on helix α-4 of the

other monomer. The second cluster is formed between Val 71 on helix α-3 with Ala 102' and Arg 105' on helix α-4. The final cluster is formed by Arg 64 and Lys 55 on β-4 and Tyr 75 on helix α-3 with Tyr 99' on helix α-4. Details of the hydrophilic contacts are shown in table 6.5 and images of the dimer interface are shown in figure 6.8. In addition to these contacts there are eleven water molecules that bridge the dimer interface and make fourteen contacts with monomer A and thirteen contacts with monomer B (table 6.6).

| Residue Monomer A | Residue Monomer B | Distance (Å) | Angle (°) |
|---|---|---|---|
| **Salt Bridges** | | | |
| Arg 64 NH1 & NH2 | Asp 96' OD2 & OD1 | 2.6 & 2.2 | 109.7 & 132.6 |
| Asp 96 OD1 & OD2 | Arg 64' NH2 & NH1 | 2.3 & 2.8 | 144.8 & 98.5 |
| **Hydrogen Bonds** | | | |
| Pro 65 O | Tyr 99' OH | 2.6 | 141.3 |
| Asp 78 OD1 | Arg 101' NE | 3.1 | 107.3 |
| Asp 78 OD1 | Arg 105' NH2 | 3.2 | 147.7 |
| Tyr 99 OH | Pro 65' O | 2.6 | 136.4 |
| Arg 101 NE | Asp 78' OD1 | 3.0 | 112.0 |
| Arg 105 NH2 | Asp 78' OD1 | 3.1 | 143.1 |

**Table 6.5.** Details of interactions at the dimer interface of *Os*GSTU1-1.

| Residue Monomer A | Distance (Å) | Water | Residue Monomer B | Distance (Å) |
|---|---|---|---|---|
| Phe 106 O | 2.9 | 10 | His 53' NE2 | 2.8 |
| Asp 109 OD2 | 2.8 | | | |
| Ala 98 O | 2.8 | 26 | Asp 78' OD1 | 2.8 |
| Arg 101 NH2 | 2.4 | 39 | Asp 78' OD2 | 2.6 |
| Arg 105 NH1 | 2.7 | | | |
| His 53 NE2 | 2.9 | 64 | Phe 106' O | 3.0 |
| | | | Asp 109' OD2 | 2.7 |
| Asp 78 OD2 | 2.7 | 66 | Arg 101' NH2 | 2.9 |
| Arg 101 NH2 | 2.9 | | Arg 105' NH1 | 3.0 |
| Asp 78 OD1 | 2.9 | 71 | Ala 98' O | 2.9 |
| Val 71 O | 3.0 | 78 | Arg 105' NH2 | 3.0 |
| Lys 55 NZ | 3.2 | 103 | Arg 113' NH2 | 2.9 |
| Arg 105 NH2 | 3.2 | 247 | Val 71' O | 3.0 |
| Arg 113 NH2 | 3.2 | 280 | Lys 55' NZ | 2.7 |
| Arg 113 NH1 | 3.3 | 308 | Asp 112' OD1 | 3.2 |

**Table 6.6.** Details of bridging water molecules at the dimer interface of *Os*GSTU1-1.
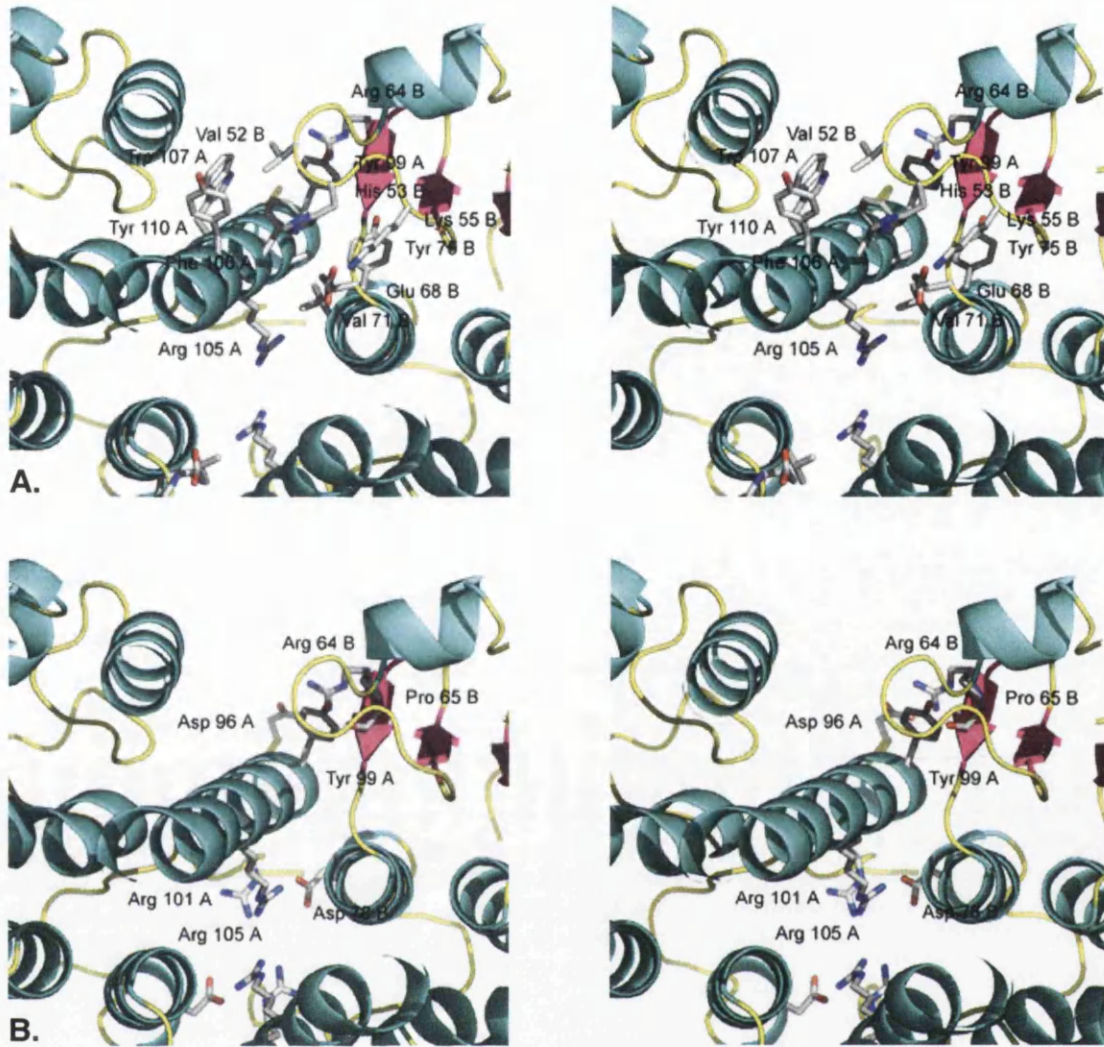
**Figure 6.8.** Dimer interface of *Os*GSTU1-1. **A.** Stereo view of hydrophobic contacts. **B.** Stereo view of hydrophilic contacts. Image generated with PyMOL (Delano, 2002).

# Chapter 7      Structure of OsGSTU1-1 with Substrates

## 7.1    Structure Solution of OsGSTU1-1 Soaks

Data for the crystals of OsGSTU1-1 soaked with CDNB, alachlor, acifluorfen, fluorodifen, fenclorim and the metolochlor-GSH conjugate were collected and processed as described in section 4.2.3. After converting the intensities to amplitudes the CCP4 program CAD was used to take the same $R_{free}$ set as the OsGSTU1-1 with GSH structure, for these datasets were isomorphous. The structures were solved using the structure of OsGSTU1-1 with GSH as a starting model and performing rigid-body refinement to optimize the position of the structure with respect to the data. REFMAC5 was used to perform ten rounds of rigid body refinement with each dimer defined as a rigid body group. This was followed by three cycles of restrained individual atom refinement with REFMAC5 after which weighted 2Fo-Fc and Fo-Fc maps were generated using FFT.

The active site of each of these structures was examined using QUANTA. In the structures that had been soaked with CDNB, alachlor, acifluorfen, fluorodifen or fenclorim no extra density was seen in the active sites, but there was clear ligand density in three out of four of the chains of the metolachlor-GSH soaked structure (figure 7.1). A model of the metolachlor-GSH conjugate was built using the program Cerius[2] (Accelrys, 2003) and energy minimised using Insight II (Molecular Simulations, 2000). A library for this conjugate was then created by running LIBCHECK (CCP4, 1994) through the Monomer Library Sketcher in the CCP4 Interface. After adding the ligands and performing one round of rebuilding and refinement it was clear from the electron density that there was an error in the stereochemistry of the chiral centre C35 (figure 7.2). This was corrected and after twelve rounds of rebuilding and refinement, and with the addition of 576 solvent molecules, three GSH molecules, three metolachlor-GSH molecules, ten glycerol molecules, two magnesium ions, and fifteen chloride ions, the $R_{work}$ had been

reduced to 16.0% and the R$_{free}$ to 26.0%. Refinement statistics are shown in table 7.1.
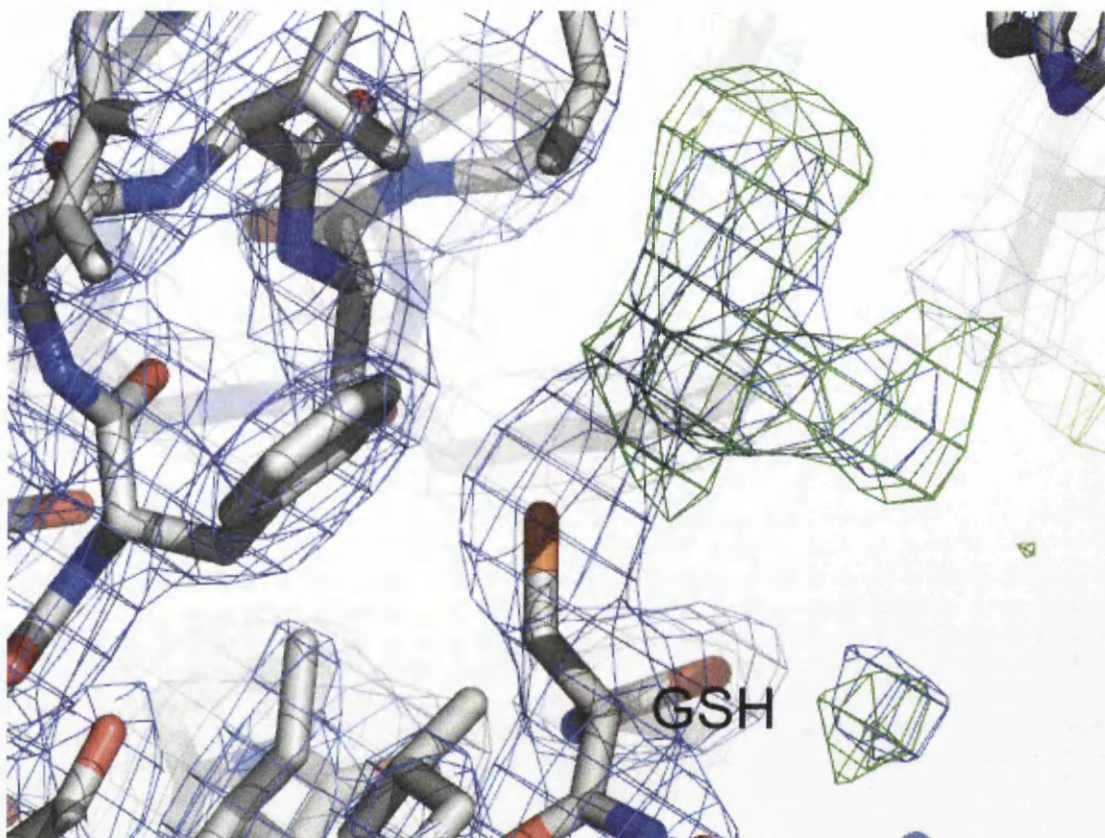


**Figure 7.1.** Active site of _Os_GSTU1 showing clear density for the metolachlor-GSH conjugate. A 2Fo-Fc map is shown in slate (1 sigma) and a Fo-Fc map is shown in green (3 sigma). Image created with PyMOL.
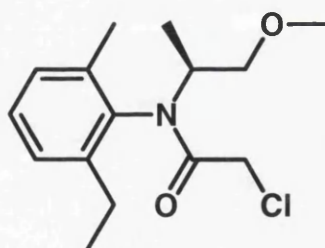


**Figure 7.2.** The correct structure of S-metolachlor. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Geometric analysis of the structure showed bond and angle deviations from ideality of 0.034Å and 2.652° respectively. Analysis using the program PROCHECK (Laskowski _et al._, 1993) showed the geometry was generally better than average for a 2.4Å structure.

| Space Group | $P2_12_12_1$ |
|---|---|
| a (Å) | 76.225 |
| b (Å) | 90.540 |
| c (Å) | 164.911 |
| Resolution Range (Å) | 35.0 – 2.4 |
| No of unique reflections in resolution range | 45151 |
| $R_{Work}$ (%)[a] | 16.0 |
| No of reflections used in $R_{free}$ | 2081 (5.0%) |
| $R_{Free}$ (%) | 26.0 |
| Total number of protein atoms | 7138 |
| Total number of solvent atoms | 576 |
| Total number of hetero atoms | 246 |
| Average B factor (Å$^2$) | |
| Main-chain atoms | 48.48 |
| Side-chain atoms | 48.96 |
| Solvent atoms | 51.37 |
| Hetero atoms | 61.71 |
| Root mean square deviation | |
| Bonds (Å) | 0.034 |
| Bond angles (°) | 2.652 |
| Ramachandran quality, % in[b] | |
| Most favoured regions | 90.0 |
| Allowed regions | 9.5 |

[a] $R = \Sigma_{hkl}( |F_{obs} - F_{calc}|) / \Sigma|F_{obs}|$

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

**Table 7.1.** Refinement statistics for *Os*GSTU1-1 with Metolachlor-GSH in $P2_12_12_1$.

In addition to the contacts described in section 5.2.1, the metolachlor-GSH conjugate is forming several hydrophobic contacts with *Os*GSTU1. These contacts consist of Phe 12 and Pro 16 from the N-terminal domain and Ser 120, Trp 175 and Phe 220 from the C-terminal domain (figures 7.3 and 7.4).

**Figure 7.3.** Stereo view of the active site of *Os*GSTU1 with the metolachlor-GSH conjugate. Residues involved in contacts are also shown. Image generated with Dino (Philippsen, 2003).



**Figure 7.4.** Diagram of the interactions of the metolachlor-GSH conjugate with *Os*GSTU1. Image generated with LIGPLOT.

## 7.2    Structure Solution of *Os*GSTU1-1 with DNP-GSH

Data for *Os*GSTU1-1 co-crystallised with the DNP-GSH conjugate was collected to 3.1Å and processed as described in section 5.2.4. A monomer of the completed *Os*GSTU1 structure with GSH was used as a search model for AMoRe using data from 10.0-4.0Å. A clear solution for two monomers in the AU was found with a correlation coefficient of 71.8% and an R factor of 36.4% (table 7.2 & figure 7.5) compared to 40.7% and 48.8% for the next highest solution.

| | Correlation Of Intensities (%) | R-factor (%) |
|---|---|---|
| **Rotation Function** | 30.4 | 56.4 |
| **Translation Function** | 34.0 | 52.2 |
| **Rigid Body Fitting** | 71.8 | 36.4 |

**Table 7.2.** Solutions found by AMoRe for *Os*GSTU1-1 with DNP-GSH.
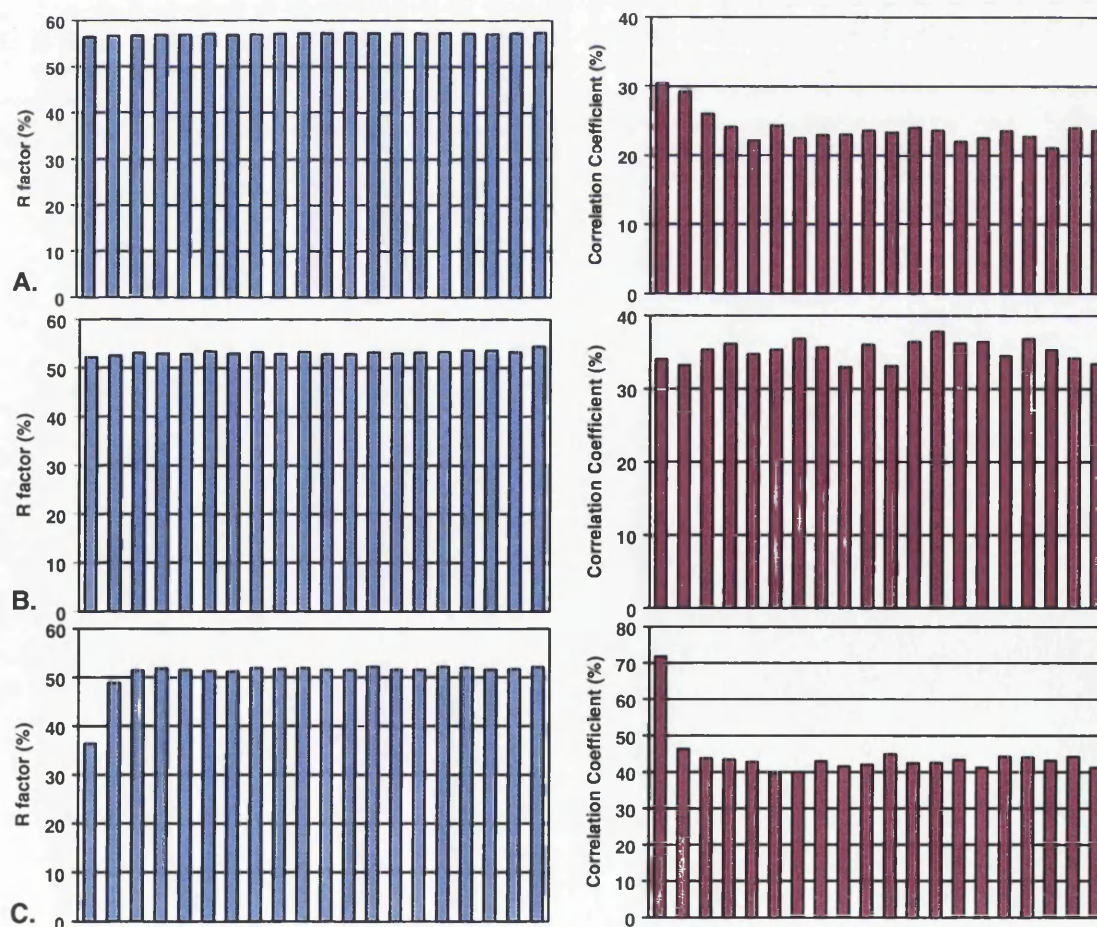


**Figure 7.5.** Statistics from each stage of AMoRe with Crystallographic R-factor shown in blue and the Correlation of Intensities shown in burgundy. From top to bottom they are statistics from: **A.** the rotation function **B.** the translation function **C.** after rigid-body refinement.

From this solution, the matrices provided could be used to place the two monomers in the unit cell using LSQKAB (CCP4, 1994). The packing of the two monomers within the unit cell was examined using the molecular graphics program SETOR (Evans, 1993). The fact that there were clear crystal contacts, and no overlapping regions in the structure visually confirmed the correct solution (figure 7.6).
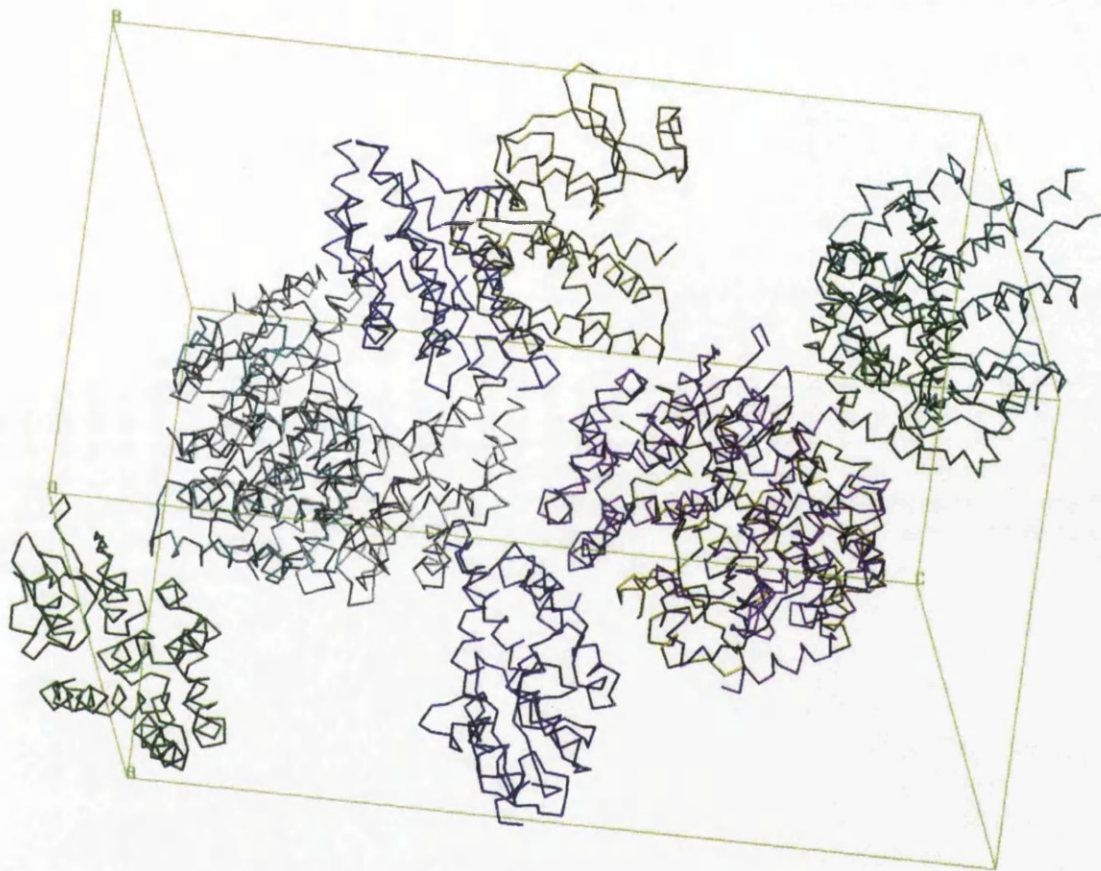


**Figure 7.6.** Packing of _Os_GSTU1-1 within the $P3_2 21$ unit cell. Image generated with SETOR (Evans, 1993).

REFMAC5 was used to perform ten rounds of rigid body refinement followed by four rounds of refinement using strict NCS restraints. Maximum likelihood weighted 2Fo-Fc and Fo-Fc maps were generated using FFT and the active sites of the _Os_GSTU1 protein subunits examined using Quanta. Clear density was seen in each active site for the DNP-GSH conjugate. A model of the DNP-GSH conjugate was built using the program Cerius$^2$ and energy minimised using Insight II. A

running LIBCHECK through the Monomer Library Sketcher in the CCP4 Interface. Three more cycles of rebuilding and refinement were carried out using NCS restraints followed by three cycles of rebuilding and refinement with restraints. During building, eighty-nine solvent molecules, two GSH molecules, two DNP-GSH molecules, three glycerol molecules, one Tris molecule, and two calcium ions were added, and the $R_{work}$ had been reduced to 17.7% and the $R_{free}$ to 30.9%. Refinement statistics are shown in table 7.3.

| | |
|---|---|
| Space Group | $P3_221$ |
| a (Å) | 76.613 |
| b (Å) | 76.613 |
| c (Å) | 146.072 |
| Resolution Range (Å) | 35.0 – 3.1 |
| No of unique reflections in resolution range | |
| $R_{Work}$ (%)[a] | 17.7 |
| No of reflections used in $R_{free}$ | 476 (5.2%) |
| $R_{Free}$ (%) | 30.9 |
| Total number of protein atoms | 3508 |
| Total number of solvent atoms | 89 |
| Total number of hetero atoms | 132 |
| Average B factor ($Å^2$) | |
| Main-chain atoms | 28.08 |
| Side-chain atoms | 27.11 |
| Solvent atoms | 23.69 |
| Hetero atoms | 47.86 |
| Root mean square deviation | |
| Bonds (Å) | 0.021 |
| Bond angles (°) | 2.125 |
| Ramachandran quality, % in[b] | |
| Most favoured regions | 87.5 |
| Allowed regions | 12.0 |

[a] $R = \Sigma_{hkl}( |F_{obs} - F_{calc}|) / \Sigma|F_{obs}|$

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

**Table 7.3.** Refinement statistics for *Os*GSTU1-1 with DNP-GSH in $P3_221$.

In addition to the contacts described in section 5.2.1, the DNP-GSH conjugate appears to be forming few other contacts with *Os*GSTU1. The main contact between the protein and CDNB seems to be a hydrogen bond from the hydroxyl of Tyr 116 to the centre of the aromatic ring of CDNB. The distance between the oxygen and the centre of the ring is 3.2Å (figures 7.7 and 7.8).



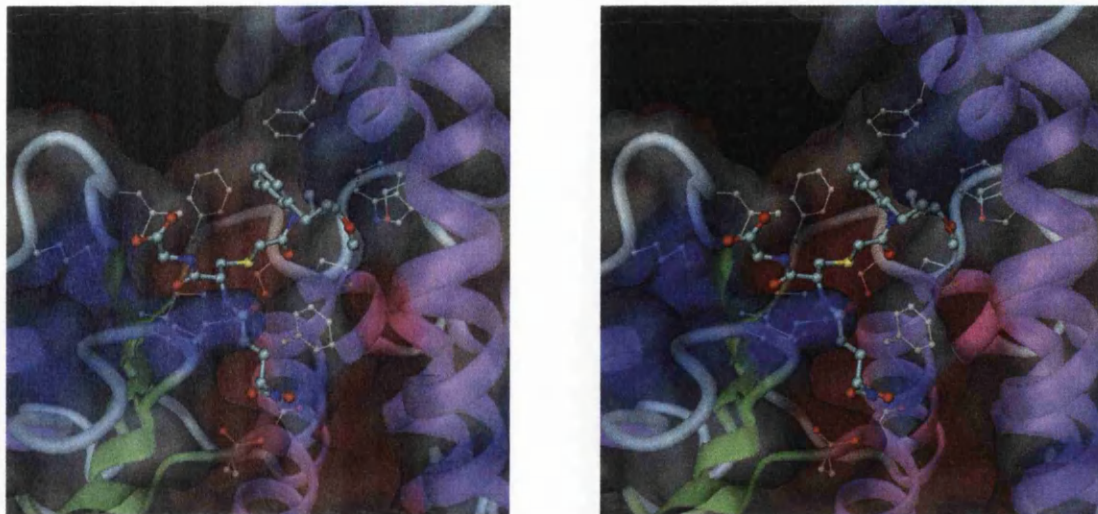**Figure 7.7.** Diagram of the interactions of the DNP-GSH conjugate with *Os*GSTU1. Image generated with LIGPLOT.

**Figure 7.8.** Stereo view of the active site of *Os*GSTU1 with the DNP-GSH conjugate. Residues involved in contacts are also shown. Image generated with Dino (Philippsen, 2003).
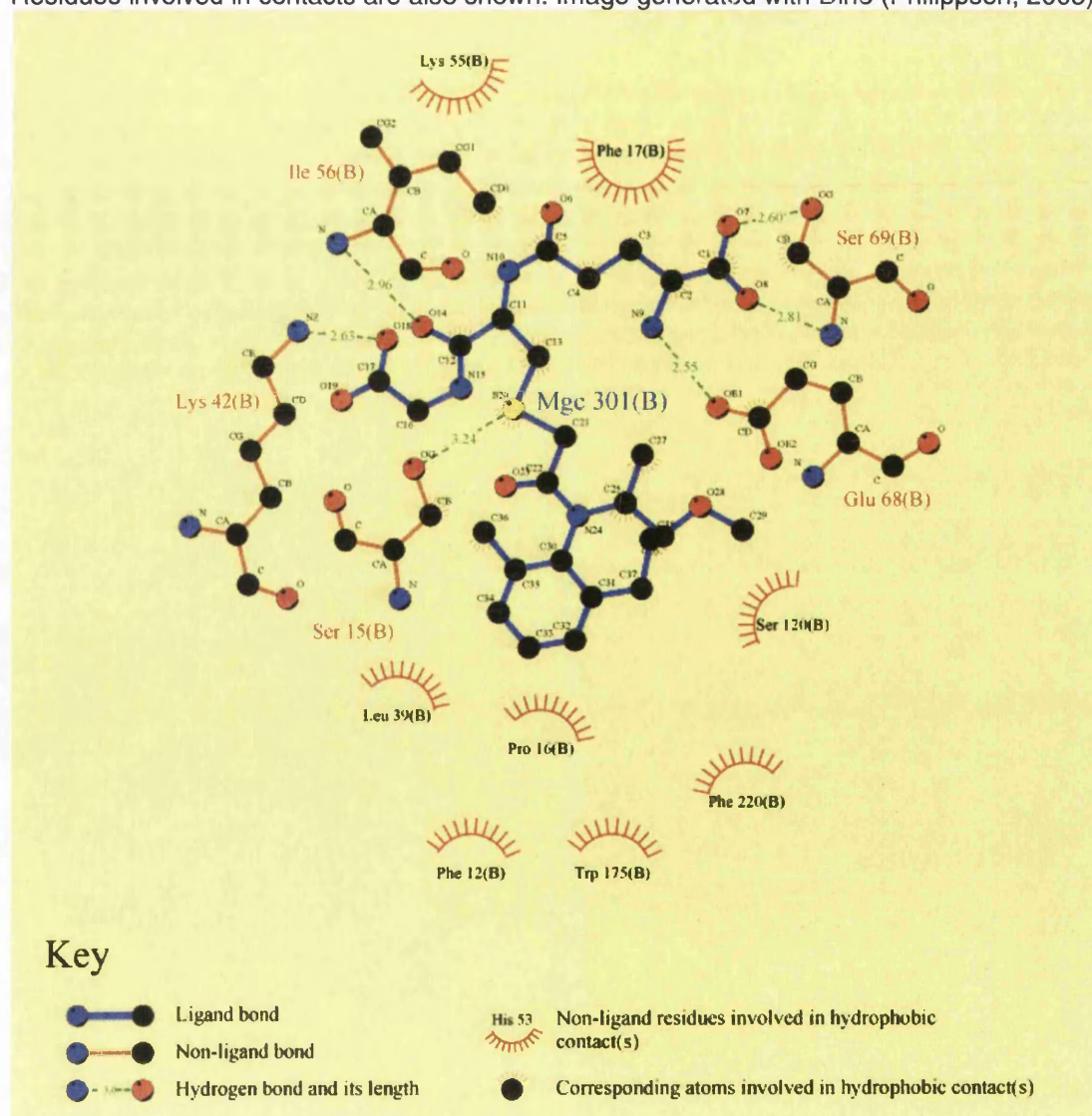
## 7.3   Structure Solution of *Os*GSTU1-1 with NTFP-GSH

Data for *Os*GSTU1-1 co-crystallised with the NTFP-GSH conjugate was collected and processed as described in section 4.2.4. As the structure was in the same space group as that of *Os*GSTU1-1 with DNP-GSH, the same $R_{free}$ set was used. Since the data for *Os*GSTU1-1 with NTFP-GSH was isomorphous with the data used to solve the structure of *Os*GSTU1-1 with DNP-GSH, the phases could be directly obtained by rigid-body refinement with the final structure of *Os*GSTU1-1 with GSH. Ten rounds of rigid body refinement, with the dimer defined as a rigid body group, were performed in REFMAC5. Following this the $R_{work}$ had been reduced to 32.4% and the $R_{free}$ to 35.7%. This was followed by three cycles of refinement with REFMAC5 and a 2Fo-Fc and a Fo-Fc map were generated using FFT.

The active site of the structure was examined using QUANTA and clear density could be seen in both monomers for the NTFP-GSH conjugate. A model of the NTFP-GSH conjugate was built using the program Cerius$^2$ and energy minimised using Insight II. A library for this conjugate was then created by running LIBCHECK through the Monomer Library Sketcher in the CCP4 Interface. After fifteen rounds of rebuilding and refinement and with the

addition of 320 solvent molecules, two GSH molecules, two NTFP-GSH molecules, seven glycerol molecules, two Tris molecules, one calcium ion, and fifteen chloride ions, the $R_{work}$ had been reduced to 16.4% and the $R_{free}$ to 27.2%. Refinement statistics are shown in table 7.4.

Geometric analysis of the structure showed bond and angle deviations from ideality of 0.022Å and 1.915° respectively. Analysis using the program PROCHECK (Laskowski *et al.*, 1993) showed the geometry was better than average for a 2.3Å structure.

| | |
|---|---|
| Space Group | $P3_221$ |
| a (Å) | 77.129 |
| b (Å) | 77.129 |
| c (Å) | 147.873 |
| Resolution Range (Å) | 35.0 – 2.4 |
| No of unique reflections in resolution range | |
| $R_{Work}$ (%)[a] | 16.4 |
| No of reflections used in $R_{free}$ | 1029 (5.3%) |
| $R_{Free}$ (%) | 27.2 |
| Total number of protein atoms | 3605 |
| Total number of solvent atoms | 320 |
| Total number of hetero atoms | 180 |
| Average B factor (Å$^2$) | |
| Main-chain atoms | 52.40 |
| Side-chain atoms | 52.61 |
| Solvent atoms | 53.25 |
| Hetero atoms | 68.80 |
| Root mean square deviation | |
| Bonds (Å) | 0.022 |
| Bond angles (°) | 1.915 |
| Ramachandran quality, % in[b] | |
| Most favoured regions | 91.4 |
| Allowed regions | 8.0 |

[a] $R = \sum_{hkl}( |F_{obs} - F_{calc}|) / \sum|F_{obs}|$.

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

**Table 7.4.** Refinement statistics for *Os*GSTU1-1 with NTFP-GSH in $P3_221$.

In addition to the contacts already described for GSH there are several contacts seen between fluorodifen and the protein. There are two hydrophobic contacts seen between the ligand and two leucine groups. One contact between Leu 39 in the N-terminal domain to the aromatic ring, and one between Leu 224 and the carbon of the trifluoromethyl group. There also again appears to be a hydrogen bond between the centre of the aromatic ring and the hydroxyl group of Tyr 116 at a distance of 3.2Å. There is also a contact between the trifluoromethyl group and the NE of Trp 123 through two water molecules. The hydrogen bonding distances for these contacts are 2.6Å, 2.7Å and 3.1Å. These contacts can be seen in figures 7.9 and 7.10.
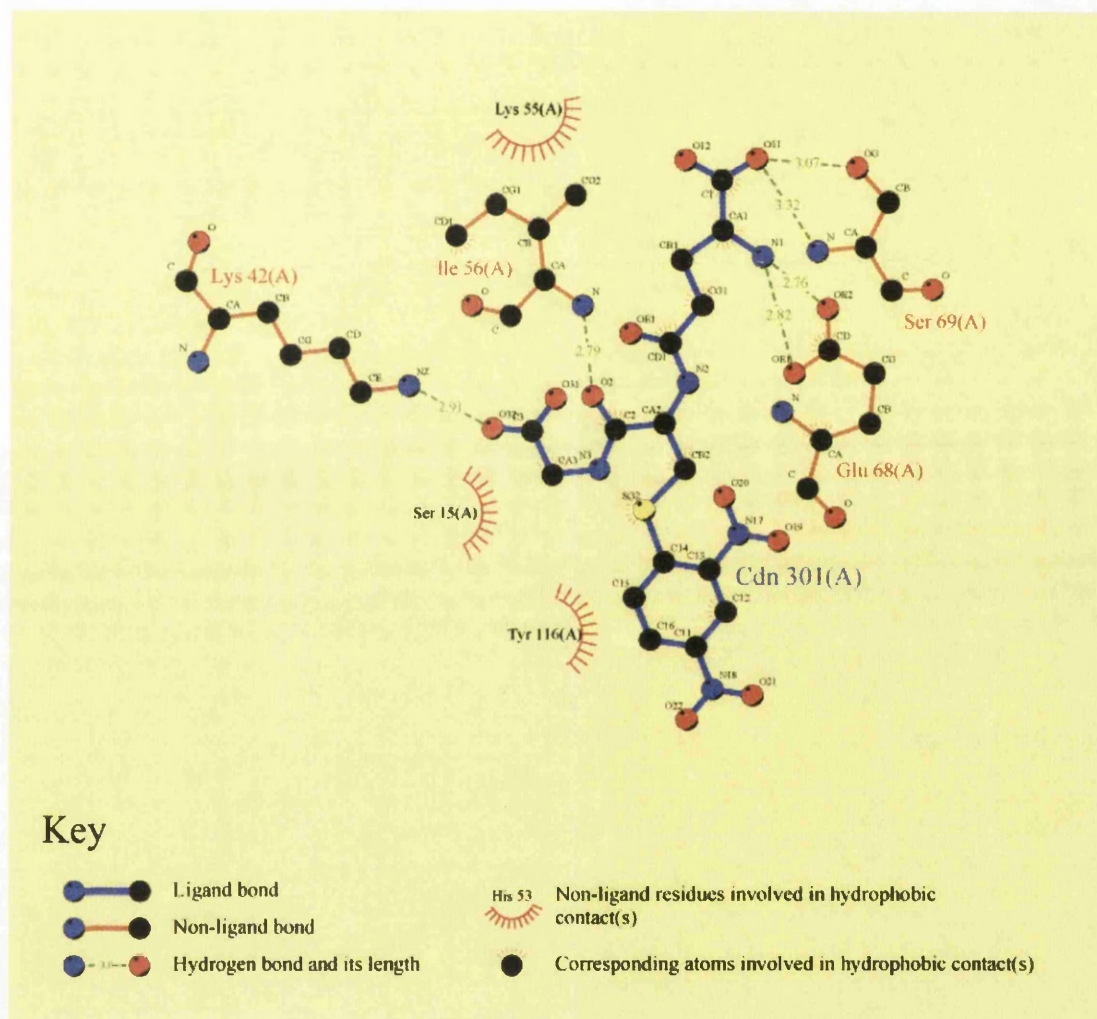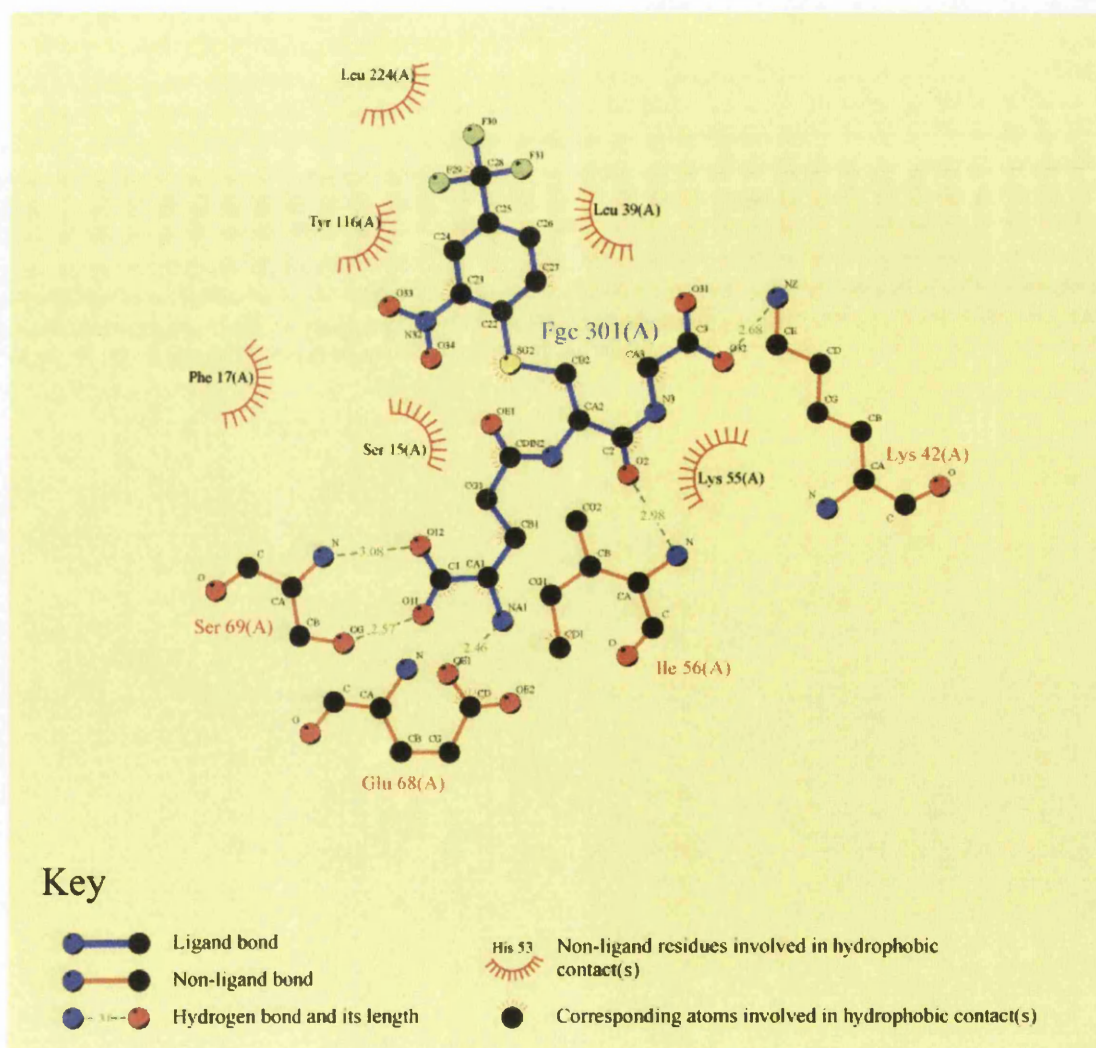


**Figure 7.9.** Diagram of the interactions of the NTFP-GSH conjugate with *Os*GSTU1. Image generated with LIGPLOT.
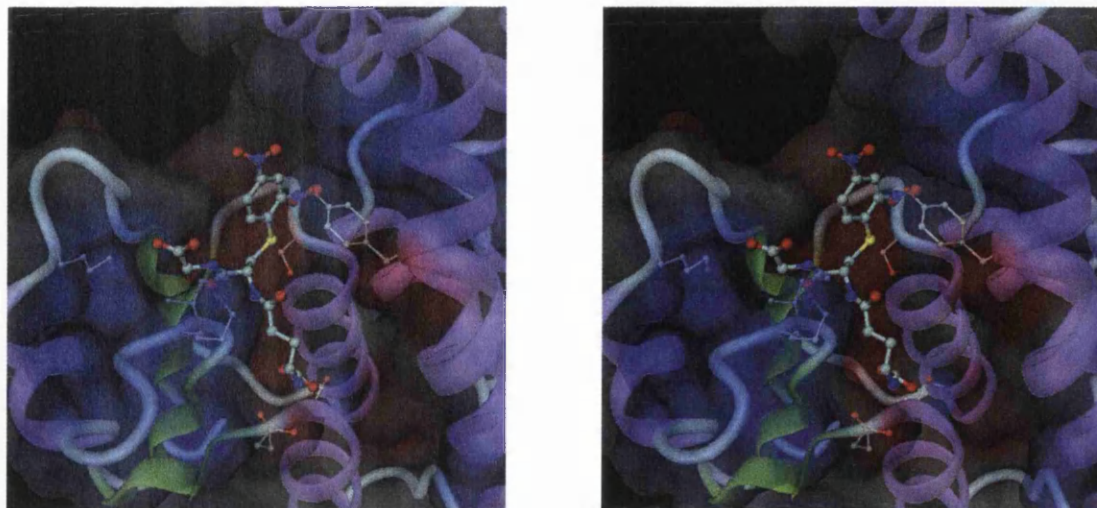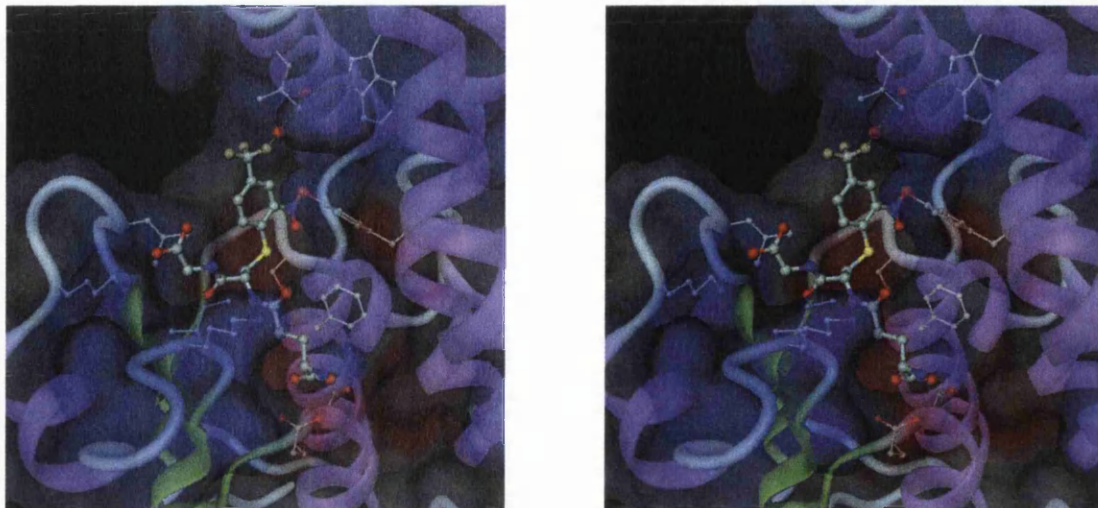
**Figure 7.10.** Stereo view of the active site of *Os*GSTU1 with the NTFP-GSH conjugate. Residues involved in contacts are also shown. Image generated with Dino (Philippsen, 2003).

## 7.4    Comparison of the Four *Os*GSTU1 Structures

If we compare the structure of *Os*GSTU1 with GSH with the three ligand bound structures (figures 7.11 & 7.12) it can be seen that a number of residues adopt different conformations to accommodate the different ligands within the active site. Looking at the metolachlor-GSH structure the main difference is seen in Tyr 116. In the GSH structure this residue is sitting at the entrance to the aromatic pocket but in the metolachlor-GSH structure this residue has swung out of position to allow the bulky ligand access to the pocket and is disordered in the structure and cannot be seen. The only other major shift in conformation is seen in Leu 224. Although this residue is not making any contacts with the ligand itself it seems to be moving to form better contacts with Trp 123 and Phe 220 which are making small movements on ligand binding.

In the NTFP-GSH structure the tyrosine at position 116 has moved slightly closer to the active site and is in a good position to form a hydrogen bond to the centre of the aromatic ring of the ligand. The top of helix α-9 has also made an approximately 2.5Å shift towards the active site bringing Leu 224 into a position where it can form a hydrophobic contact with the carbon of the trifluoromethyl group on the ligand. Lys 227 has also made a large change in

conformation, by swinging down towards the active site, but does not seem to be making any sort of contact with the ligand at all.
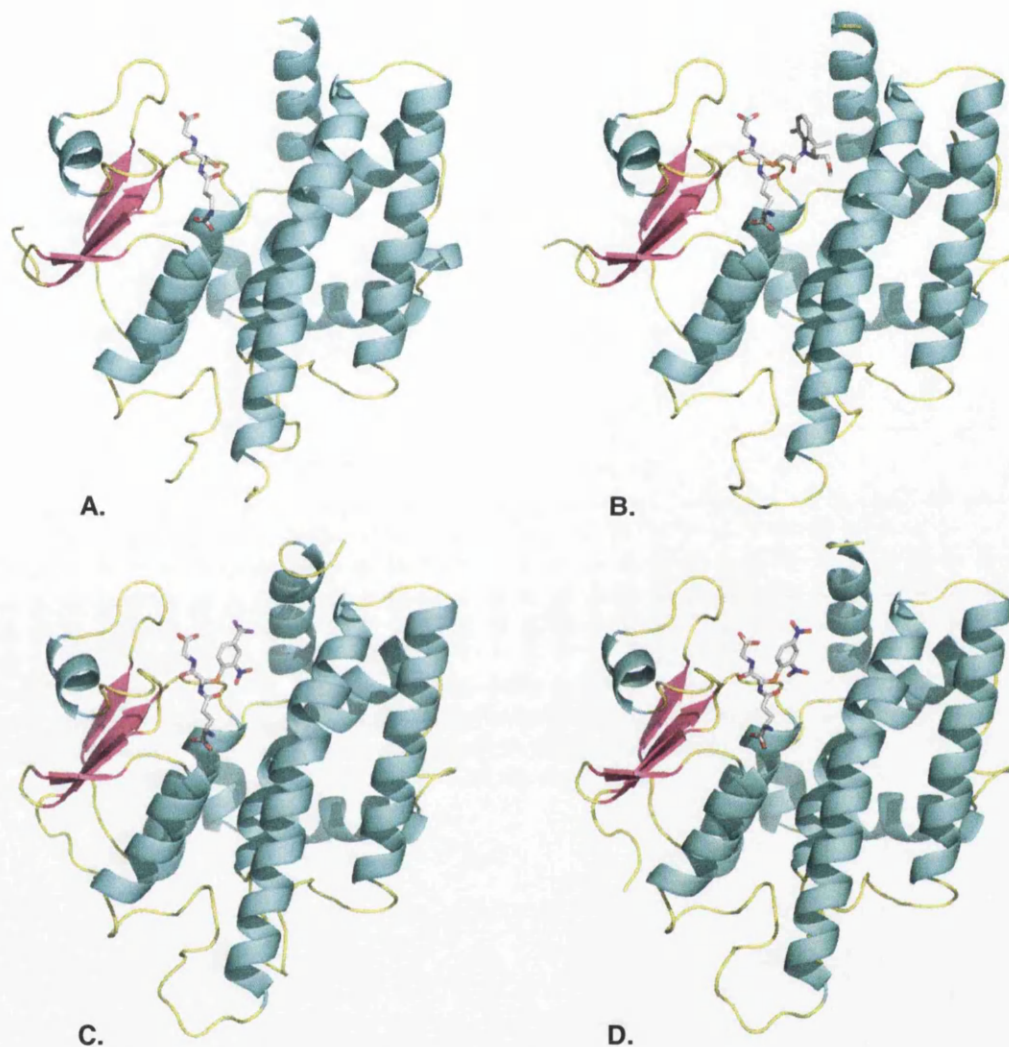


**Figure 7.11.** Diagram of *Os*GSTU1 with different ligands. **A.** GSH **B.** Metolachlor-GSH conjugate **C.** NTFP-GSH conjugate **D.** DNP-GSH conjugate Diagram generated with PyMOL (Delano, 2002).

In the structure with DNP-GSH bound, which is similar in size and structure to NTSP-GSH, the tyrosine at position 116 is also in a position to form a hydrogen bond to the centre of aromatic ring in the ligand. In this structure the top of helix α-9 is seen to make a 3.5Å shift towards the active site but in this case no further contacts can be seen between this helix and the ligand. Although the density for this structure was too poor to build in the side chain for Lys 227 the helix is in such a position that it would be possible for the side

chain of this residue to sit in hydrogen bonding distance from the nitro group of the ligand.
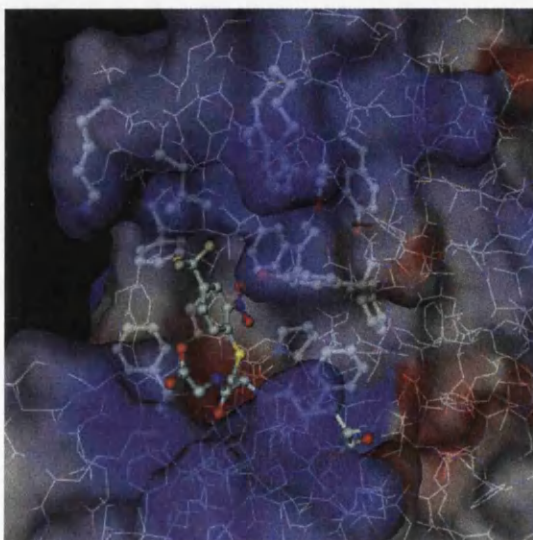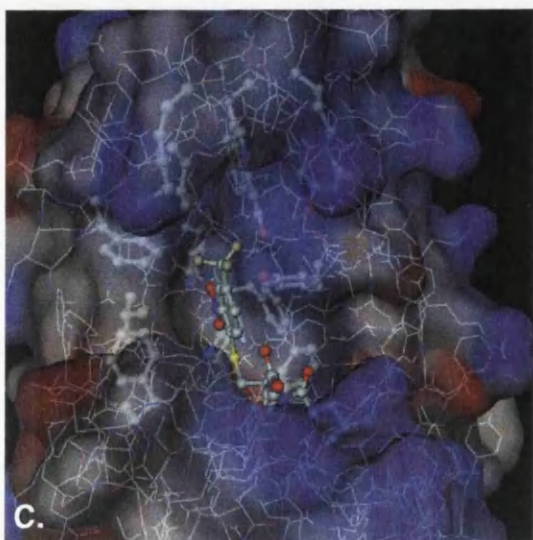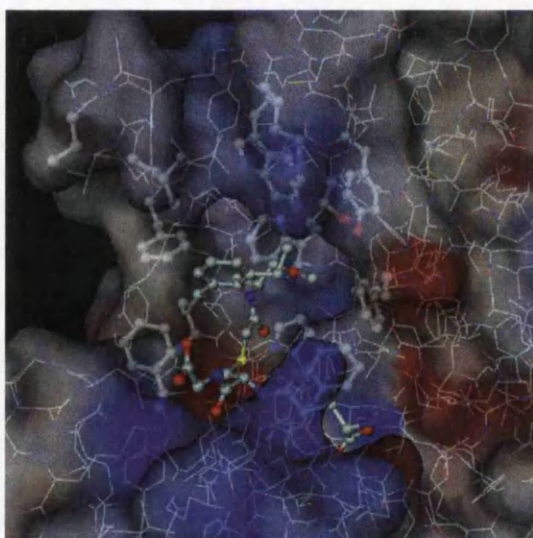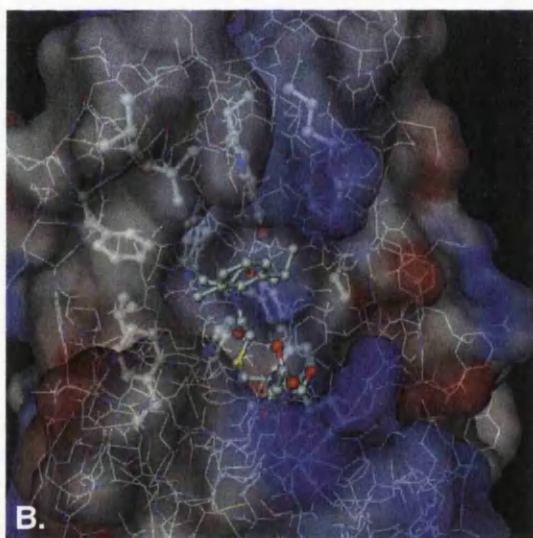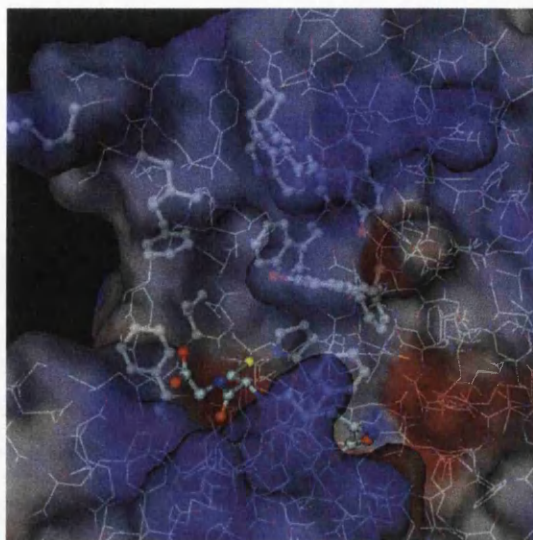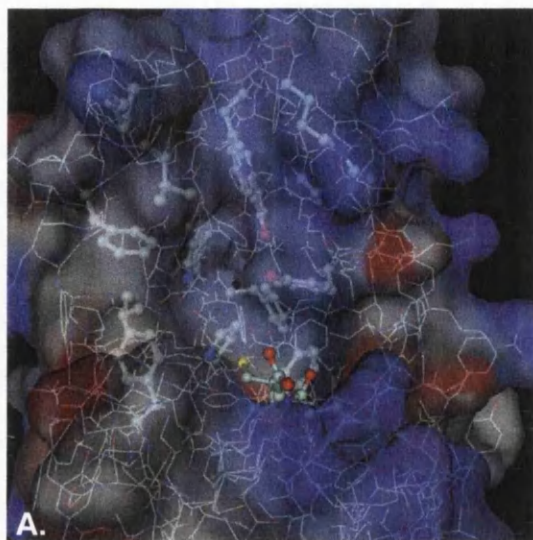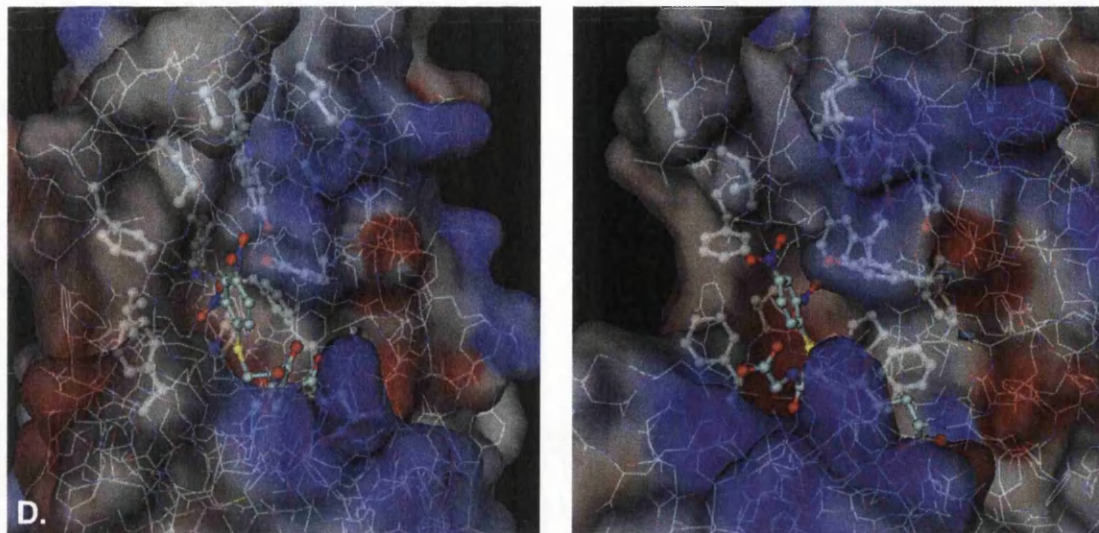
**Figure 7.12.** Images of the active site of *Os*GSTU1 with different ligands and H-site residues highlighted. Each of the structures H-site is shown in two different orientations. **A.** GSH. **B.** Metolachlor-GSH. **C.** NTFP-GSH. **D.** DNP-GSH. Images generated with Dino (Philippsen, 2003).

## 7.5    Chloracetanilide Herbicide Binding

The H-site of *Os*GSTU1 has two lobes, 'A' which is surrounded by Phe 12 & 220, Trp 13, and Val 14, and 'B' which is surrounded mainly by Pro 16, Tyr 116, and Trp 175 (figure 7.13 A). As can be seen in the structure of *Os*GSTU1 with metolochlor-GSH, Tyr 116 can swing away from the active site, opening up lobe 'B' to accommodate large, non-planar ligands (figure 7.13 B). The side chain of Tyr 116 is in fact disordered in the structure with metolachlor-GSH bound and plays no part in the binding of chloracetanilide herbicides. Figure 7.14 shows a sequence alignment of *Os*GSTU1 with four Tau class GSTs from soybean highly active against chloracetanilide herbicides (McGonigle *et al.*, 2000). None of these enzymes have a tyrosine residue at this position confirming that it is not involved in GSH conjugation of chloracetanilides. Instead the tyrosine acts in *Os*GSTU1 as a gatekeeper to lobe 'B' restricting the types of substrate that can bind.

The structure of *Os*GSTU1 with metolachlor-GSH is a good model of chloracetanilide binding as there is no significant conformational shift in the herbicide structure during conjugation. The reaction proceeds by a simple Sn2
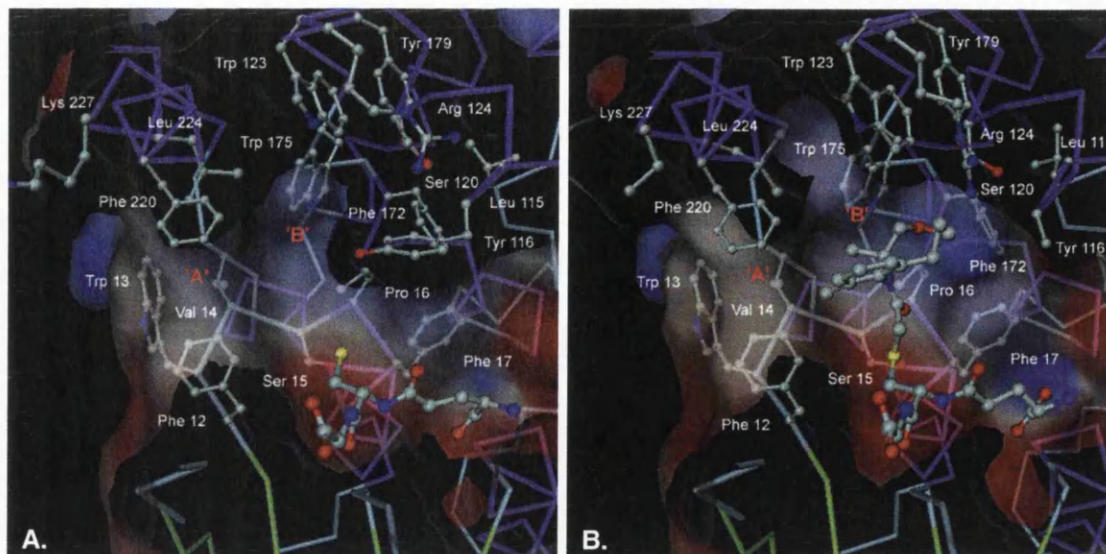
**Figure 7.13.** Cutaway diagrams of *Os*GSTU1 active site. **A.** *Os*GSTU1 with GSH. **B.** *Os*GSTU1 with metolachlor-GSH. Images generated with Dino (Philippsen, 2003).

substitution mechanism with a chloride ion released as the leaving group. That few of the residues interacting with the herbicide are conserved across the sequences of Tau class GSTs with high activity for chloracetanilides suggests that the position of lobe 'B' relative to the G-site may be more important than the exact residues forming lobe 'B'.
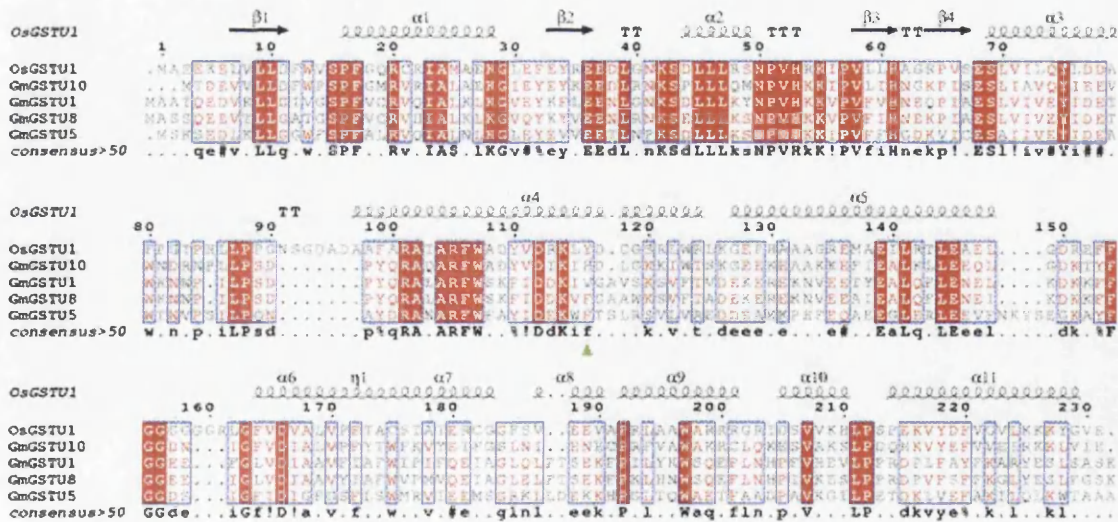


**Figure 7.14.** Alignment of the sequence of *Os*GSTU1 with those of four Tau class GSTs from soybean highly active in chloracetanilide detoxification. The position of residue 116 is indicated with a green arrow. Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

## 7.6    Diphenyl Ether Herbicide Binding

Using the structure of *Os*GSTU1 with GSH, and the structure of *Ta*GSTU4 (PDB ID 1GWC; Thom *et al.*, 2002), the structure of *Zm*GSTU2 has been modelled using SWISS-MODEL, and a molecule of GSH and a molecule of fluorodifen docked into the active site (figure 7.15 B; Dixon *et al.*, 2003; Appendix C). *Zm*GSTU2 has very similar active site architecture as *Os*GSTU1 and in this structure the model of the 4-nitrophenyl moiety of fluorodifen is sitting in lobe 'A', Tyr 110 has moved away from the H-site and the 2-nitro-4-(trifluoromethyl)-phenyl moiety is sitting in lobe 'B' in good geometry for nucleophilic attack from the cysteinyl moiety of GSH. In the crystal structure of *Os*GSTU1 with NTFP-GSH (figure 7.15 A) the 2-nitro-4-(trifluoromethyl)-phenyl moiety is partially occupying lobe 'B', with the side chain of Tyr 116 occupying the rest of the pocket. As a result the detailed orientation of this group is different, the nitro group on the ring is facing in towards the pocket as opposed to the trifluoromethyl group, and Tyr 116 hydroxyl is forming a hydrogen bond to the centre of the aromatic ring, while in the model it hydrogen bonds with the nitro group.
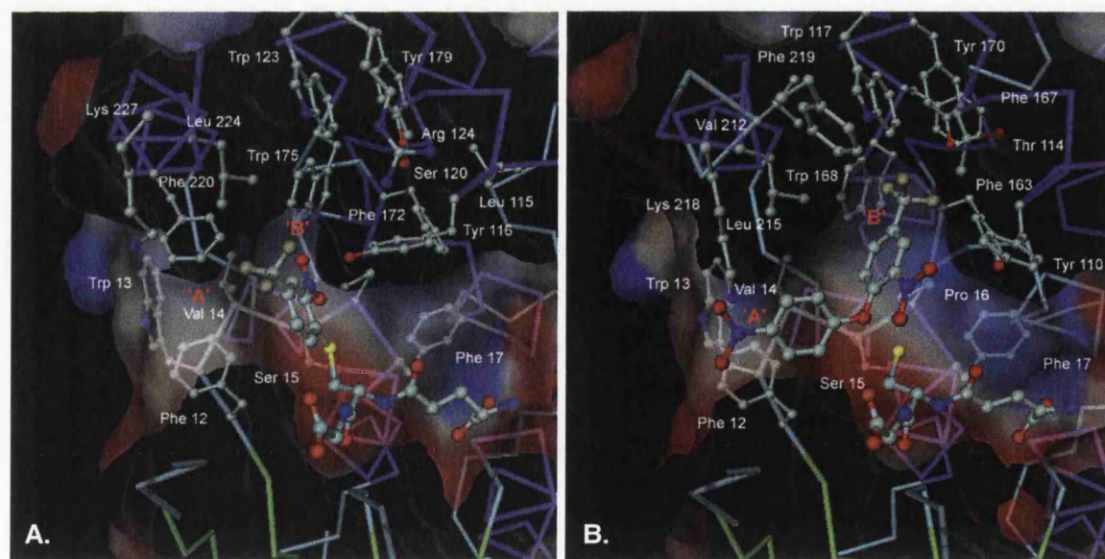


**Figure 7.15.** Cutaway diagrams of Tau class GST active sites. **A.** *Os*GSTU1 with NTFP-GSH. **B.** Predicted structure of *Zm*GSTU2 with fluorodifen and GSH docked. Images generated with Dino (Philippsen, 2003).

The differences between the model and the structure are to be expected, as the herbicide must be oriented so that the sulphur of the thiolate anion of GSH can attack from outside the plane of the ring. The reaction causes a change in the conformation of the molecule, as in the product the glutathione sulphur is in the plane of the aromatic ring. It is conceivable that this conformational change is associated with product release. This raises the question whether the NTFP-GSH conjugate structure can inform us of the correct orientation of the diphenyl ether substrate.

If the model of fluorodifen binding is correct then only one of the susceptible ether carbons is suitably oriented for GSH conjugation. The other ring sits flat in lobe 'A' where the carbon atom is inaccessible. This said, examining the chemical structures of three common diphenyl ether herbicides and their products (figure 7.16) it seems unlikely that the phenyl ring to be attacked must sit in lobe 'B'. For acifluorfen and fomesafen the bulky substituents on this ring could not occupy this pocket for steric reasons, however in both cases these substituents could face out of the pocket.



**Fluorodifen**              **Acifluorfen**              **Fomesafen**
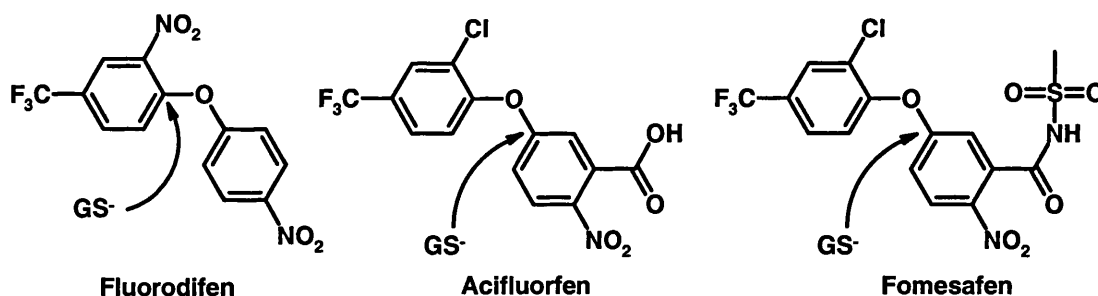
**Figure 7.16.** The structure of three diphenyl ether herbicides. The point of attack of GS⁻ on each of the herbicides is indicated. Image generated with ISIS/Draw (MDL Information Systems, Inc).

Comparing the conjugation site on each herbicide it can be seen that the carbon that is being attacked is the one either side of the ether moiety, and importantly, is always the most electrophilic (and hence reactive) of the pair which is governed by the electron withdrawing substituents on the rings. If in all three cases the trifluoromethyl-substituted ring occupies lobe 'B' then the

GSH thiolate can attack either of the two ether carbons on the herbicide. This would suggest that our model of fluorodifen binding is incorrect.

Alternatively, as these compounds are not natural substrates for the enzyme it is possible that diphenyl ether herbicides can bind with either of the aromatic rings within lobe 'B'. If this was the case then we would expect to see a mixture of reaction products, however this is not observed (Lamoureux *et al.*, 1993).

From the crystal structure of the NTFP-GSH conjugate we see that the nitro group partially occupies lobe 'B'. That a chloro group in both acifluorfen and fomesafen replaces this could explain why the structurally similar aromatic ring is not the site of conjugation, as this group may no longer bind to lobe 'B'. Clearly the structures do not fully explain the observed data, however they suggest a number of alternative solutions, which can be tested experimentally.

# Chapter 8        Structure of *Os*GSTU4-4

## 8.1    Structure Solution of *Os*GSTU4-4 with GSH

Data for *Os*GSTU4-4 with GSH was collected to 2.3Å and was processed as described in section 2.4.5. Comparison of the amino acid sequence with those of *Os*GSTU1 and *Ta*GSTU4 (figure 8.1) was used to identify the best search model for the *Os*GSTU4-4 structure. The sequence of *Os*GSTU4 shares 43% identity (56% similarity) with *Os*GSTU1 and 68% identity (76% similarity) with *Ta*GSTU4 clearly indicating that the latter structure would be a better phasing model. A monomer of the complete *Ta*GSTU4 structure was used as a search model in the molecular replacement program AMoRe using data between 10.0-4.0Å. A clear solution for one monomer in the AU was found with a correlation coefficient of 50.5% and an R factor of 57.0% (table 8.1 and figure 8.2) compared to 37.8% and 62.3% for the next highest solution.
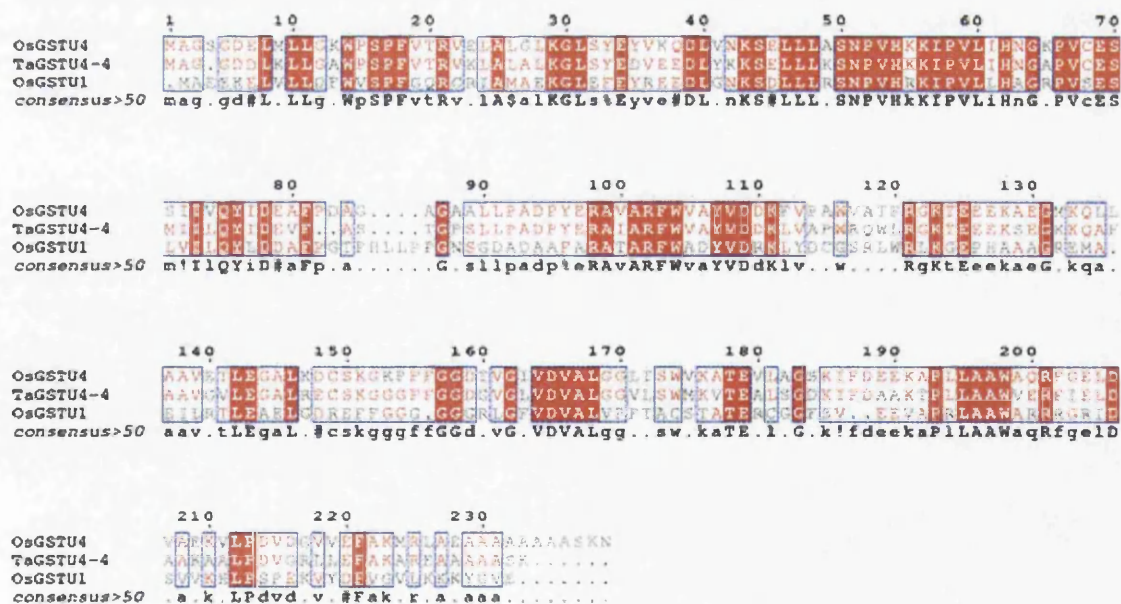


**Figure 8.1.** Alignment of the sequence of *Os*GSTU4 with *Os*GSTU1 and *Ta*GSTU4. Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

| | Correlation Of Intensities (%) | R-factor (%) |
|---|---|---|
| Rotation Function | 13.1 | 67.1 |
| Translation Function | 31.8 | 63.8 |
| Rigid Body Fitting | 50.5 | 57.0 |

**Table 8.1.** Solutions found by AMoRe for *Os*GSTU4-4.
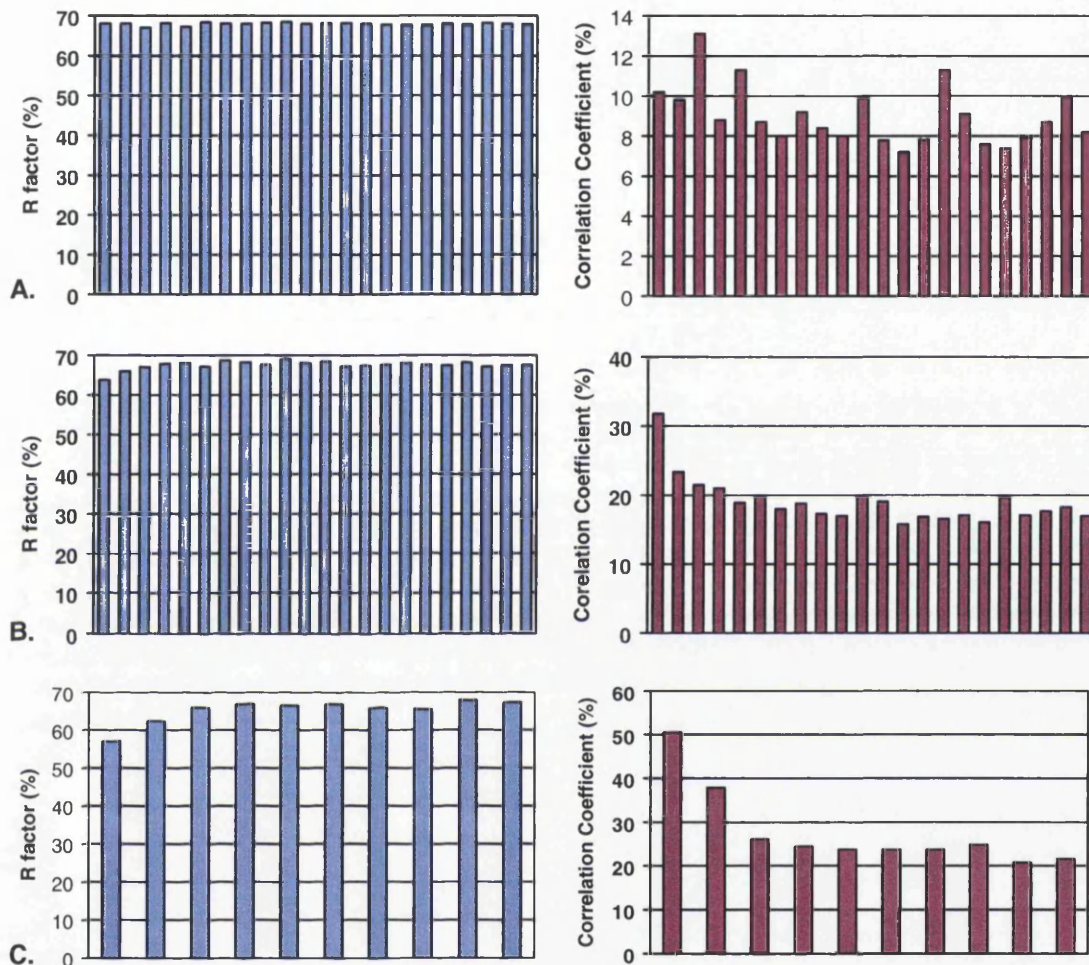


**Figure 8.2.** Statistics from each stage of AMoRe with Crystallographic R-factor shown in blue and the Correlation of Intensities shown in burgundy. From top to bottom they are statistics from: **A.** the rotation function **B.** the translation function **C.** after rigid-body refinement.

| | | | |
|---|---|---|---|
| | 0.00215 | 0.61438 | 0.78901 |
| Rotation Matrix | -0.99934 | -0.02726 | 0.02395 |
| | 0.03622 | -0.78854 | 0.61392 |
| Translation | -1.85 | 70.17 | 31.33 |

The packing of the probe structure in the unit cell was scrutinized using the graphics program SETOR (Evans, 1993). The absence of overlapping regions of structure and clear crystal contacts visually confirmed that the solution was correct (figure 8.3).
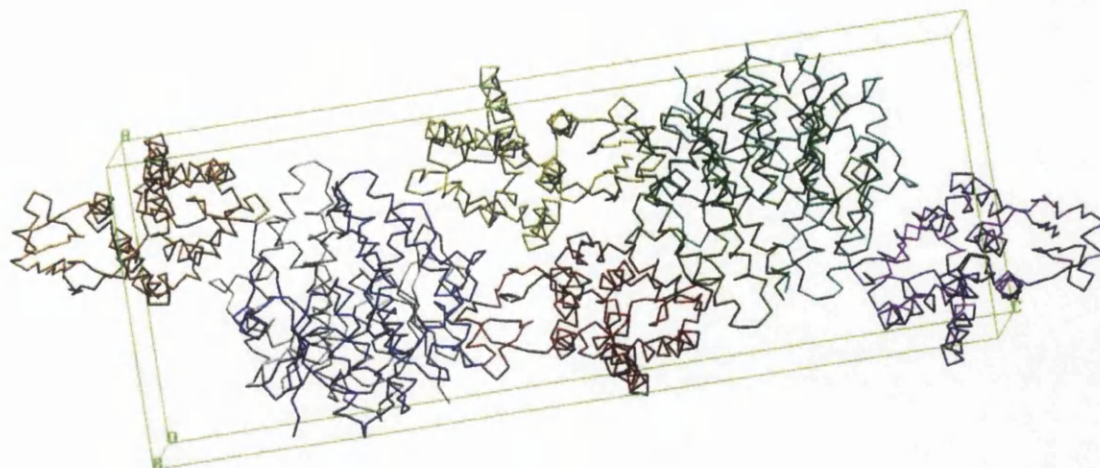
**Figure 8.3.** Packing of *Os*GSTU4-4 within the unit cell. Image generated with SETOR (Evans, 1993).

REFMAC5 was used to perform ten cycles of rigid body refinement with each domain of the monomer defined as a rigid body group. This was followed by three rounds of restrained refinement with REFMAC5 after which weighted 2Fo-Fc and Fo-Fc maps were generated from the resultant model using FFT. After this, fifteen rounds of rebuilding using QUANTA and refinement with Refmac5 were carried out. During building 198 solvent molecules, two molecules of GSH, two molecules of Tris, four calcium ions, and five chloride ions were built into the structure until the $R_{work}$ had been reduced to 16.6% and the $R_{free}$ to 24.6%. Refinement statistics are shown in table 8.2.

Analysis of the geometry of the structure showed bond and angle deviations from ideality of 0.018Å and 1.642° respectively. Analysis using the program PROCHECK (Laskowski *et al.*, 1993) showed the geometry was better than average for a 2.3Å structure.

| Space Group | P4₁22 |
|---|---|
| a (Å) | 56.092 |
| b (Å) | 56.092 |
| c (Å) | 157.152 |
| Resolution Range (Å) | 32.0 – 2.3 |
| No of unique reflections in resolution range | |
| $R_{Work}$ (%)[a] | 16.6 |
| No of reflections used in $R_{free}$ | 1120 (9.8%) |
| $R_{free}$ (%) | 24.6 |
| Total number of protein atoms | 1766 |
| Total number of solvent atoms | 198 |
| Total number of hetero atoms | 65 |
| Average B factor (Å²) | |
| Main-chain atoms | 40.98 |
| Side-chain atoms | 39.64 |
| Solvent atoms | 48.94 |
| Hetero atoms | 70.01 |
| Root mean square deviation | |
| Bonds (Å) | 0.018 |
| Bond angles (°) | 1.642 |
| Ramachandran quality, % in[b] | |
| Most favoured regions | 89.7 |
| Allowed regions | 10.3 |

[a] $R = \Sigma_{hkl}( |F_{obs} - F_{calc}|) / \Sigma|F_{obs}|$

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

**Table 8.2.** Refinement statistics for OsGSTU4-4 with GSH in P4₁22.

## 8.2    Structure of OsGSTU4-4 with GSH

The OsGSTU4-4 structure is a homo-dimer, with each molecule composed of two domains. The N-terminal domain (residues 1-81) is a α/β domain and contains the GSH binding site. There is then a linker region (residues 82-94) that connects the N-terminal domain to the C-terminal domain (residues 95-239), which is composed of six helices and contains the hydrophobic binding pocket. The structure of OsGSTU4-4 is shown in figure 8.4. The most notable difference to the other Tau class GST structures is the long C-terminal

extension to helix α-9, which is twenty-six residues long in *Os*GSTU4 and contains a chain of eight alanine residues, in comparison to nineteen residues in *Os*GSTU1 and twenty residues in *Ta*GSTU4.
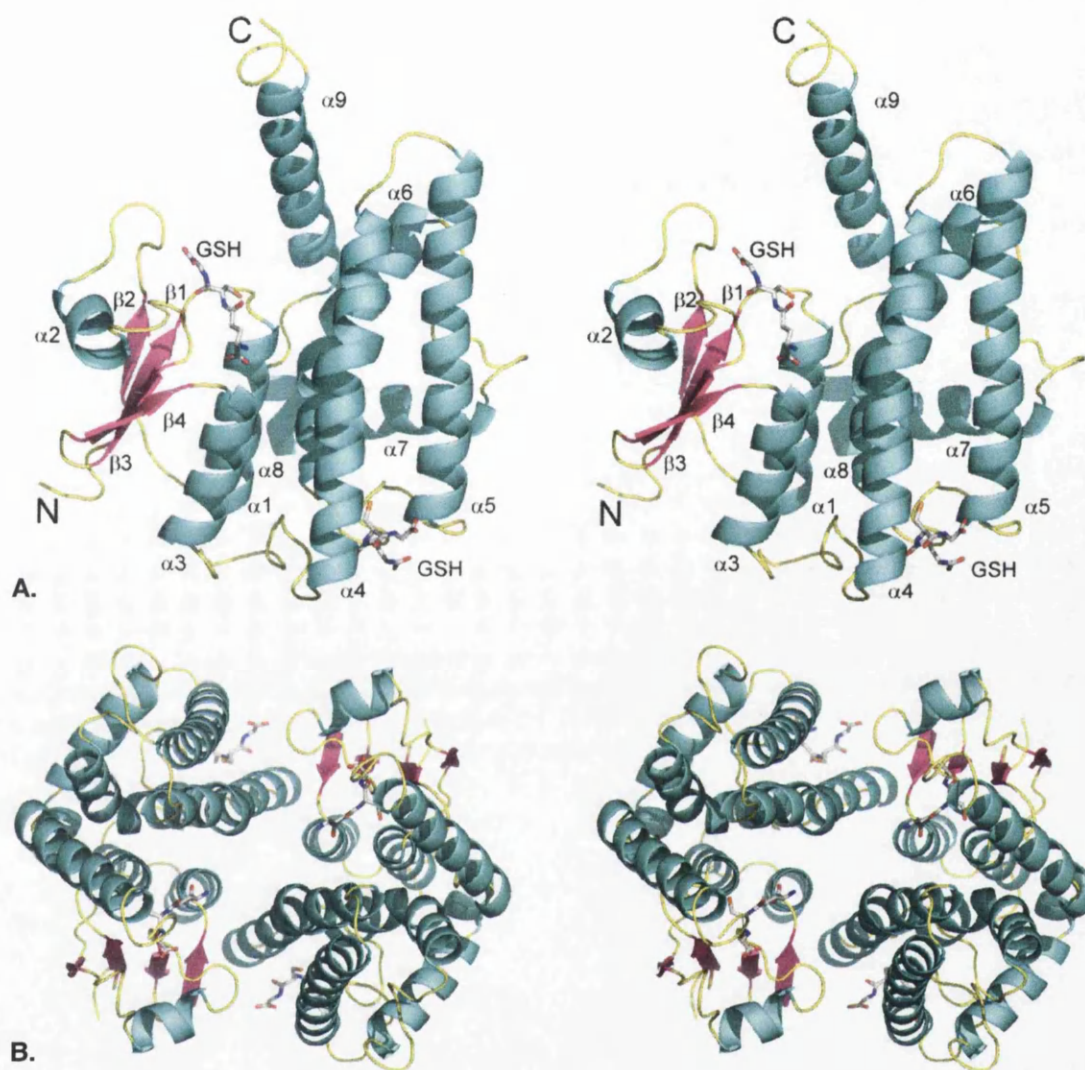


**Figure 8.4.** Stereo images of the monomeric and dimeric structure of *Os*GSTU4-4 with GSH. Images generated with PyMOL (Delano, 2002).

### 8.2.1  N-Terminal Domain

The N-terminal domain of *Os*GSTU4 is similar to that of *Os*GSTU1 described in section 6.2.1 composed of a four-strand β-sheet, of strand order β-2 (34-37) β-1 (8-12) β-3 (59-62) β-4 (65-68), with strand β-3 running antiparallel to the others. This is flanked by three α-helices, helix α-1 (17-29) and helix α-3 (70-81) below the plane of the sheet, and helix α-2 (45-51) above the sheet facing the solvent (figure 8.4 A). An interesting difference seen between this

structure and that of the other Tau class GST structures is seen in the short loop that connects strand β-1 to helix α-1. In this loop the proline residue at position 15 is in the less favourable *cis* conformation. This proline residue is seen in *Ta*GSTU4 but is in the *trans* conformation.

The G-site of *Os*GSTU4 is located in the cleft formed between the two motifs in the N-terminal domain. The structure of *Os*GSTU4 contains a molecule of GSH bound to the G-site and forms a number of favourable interactions with the protein (figure 8.5), similar to those seen in *Os*GSTU1. The GSH molecule interacts with the side-chains of Lys 43 and Glu 69, the side-chain hydroxyl and main-chain nitrogen of Ser 70, and the main-chain nitrogen and carbonyl oxygen of Ile 57. Hydrophobic interactions are also seen with Phe 18 and Lys 56. The sulphur moiety of GSH, as in the other Tau class GST structures, is in close proximity to the side-chain hydroxyl of Ser 16, which is known to help stabilise the formation of GS$^-$. In addition to these contacts, two water molecules are seen to form contacts with both the protein and GSH. The first water molecule makes a contact with the O11 atom of GSH and with the main-chain carbonyl oxygen of Pro 58 and the main-chain nitrogen of Glu 69. The second water molecule forms contacts with the O12 atom of GSH and with the main-chain nitrogen of Ser 71.

There then follows a thirteen residue linker region (82-94), of sequence PDAGAGAALLPAD, that connects the N-terminal domain to the C-terminal domain. This is two residues shorter than the linker in *Os*GSTU1 (FPGTPHLLPPANSGA), which contains four prolines, but three residues longer than the linker region of *Ta*GSTU4 (ASTGPSLLPA), which also contains two prolines. All three linkers contain a LLPA motif (LLPPA in *Os*GSTU1), which is fairly well conserved throughout Tau class GSTs.
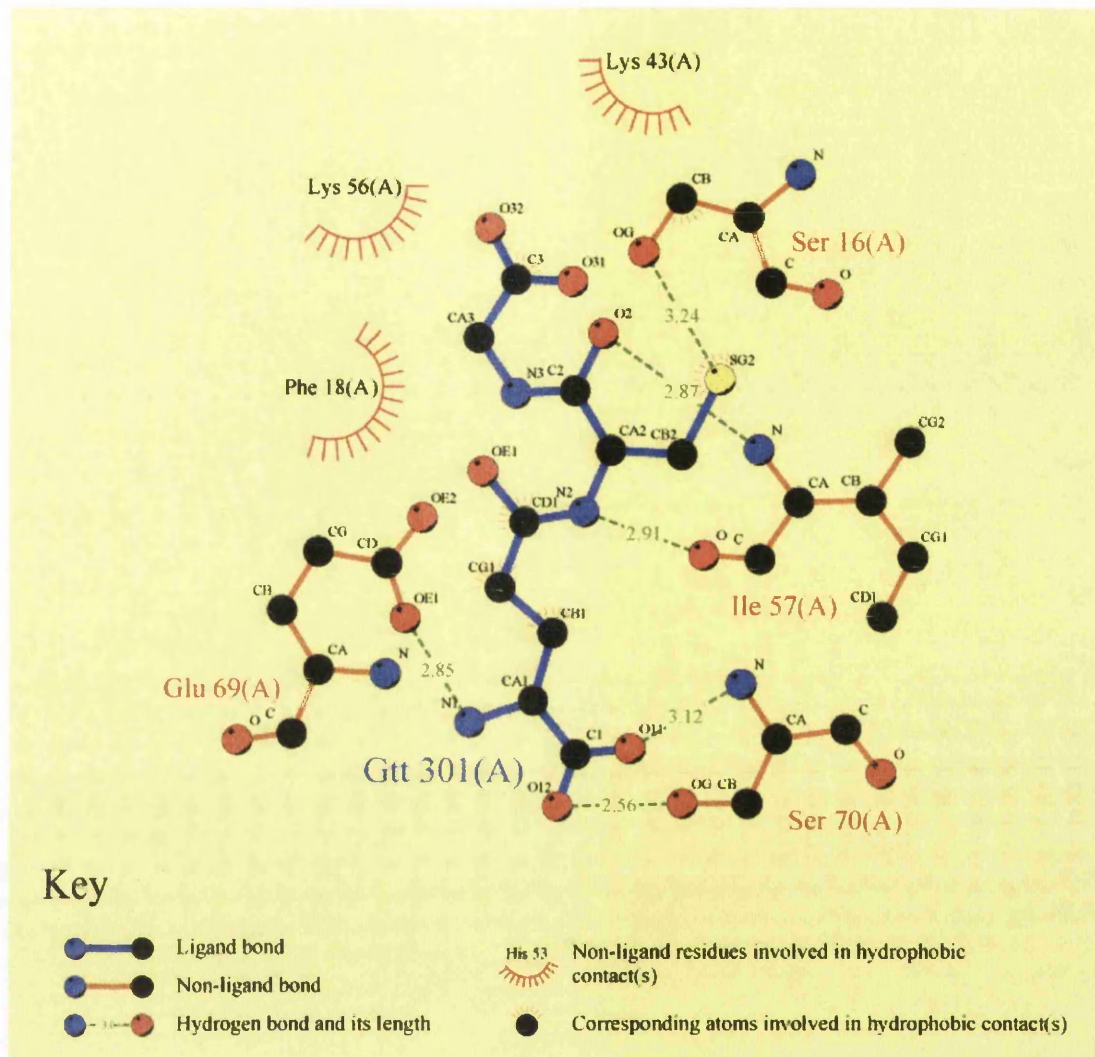
**Figure 8.5.** Diagram of interactions between GSH and the G-site of *Os*GSTU4. Image generated with LIGPLOT (Wallace *et al.*, 1995).

### 8.2.2 C-Terminal Domain

The C-terminal domain (95-239) of *Os*GSTU4, like that of the other Tau class GST structures, is composed of six, partially interrupted α-helices connected by a variety of loops. The first two helices (α-4 95-120, α-5 125-150) are relatively long and helix α-4 is broken at position 111. Helix α-5 is connected to helix α-6 (163-182) by a thirteen residue long flexible loop. Helix α-6 is distorted by two glycine residues (169 & 170) and along with helices α-4 and α-5 forms a three-helix bundle that forms the core of the C-terminal domain. Helix α-7 (193-204) runs almost perpendicular to helices α-5 and α-6, and is linked by one residue to the short helix α-8 (206-211) which runs at a 90° angle to helix α-7. Helix α-9 (215-233) crosses the top of helices α-5 and α-6

at around a 65° angle and it is this area that contains most of the hydrophobic-binding pocket (H-site). The proposed H-site of *Os*GSTU4 lies in a fairly open pocket between helices α-4, α-6 and α-9 and is lined with aromatic and other hydrophobic residues. These consist of Val 113, 117 & 219, Trp 116 & 174, phenylalanine 120 & 221, Leu 171 and Ala 222. A few residues from the N-terminal domain also seem to form part of the H-site, namely Pro 15 & 17, Phe 18 and Leu 40. Although the H-site is mostly hydrophobic there are a few charged hydrogen-bond donors in Lys 13 and Arg 121 & 225.

### 8.2.3 Interactions Between Domains

The interface between the N-terminal and C-terminal domains has a buried surface area of 1759Å$^2$ and is dominated by van der Waals interactions between hydrophobic and aromatic residues. The main residues involved in these interactions are Lys 13, Trp 14, Pro 17, Phe 18, Thr 20, Glu 23, Leu 24 & 28, and Val 74 in the N-terminal domain, and Val 161, 166 & 207, Ile 163, Gly 170, Leu 171, Pro 213, and Phe 221 in the C-terminal domain. There are also five hydrogen bonds between the two domains formed by four residues in the N-terminal domain to four residues in the C-terminal domain. These residues are Pro 15, Arg 21, Ser 71, Gln 74, Arg 102, Asp 109, Ala 167 and Trp 174. These interactions are detailed in table 8.3.

| Residue N-term domain | Residue C-term domain | Distance (Å) | Angle (°) |
|---|---|---|---|
| **Hydrogen Bonds** | | | |
| Pro 15 O | Trp 174 NE1 | 2.9 | 130.0 |
| Arg 21 NH2 | Asp 109 OD1 | 2.7 | 142.7 |
| Arg 21 NE | Ala 167 O | 2.9 | 119.3 |
| Ser 71 OG | Asp 109 OD2 | 2.6 | 122.0 |
| Gln 75 OE1 | Arg 102 NE | 3.1 | 124.6 |

**Table 8.3.**   Details of the interactions between the N-terminal and C-terminal domains of *Os*GSTU4.

### 8.2.4 Interactions at the Dimer Interface

The functional biological unit of *Os*GSTU4-4 is a homodimer. This is generated crystallographically from the monomer in the AU by a rotation along axis *a* (symmetry operation *X*, -*Y*, ½-*Z*). The dimer buries a solvent accessible surface area of 2177Å$^2$, which is comparable to the buried surface area of

*Ta*GSTU4-4 (2096Å², Thom *et al.*2002), but smaller than the buried surface area of OsGSTU1-1 (2703Å²). The dimer interface of OsGSTU4-4 has several hydrophilic (32%) and hydrophobic (68%) interactions. There are four hydrogen bonds formed, between Pro 66 on strand β-4 to Tyr 96' on helix α-4, and between Glu 79 on helix α-3 to Arg 102' on helix α-4.

There are nine residues on each monomer involved in van der Waals contacts and form two clusters. The first cluster is formed by Val 53 and His 54 on the loop between helix α-2 and strand β-3, and Ile 72 on helix α-3 of one monomer, and Phe 103', Tyr 107' and Lys 111' on helix α-4 of the second monomer. The second cluster is formed between Lys 65 on strand β-4 and Tyr 76 on helix α-3 of one monomer, and Tyr 96' on helix α-4 of the second monomer (table 8.4). In addition to these contacts there are nine water molecules (four waters and their symmetry related molecules and one water at a special position) bridging the dimer interface and making nine contacts with each monomer (table 8.5).

| Residue Monomer A | Residue Monomer B | Distance (Å) | Angle (°) |
|---|---|---|---|
| **Hydrogen Bonds** | | | |
| Pro 66 O | Tyr 96' OH | 2.7 | 137.6 |
| Glu 79 OE2 | Arg 102' NH2 | 3.2 | 108.3 |
| Tyr 96 OH | Pro 66' O | 2.7 | 137.6 |
| Arg 102 NH2 | Glu 79' OE2 | 3.2 | 108.3 |

**Table 8.4.** Details of the interactions at the dimer interface of OsGSTU4-4.

| Residue Monomer A | Distance (Å) | Water | Residue Monomer B | Distance (Å) |
|---|---|---|---|---|
| Glu 79 OE2 | 2.8 | 2 | Pro 95' O | 2.8 |
| His 54 NE2 | 3.0 | 5 | Phe 103' O | 3.0 |
| Iso 72 O | 3.1 | 12 | Arg 102' NH2 | 2.9 |
| Gln 75 NE2 | 3.0 | 44* | Gln 75' NE2 | 3.0 |
| Asp 110 OD1 | 3.0 | 104 | Asp 110' OD2 | 2.8 |

**Table 8.5.** Details of bridging water molecules at the dimer interface of OsGSTU4-4.

## 8.3   Comparison of Tau Class Structures

A superposition of the three Tau class GST structures known was performed showing a RMSD of 1.165 (over 199 residues) for *Os*GSTU1 and *Os*GSTU4, 1.172 (over 191 residues) for *Os*GSTU1 and *Ta*GSTU4, and 0.941 (over 212 residues) for *Os*GSTU4 and *Ta*GSTU4 (figure 8.6). The N-terminal domains are seen to superimpose well, as would be expected due to their fairly high sequence similarity (figure 8.7). The C-terminal domains, which have much lower sequence similarity, and which contains the H-site, superimpose less well than the N-terminal domains. The major shifts can be seen in the tops of helices α-4 and α-5, which are shifted away from the N-terminal domain in the rice enzymes, and in helices α-6 and α-9 which are shifted away in *Os*GSTU4. It is this region, formed between helices α-4, α-6 and α-9 that contains the H-site
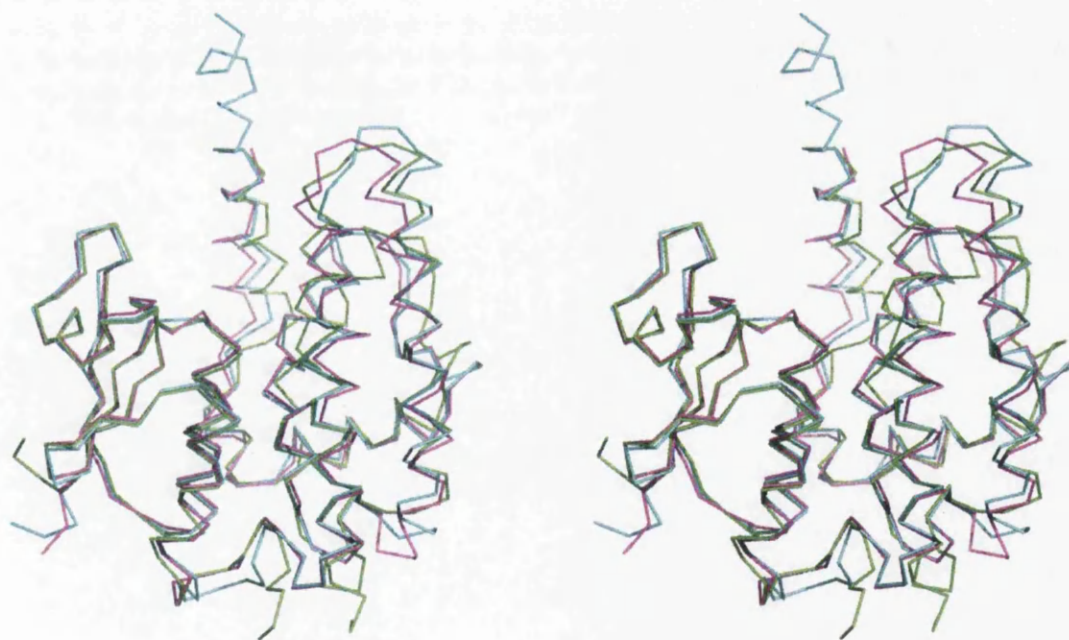


**Figure 8.6.** Stereo diagram of the Cα traces of *Os*GSTU1 (green), *Os*GSTU4 (cyan), and *Ta*GSTU4 (magenta). Image generated with PyMOL (Delano, 2002).
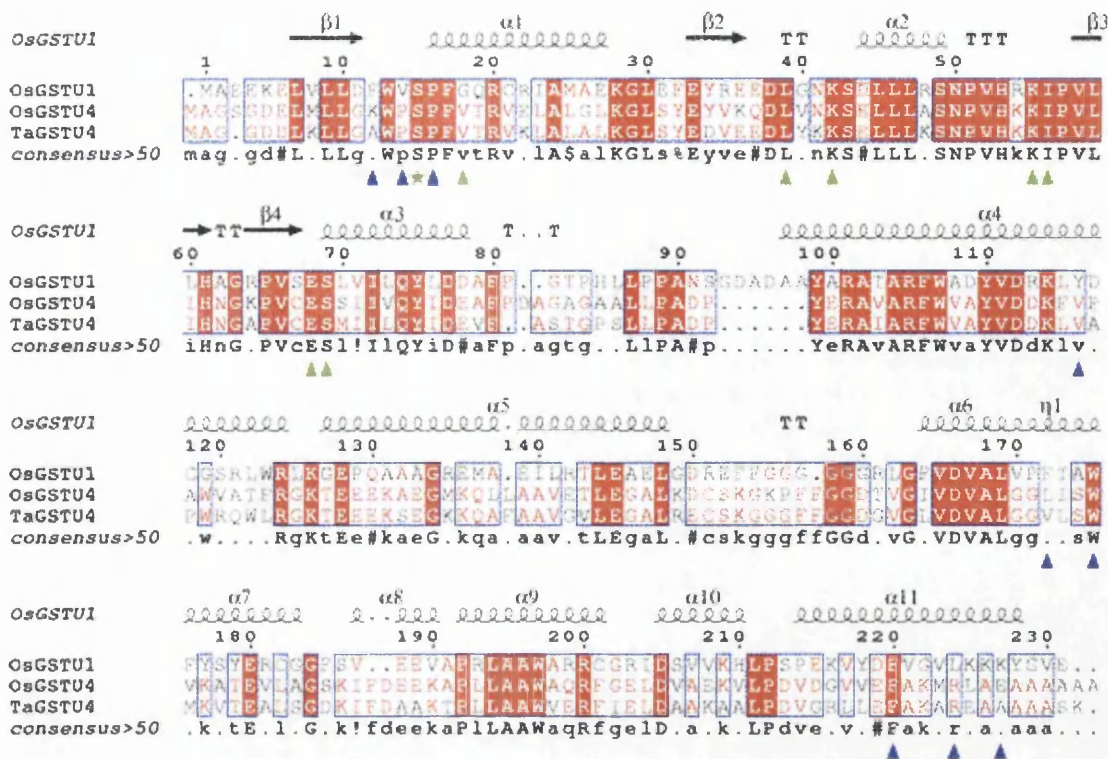
**Figure 8.7.** Sequence alignment of the three Tau class GSTs whose structure has been solved and showing secondary structure elements of *Os*GSTU1. ★ Active site serine. ▲ Proposed G-site residues. ▲ Main residues of the roposed H-site of *Os*GSTU1. Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

Working from the N-terminal to the C-terminal the first major difference between the three Tau class structures is seen at position 11 on strand β-1 (positions refer to *Os*GSTU1). In both *Os*GSTU4 and *Ta*GSTU4 this is a glycine residue but in *Os*GSTU1 this is replaced with an aspartate, which forms a salt bridge with Arg 22. As this is a buried residue, this seems a large change but a complementary change is seen at position 18 where a valine in *Os*GSTU4 and *Ta*GSTU4 is replaced with a glycine in *Os*GSTU1. There is no significant change in the main-chain position at either of these residues despite the change in the buried residues (figure 8.8). In a sequence comparison of 107 Tau class proteins from plants (rice, maize, soybean & *Arabidopsis*; appendix B) it can be seen that with very few exceptions the residue at position 11 is either aspartate (16%) or glycine (78%). If the residue at position 11 is an aspartate then the residue at 18 is always a glycine (except *At*GSTU28 where it is an alanine), and it is also seen that there is an

arginine at position 22 (except *At*GST26 & 28 where it is a lysine) forming an electrostatic interaction and hydrogen bonding to the aspartate. However if the residue at 11 is glycine then there is room for more variation at 18. While valine is preferred (49%), alanine (18%), serine (15%) and threonine (8%) are all fairly common with five other amino acids seen in the remaining five sequences. There is also a large variance at position 22 with 48% being arginine or lysine, 24% aspartate or glutamate, 10% glutamine, 8% valine, leucine or isoleucine and 6% others. This is interesting, as in *Os*GSTU1, where this residue is an arginine it forms two hydrogen bonds to the main chain oxygen atoms in the loop containing the active site residue, Ser 15, stabilizing its conformation.
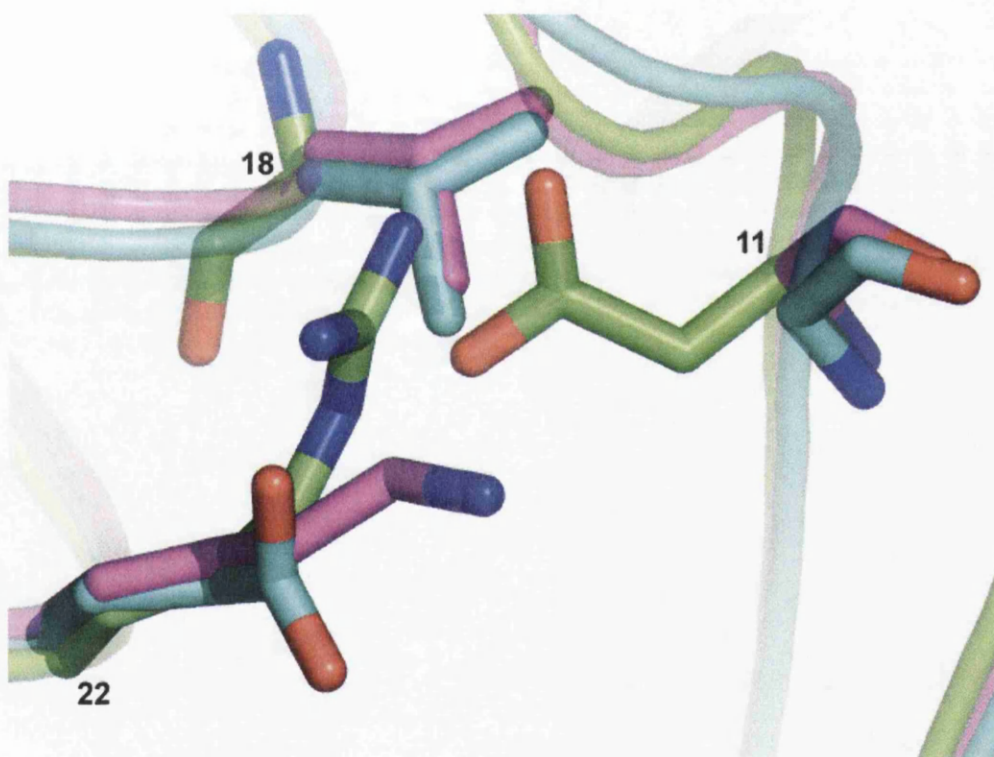


**Figure 8.8.** Variation in residues 11, 18 and 22 in Tau class GSTs. *Os*GSTU1 is shown in green, *Os*GSTU4 in cyan and *Ta*GSTU4 in magenta. Image generated with PyMOL (Delano, 2002).

Another important residue is Trp 13 on the loop between strand β-1 and helix α-1, which is conserved in 78% of the sequences. In the *Os*GSTU1 structure this is followed by a valine while in the *Os*GSTU4 and *Ta*GSTU4 structures it is followed by a proline (figure 8.9). In a comparison of *Os*GSTU1 and *Ta*GSTU4 it can be seen that the presence of the proline – which stacks

against the side-chain of the tryptophan – pushes the tryptophan side-chain further away. In the GSTU4 structures it is a proline following the tryptophan but in *Os*GSTU4 the peptide bond between them is *cis*. There is also a 180° flip around the Cβ atom between the two GSTU4 structures with an accompanying change from the *gauche⁻* rotamer in *Os*GSTU4 to the *trans* rotamer in *Ta*GSTU4. In *Os*GSTU1 and *Ta*GSTU4 there is a glutamate at position 36, which forms a hydrogen bond to the main-chain nitrogen of the tryptophan.
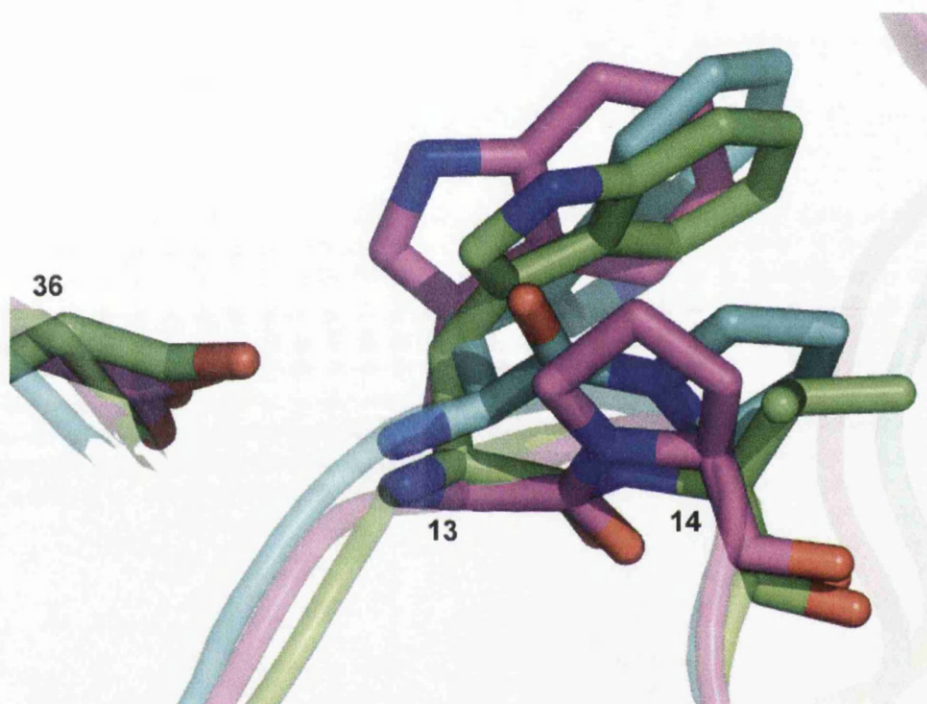


**Figure 8.9.** Variation in residues 13, 14 and 36 in Tau class GSTs. *Os*GSTU1 is shown in green, *Os*GSTU4 in cyan and *Ta*GSTU4 in magenta. Image generated with PyMOL (Delano, 2002).

There is a conserved asparagine and proline pair at positions 50 and 51 on the end of helix α-2. The side-chain nitrogen of the Asn 50 is forming a hydrogen bond to the main-chain carbonyl of the active site residue, Lys 55 and Pro 51 seems to help maintain the conformation of the loop between helix α-2 and strand β-3 (figure 8.10).
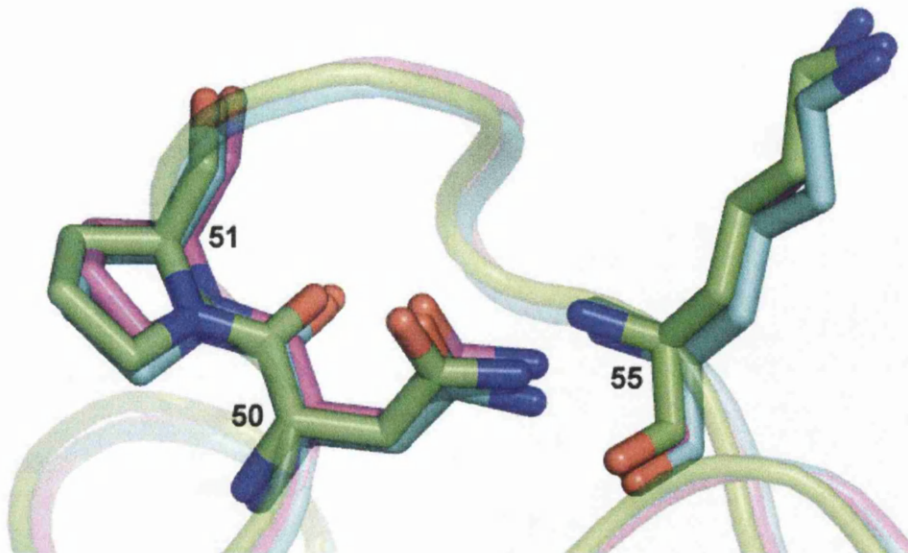
**Figure 8.10.** Conservation in residues 50, 51 & 55 in Tau class GSTs. *Os*GSTU1 is shown in green, *Os*GSTU4 in cyan and *Ta*GSTU4 in magenta. Image generated with PyMOL (Delano, 2002).

### 8.3.1  N-Terminal Domain

As described in sections 6.2.1 and 8.2.1 there are certain residues that are well conserved in the N-terminal domain for the binding of GSH. The first of these residues is Ser 15, which is completely conserved throughout these Tau class GSTs, although in other classes of GSTs this can be a tyrosine, or in the cases of omega and lambda class GSTs, this residue is a cysteine. The following residue is a proline in 81% of sequences, and in 94% of Tau class GST sequences there is a phenylalanine or a tyrosine at position 17, which forms a hydrophobic contact with GSH. The residue at position 39 is always a residue capable of forming a hydrophobic contact to GSH but with more variation – in 90% of Tau class sequences it is a leucine, isoleucine or valine. At position 42 there is almost always a lysine residue forming a hydrogen bond to the O32 atom of GSH, but in a very few cases this is replaced by an arginine. There is another well-conserved lysine residue at position 55 (85%) that is forming a hydrophobic contact with GSH. At position 56 the residue is in general either an isoleucine or a valine followed by a completely conserved *cis*-proline. This proline has to be in the *cis* conformation to allow the preceding residue to form an anti-parallel β-sheet interaction with the cysteinyl moiety of GSH (Dirr *et al.*, 1994; Wilce & Parker, 1994). This *cis*-proline is seen in an equivalent position in the structures of thioredoxins and

glutaredoxins, which are thought to be the ancestors of GSTs (Eklund *et* al., 1991; Eklund *et al.*, 1992). There is a conserved glutamate and serine pair at positions 68 and 69 that between them form three hydrogen bonds to GSH.

## 8.3.2 C-Terminal Domain

Compared to the G-site, the H-site of Tau class GSTs is much less well conserved. The regions in the C-terminal domain that are fairly well conserved are not in the sequences contributing to the H-site. These regions therefore must be structural, for instance most of helix α-4 is fairly well conserved from position 99 with the sequence YQ/ER/KAXARFWAXY/FI/VD/EXK. There is also a well-conserved LE pair at positions 144 and 145 in helix α5. The bottom of helix α6 is conserved with the sequence GY/FV/LDI/V but there is much less conservation towards the top of the helix where several H-site residues are located.

## 8.3.3 Active Site

If we compare the active sites of the three Tau class structures that have been solved we can see that they are very different from one another (figure 8.11) although there are a few residues conserved between the three proteins. The first of these residues are Pro 16 and Phe 17, which are conserved in 81% and 94% (as phenylalanine or tyrosine) of sequences respectively. The tryptophan at position 175 is at the centre of the H-site and is conserved in 71% of Tau class GST sequences. The phenylalanine at position 220 is also conserved between these three proteins and is seen as a phenylalanine or tyrosine in 74% of the 107 Tau sequences that have been aligned (appendix B).

If we look at the individual proteins we can see that the H-site of *Os*GSTU4 (figure 8.11 A) is large, open, and flat and lined mostly with aromatic and hydrophobic residues. It does have some hydrogen bonding partners available in Arg 121 & 225 and Lys 13. These arginines are oriented in such a way that they can move over the active site and possibly provide partners to π-stack against aromatic rings in a ligand.
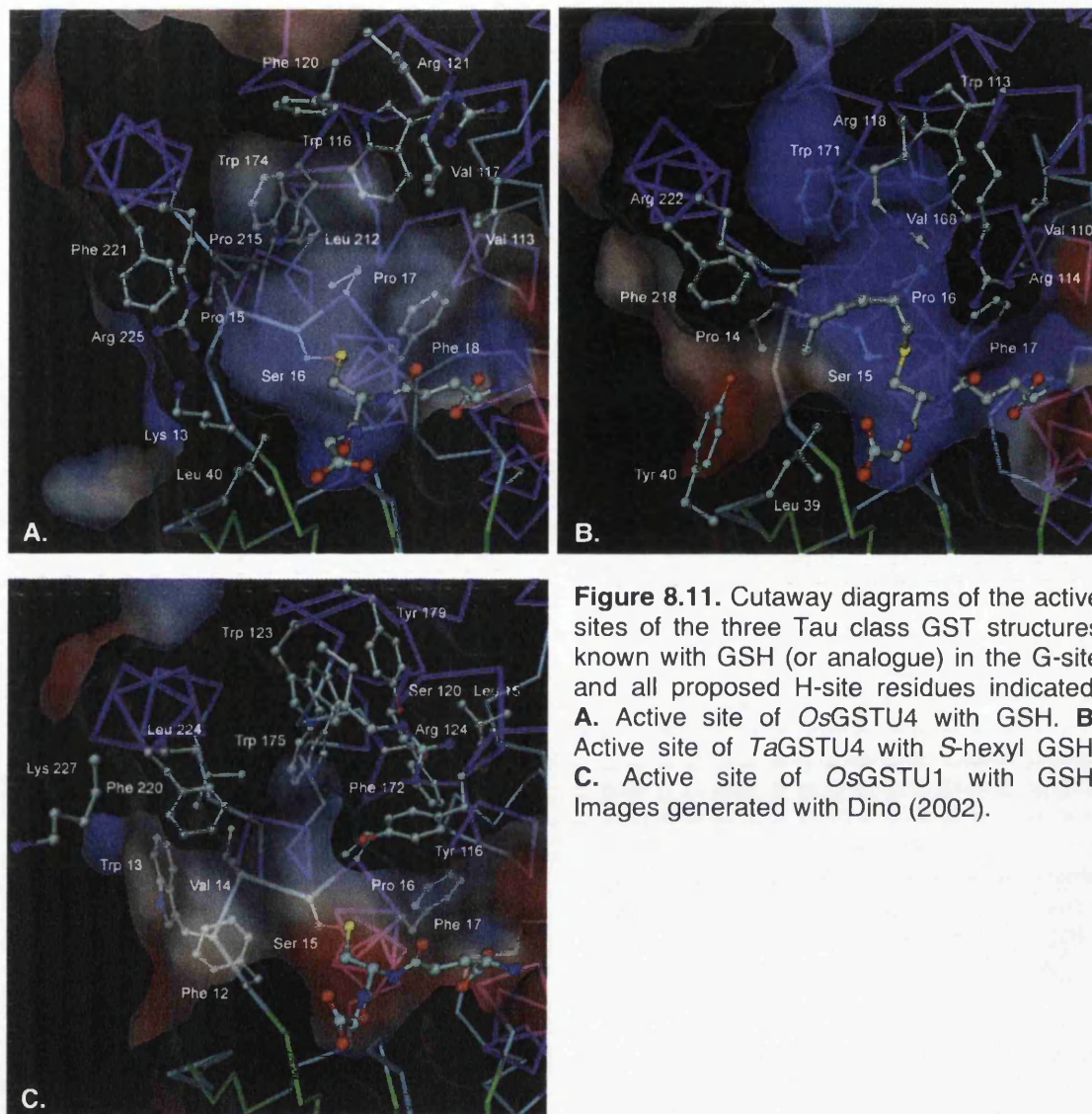
**Figure 8.11.** Cutaway diagrams of the active sites of the three Tau class GST structures known with GSH (or analogue) in the G-site and all proposed H-site residues indicated. **A.** Active site of *Os*GSTU4 with GSH. **B.** Active site of *Ta*GSTU4 with *S*-hexyl GSH. **C.** Active site of *Os*GSTU1 with GSH. Images generated with Dino (2002).

The active site of *Ta*GSTU4 (figure 8.11 B) has some similarities to that of *Os*GSTU4 with two conserved tryptophans (113 & 171) and a valine (110) again forming the basis of the active site. But unlike *Os*GSTU4 the left hand side of the H-site is more open and it forms a two-lobe structure like that of *Os*GSTU1 (figure 8.11 C). In the case of *Ta*GSTU4 there are two arginine (114 & 118) at the top of helix α-4 and one towards the top of helix α-9 (222) that are sitting over the top of the active site. Together they make the active site fairly basic but could swing out of position to allow access to the H-site and could provide hydrogen-bonding partners or could move back over the ligand to form π-stacking interactions with hydrophobic, aromatic molecules.

The active site of *Os*GSTU1 is quite different from that of the other Tau class enzymes structures, and has been described in detail in sections 7.4 and 7.5. It is less open than the others and contains a deep pocket that is lined with aromatic residues. It has a tyrosine residue at position 116 which could either swing over to close the pocket, or if it swings the other way can open up the pocket and allow a larger molecule access.

# Chapter 9     GSH Induced Conformational Shift in *Zm*GSTF1 and *At*GSTT1

## 9.1   Structural Flexibility in Glutathione *S*-Transferases

Phi, Alpha and Pi class GSTs have all been postulated to undergo conformational changes upon ligand binding. For example, in the structure of *Zm*GSTF2-2 (PDB ID 1AW9; Neuefeind *et al.*, 1997a) there is a highly flexible region in the N-terminal domain between residues 33 and 44, with no clear electron density for this region. In the structure of the homologous protein *Zm*GSTF1-1 with lactoylglutathione tightly bound (PDB ID 1AXD; Neuefeind *et al.*, 1997b), this region is fixed in place and clearly defined. Analysis of the structure shows no crystal contacts in this region suggesting that GSH binding is causing this region to adopt a rigid conformation. Lys 41 of this loop binds to the glycinyl carboxylic acid of GSH, and it has been postulated, before any structures were solved, that Phi class GSTs undergo an induced fit mechanism (Ketterer *et al.*, 1988). Kinetic analysis of *Zm*GSTF1-1 has shown that there is communication between the G-site and the H-site with GSH binding increasing the affinity of the enzyme for CDNB with a coupling constant of $\alpha = 0.4$ for the Kd of CDNB, and this is postulated to be due to the ordering of Phe 35 in the same loop, which forms part of the H-site (Labrou *et al.*, 2001).

In the human Pi class GST, *Hu*GSTP1-1 a similar ordering of the residues 36-48, which contains helix $\alpha$-2, is seen (figure 9.1; Hitchens *et al.*, 2001). In the apo-forms of the enzyme (PDB IDs 14GS & 16GS; Oakley *et al.*, 1998) no discernible electron density can be seen for this region, but in the structures with GSH or inhibitors bound (PDB IDs 5GSS, 6GSS, 7GSS, 8GSS, 9GSS & 10GS; Oakley *et al.*, 1997) this region is again seen to adopt a rigid conformation. Again, no crystal contacts are present to fix the conformation of this loop region. NMR studies of the enzyme in apo-form and with GSH and *S*-hexyl-GSH bound show similar results and suggest that between −4 and −6 kJ

M$^{-1}$ of the free energy of GSH binding is used in the ordering of helix α-2 (Hitchens _et al._, 2001). In fact, in the apo-form, a disulfide bond can be formed between Cys 48 in this loop region and Cys 120 in the C-terminal domain, which are separated by 18Å in the ligand bound forms, again indicating the large range of movement available to this loop in the absence of GSH (Ricci _et al._, 1991).
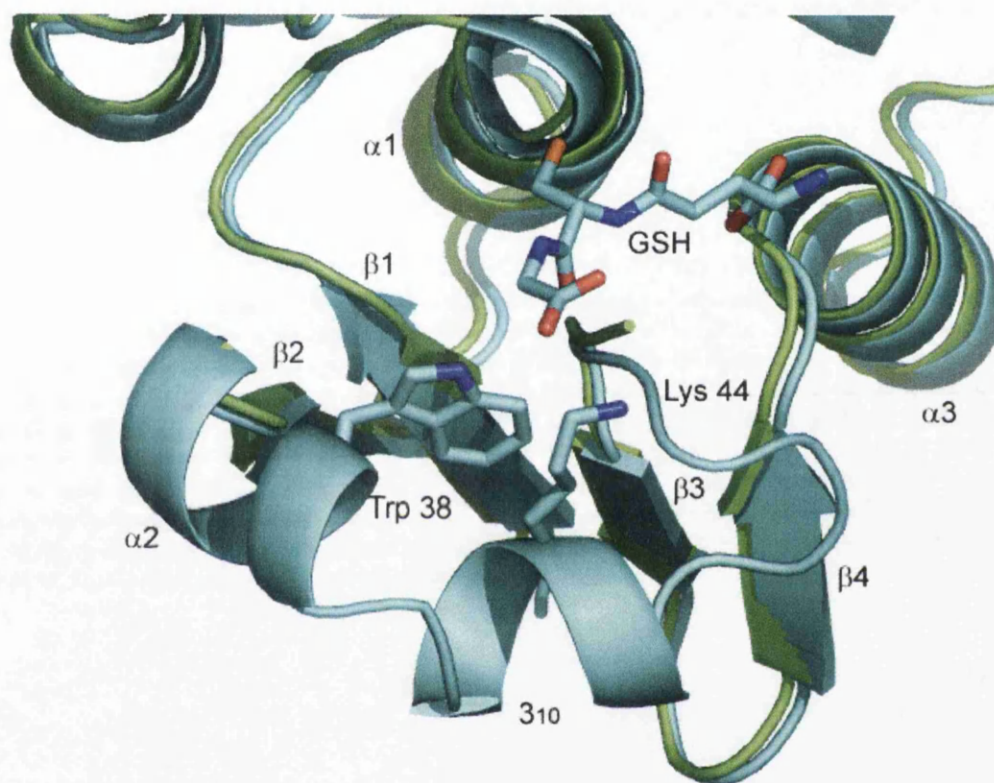


**Figure 9.1.** Comparison of apo-form _Hu_GSTP1-1 (green) and _Hu_GSTP1-1 with GSH bound (blue). Image generated with PyMOL (Delano, 2002).

Disorder-to-order transitions are also seen in the crystal structures of human Alpha class GST, _Hu_GSTA1-1 upon ligand binding (Sinning _et al._, 1993; Cameron _et al._, 1995). However, in this case no conformational changes are seen in the N-terminal domain or most of the C-terminal domain but in the apo-form enzyme there is no electron density discernable for the terminal α-helix. In the ligand bound forms of the protein this helix forms a lid over the H-site and providing a highly hydrophobic environment for the ligand (Sinning _et al._, 1993), and has

## 9.2    Further Refinement of *Zm*GSTF1-1 Apo-Form Structures

Previous to this project Russell Thom (2001b) had crystallized, collected data and solved the structure of the Phi class GST from maize, *Zm*GSTF1-1, in two crystal forms. Both these structures are in the apo form. The structures had been built and partially refined with an $R_{work}$ of 23.6% and $R_{free}$ of 27.8% for the $P2_1$ structure and $R_{work}$ of 23.4% and $R_{free}$ of 29.9% for the P432 structure. The structure of this enzyme has previously been reported in complex with lactoylglutathione (PDB ID 1AXD; Neuefeind *et al.*, 1997b) and an atrazine-GSH conjugate bound in the active site (PDB ID 1BYE; Prade *et al.*, 1998).

During this project the rebuilding and refinement of the two structures of *Zm*GSTF1-1 were completed with a view to investigating if any GSH induced conformational change occurs by comparison of all the *Zm*GSTF1-1 structures

### 9.2.1    Refinement of *Zm*GSTF1-1 in $P2_1$

The structure of *Zm*GSTF1-1 in $P2_1$ was rebuilt using QUANTA (Molecular Simulations Inc., 1986) and refined using REFMAC5 (Murshudov *et al.*, 1996). After five cycles of rebuilding and refinement, with the addition of 253 solvent molecules the $R_{work}$ had been reduced to 18.8% and the $R_{free}$ to 27.8% from 23.6% and 27.8% respectively. Refinement statistics are shown in table 9.1.

### 9.2.2    Refinement of *Zm*GSTF1-1 in P432

The structure of *Zm*GSTF1-1 in P432 was rebuilt using QUANTA (Molecular Simulations Inc., 1986) and refined using REFMAC5 (Murshudov *et al.*, 1996). After five cycles of rebuilding and refinement, with the addition of 38 solvent molecules and one chloride ion the $R_{work}$ had been reduced to 18.1% and the $R_{free}$ to 27.1% from 23.4% and 29.9% respectively. Refinement statistics are shown in table 9.1.

| Space Group | P2₁ | P432 |
|---|---|---|
| a (Å) | 52.94 | 116.505 |
| b (Å) | 75.46 | 116.505 |
| c (Å) | 106.98 | 116.505 |
| $\beta$ (°) | 100.05 | |
| Resolution Range (Å) | 25.0 – 2.8 | 25.0 – 2.85 |
| No of unique reflections in resolution range | 21345 | 7136 |
| $R_{Work}$ (%)[a] | 18.8 | 18.1 |
| No of reflections used in $R_{free}$ | 1010 (5.1%) | 325 (5.1%) |
| $R_{Free}$ (%) | 27.8 | 27.1 |
| Total number of protein atoms | 6647 | 1653 |
| Total number of solvent atoms | 253 | 38 |
| Total number of hetero atoms | 0 | 1 |
| Average B factor (Å²) | | |
| Main-chain atoms | 23.93 | 49.82 |
| Side-chain atoms | 24.43 | 50.36 |
| Solvent atoms | 24.09 | 51.39 |
| Hetero atoms | N/A | 69.28 |
| Root mean square deviation | | |
| Bonds (Å) | 0.022 | 0.024 |
| Bond angles (°) | 2.076 | 2.442 |
| Ramachandran quality, % in[b] | | |
| Most favoured regions | 91.5 | 81.6 |
| Allowed regions | 8.0 | 17.9 |

[a] $R = \sum_{hkl}( |F_{obs} - F_{calc}|) / \sum|F_{obs}|$

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

**Table 9.1.** Refinement statistics for ZmGSTF1-1 in P2₁ and P432.

## 9.3   Structure of ZmGSTF1-1

The ZmGSTF1-1 structure is a homo-dimer, with each molecule composed of two domains. The N-terminal domain (residues 1-78) is a $\alpha/\beta$ domain and contains the GSH binding site. There is then a linker region (residues 79-87) that connects the N-terminal domain to the C-terminal domain (residues 88-214), which is composed of six helices and contains the hydrophobic binding pocket. The structure of ZmGSTF1-1 is shown in figure 9.2.

terminal domain (residues 88-214), which is composed of six helices and contains the hydrophobic binding pocket. The structure of *Zm*GSTF1-1 is shown in figure 9.2.
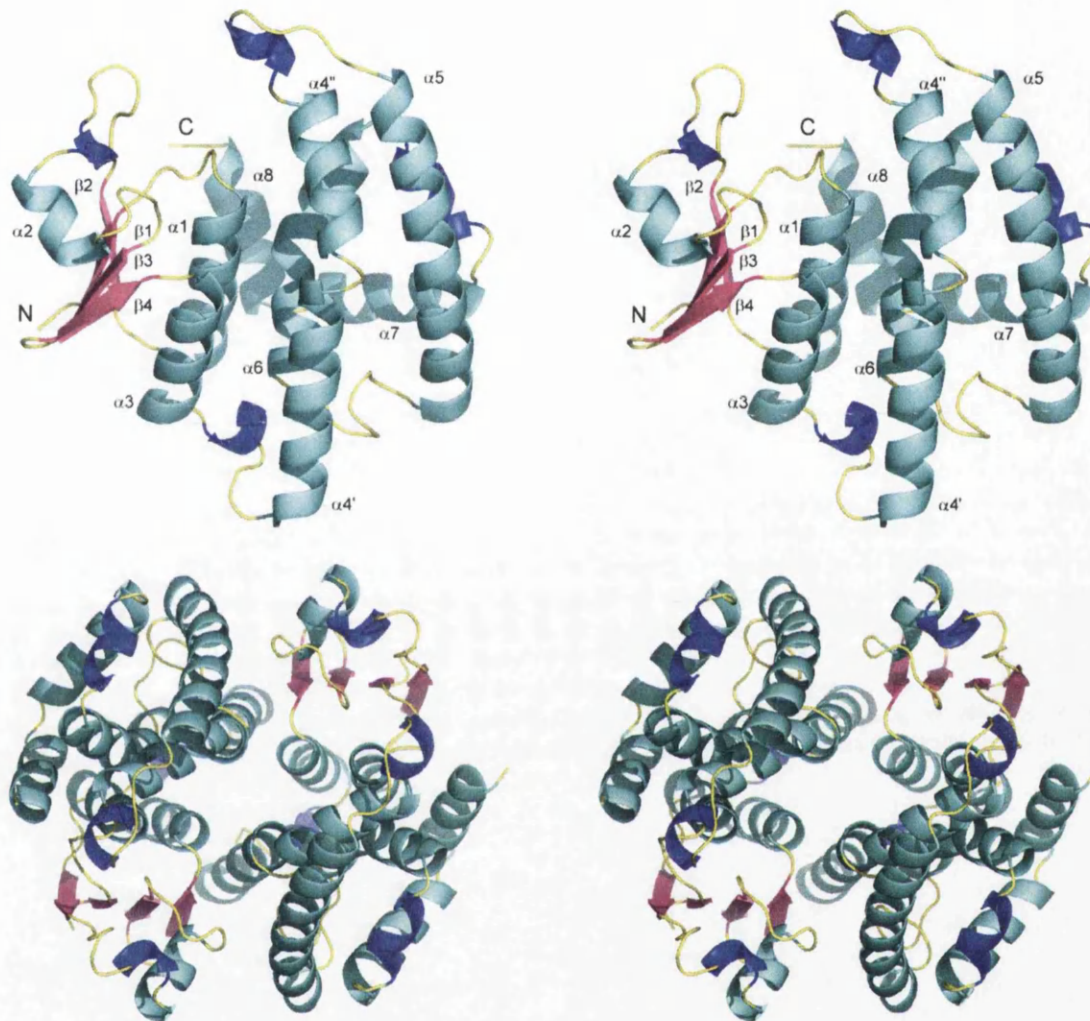


**Figure 9.2.** Stereo views of the *Zm*GSTF1-1 monomer and dimer represented as a ribbon. Images generated with PyMOL (Delano, 2002).

### 9.3.1   Comparison of *Zm*GSTF1 Structures

After completing the structures of two apo-form enzymes they were compared with the structures of the two ligand-bound forms. After superimposing the four chains of the $P2_1$ structure, the one chain of the P432 structure, the four chains of the atrazine-GSH structure, and the two chains of the lactoylglutathione structure it appeared that there was evidence for a conformational shift on ligand binding (figure 9.3). The major shift is in the top half of helix α-4 and in helix α-5, which have

moved away from the N-terminal domain, opening up the active site. By examining the crystal structures it can be seen that this shift region is not involved in any crystal contacts.
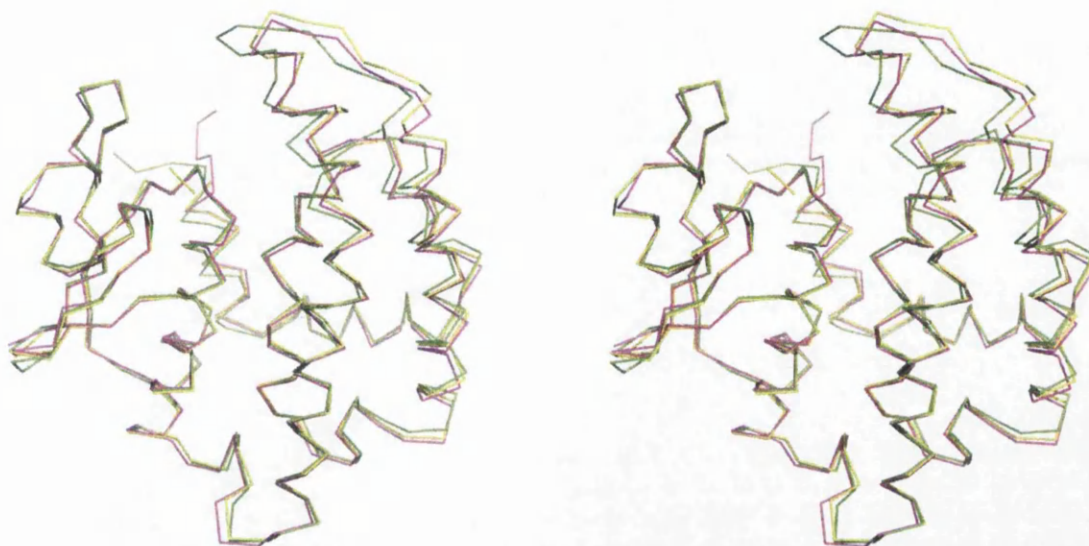


**Figure 9.3.** Stereo diagram of the Cα traces of three structures of *Zm*GSTF1 superimposed on the N-terminal domain. The structure of the P432 apo-form of *Zm*GSTF1 is shown in green, *Zm*GSTF1 with lactoylglutathione in yellow, and *Zm*GSTF1 with atrazine-glutathione in pink. Image generated with PyMOL (Delano, 2002).

The program LSQKAB (CCP4, 1994; Kabsch, 1976) was used to compare the Cα coordinates of the apo-form enzymes with the ligand-bound enzymes (figure 9.4). In the comparison between the apo-form enzyme and that of the lactoylglutathione bound structure there is a large RMSD in the loop region containing helix α-2, as would be expected from the comparison of this structure with that of *Zm*GSTF2 (Neuefeind *et al.*, 1997a), and in the following region up to residue 66 (strands β-2 and β-3). The shift in the loop region is also seen in the structure with atrazine-glutathione bound but less so in the following region. The largest shift in the structures is seen at Gly 123, which is right at the top of the loop between helices α-4 and α-5 where the major conformational shift is seen between the structures in figure 9.3.
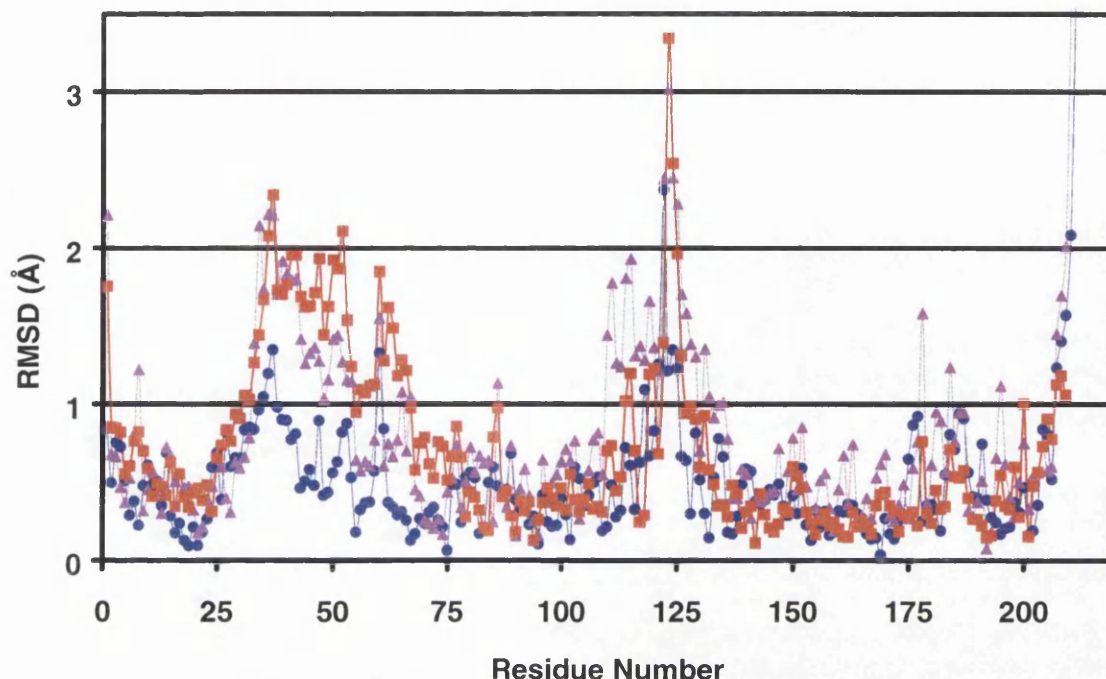
**Figure 9.4.** RMSD of the Cα coordinates of the *Zm*GSTF1 structures. ● RMSD between the two apo-form structures. ▲ RMSD between apo-form *Zm*GSTF1 and *Zm*GSTF1 with atrazine-glutathione bound. ■ RMSD between apo-form *Zm*GSTF1 and *Zm*GSTF1 with lactoylglutathione bound.

### 9.3.2   Active Site of *Zm*GSTF1

The G-site of *Zm*GSTF1 is located in the cleft between the βαβ motif and the ββα motif of the N-terminal domain. The principle residues of the G-site are Ser 11 & 67, Asn 13, His 40, Lys 41, Gln 53, Val 54, and Glu 66. Upon binding of GSH small changes are seen in these residues, with Asn 13, His 40, and Lys 41 changing conformation to form contacts with the ligand (figure 9.5).

The H-site of *Zm*GSTF1 is located in the cleft between the N-terminal domain and the C-terminal domain. In the H-site, several changes are seen upon ligand binding. In the structure containing lactoylglutathione Phe 35 has moved round away from the G-site, and Met 10 has moved in closer to the H-site. No density is seen in the structure for the side chain of Ile 118. When atrazine-GSH is bound Phe 35 has moved even further round to π-stack with the aromatic ring in the ligand. Although the orientation of Trp 12 is fairly constant in the three structures, the loop containing Ile 118 is drawn in closer to the H-site to interact with atrazine.
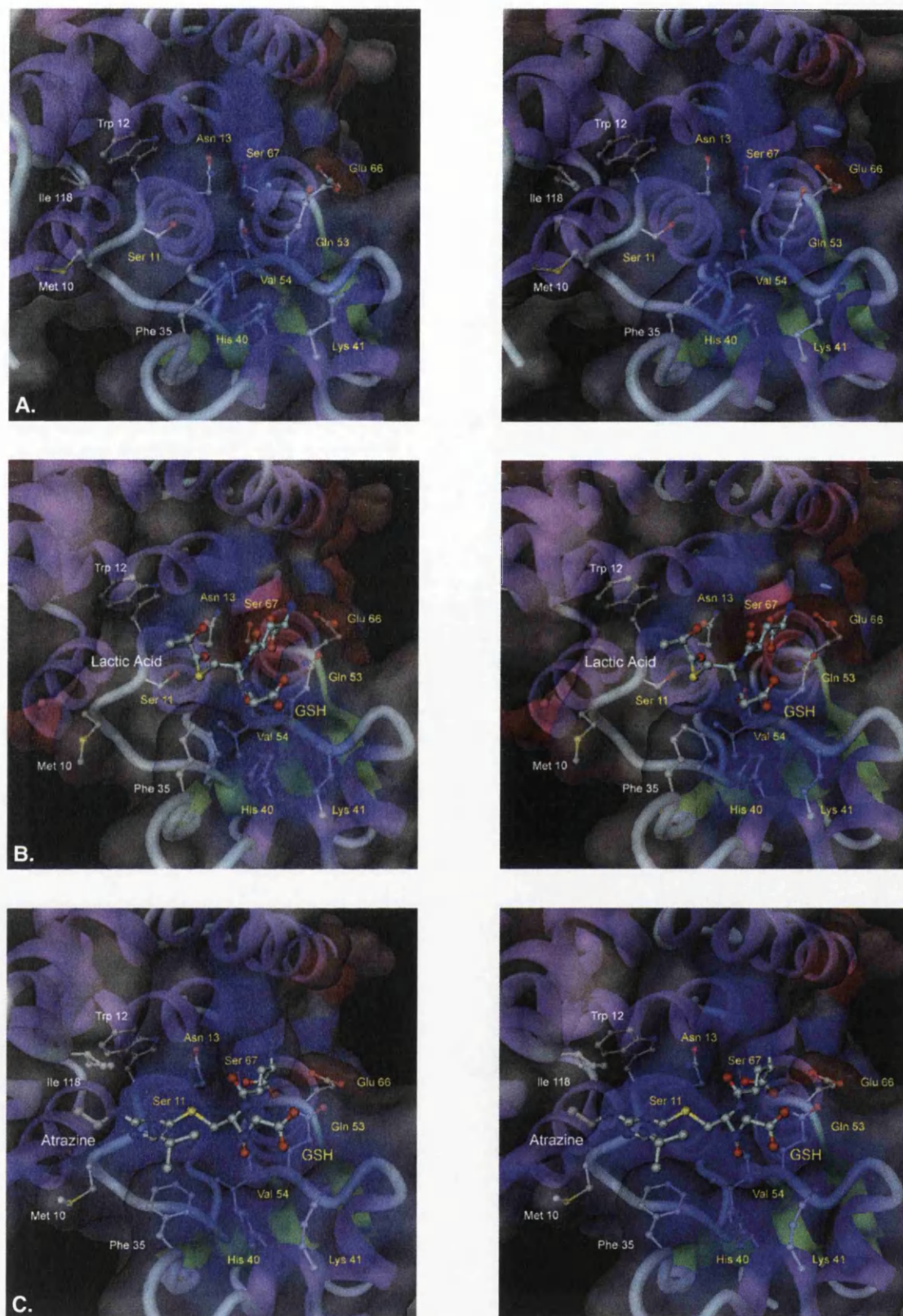
**Figure 9.5.** Active site of *Zm*GSTF1 with H-site residues labelled in white and G-site residues labelled in yellow. **A.** Apo form *Zm*GSTF1. **B.** *Zm*GSTF1 with lactoylglutathione (PDB ID 1AXD, Neuefeind *et al.*, 1997b). **C.** *Zm*GSTF1 with atrazine-glutathione conjugate (PDB ID 1BYE, Prade *et al.*, 1998). Images generated with DINO (2003).

To further investigate any conformational shifts in *Zm*GSTF1 upon GSH binding, ultraviolet difference spectroscopy and circular dichroism were employed.

### 9.3.2   Ultraviolet Difference Spectroscopy

Proteins absorb in the ultraviolet range of the spectrum due to the presence of tyrosine and tryptophan residues. The absorbance of these residues is altered from that of free amino acids due to the environment in which they sit in the protein. If upon ligand binding a conformational change occurs in the protein then the environment of these residues will be altered causing a change in the absorbance spectra of the protein. UV difference spectroscopy allows you to measure the changes in the UV spectra, and therefore examine any conformational change occurring upon ligand binding.

The UV difference spectroscopy was carried out on a Jasco V-550 UV/VIS spectrophotometer using 1.0cm quartz cuvettes thermostated to 25°C. A scan was performed from 240-400nm at a scan speed of 50nm min$^{-1}$ with the slits set to 3nm. A scan was carried out with 850µl of potassium phosphate buffer, pH 7.0 in each cuvette to provide a baseline scan. 150µl of 5mg ml$^{-1}$ *Zm*GSTF1-1 was added to the cuvettes and a further baseline scan was run. 16µl of 10mM GSH was added to one cuvette and 16µl of buffer to the second cuvette and the samples were scanned. A final scan of buffer vs 16µl ml$^{-1}$ GSH was run and subtracted from the spectra to remove any effect on absorbance by the GSH. The resultant spectrum is shown in figure 9.6. This clearly shows an alteration in UV absorption by the protein on binding of GSH.
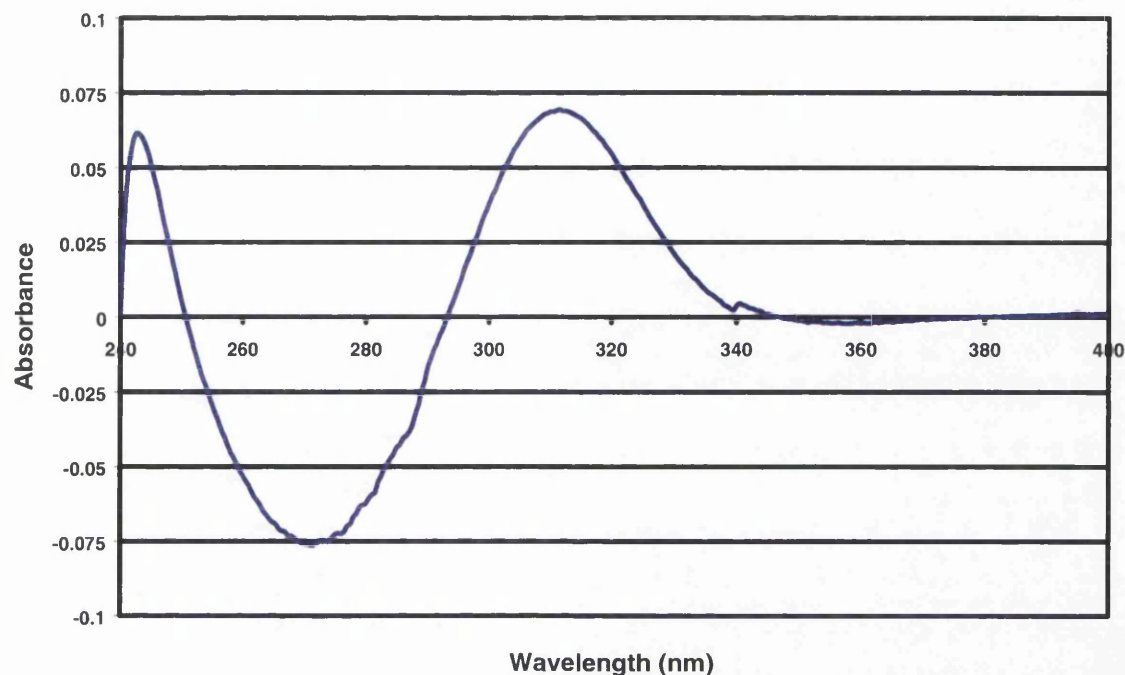
**Figure 9.6.** UV difference spectrum of *Zm*GSTF1-1 on addition of GSH.

### 9.3.3  Circular Dichroism

CD is a technique that utilises the differential absorption of right and left circularly polarized light by a protein and can be used to determine to relative proportions of secondary structure elements. Structural changes in the protein can also be seen due to alteration of the characteristic spectrum. Near-UV CD of *Zm*GSTF1-1 and *Zm*GSTF1-1 with GSH was carried out by Dr. Sharon Kelly of the Scottish Circular Dichroism Facility using a Jasco J600 Spectropolarimeter with 0.5cm quartz cuvettes. The resultant spectra can be seen in figure 9.7. The spectrum of *Zm*GSTF1-1 is shown in pink and the spectrum of *Zm*GSTF1-1 with GSH (corrected for absorbance by GSH) is shown in blue. From this it can be clearly seen that a conformational shift is occurring on GSH binding.
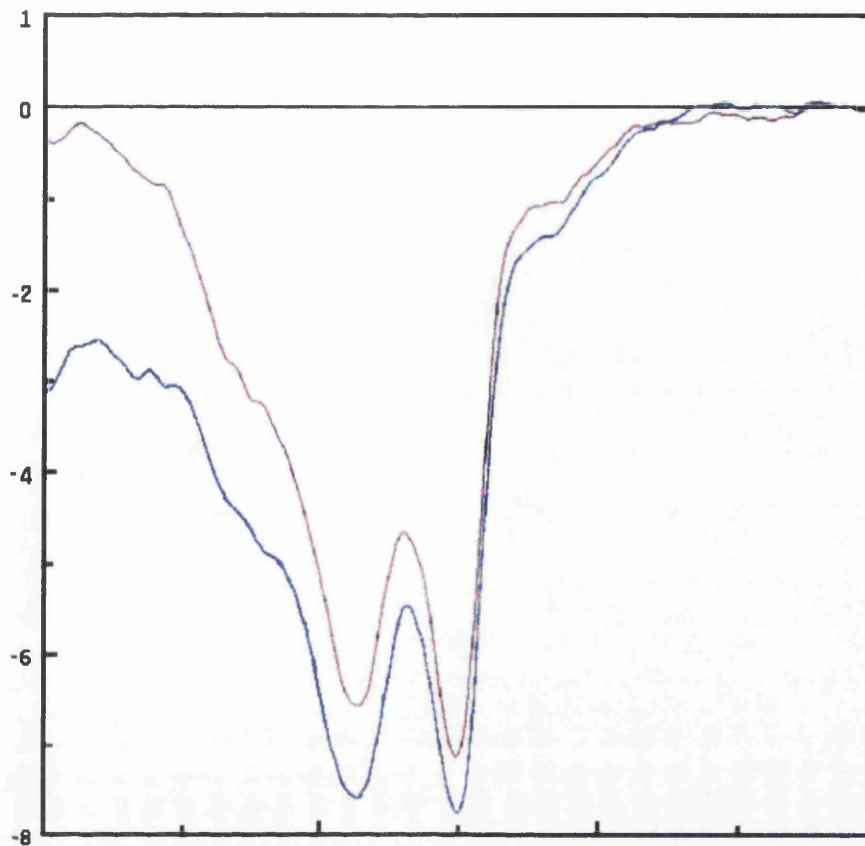
**Figure 9.7.** Near-UV CD spectra of *Zm*GSTF1-1 (pink) and *Zm*GSTF1-1 with GSH (blue).

When these experiments were repeated using further batches of *Zm*GSTF1-1 no conformational shift was seen using either of the techniques. It is unclear why this should be the case. We suspect that the presence of a number of free thiol groups, which are readily oxidized, can alter the protein and therefore affect its structure and properties.

## 9.4   *At*GSTT1-1

*At*GSTT1-1 is a Theta class GST from *A. thaliana* whose structure was solved by Russell Thom (2001b). The crystal structure of this enzyme contains four monomers and large differences can be seen between them. In the crystal structure, chain A superimposes well with chain C, and chain B with chain D but comparing A with D (figure 9.8) we can see that a large shift is occurring at the top of helix α-4 and helix α-5. Also, in

the crystal structure the last twelve to sixteen residues are not seen in each chain, which in the human enzyme, _Hu_GSTT2-2 (PDB IDs 1LJR, 2LJR, & 3LJR; Rossjohn _et al._, 1998b), forms a loop and helix, which sits over the active site.
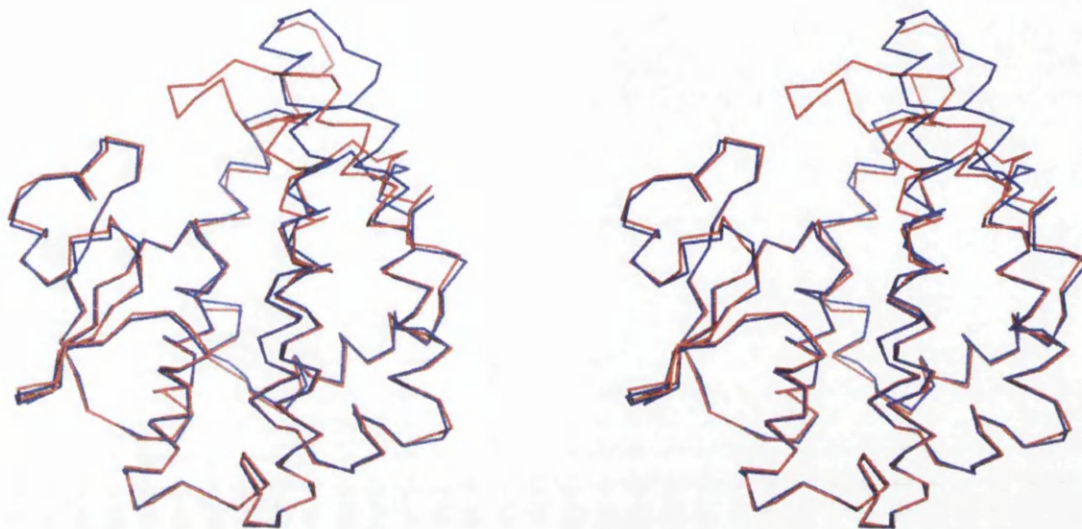


**Figure 9.8.** Stereo diagram of the superimposed Cα trace of chain A (red) and chain D (blue) of the crystal structure of _At_GSTT1-1. Image generated with PyMOL (Delano, 2002).

The _At_GSTT1-1 structure contains no ligands bound but it was decided to see if, like _Zm_GSTF1-1 spectroscopic methods could detect if any conformational shift in the loop, or disorder-to-order transition would occur on GSH binding. UV difference spectroscopy and near-UV CD were carried out as described above, with far-UV CD being carried out by Dr. Sharon Kelly of the Scottish Circular Dichroism Facility using a Jasco J600 Spectropolarimeter with 0.02cm quartz cuvettes.

After running the UV difference spectrum, and correcting for cuvette and buffer effects no real conformational shift could be detected in the spectrum (figure 9.9). Although there is some apparent signal here, it is at the level of noise and therefore suggests that there is no conformational shift, or disorder-to-order transition in _At_GSTT1-1 on GSH binding.
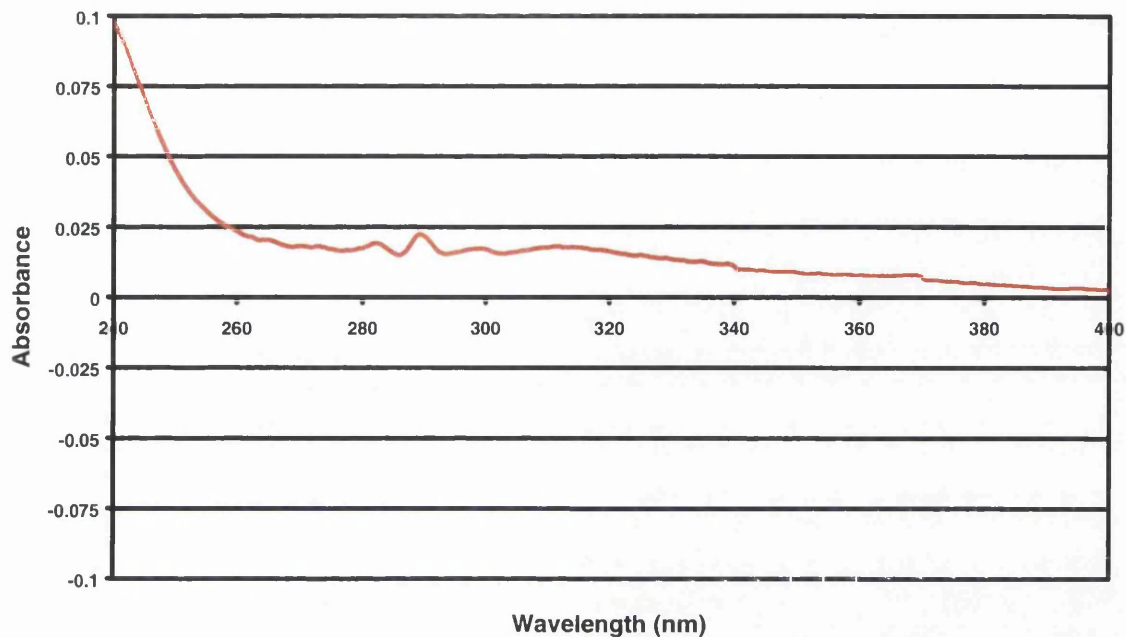
**Figure 9.9.** UV difference spectrum of *At*GSTT1-1 on addition of GSH.

Examining the near-UV and far-UV CD spectra it appears that the spectra of *At*GSTT1-1 superimpose well with the spectra of *At*GSTT1-1 with GSH further confirming that no conformational change was occurring on binding of GSH (figure 9.10).

After seeing no evident conformational changes upon GSH binding from the spectroscopic work, the crystal structure of *At*GSTT1-1 was re-examined (figure 9.11). From this it can be seen that chains A and C, which are in the same conformation have the top of the loop between helices α-4 and α-5 stacking against each other which pushes them towards a 'closed' conformation. However, in chains B and D the monomers are stacking closer together with the H-sites interacting and therefore keeping this loop in the 'open' conformation. The ability of these regions of GST structure to be affected by crystal contacts has been observed in numerous cases. In the *At*GSTT1-1 structure the most likely explanation for the observed change in structure is due to crystal contacts
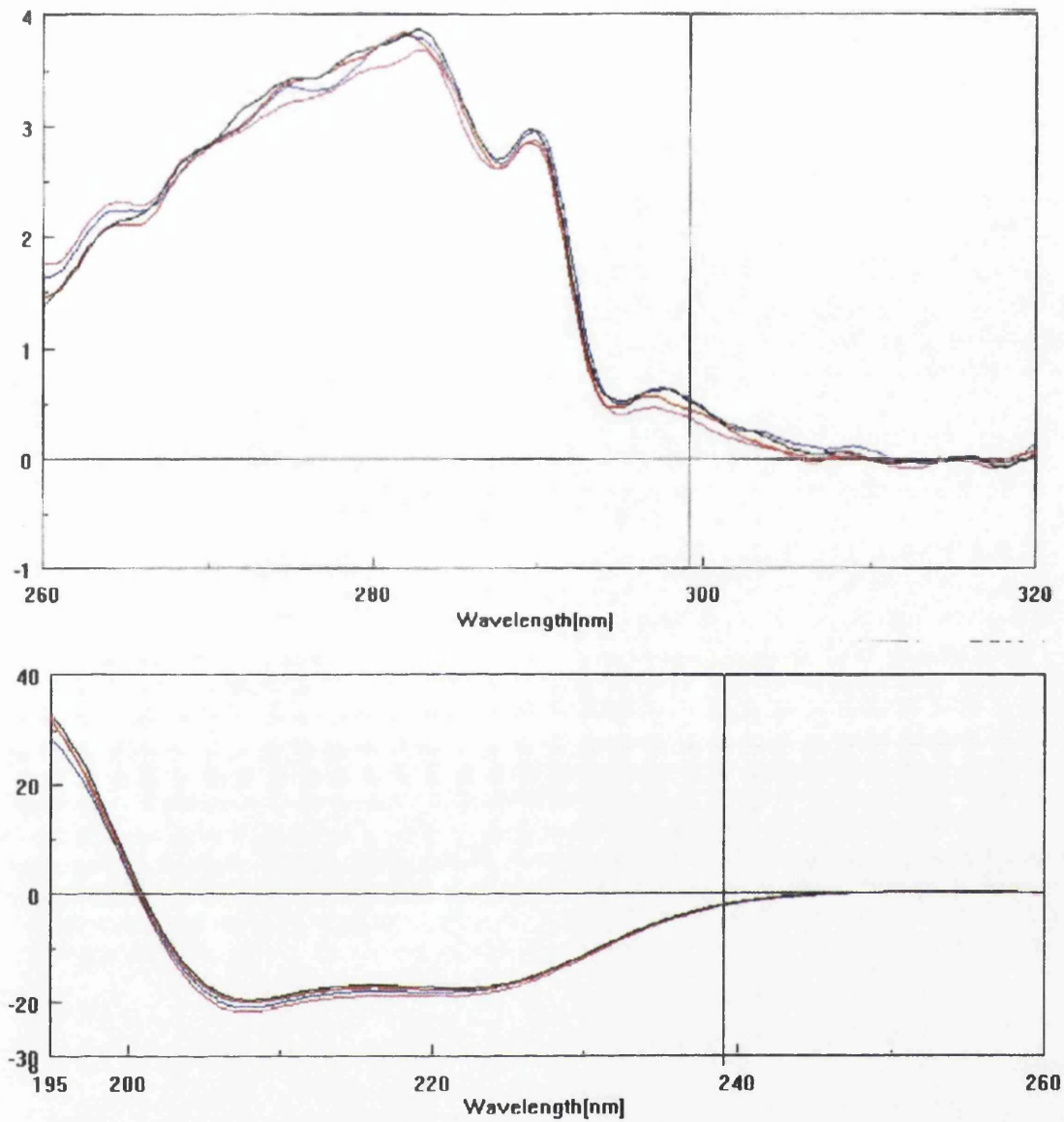
**Figure 9.10.** Near-UV (top) and far-UV CD spectra of *At*GSTT1-1 (pink, blue) and *At*GSTT1-1 with GSH (red, black).
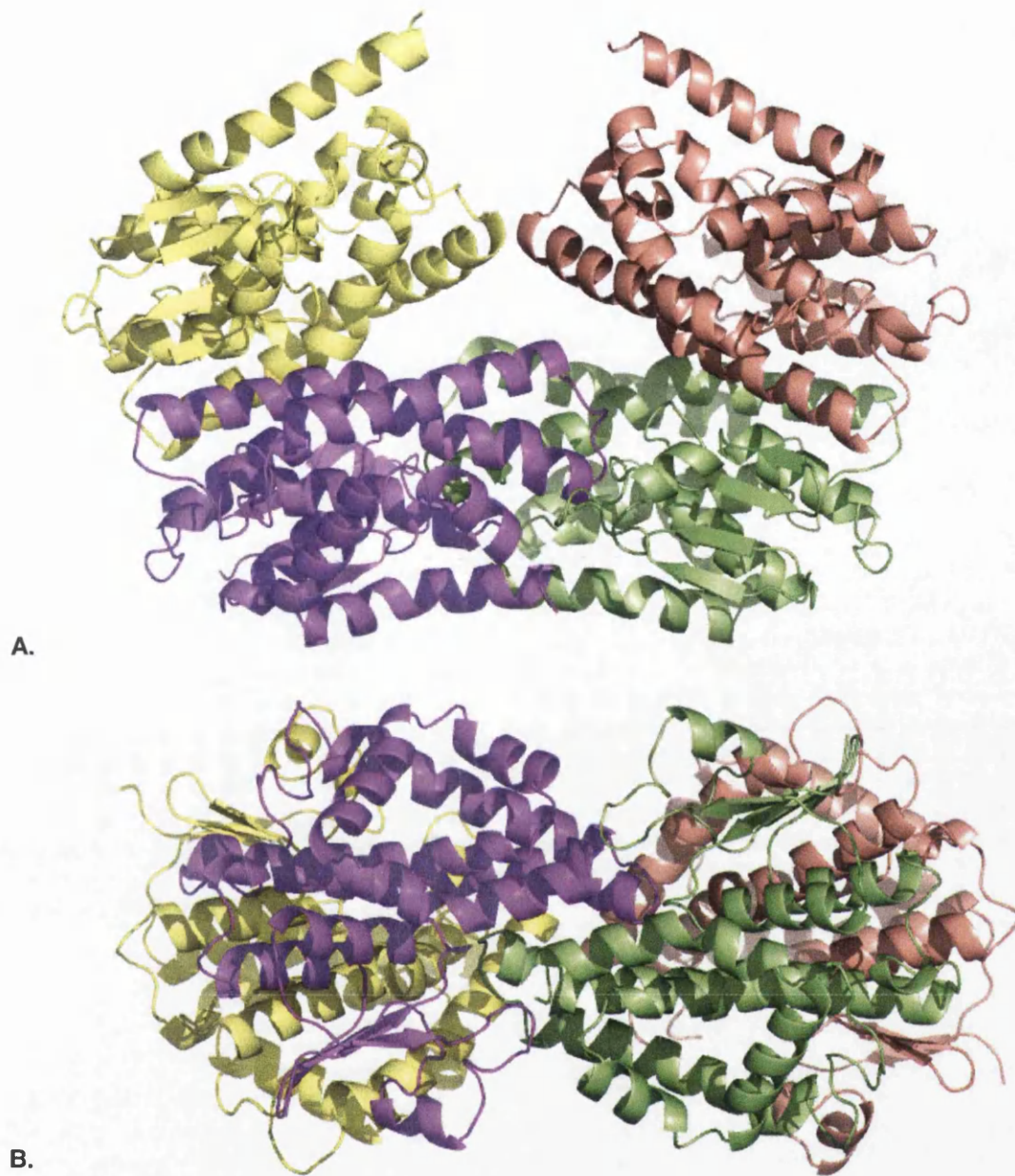
**Figure 9.11.** The asymmetric unit of *At*GSTT1-1. **A.** Shown along axis *a*. **B.** Shown along axis *b*. Chain A is shown in yellow, chain B in pink, chain C in salmon, and chain D in lime. Image generated with PyMOL (Delano, 2002).

# Chapter 10     Discussion

## 10.1  GST Phylogenies

In 2004, Soranzo and co-workers identified 61 genes for GSTs in rice by screening the rice dbEST division of the GenBank sequence database. Of these genes identified, forty were assigned as Tau class, sixteen Phi class, three Zeta class and two Theta class. One microsomal type GST sequence was also identified. By searching the GenBank database with the sequences of Lambda class GSTs and DHARs from *Arabidopsis thaliana* (Dixon *et al.*, 2002a) using a Phi- and Psi-Blast search (Altschul *et al.*, 1997) two Lambda class GSTs and two DHARs have been identified in the rice genome (figure 10.1).



**Figure 10.1.** Phylogenetic tree of GSTs found in rice. The sequences used to generate this tree are listed in appendix D. Sequence alignment performed with ClustalW (Higgins *et al.*, 1994) and the tree was generated using JalView (Clamp *et al.*, 2004).

The fact that there are only a few copies of the Zeta, Theta, Lambda and DHAR class GSTs indicates that there has been little pressure for these

genes to duplicate. These proteins are generally involved in primary cellular metabolism, so there would be little need for many copies of these proteins. The Tau and Phi classes however are involved in detoxification of xenobiotics by GSH conjugation, and in protecting the plants against environmental stresses. Due to the sessile nature of plants there has been large pressure for these genes to duplicate to protect the plants from a wide range of toxins that may be present in their environment.

Looking at a phylogenetic tree of maize GSTs it can be seen that this trend holds out in other species (figure 10.2). McGonigle *et al.* (2000) identified forty-two GST sequences by Blast searching the NCBI maize EST database. Here the sequences found by McGonigle *et al.* have been renamed to follow the standard class nomenclature and included a Theta class sequence identified by database searches. The fact that there are less sequences than are seen in rice does not imply that there are less sequences in maize, rather that the sequence of the genome is not yet completed.



**Figure 10.2.** Phylogenetic tree of GSTs found in maize. The sequences used to generate this tree are listed in appendix D. Sequence alignment performed with ClustalW (Higgins *et al.*, 1994) and the tree was generated using JalView (Clamp *et al.*, 2004).

**145**

Similarly in soybean, twenty-six GSTs were identified by McGonigle *et al.* (2004), twenty Tau class, one Zeta class, one lambda class, and four Phi class GSTs. By examining the sequences identified by McGonigle and co-workers one of the Phi class GST sequences has been reassigned as a Theta class GST. Searching of the GenBank databases has identified a further four Phi class GSTs, a Theta class and a DHAR (figure 10.3).



**Figure 10.3.** Phylogenetic tree of GSTs found in soybean. The sequences used to generate this tree are listed in appendix D. Sequence alignment performed with ClustalW (Higgins *et al.*, 1994) and the tree was generated using JalView (Clamp *et al.*, 2004).

The *A. thaliana* GST gene family has been extensively characterized by Wagner *et al.* (2002) who identified forty-seven sequences – twenty-eight Tau class, fourteen Phi class, three Theta class and two Zeta class. Dixon *et al.* (2002a) identified a further five sequences – two Lambda class and three DHAR sequences (figure 10.4).
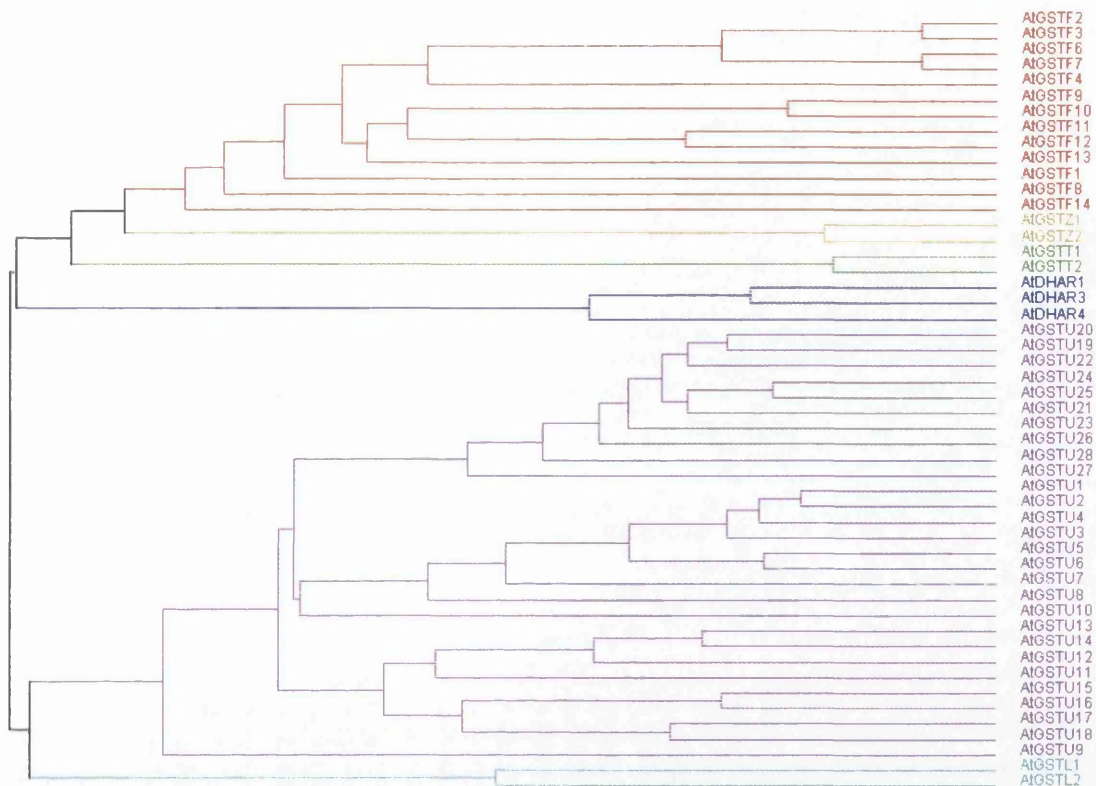
**Figure 10.4.** Phylogenetic tree of GSTs found in *Arabidopsis thaliana*. The sequences used to generate this tree are listed in appendix D. Sequence alignment performed with ClustalW (Higgins *et al.*, 1994) and the tree was generated using JalView (Clamp *et al.*, 2004).

In contrast, the wheat superfamily of GSTs has not been very well characterized. Previously only eight Tau class, ten Phi class, one Lambda class, and one Zeta class GST had been identified. By searching The Institute for Genomic Research's (TIGR) wheat genome database ten more Tau class, twelve Phi class, one theta class, one Lambda class, and a DHAR have been identified (figure 10.5).
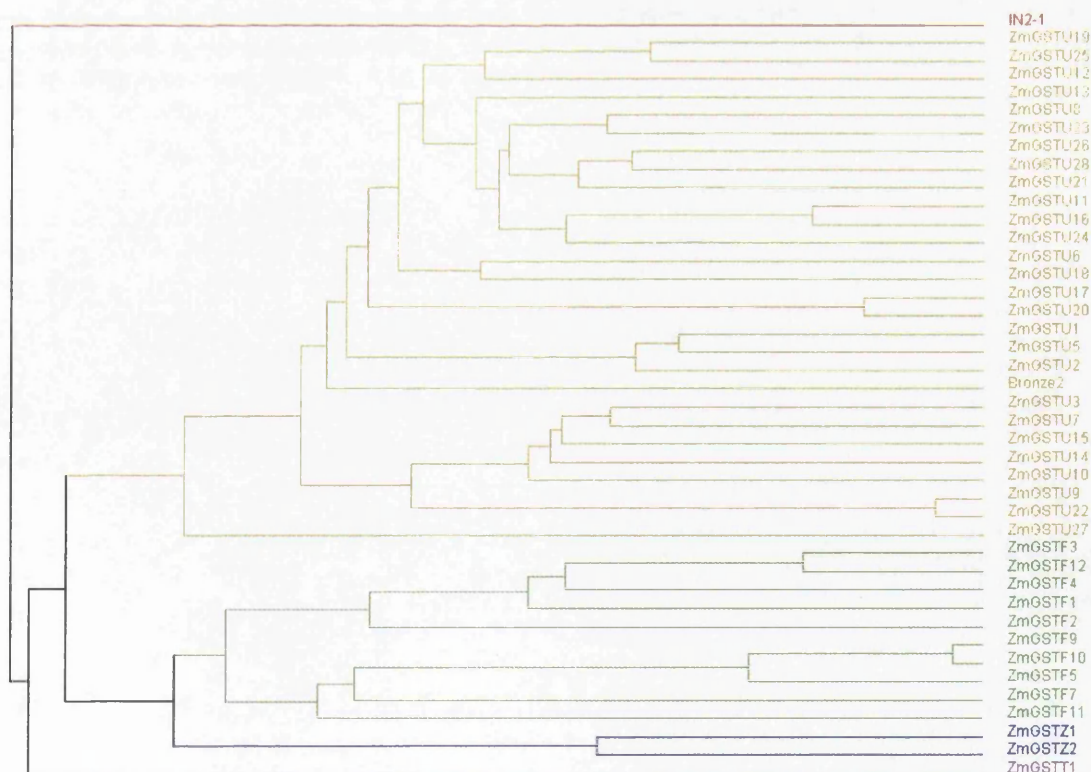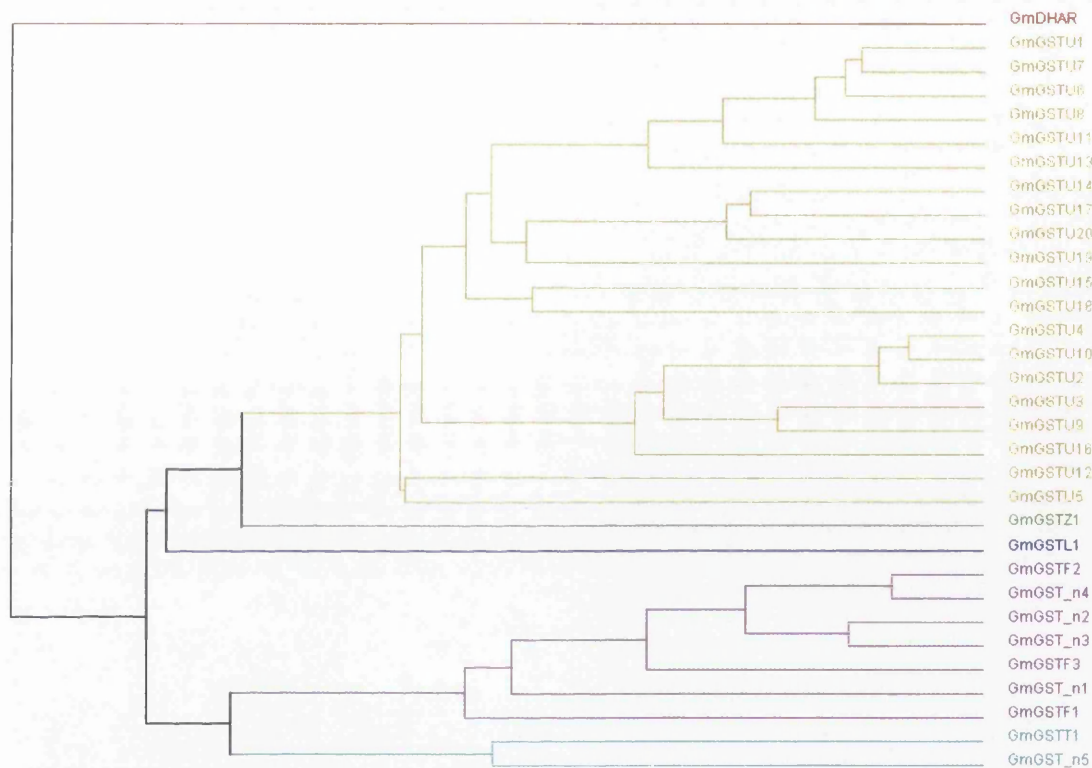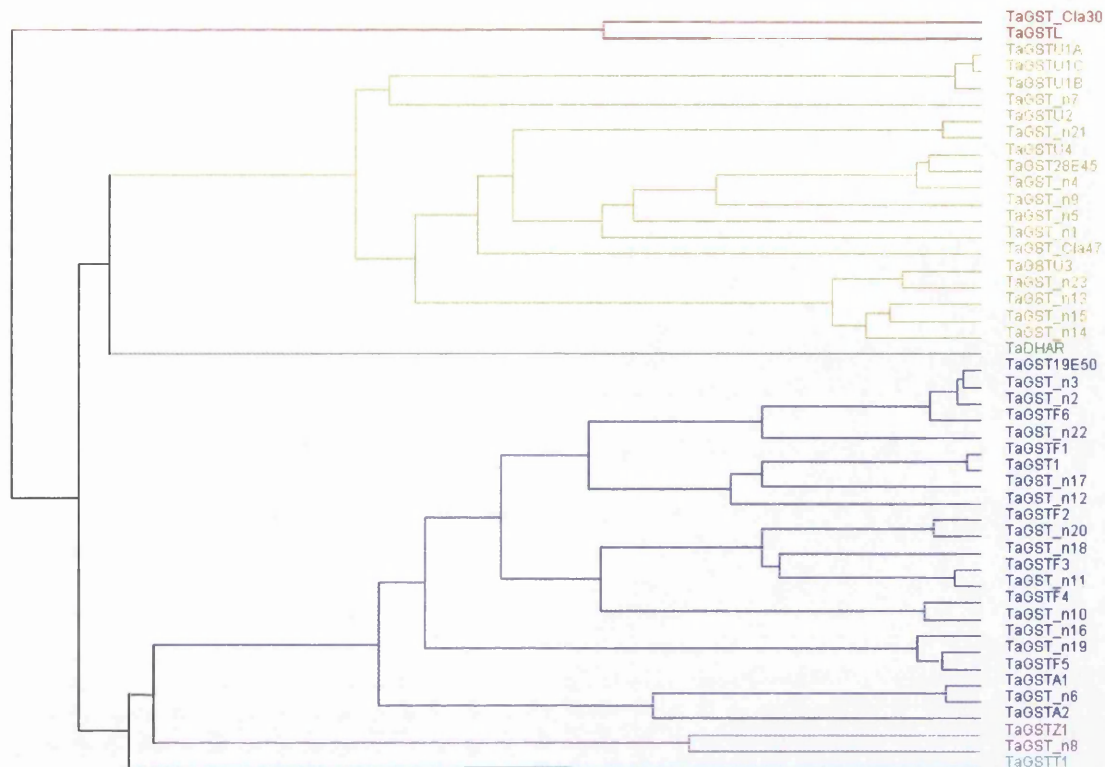
**Figure 10.5.** Phylogenetic tree of GSTs found in wheat. The sequences used to generate this tree are listed in appendix D. Sequence alignment performed with ClustalW (Higgins *et al.*, 1994) and the tree was generated using JalView (Clamp *et al.*, 2004).

## 10.2 GST Classes

The Tau class of GST enzymes is the largest class seen in the different plant species studied. Despite extensive duplication events, with up to twenty-three genes clustered together on chromosome ten of rice, the individual genes can share as little as 17% sequence identity within an individual species (Soranzo *et al.*, 2004). An alignment of Tau class GST sequences from representative plant species shows, as would be expected that the N-terminal domain of the sequences are more highly conserved than the C-terminal domain (figure 10.6). The conserved residues in the N-terminal domain are either involved in GSH binding or are structural residues. None of the residues that are conserved in the C-terminal domain are involved in ligand binding. Most of these residues are structural and are involved in the three-helix bundle at the centre of the domain and at the dimer interface.

**Figure 10.6.** Sequence alignment of Tau class GSTs from representative species with secondary structure elements of wheat Tau4 indicated (PDB ID 1GWC; Thom *et al.*, 2002). ★ Active site serine. ▲ Proposed G-site residues. ▲ Proposed H-site residues. The sequences used are (species name; GenBank® accession number): Wheat Tau4 (*Triticum aestivum*; AY013753), Rice Tau4 (*Oryza sativa*; AF309378), Maize Tau11 (*Zea mays*; AF244690), *Arabidopsis* Tau11 (*Arabidopsis thaliana*; AAF79859), Soybean Tau1 (*Glycine max*; J03917), *Arabidopsis* Tau1 (*Arabidopsis thaliana*; AAQ22631), Wheat Tau1 (*Triticum aestivum*; CAC94001), Maize Tau1 (*Zea mays*; Y12862), Rice Tau1 (*Oryza sativa*; AF050102), Soybean Tau4 (*Glycine max*; AF048978). Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

The Phi class of GST sequences is the second largest class in plant species. Phi class GSTs have also undergone extensive duplication events with two major gene clusters seen on chromosomes one and three in the rice genome (Frova, 2003). Similar to the Tau class of GSTs, the Phi class have higher sequence similarity in the N-terminal domain, and the conserved residues in the C-terminal domain are mostly structural residues with none contributing to the H-site (figure 10.7).
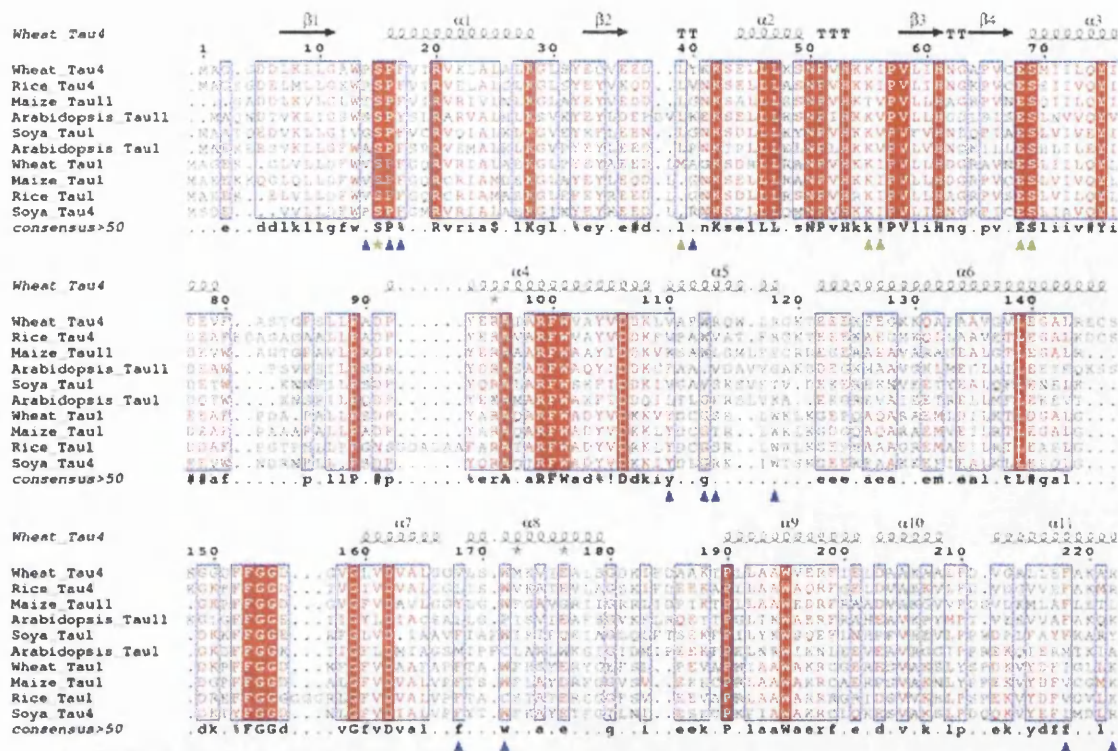
**Figure 10.7.** Sequence alignment of Phi class GSTs from representative species with secondary structure elements of maize Phi1 indicated (PDB ID 1AXD; Neuefeind *et al.*, 1997b). ★ Active site serine. ▲ Proposed G-site residues. ▲ Proposed H-site residues. The sequences used are (species name; GenBank® accession number): Maize Phi1 (*Zea mays*; M16901), Wheat Phi1 (*Triticum aestivum*; AJ440796), Rice Phi2 (*Oryza sativa*; AF062403), Rice Phi8 (*Oryza sativa*; NM_193836), Wheat Phi5 (*Triticum aestivum*; AJ440794), *Arabidopsis* Phi2 (*Arabidopsis thaliana*; AL161494), Soybean Phi3 (*Glycine max*; AF243378), *Arabidopsis* Phi1 (*Arabidopsis thaliana*; X68304). Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

The Zeta class of GSTs is one of the oldest classes and members are seen throughout all plant and animal species. Only a few copies of Zeta class genes are found in each organism and they are involved in aromatic amino acid degradation where they catalyse the isomerization of MAA to FAA (Fernández-Cañón & Peñalva, 1998). Zeta class sequences are more highly conserved than those of Phi and Tau class GSTs since there has been very little selective pressure on the genes to diverge. Unlike the Tau and Phi class, some of the conserved residues within the C-terminal domain are in the proposed H-site (figure 10.8; Polekhina *et al.*, 2001).
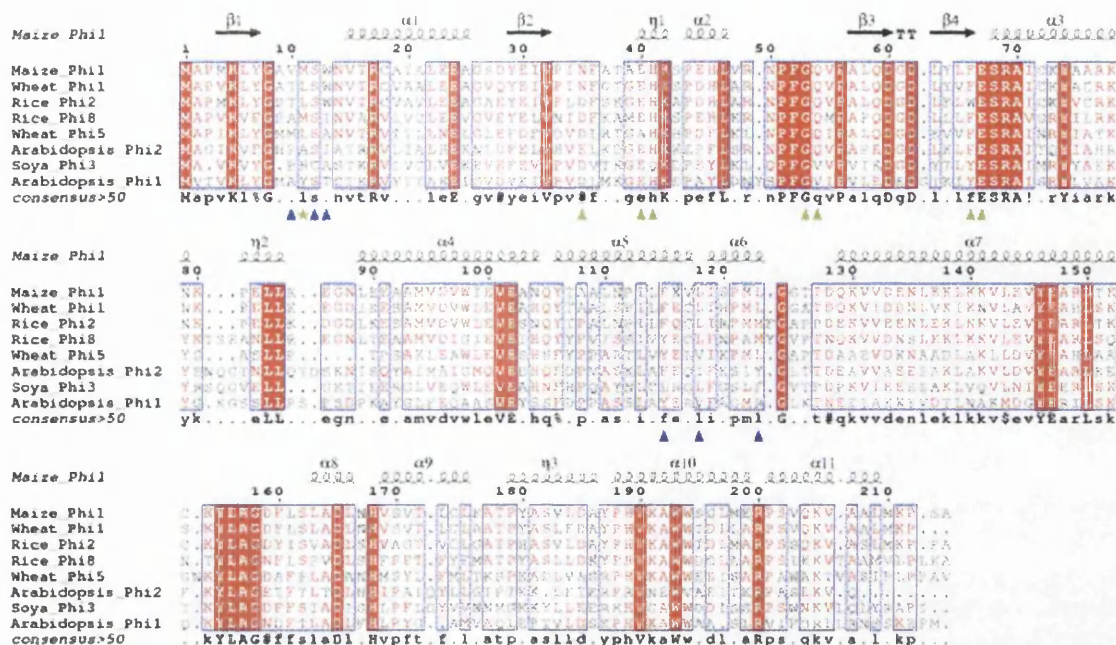
**Figure 10.8.** Sequence alignment of Zeta class GSTs from representative species with secondary structure elements of Human Zeta indicated (PDB ID 1FW1; Polekhina *et al.*, 2001). ★ Proposed active site serine. ▲ Proposed G-site residues. ▲ Proposed H-site residues. The sequences used are (species name; GenBank® accession number): Human Zeta (*Homo sapiens*; AC007954), Mouse Zeta (*Mus musculus*; NM_010363), Chicken Zeta (*Gallus gallus*; XM_421288), Frog Zeta (*Xenopus laevis*; BC087520), Zebrafish Zeta (*Danio rerio*; NM_001002481), Fruit Fly Zeta (*Drosophila melanogaster*; AY060732), Nematode Zeta (*Caenorhabditis elegans*; NM_077561), Arabidopsis Zeta (*Arabidopsis thaliana*; AY288155), Rice Zeta (*Oryza sativa*; AF309381). Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

In the Theta class of GSTs, which is found throughout all organisms, and is thought to be one of the ancestral GSTs, there is less sequence conservation seen between the enzymes than in the Zeta class (figure 10.9). Despite being involved in the detoxification of xenobiotics and having very high GST-GPOX activity there are only a few copies of Theta class genes in each organism, suggesting that unlike Tau and Phi classes there has been little pressure on the enzymes to diverge. What is unusual in this class is that not all of the G-site residues are conserved between members.
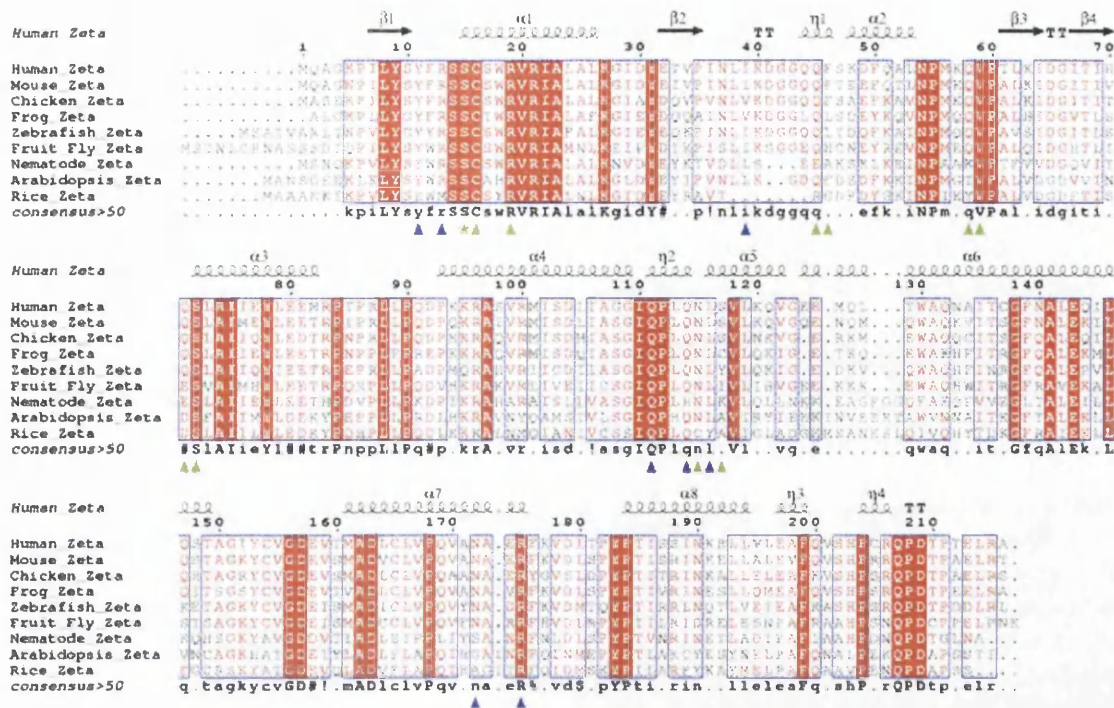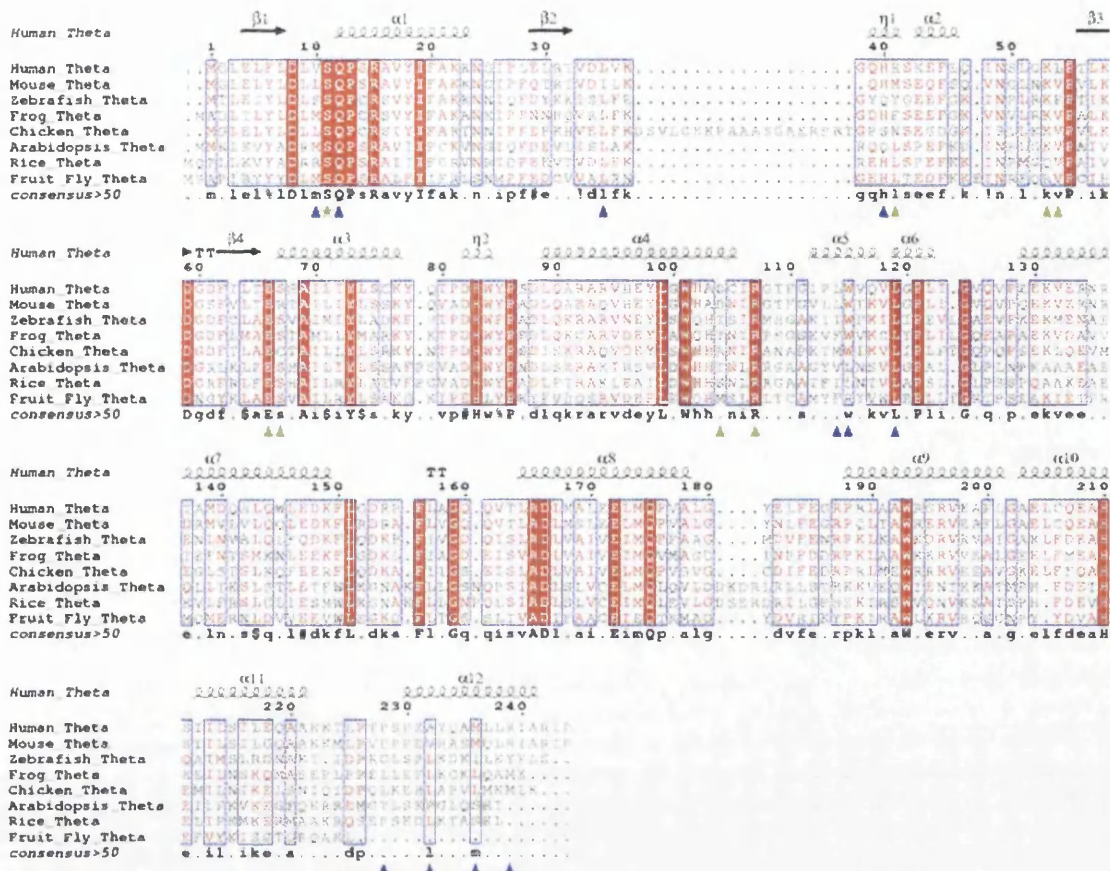
**Figure 10.9.** Sequence alignment of Theta class GSTs from representative species with secondary structure elements of Human Theta indicated (PDB ID 1LJR; Rossjohn *et al.*, 1998b). ✭ Proposed active site serine. ▲ Proposed G-site residues. ▲ Proposed H-site residues. The sequences used are (species name; GenBank® accession number): Human Theta (*Homo sapiens*; BC071700), Mouse Theta (*Mus musculus*; ), Zebrafish Theta (*Danio rerio*; BC056725), Frog Theta (*Xenopus laevis*; BC072203), Chicken Theta (*Gallus gallus*; NM_205365), Arabidopsis Theta (*Arabidopsis thaliana*; AJ131580), Rice Theta (*Oryza sativa*; AF402793), Fruit Fly Theta (*Drosophila melanogaster*; AE003571). Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

Lambda class GSTs are a plant specific class that are unusual in that they are monomeric and have very high sequence similarity between species (figure 10.10). In these enzymes the active site residue is a cysteine, allowing the enzyme to form a mixed disulphide with GSH. This confers thiol transferase activity on these enzymes and they are thought to be involved in REDOX homeostasis (Dixon *et al.*, 2002a). Plant species generally contain two Lambda class GSTs, one of which is cytosolic and one of which is targeted to the chloroplast, and in *Arabidopsis* the cytosolic enzyme is substantially induced under a variety of oxidizing conditions (Dixon *et al.*, 2002a)

**Figure 10.10.** Sequence alignment of Lambda class GSTs from representative species. The sequences used are (species name; GenBank® accession number): Maize Lambda (*Zea mays*; X58573), Wheat Lambda (*Triticum aestivum*; CAA76758), Rice Lambda (*Oryza sativa*; AC084405), *Arabidopsis* Lambda (*Arabidopsis thaliana*; NM_120356), Soybean Lambda (*Glycine max*; AF249913). Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

Microsomal GSTs are members of the MAPEG superfamily of proteins (Jakobsson *et al.*, 2000) and have been identified in bacteria, fungi, plants, and mammals (Jakobsson *et al.*, 1999). The functional unit of microsomal GSTs is trimeric and sits in the microsomal membrane with each monomer composed of a four helix bundle with the active site sitting on the cytosolic side of the membrane (Holm *et al.*, 2002). The homo-trimer binds only one molecule of GSH and the three subunits interact in a functional manner (Sun & Morgenstern, 1997; Svensson *et al.*, 2000). The proposed role of microsomal GSTs is in protection of the membranes from lipid peroxidation and in protection of the cell from lipophilic xenobiotics. Knockout experiments in *Drosophila melanogaster* have shown that flies lacking this protein exhibit a reduced lifespan (Toba & Aigaki, 2000).
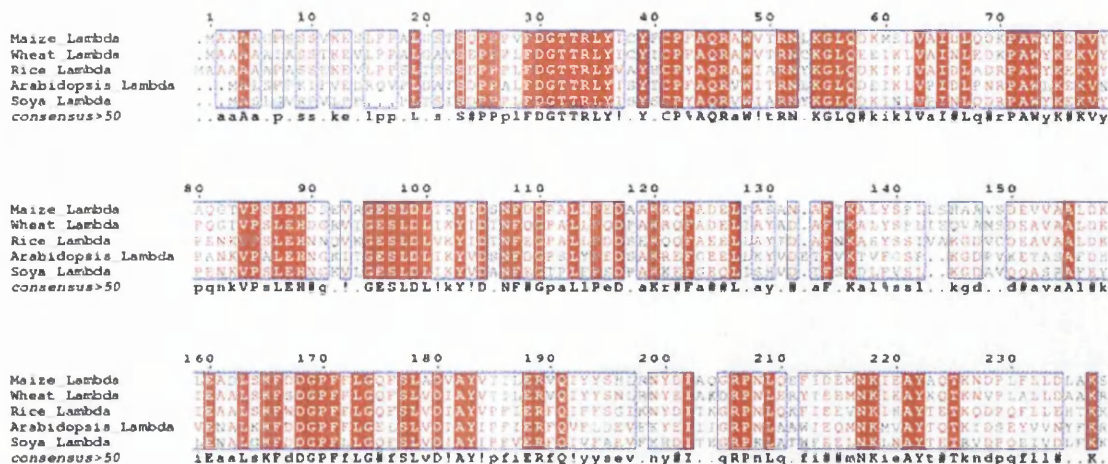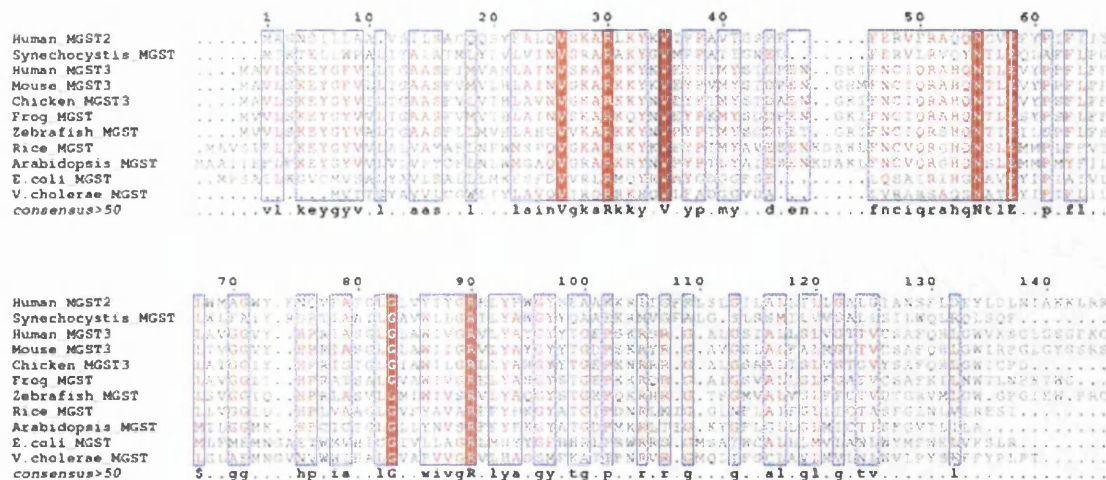
**Figure 10.11.** Sequence alignment of microsomal GSTs from representative species. The sequences used are (species name; GenBank® accession number): Human MGST2 (*Homo sapiens*; U77604), *Synechocystis* MGST (*Synechocystis sp.*; BAA17849), Human MGST3 (*Homo sapiens*; AF026977), Mouse MGST3 (*Mus musculus*; BC029669), Chicken MGST3 (*Gallus gallus*; XM_422247), Frog MGST (*Xenopus laevis*; BC087365), Zebrafish MGST (*Danio rerio*; NM_213427), Rice MGST (*Oryza sativa*; XM_469492), *Arabidopsis* MGST (*Arabidopsis thaliana*; NM_105255), *E. coli* MGST (*Escherichia coli*; AAC74939), *V. cholerae* MGST (*Vibrio cholerae*; AE004346). Sequence alignment performed with MultAlin (Corpet, 1988) and image generated with ESPript (Gouet, *et al.*, 1999).

## 10.3  Activity of GSTs

During this project the enzymes *Os*GSTU1, *Os*GSTU4, *Os*GSTF5, *Zm*GSTU4 and An9 were assayed against CDNB, the herbicide safener fenclorim, and a number of herbicides (section 2.6.). Unfortunately, due to problems with the experimental apparatus only the results for the assays with CDNB, fenclorim, fluorodifen, and fenoxaprop were interpretable (table 10.1). From this data it can be seen that *Zm*GSTU4 and An9 have very low activity towards CDNB. This is to be expected, as the primary role of these proteins is anthocyanin sequestration in the vacuole of maize and petunia respectively. The pH dependence of enzyme activities towards these substrates can also be seen by the fenclorim assays. *Os*GSTU1 and *Zm*GSTU4 performed better at pH 9.5, while *Os*GSTU4 and An9 performed better at pH 6.8.

| Substrates | Specific activity (nanokatals mg$^{-1}$) | | | | |
|---|---|---|---|---|---|
| | *Os*GSTU1 | *Os*GSTU4 | *Zm*GSTU4 | *Os*GSTF5 | An9 |
| CDNB | 105.4 | 603.1 | 4.2 | 18.2 | 4.0 |
| Fenclorim pH 6.8 | 0.008 | 0.015 | 0.001 | 0.001 | 0.004 |
| Fenclorim pH 9.5 | 0.014 | 0.010 | 0.005 | 0.001 | 0.002 |
| Fluorodifen | 0.072 | 0.020 | 0.002 | ND | 0.001 |
| Fenoxaprop | 0.007 | 0.034 | ND | 0.002 | ND |

**Table 10.1.** Specific activities of recombinant GSTs to xenobiotic compounds. Tau class GSTs are highlighted in purple and Phi class GSTs in blue.

The activities of these proteins were compared with those of several other Tau and Phi class GSTs that had been characterized (tables 10.2 and 10.3). Unfortunately, the enzymes have been characterized against different substrates so it is not always possible to make direct comparisons. All the enzymes have been assayed against CDNB however. For both Phi and Tau class GSTs there is a large range of activities seen against CDNB. In the Tau class activities range from 4.2 nkats mg$^{-1}$ for *Zm*GSTU4-4 up to 8140 nkats mg$^{-1}$ for *Ta*GSTU2-2 (Thom *et al.*, 2002). In the Phi class activities range from 4.0 nkats mg$^{-1}$ for An9 up to 2519 nkats mg$^{-1}$ for *Ta*GSTF1-1 (Cummins *et al.*, 2003).

Fluorodifen is one herbicide that has been assayed against most of the enzymes studied. The enzyme that shows the highest activity against fluorodifen is the Phi class GST, *Os*GSTF3-3 at 14.3 nkats mg$^{-1}$ (Cho & Kong, 2005). For all the other enzymes the activity is much lower or non-detectable. Interestingly, *Os*GSTU4-4 and *Ta*GSTU4-4 (Thom *et al.*, 2002), which have high sequence similarity, have almost identical activities against both CDNB and fluorodifen. *Os*GSTU1-1 shares high sequence similarities with ZmGSTU1-1 and ZmGSTU2-2 (Dixon et al., 1999), but in this case its activity against both substrates is lower, although by a similar ratio in each case.

Metolachlor has also been assayed against a wide range of enzymes. In this case two Phi class enzymes show high activity at 187.8 nkats mg$^{-1}$ for *Os*GSTF3-3 and 171.7 nkats mg$^{-1}$ for *Os*GSTF1-1 (Deng & Hatzios, 2002)

with all other Phi class and all the Tau class showing low or no activity. The profile of activity against fenoxaprop is broadly similar in both Phi and Tau class with TaGSTU4-4 having the highest activity at 1.06 nkats mg$^{-1}$. Although its activity against CDNB and fluorodifen is identical to that of OsGSTU4-4 its activity against fenoxaprop is approximately thirty times higher.

Although some of these herbicides appear to be detoxified at a higher rate in one class of enzyme than the other, this only applies to one or two members of the class in general. In some cases no real preference can be seen between the two classes suggesting that traditional views that herbicide detoxification is class specific may not be entirely true. To investigate this further it will be necessary to obtain a more complete characterization of both Phi and Tau class enzymes (which was not possible due to time constraints in this thesis). It would have been useful to include the data of McGonigle et al., (2000), however the values obtained were not consistent with assay results obtained by other researchers and were therefore omitted. In the future the ability to clone and over-express large numbers of GST ESTs will permit the combination of chemical and structural characterization which should enhance our understanding of the currently confusing diverse herbicide detoxification activity displayed by these enzymes.

| Substrates | Specific activity (nanokatals mg$^{-1}$) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OsU1 | OsU4 | ZmU1* | ZmU2* | ZmU3* | ZmU4 | TaU1[†] | TaU2[†] | TaU3[†] | TaU4[†] | AtU19[‡] | GmU1[§] | GmU2[§] | GmU3[§] |
| CDNB | 105.4 | 603.1 | 300 | 600 | 70 | 4.2 | 1970 | 8140 | 30 | 657 | 295 | 65.4 | 384.6 | 107.4 |
| Fenclorim pH 6.8 | 0.08 | 0.015 | – | – | – | 0.001 | – | – | – | – | – | – | – | – |
| Fenclorim pH 9.5 | 0.14 | 0.010 | – | – | – | 0.005 | – | – | – | – | – | – | – | – |
| Fluorodifen | 0.072 | 0.020 | 0.36 | 0.52 | 0.07 | 0.002 | 2.00 | 0.04 | ND | 0.03 | – | – | – | – |
| Acifluorfen | – | – | – | – | – | – | – | – | – | – | – | 0.04 | 0.02 | ND |
| Fomesfen | – | – | – | – | – | – | – | – | – | – | – | 0.14 | 0.05 | ND |
| Metolachlor | – | – | 0.23 | ND | 0.15 | – | 0.13 | ND | ND | 0.43 | 0.2 | 0.09 | 0.05 | 0.05 |
| Alachlor | – | – | 0.09 | ND | 0.33 | – | – | – | – | – | – | – | – | – |
| Fenoxaprop | 0.007 | 0.034 | – | – | – | ND | ND | 0.03 | 0.06 | 1.06 | – | – | – | – |
| Atrazine | – | – | ND | ND | ND | – | – | – | – | – | – | – | – | – |
| Chlorimuron Ethyl | – | – | – | – | – | – | – | – | – | – | – | 0.02 | 0.01 | ND |
| Ethacrynic Acid | – | – | 30 | ND | – | – | – | – | – | – | 2.6 | – | – | – |

**Table 10.2.** Specific activities of Tau class GSTs. ND = No significant activity detected. *Dixon *et al.*, 1999. [†]Thom *et al.*, 2002. [‡]DeRidder *et al.*, 2002. [§]Andrews *et al.*, 2005.

| Substrates | Specific activity (nanokatals mg$^{-1}$) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OsF1[*] | OsF3[†] | OsF5 | ZmF1[‡] | ZmF2[‡] | TaF1[§] | TaF2[§] | TaF3[§] | TaF4[§] | TaF5[§] | TaF6[§] | An9 |
| CDNB | 481.7 | 127 | 18.2 | 2450 | 180 | 2519 | 407 | 187 | 237 | 174 | 980 | 4.0 |
| Fenclorim pH 6.8 | 4 (pH 6.5) | – | 0.001 | – | – | – | – | – | – | – | – | 0.004 |
| Fenclorim pH 9.5 | – | – | 0.001 | – | – | – | – | – | – | – | – | 0.002 |
| Fluorodifen | – | 14.3 | ND | ND | ND | 0.01 | 0.14 | 0.1 | 0.03 | 0.03 | 0.04 | 0.001 |
| Acifluorfen | – | – | – | – | – | – | – | – | – | – | – | – |
| Fomesfen | – | – | – | – | – | – | – | – | – | – | – | – |
| Metolachlor | 171.7 | 187.8 | – | ND | ND | 0.05 | 0.32 | 0.51 | 0.06 | 0.04 | 0.04 | – |
| Alachlor | – | 112.7 | – | 0.35 | 0.07 | – | – | – | – | – | – | – |
| Fenoxaprop | – | – | 0.002 | – | – | 0.09 | 0.05 | 0.09 | 0.04 | 0.07 | 0.01 | ND |
| Atrazine | – | 36.3 | – | 0.048 | ND | – | – | – | – | – | – | – |
| Chlorimuron Ethyl | – | – | – | – | – | – | – | – | – | – | – | – |
| Ethacrynic Acid | 17.5 | 7.8 | – | 48 | 6 | ND | ND | ND | ND | ND | 1.4 | – |

**Table 10.3.** Specific activities of Phi class GSTs. ND = No significant activity detected. *Deng & Hatzios, 2002. †Cho & Kong, 2005. ‡Dixon et al., 1999. §Cummins et al., 2003.