

Received Date : 13-Oct-2016

Revised Date : 25-Nov-2016

Accepted Date : 25-Nov-2016

Article type : Research Article

Structural evidence for Arabidopsis glutathione transferase AtGSTF2 functioning as a transporter of small organic ligands

Laziana Ahmad,^{a,b} Elizabeth L. Rylott,^b Neil C. Bruce,^b Robert Edwards^{c*} and Gideon Grogan^{a*}

^aYork Structural Biology Laboratory, Department of Chemistry and ^bCentre for Novel Agricultural Products, Department of Biology, and University of York, YO10 5DD York, U.K.

^cSchool of Agriculture, Food & Rural Development, Agriculture Building, Newcastle University, NE1 7RU, U.K.

* corresponding authors

Professor Robert Edwards: Tel: 44 191 208 6869; Fax: 44 191 222 7811; e-mail: robert.edwards@ncl.ac.uk

Professor Gideon Grogan: Tel: 44 1904 328256; Fax: 44 1904 328266; e-mail: gideon.grogan@york.ac.uk

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2211-5463.12168

FEBS Open Bio (2016) © 2016 The Authors. Published by FEBS Press and John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract

Glutathione transferases (GSTs) are involved in many processes in plant biochemistry, with their best characterised role being the detoxification of xenobiotics through their conjugation with glutathione. GSTs have also been implicated in non-catalytic roles, including the binding and transport of small heterocyclic ligands such as indole hormones, phytoalexins and flavonoids. Although evidence for ligand binding and transport has been obtained using gene deletions and ligand binding studies on purified GSTs, there has been no structural evidence for the binding of relevant ligands in non-catalytic sites. Here we provide evidence of non-catalytic ligand binding sites in the phi class GST from the model plant *Arabidopsis thaliana*, AtGSTF2, revealed by X-ray crystallography. Complexes of the AtGSTF2 dimer were obtained with indole-3-aldehyde, camalexin, the flavonoid quercetrin and its non-rhamnosylated analog quercetin, at resolutions of 2.00, 2.77, 2.25 and 2.38 Å respectively. Two symmetry-equivalent binding sites (L1) were identified at the periphery of the dimer, and one more (L2) at the dimer interface. In the complexes, indole-3-aldehyde and quercetrin were found at both L1 and L2 sites, but camalexin was found only at the L1 sites and quercetin only at the L2 site. Ligand binding at each site appeared to be largely determined through hydrophobic interactions. The crystallographic studies support previous conclusions made on ligand binding in non-catalytic sites by AtGSTF2 based on isothermal calorimetry experiments (Dixon *et al.*, *Biochem. J.*, 2011, **438**, 63-70) and suggest a mode of ligand binding in GSTs commensurate with a possible role in ligand transport.

Keywords: *Arabidopsis thaliana*; GST, glutathione plant metabolism, ligand transport, flavonoids, structural biology, AtGSTF2

Abbreviations: GST, Glutathione transferase; ITC, isothermal calorimetry; LB, Luria-Bertani; IPTG, isopropyl β -D-1-thiogalactopyranoside; PTPC, propanoic acid, cacodylate, bis-tris propane system; PEG, polyethylene glycol; DMSO, dimethyl sulfoxide; PDB, Protein Databank; CDNB, 1-chloro-2,4-dinitrobenzene; MAPEG, Membrane-Associated Protein in Eicosanoid and Glutathione metabolism.

Introduction

Glutathione transferases (GSTs; E.C. 2.5.1.18) are a large group of enzymes with a major role in the detoxification of xenobiotic compounds [1-3]. GSTs promote the conjugation of the tripeptide glutathione (GSH) to an electrophilic centre within an acceptor molecule by deprotonating the GSH thiol, lowering the pK_a from 8.7 to 6.2, so as to form a thiolate of high nucleophilic reactivity [4]. Plant GSTs have been of particular interest in recent years [5], due to their role in detoxifying xenobiotics [6], including trinitrotoluene [7] and herbicides [8]. In *Arabidopsis thaliana* (*At*), a model species for plant genetic studies, 54 soluble GSTs plus one MAPEG (Membrane-Associated Protein in Eicosanoid and Glutathione metabolism) have been identified [9]. The soluble enzymes have been classified into seven distinct groups on the basis of their sequence identity: phi (F), tau (U), theta (T), zeta (Z), lambda (L), dehydroascorbate reductase (DHAR) and tetrachlorohydroquinone dehalogenase (TCHQD) [9]. A number of biochemical roles have been attributed to some of these groups. For example, in the zeta class, *At*GSTZ1 has been shown to have identical roles to its human homologue (*Hs*GSTZ1) with respect to tyrosine and phenylalanine catabolism [10]. Many of the other classes of GSTs have less well-defined functions, though members of the theta, tau and phi classes exhibit GSH-dependent peroxidase activity toward organic hydroperoxides [11].

Paradoxically, while GSTs have a conserved ability to bind GSH, the only clearly established role for GSTs demonstrated *in planta* is in anthocyanin biosynthesis, where GST mediated conjugation does not appear to be required. First demonstrated in the maize bronze-2 mutant (*ZmGSTF4*) [12], this phi-class GST was proposed to catalyse the conjugation of cyanidin-3-*O*-glucoside with GSH. However, glutathionylated anthocyanins have not been identified in plant cells. Furthermore, a phi class GST from *Petunia*, named AN9, was also shown to be involved in anthocyanin biosynthesis, but this was not dependent on conjugating activity toward these pigments *in vitro* [12]. To explain the function of these tau class GSTs in flavonoid metabolism, it has been suggested that they function as carrier proteins, facilitating sequestration of anthocyanins into the vacuole [13]. In support of this hypothesis, recent studies on the cytoplasmic and tonoplast-localised Arabidopsis *AtGSTF12* (TT19) have shown the protein can directly bind cyanidin and cyanidin-3-*O*-glucoside [14].

The Arabidopsis phi class *AtGSTF2* has been the subject of several ligand binding studies, following the observation that the protein bound both indole-3-acetic acid (IAA) and 1-*N*-naphthylphthalamic acid (NPA), an endogenous flavonoid regulator of auxin transport [15]. It was shown that NPA competed for binding with the flavonoids quercetin and kaempferol, strongly suggesting that these ligands bound to the same site in *AtGSTF2*. Later studies showed that purified recombinant *AtGSTF2* bound a range of heterocyclic compounds, including the flavonoid quercetrin (quercetin-3-*O*-rhamnoside), the indoles camalexin, harmaline, norharmaline and indole-3-aldehyde and the flavin lumichrome [16]. These binding interactions were not disrupted by the addition of GSH, and no conjugation to the ligands was observed. Furthermore the binding of harmaline and lumichrome caused changes in the catalytic GST activity of the enzyme toward the model substrate 1-chloro-2,4-dinitrobenzene (CDNB), which was suggestive of allosteric interactions that occurred at a

different site(s) to the 'G'- and 'H'-sites used for the conjugation reaction. X-ray crystallographic evidence for the ability of GSTs to employ a distinct 'L' (Ligand) site for ligand transport, separate from the GSH-conjugation site, has been previously provided by crystal structures of GSTs from organisms including the parasitic worm *Schistosoma japonica* [17] and human GSTs [18, 19, 20]. For example, in the human GSTO1 (hGSTO1), the dye Cibacron Blue and other ligands were found to bind in the hydrophobic 'H' site near, but not overlapping with, the 'G'-site [19]. A further 'L'-site in hGSTO1, again distinct from the GSH site, and in which the aromatic moiety of *S*-(4-nitrophenacyl)glutathione was bound, was found buried more deeply within the dimer interface [20]. Additionally, the ligand 4-(nitrophenol) methanethiol, thought to be a breakdown product of *S*-(*p*-nitrobenzyl)-glutathione was reported to bind to a peripheral hydrophobic binding site in the tau class GST *GmGSTU4-4* from *Glycine max* [21]. Furthermore, mutagenesis studies have suggested the presence of an 'L'-site in a phi GST from *Zea mays* (*ZmGSTF1*) that overlapped with the 'G'- and 'H'-site [22]. Despite the *in vitro* evidence for small molecule binding by members of the phi class of GSTs in Arabidopsis, few other structural insights into these interactions have yet been reported. In order to obtain further insight into the ligand transport properties of plant GSTs, we now report X-ray crystallographic studies conducted with *AtGSTF2* in the presence of a range of ligands. Three structurally distinct ligands out of six of those identified as binding partners for *AtGSTF2* in previous studies were selected for study: namely indole-3-aldehyde **1**, camalexin **2** and quercetrin **3** (**Figure 1**). The non-rhamnosylated derivative of quercetrin, quercetin **4**, was also used as a ligand. The results, in combination with ITC studies previously reported [16], provide evidence of previously unidentified ligand binding sites in *AtGSTF2*, knowledge of which will be important in understanding the involvement of these proteins in the binding and transport of small molecules in various plant physiological processes.

Materials and Methods

Gene expression and protein purification

The pET24b vector containing the *AtGSTF2* gene, as prepared by Dixon and co-workers [16], was used to transform *Escherichia coli* Tuner (DE3) cells (Novagen) that also contained the pRARE plasmid from Rosetta (Novagen). Transformants were grown on Luria Bertani (LB) agar supplemented with 100 $\mu\text{g mL}^{-1}$ of kanamycin and 50 $\mu\text{g mL}^{-1}$ of chloramphenicol at 37 °C. A single colony of a plate grown overnight was used to inoculate 4 x 5 mL of LB broth. These starter cultures were grown overnight at 37 °C with shaking at 180 r.p.m. and were then used to inoculate LB broth (4 x 500 mL cultures) in which cells were grown until the optical density (OD_{600}) of the culture had reached approximately 0.6. At this point the expression of *AtGSTF2* was induced by the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG, final concentration of 1 mM). The cultures were then incubated at 20 °C in an orbital shaker overnight at 180 r.p.m. After approximately 18 h growth, the cells in each case were harvested by centrifugation at 4225 x g for 15 min in a Sorvall RC5B Plus centrifuge and were then resuspended in Tris/HCl buffer pH 7.5 (100 mL, 20 mM, henceforth referred to as 'buffer'). Cells were disrupted by ultrasonication for 3 x 30 s bursts at 4 °C with 1 min intervals and the soluble and insoluble material fractions separated by centrifugation at 26892 x g for 30 min. The supernatant, containing the soluble *AtGSTF2*, was loaded onto a 10 mL GSH sepharose 4B (GE healthcare). Column fractions were analysed by SDS PAGE and the fractions containing purified proteins were pooled and concentrated using a 10 kDa cut-off Centricon® filter membrane. Concentrated protein was loaded onto an S75 Superdex™ gel filtration column that had been equilibrated with buffer also including addition of 150 mM NaCl. Fractions containing pure protein were pooled and stored at -20 °C.

Protein crystallisation

Ligands **1** – **4** were purchased from Sigma Aldrich. Pure AtGSTF2 was pre-incubated with ligands **1**, **2**, **3** or **4** prior to crystallisation experiments, at concentrations of either 5 or 10 mM for 1 h, followed by microcentrifugation at 16,300 x g to remove any insoluble precipitates resulting from complexation. Ligand-complexed proteins were then subjected to crystallisation trials using a Mosquito® ROBOT (TTP LabTech) and a range of commercially available crystallisation screens in 96-well plate sitting drop format, in which each drop consisted of 150 nL protein and 150 of precipitant reservoir solution. Crystals of AtGSTF2 in complex with indole-3-aldehyde **1** and camalexin **2** were obtained in 0.2 M sodium acetate and 20% (w/v) PEG 3350. Initial crystals observed for the complex of AtGSTF2 with quercetrin **3** and AtGSTF2 with quercetin **4** were in 0.1 M PTPC (propanoic acid, cacodylate, bis-tris propane system) and 15% (w/v) PEG 1.5K at pH 7.0. In all cases a protein concentration of 10 mg mL⁻¹ was employed. Larger crystals for diffraction analysis were obtained using the hanging drop vapour diffusion method in 24-well plate Linbro dishes, with 2 µL drops consisting of a 1:1 ratio of mother liquor to protein solution. The best crystals of AtGSTF2-indole-3-aldehyde **1** and AtGSTF2-camalexin **2** complexes were obtained in drops containing 0.2 M sodium acetate and 20% (w/v) PEG 3350 with 1% (v/v) *n*-propanol. For AtGSTF2-quercetrin **3** and AtGSTF2-quercetin **4** complexes, the best crystals were obtained from drops using the same conditions employed in the Mosquito® screen. Prior to analysis on in-house X-ray equipment, the crystals were washed with the mother liquor solution containing 20% (v/v) ethylene glycol as cryoprotectant and the appropriate ligand at the crystallisation concentration, followed by flash-cooling in liquid nitrogen. Crystals were tested for diffraction using a Rigaku Micromax-007HF fitted with Osmic multilayer optics and a MARRESEARCH MAR345 imaging plate detector. Those

crystals that diffracted to a resolution of equal to, or better than, 3 Å resolution were retained for dataset collection at the synchrotron.

Data collection, structure solution, model building and refinement

Complete datasets described in this report were collected at Diamond Light Source, Didcot, Oxfordshire, U.K. Complexes with **1** and **2** were collected on beamline I04-1 and complexes with **3** and **4** on beamline I03. Data were processed and integrated using XDS [23] and scaled using SCALA [24] included in the Xia2 processing system [25]. Data collection statistics are given in **Table 1**. Complexes of *At*GSTF2-indole-3-aldehyde **1**, *At*GSTF2-quercetrin **3** and *At*GSTF2-quercetin **4** were each in space group $P2_12_12_1$, with six molecules in the asymmetric unit, constituting a trimer of dimers. The crystals of complex of *At*GSTF2-camalexin **2** were in space group $P1$ with 24 molecules in the asymmetric unit, consisting of four trimers of dimers. The structure of each complex was solved using MOLREP [26], using a monomer of *At*GSTF2 (PDB code 1GNW; 100% sequence identity) as the model. The solvent content in the *At*GSTF2-indole-3-aldehyde **1**, *At*GSTF2-quercetrin **3** and *At*GSTF2-quercetin **4** complexes was 42% and in the *At*GSTF2-camalexin **2** complex was 47%. The structures were built and refined using iterative cycles using Coot [27] and REFMAC [28], employing local NCS restraints in the refinement cycles. Following building and refinement of the protein and water molecules, clear residual density was observed in the omit maps at the dimer interfaces within the larger hexameric complexes. In each case, these could be successfully modelled as the ligands that had been used for co-crystallisation. Ligands and associated refinement libraries were prepared using PRODRG [29]. The complex with **1** featured three molecules of **1** per dimer, with two at the **L1** and one at the **L2** sites. The complex with **2** featured two molecules per dimer, one at each of the **L1** sites, and while there was some density at the **L2** site, the occupancy was not considered sufficiently

substantial to model the camalexin ligand. The complex with quercetrin **3** featured three ligands in the dimer, at **L1** and **L2**, although, in two of the dimers, sufficient density was only observed for two ligands to be modelled. In addition, the rhamnose moiety of one of the seven ligands, at the **L2** site between subunits 'C' and 'D', could not be modelled. The complex with quercetin **4** featured one ligand per dimer, at the **L2** site in each case. The final structures exhibited % R_{cryst} and R_{free} values of 19.9/23.4 (complex with **1**); 25.0/28.4 (complex with **2**); 21.4/25.2 (complex with **3**) and 20.4/24.6 (complex with **4**). All structures were finally validated upon deposition at the PDB. Refinement statistics for all structures are presented in **Table 1**. The Ramachandran plot for the complex with **1** showed 98.4% of residues to be situated in the most favoured regions, 1.0% in additional allowed and 0.6% residues in outlier regions. For the complex with **2**, the corresponding values were 96.5%, 3.0% and 0.5%. For the complex with **3**, the corresponding values were 98.0%, 1.3% and 0.7%. For the complex with **4**, the corresponding values were 98.2%, 1.0% and 0.8%. Coordinates and structure factors for *At*GSTF2 complexes with indole-3-aldehyde **1**, camalexin **2**, quercetrin **3** and quercetin **4** have been deposited in the Protein Databank (PDB; <http://www.rcsb.org/pdb/>) with accession codes 5a4u, 5a5k, 5a4w and 5a4v respectively.

RESULTS

Crystal Structures of *At*GSTF2 in complex with ligands **1** – **4**.

The crystal structure of *At*GSTF2 has been published previously in complex with *S*-hexylglutathione (1GNW) [30] and also the glutathione conjugate of the herbicide FOE-4053 (1BX9) [31]. In the crystal structure 1GNW, protomers of *At*GSTF2 are found in a classical dimeric association. In order to determine crystallisation conditions for complex formation between *At*GSTF2 and the heterocyclic ligands **1-4**, fresh crystallisation screens were

performed with the protein pre-incubated with 10 mM ligand (or 5 mM ligand in the case of the less soluble quercetin). Crystal complexes were obtained in each case.

The statistics for data collection and refinement are shown in **Table 1**. As determined with the glutathione-conjugate complex structure 1GNW [30], *At*GSTF2 structures featured dimers in each ligand complex structure, with different numbers of monomers observed in the asymmetric unit depending on the space group. In the case of complexes formed with **1**, **3** and **4**, crystals grew in the $P2_12_12_1$ space group, with six monomers in the asymmetric unit. In the case of camalexin **2**, the space group was $P1$, with twenty-four monomers found. After building the peptide backbone, side chains and water molecules, clear residual density for each complex was observed in omit maps at an electron density level of 3σ that could be modelled as the relevant ligand in each case.

Figure 2 shows the structure of the *At*GSTF2 dimer as observed in the complex with indole-3-aldehyde **1**, with the selected alpha-helices labelled for ease of reference. Secondary structure analysis on the PDB server (<http://www.rcsb.org/pdb/>) shows that each monomer contains twelve α -helices ($\alpha 1$: residues 13-24; $\alpha 2$: 46-49; $\alpha 3$: 68-78; $\alpha 4$: 94-109; $\alpha 5$: 112-122; $\alpha 6$: 124-127; $\alpha 7$: 134-157; $\alpha 8$: 168-171; $\alpha 9$: 174-180; $\alpha 10$: 186-190; $\alpha 11$: 193-204 and $\alpha 12$: 206-211) two 3_{10} helices (36-38 and 40-42) and four β -strands ($\beta 1$: 4-8; $\beta 2$: 30-33; $\beta 3$: 57-60 and $\beta 4$: 63-66). For the complex with **1**, representative monomer pairs in a dimer superimposed with an rmsd of 0.44 Å over 209 C-alpha atoms, with no significant differences in amino acid side chain positions. For complexes with **2**, **3** and **4**, the rmsd values were 0.13, 0.37 and 0.35 Å respectively. The ligand **1** was observed in three locations in the dimer (**Figure 3-I**). Two of these sites, each named **L1**, were symmetry equivalent and located in a hydrophobic binding pocket formed between helices α -4 and α -7 and the loop region between Lys159 and Glu164 in each subunit

of the dimer. The other site **L2**, was found at the base of the dimer interface, with contributions from helix α -3 of one monomer and α -4 of its neighbour. None of the new ligand binding sites was close in space to the glutathione binding GSX site at the concave head of the dimer (**Figure 3-V**). For the complex with camalexin **2**, ligand density was only observed at the **L1** sites, with two ligands bound per dimer (**Figure 3-II**). By contrast, the much larger quercetrin **3**, with the pendant rhamnose, was observed at **L1** and **L2** sites in the complex structure (**Figure 3-III**), with three ligands in the dimer. **Figure 4** illustrates surface representations of this complex. In contrast, the non-rhamnosylated flavonol quercetin **4** only displayed ligand density in the **L2** site at the dimer interface (**Figure 3-IV**), with just one ligand per dimer.

Characterisation of the Ligand Binding Sites

Detail of the ligand binding site **L1** from the complex with indole-3-aldehyde **1** is shown in **Figure 5a**, with representative electron density from the relevant maps. Ligand **1** is bound in a hydrophobic pocket formed by Val150, Tyr151, Ile102, Val106 and Arg154 of one subunit ('A') and Phe52 and Phe66 ('B') of its neighbour. The plane of the aromatic ligand is sandwiched between the side-chain of Arg154 and a hydrophobic shelf formed from (A)Ile102, (A)Val106 and (B)Phe52 and (B)Phe66. The aldehyde moiety is at a distance of 3.2 Å from the peptidic carbonyl of Leu161, 3.7 Å from the backbone carbonyl of Ile99, and 4.0 Å from the side-chain hydroxyl of (A)Thr169. The indole nitrogen was not observed to make hydrogen bonding contact with any side chains in these sites.

The binding of indole-3-aldehyde **1** in the **L2** site was almost entirely characterised by hydrophobic interactions (**Figure 5b**). The plane of the bicyclic indole is sandwiched between the side chain of (A)His77 and (B)Tyr97; the heterocyclic nitrogen is 4.1 Å from the

side chain of (A)Gln73. The **L2** site is symmetrical, owing to its location at the 2-fold axis of the monomer interface, but ligand density for **1** was much less substantial at the putative reciprocal binding site formed by (B)His77, (A)Tyr97 and (B)Gln73, and the ligand was not successfully modelled here.

As with **1**, the plane of the camalexin ligand again lies within the hydrophobic pocket formed by the side chains of binding site **L1** (**Figure 6**); The aromatic rings of **1** and **2** can be superimposed from their complex structures, but, being a larger ligand, the thiazole ring of **2** is observed beneath the guanidinium group of Arg154, and projects more than **1** toward the periphery of the dimer. The indole ring is rotated approximately 60° relative to the orientation observed with **1**, bringing the indole nitrogen within a distance of 4.4 Å of the backbone carbonyl of Val150.

The rhamnosylated flavonoid quercetrin **3** is the largest of the four ligands for which a complex was obtained. In the **L1** sites, the resorcinol ring of the flavone occupies the equivalent site to the benzene ring of **1** (**Figure 7a**). The OAC hydroxyl is 2.5 Å from water molecule, which, in turn, is 2.9 Å from the phenolic hydroxyl of Tyr151. The planar bicyclic chromanone system is stacked between Arg154 and the hydrophobic shelf, as with **1**, but the side chain of Arg154 is shifted relative to the **1** complex owing to the presence of the catechol substituent on the chromanone system. The plane of the catechol ring substituent is rotated approximately 45° relative to the chromanone. The OAE catechol hydroxyl is 3.9 Å from the guanidinium group of the displaced Arg154 side chain. The rhamnose sugar assumes a conformation parallel to that of the catechol ring, with hydroxyl groups O2 and O3 3.1 Å and 3.3 Å respectively from with the backbone carbonyl group of Ser48 in the 'B' monomer at the dimer interface.

Both quercetrin **3** and the non-rhamnosylated quercetin **4** are found within the **L2** site. In the complex with **4**, the tricyclic flavone superimposes exactly with that of the rhamnosylated **3**. As a consequence of the two-fold symmetry of this site, there is some evidence for these larger ligands being bound in reciprocal conformations. The most convincing refinement has the chromanone ring stacked between (A)His77 and (B)Tyr97. The OAC atom of the resorcinol moiety is 3.5 Å from the side-chain amide of (A)Gln73 and the OAF at a distance of 3.4 Å from the backbone carbonyl of (B)Ile94. The catechol ring is stacked between (A)Tyr97 and (B)His77. In contrast to quercetrin binding in the **L1** site, the three rings of the flavone system are co-planar in the **L2** site (**Figure 7b**). In the complex with quercetrin **3**, the rhamnose occupies a site at the periphery of the dimer, with the endocyclic oxygen 2.9 Å from the backbone carbonyl of (B)Ser91. The O2 hydroxyl of rhamnose is also 3.0 Å from the backbone carbonyl of (B)Lys92.

Discussion

The data reported herein represent the first structures of a plant GST complexed with natural products, through selective hydrophobic interactions localized to two newly-identified ligand-binding sites **L1** and **L2**. These were remote from the active site of the enzyme more classically associated with interactions with xenobiotics, and their glutathionylated derivatives, formed following conjugation (GSX sites in **Figure 3-V**) [30]. Each site is also distinct from the peripheral hydrophobic site previously described for 4-(nitrophenol) methanethiol in the tau class GST *GmGSTU4-4* from *Glycine max* [21]. The residues forming the binding site **L1** do not appear to be well-conserved among plant GSTs for which the structures have been determined, featuring neither in zeta (1E6B) [10] or tau (1GWC) [32] plant GSTs; indeed in 1GWC a tryptophan residue W101, which superimposes with Gly103 in *AtGSTF2*, occupies the **L1** site. This Trp is also conserved in the phi GST F1 (4RI6) from

poplar [33]. In the L2 site, the hydrophobic residues His77 and Trp97 that form the hydrophobic pocket binding the aromatic ligands are again not conserved in 1E6B or 1GWC, being replaced by Glu and Arg residues respectively, and Asp and Lys in 4RI6.

The classes of heterocyclic ligands bound within the *At*GSTF2 structures reported herein represent important types of biologically active plant secondary metabolites derived from indoles and polyphenols respectively. The selectivity of these binding interactions is suggestive of a physiological function. Roles associated with interactions located away from the active site are most likely related to sequestration and transport of biologically active ligands. A role for these interactions in the allosteric activation of *At*GSTF2 has also been suggested, based on an observed increase in k_{cat} for the conjugation of the model compound 1-chloro-2,4-dinitrobenzene (CDNB) with GSH in the presence of harmaline [16], however, no such activation was observed for ligands **1-4** in the present study (data not shown) and very little change in protein structure was observed when the ligand complexes obtained herein were compared with structures of the *apo*-protein. While we recognize the possibilities of ligand binding as a crystallographic artefact at the ligand concentrations used in this study, both binding constants and enthalpy of complex formation values determined by ITC were presented by Dixon [16], and suggest agreement with the structural observations. Quercetin **3**, which has the most polar functionality of ligands **1** to **4**, binds in both **L1** and **L2** sites, and displays most hydrogen bonding interactions with the protein, gave a ΔH value of $-21.1 \text{ kcal mol}^{-1}$ and a K_a of $0.16 \mu\text{M}^{-1}$ in that work. Indole-3-aldehyde **1**, which also binds in both **L1** and **L2** sites, but makes fewer interactions, gave a less negative ΔH value of $-13.7 \text{ kcal mol}^{-1}$ but a comparable value for the K_a , of $0.09 \mu\text{M}^{-1}$. Camalexin, which was observed to bind only in the **L1** site, gave the least negative ΔH value of $-9.3 \text{ kcal mol}^{-1}$, and a higher affinity constant of $0.84 \mu\text{M}^{-1}$.

The current study gives a structural basis for *At*GSTF2 being formally identified as an auxin binding protein [15] as well as explaining how the interactions with bioactive indoles are directly affected by competitive binding at the same L sites by specific flavonols [16, 34]. Plant secondary metabolites are of great industrial importance and understanding the specificity behind GST-ligands and how and where they are transported, would answer important biological questions, and contribute towards the genetic modification of plants for biotechnological applications.

Author Contributions

L.A. and E.L.R. performed experiments; E.L.R., N.C.B., R.E. and G.G. designed experiments and supervised the work; R.E. and G.G. wrote the manuscript, with contributions from other authors.

Acknowledgements

We thank Majlis Amanah Rakyat (MARA), Malaysia for a studentship to L.A. We also thank the Diamond Light Source for access to beamlines I03 and I04-1 under proposal number mx-9948.

References

- 1 Eaton, D.L. and Bammler, T.K. (1999) Concise review of the glutathione *S*-transferases and their significance to toxicology. *Toxicol. Sci.*, **49**, 156-164
- 2 Salinas, A.E. and Wong, M.G. (1999) Glutathione *S*-transferases-a review. *Curr. Med.Chem.*, **6**, 279-309

- 3 Sheehan, D., Meade, G., Foley V.M. and Dowd, C.A. (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.*, **360**, 1-16
- 4 Labrou, N.E., Mello, L.V. and Clonis, Y.D. (2001) Functional and structural roles of the glutathione-binding residues in maize (*Zea mays*) glutathione S-transferase I. *Biochem. J.*, **358**, 101-110
- 5 Edwards, R., Dixon, D.P. and Walbot, V. (2000) Plant glutathione S-transferases: enzymes with multiple functions in sickness and health. *Trends Plant Sci.*, **5**, 193-198
- 6 Cummins, I, Dixon, D.P., Freitag-Pohl, S., Skipsey, M. and Edwards, R. (2011) Multiple roles for plant glutathione transferases in xenobiotic detoxification. *Drug Metab. Rev.* **3**, 266-280
- 7 Gunning, V., Tzafestas, K., Sparrow, H., Johnston, E.J., Brentnall, A.S., Potts, J.R., Rylott, E.L. and Bruce, N.C. (2014), Arabidopsis glutathione transferases U24 and U25 exhibit a range of detoxification activities with the environmental pollutant and explosive, 2,4,6-trinitrotoluene., *Plant Physiol.* **165**, 854-865
- 8 Cummins, I., Cole, D.J. and Edwards, R. (1997) Purification of multiple glutathione transferases involved in herbicide detoxification from wheat (*Triticum aestivum* L.) treated with the safener fenchlorazole-ethyl. *Pestic. Biochem. Phys.* **59**, 35-49
- 9 Dixon, D.P. and Edwards, R. E. (2010) ‘Glutathione transferases’ in ‘The Arabidopsis Book’. **8**, e0131. doi: 10.1199/tab.0131. American Society of Plant Biologists.
- 10 Thom, R., Dixon, D.P., Edwards, R., Cole, D.J. and Laphorn, A.J. (2001) The structure of a zeta class glutathione S-transferase from *Arabidopsis thaliana*:

characterisation of a GST with novel active-site architecture and a putative role in tyrosine catabolism. *J.Mol. Biol.*, **308**, 949-962

- 11 Kili, K.G., Atanassova, N., Vardanyan, A., Clatot, N., Al-Sabarna, K., Kanellopoulos, P.N. Makris, A.M. and Kampranis, S.C. (2004) Differential roles of tau class glutathione *S*-transferases in oxidative stress *J. Biol. Chem.* **279**, 24540-24551
- 12 Alfenito, M.R., Souer, E., Goodman, C.D., Buell, R., Koes, R. and Walbot, V. (1998) Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione *S*-transferases. *Plant Cell*, **10**, 1135-1149
- 13 Dixon, D., Skipsey, M. and Edwards, R. (2010) Roles for glutathione transferases in plant secondary metabolism. *Phytochemistry*, **71**, 338-350
- 14 Sun, Y., Li, H. and Huang, J-R. (2012) Arabidopsis TT19 Functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Molecular Plant*, **5**, 387-400
- 15 Smith, A.P., Nourizadeh, S.D., Peer, W.A., Xu, J., Bandyopadhyay, A., Murphy, A.S. and Goldsbrough, P.B. (2003) Arabidopsis AtGSTF2 is regulated by ethylene and auxin, and encodes a glutathione *S*-transferase that interacts with flavonoids. *Plant J.*, **36**, 433-442
- 16 Dixon, D.P., Sellars, J.D. and Edwards, R. (2011) The Arabidopsis phi class glutathione transferase AtGSTF2: binding and regulation by biologically active heterocyclic ligands. *Biochem. J.*, **438**, 63-70
- 17 McTigue, M.A., Williams, D.R. and Tainer, J.A. (1995) Crystal structures of a schistosomal drug and vaccine target: glutathione *S*-transferase from *Schistosoma*

japonica and its complex with the leading antischistosomal drug praziquantel. J. Mol. Biol., **246**, 21-27

18 Ji, X., Tordova, M., O'Donnell, R., Parsons, J.F., Hayden, J.B., Gilliland, G.L. and Zimniak, P. (1997) Structure and function of the xenobiotic substrate-binding site and location of a potential non-substrate-binding site in a class π glutathione S-transferase Biochemistry, **36**, 9690-9702

19 Oakley, A.J., Lo Bello, M., Nuccetelli, M., Mazzetti, A.P. and Parker, M.W. (1999) The ligandin (non-substrate) binding site of human π class glutathione transferase is located in the electrophile binding site (H-site) 1. J. Mol. Biol. **291**, 913-926

20 Brock, J., Board, P.G. and Oakley, A.J. (2013) Structural insights into omega-class glutathione transferases: A snapshot of enzyme reduction and identification of a non-catalytic ligandin site. PLoS ONE, **8**, e60324

21 Axarli, I., Dhavala, P., Papegeorgiou, A.C. and Labrou, N.E. (2009) Crystallographic and Functional Characterization of the Fluorodifen-inducible Glutathione Transferase from *Glycine max*, reveals an active site topography suited for diphenylether herbicides and a novel L-site. J. Mol. Biol. **385**, 984-1002.

22 Axarli, I.A., Rigden, D.J. and Labrou, N.E. (2004) Characterization of the ligandin site of maize glutathione S-transferase I. Biochem. J., **382**, 885-893

23 Kabsch, W. (2010) XDS. Acta Crystallogr. Sect. D. Biol. Crystallogr. **66**, 125-132

24 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr. Sect. D. Biol. Crystallogr. **62**, 72-82

- 25 Winter. G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.*, **3**, 186-190
- 26 Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.*, **30**, 1022-1025
- 27 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect D: Biol. Crystallogr.*, **60**, 2126-2132
- 28 Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallogr. Sect. D. Biol. Crystallogr.* **53**, 240-255
- 29 Schüttelkopf, A.W. and van Aalten, D.M.F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta. Crystallogr. Sect. D. Biol. Crystallogr.*, **60**, 1355-1363
- 30 Reinemer, P., Prade, L., Hof, P., Neufeind, T., Huber, R., Zettl, R., Palme, K., Schell, J., Koelln, I., Bartunik, H.D. and Bieseler, B. (1996) Three-dimensional structure of glutathione S-transferase from *Arabidopsis thaliana* at 2.2 Å resolution: structural characterization of herbicide-conjugating plant glutathione S-transferases and a novel active site architecture. *J. Mol. Biol.* **255**, 289-309
- 31 Prade, L., Huber, R. and Bieseler, B. (1998) Structures of herbicides in complex with their detoxifying enzyme glutathione S-transferase – explanations for the selectivity of the enzyme in plants, *Structure*, **6**, 1445-1452

- 32 Thom, R., Cummins, A., Dixon, D.P., Edwards, R., Cole, D.J. and Laphorn, A.J. (2002) Structure of a tau class glutathione-S-transferase from wheat active in herbicide detoxification. *Biochemistry*, **41**, 7008-7020.
- 33 Pégeot, H., Koh, C. S., Petre, B., Mathiot, S., Duplessis, S., Hecker, A. and Rouhier, N. (2014). The poplar phi class glutathione transferase: Expression, activity and structure of GSTF1. *Front, Plant Sci.*, **5**, 712.
- 34 Zhao, J. (2015) Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.*, **20**, 576-585.

Table 1. Data Collection and Refinement Statistics for AtGSTF2 in complex with indole-3-aldehyde **1**, camalexin **2**, quercetrin **3** and quercetin **4**. Numbers in brackets refer to data for highest resolution shells.

	Complex with indole-3- aldehyde 1	Complex with camalexin 2	Complex with quercetrin 3	Complex with quercetin 4
Beamline	Diamond I03	Diamond I03	Diamond I04-1	Diamond I04-1
Wavelength (Å)	0.97625	0.97625	0.92000	0.92000
Resolution (Å)	94.41-2.00 (2.05-2.00)	87.58-2.77 (2.84-2.77)	59.09-2.25 (2.31-2.25)	59.59-2.38 (2.44-2.38)
Space Group	<i>P2₁2₁2₁</i>	<i>P1</i>	<i>P2₁2₁2₁</i>	<i>P2₁2₁2₁</i>
Unit cell (Å)	a = 87.86; b = 94.41; c = 152.38	a = 97.10; b = 113.72; c = 132.02	a = 87.35; b = 93.57; c = 152.42	a = 88.03; b = 94.83; c = 153.20
	$\alpha = \beta = \gamma = 90^\circ$	$\alpha = 83.7^\circ$ $\beta =$ 79.5° $\gamma = 65.9^\circ$	$\alpha = \beta = \gamma = 90^\circ$	$\alpha = \beta = \gamma = 90^\circ$
No. of molecules in the asymmetric unit	6	24	6	6
Unique	86285 (6308)	126932 (9333)	60022 (4365)	52158 (3838)

reflections

Completeness (%)	100.0 (100.0)	98.6 (98.1)	100.0 (100.0)	99.9 (100)
R_{merge} (%)	0.08 (0.63)	0.10 (0.72)	0.09 (0.72)	0.11 (0.68)
$R_{\text{p.i.m.}}$	0.04 (0.34)	0.10 (0.72)	0.04 (0.32)	0.07 (0.41)
Multiplicity	8.1 (8.4)	2.2 (2.2)	6.8 (7.1)	6.7 (7.0)
$\langle I/\sigma(I) \rangle$	17.2 (3.3)	6.8 (1.8)	16.5 (3.2)	14.9 (2.8)
$\text{CC}_{1/2}$	1.00 (0.89)	0.99 (0.74)	1.00 (0.88)	1.00 (0.84)
Overall B factor from Wilson plot (\AA^2)	28	35	30	21
$R_{\text{cryst}}/R_{\text{free}}$ (%)	19.9/23.4	25.0/28.4	21.4/25.2	20.4/24.6
r.m.s.d 1-2 bonds (\AA)	0.017	0.014	0.012	0.012
r.m.s.d 1-3 angles ($^\circ$)	1.85	1.98	1.69	1.46
Avg main chain B (\AA^2)	32	49	37	34
Avg side chain B (\AA^2)	35	51	40	37
Avg water B (\AA^2)	33	29	37	33
Avg ligand B (\AA^2)	26	53	51	44

Figure 1. Ligands Used in this Study. **1** = Indole-3-aldehyde; **2** = Camalexin; **3** = Quercetrin; **4** = Quercetin.

Figure 2. Structure of *At*GSTF2 dimer. The figure is derived using the complex with indole-3-aldehyde and shows selected helices and location of ligand binding sites **L1** and **L2** labelled for ease of reference.

Figure 3. Structure of dimers 'A/B' from ligand complex structures of *At*GSTF2 and showing location of ligands in binding sites **L1** and **L2**. **I**: Complex with Indole-3-aldehyde **1**; **II**: Complex with Camalexin **2**; **III**: Complex with Quercetrin **3**; **IV**: Complex with Quercetin **4**; **V**: 1GNW, an *At*GSTF2 complex with 2 molecules of *S*-hexyl glutathione 'GSX', showing the GSH conjugation site [30].

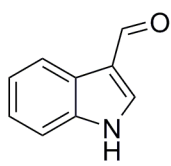
Figure 4. Electrostatic surface views of *At*GSTF2. **a**: Same view as Figure 2 **V**, in complex with 2 molecules of *S*-hexyl glutathione (PDB code 1GNW [30]); **b**: In complex with quercetrin **3**, rotated 90°, and revealing ligand binding site **L1**; **c**: In complex with quercetrin **3**, rotated 180°, and revealing ligand binding site **L2**.

Figure 5. **a**: Binding of indole-3-aldehyde **1** in the **L1** site. **b**: Binding of indole-3-aldehyde **1** in the **L2** site. Backbone and side chains of monomers **A** and **B** of a dimer of *At*GSTF2 are shown in ribbon and cylinder format in blue and gold respectively. Indole-3-aldehyde **1** is shown in ball-and-stick format with the carbon atoms in grey. Electron density map is shown in blue and corresponds to the $F_o - F_c$ omit map contoured at a level of 3σ , which was obtained from refinement prior to the building of the ligand(s). Ligand atoms from the ligand complex

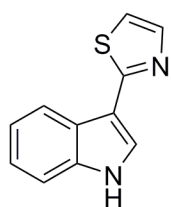
structures have been added afterwards for clarity. Selected distances, given in Ångstroms, between protein and ligand atoms are indicated as bold dashed lines.

Figure 6. Binding of camalexin **2** in the **L1** site. Backbone and side chains of monomers **A** and **B** of a dimer of *At*GSTF2 are shown in ribbon and cylinder format in blue and gold respectively. Camalexin **2** is shown in ball-and-stick format with the carbon atoms in grey. Electron density map is shown in blue and corresponds to the F_o-F_c omit map contoured at a level of 3σ , which was obtained from refinement prior to the building of the ligand(s). Ligand atoms from the ligand complex structures have been added afterwards for clarity. Selected distances, given in Ångstroms, between protein and ligand atoms are indicated as bold dashed lines.

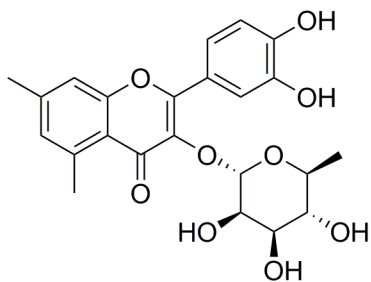
Figure 7. **a**: Binding of quercetrin **3** in the **L1** site. **b**: Binding of quercetrin **3** in the **L2** site. Backbone and side chains of monomers **A** and **B** of a dimer of *At*GSTF2 are shown in ribbon and cylinder format in blue and gold respectively. Quercetrin **3** is shown in ball-and-stick format with the carbon atoms in grey. Electron density map is shown in blue and corresponds to the F_o-F_c omit contoured at a level of 3σ , which was obtained from refinement prior to the building of the ligand(s). Ligand atoms from the ligand complex structures have been added afterwards for clarity. Selected interactions between protein and ligand are indicated as bold dashed lines with distances given in Ångstroms.



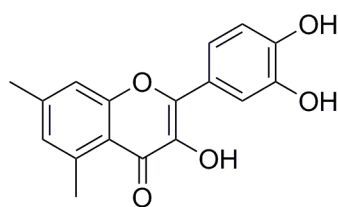
1



2



3



4

