

# The catalytic role of glutathione transferases in heterologous anthocyanin biosynthesis

Received: 14 September 2022

Accepted: 1 August 2023

Published online: 31 August 2023

Check for updates

Michael Eichenberger<sup>1</sup>, Thomas Schwander<sup>1</sup>, Sean Hüppi<sup>1,2</sup>, Jan Kreuzer<sup>1</sup>, Peer R. E. Mittl<sup>3</sup>, Francesca Peccati<sup>4</sup>, Gonzalo Jiménez-Osés<sup>4,5</sup>, Michael Naesby<sup>6</sup> & Rebecca M. Buller<sup>1</sup>✉

Anthocyanins are ubiquitous plant pigments used in a variety of technological applications. Yet, after over a century of research, the penultimate biosynthetic step to anthocyanidins attributed to the action of leucoanthocyanidin dioxygenase has never been efficiently reconstituted outside plants, preventing the construction of heterologous cell factories. Through biochemical and structural analysis, here we show that anthocyanin-related glutathione transferases, currently implicated only in anthocyanin transport, catalyse an essential dehydration of the leucoanthocyanidin dioxygenase product, flavan-3,3,4-triol, to generate cyanidin. Building on this knowledge, introduction of anthocyanin-related glutathione transferases into a heterologous biosynthetic pathway in baker's yeast results in >35-fold increased anthocyanin production. In addition to unravelling the long-elusive anthocyanin biosynthesis, our findings pave the way for the colourants' heterologous microbial production and could impact the breeding of industrial and ornamental plants.

Anthocyanins (**1**) are soluble pigments almost universally distributed amongst angiosperms (flowering plants), where they are responsible for the red, purple and blue colours of most flowers, fruits and leaves<sup>1,2</sup>. Due to their diverse colour expression and many proposed health effects<sup>3</sup>, their industrial adoption as natural colourants and bioactive compounds in the food, nutraceutical and cosmetic industries is progressively growing<sup>4,5</sup>. Currently, large-scale production of the pigments is achieved through extraction from plant raw materials with high contents of stable anthocyanins (**1**), for example, purple sweet potato, black carrot or red cabbage<sup>5</sup>, despite the associated issues of sustainability and supply. Thus, the construction of microbial cell factories for anthocyanin production is a field of great industrial interest; however, commercially viable product titres have not yet been reached, partly due to issues arising from the late steps of the biosynthetic pathway<sup>6,7</sup>.

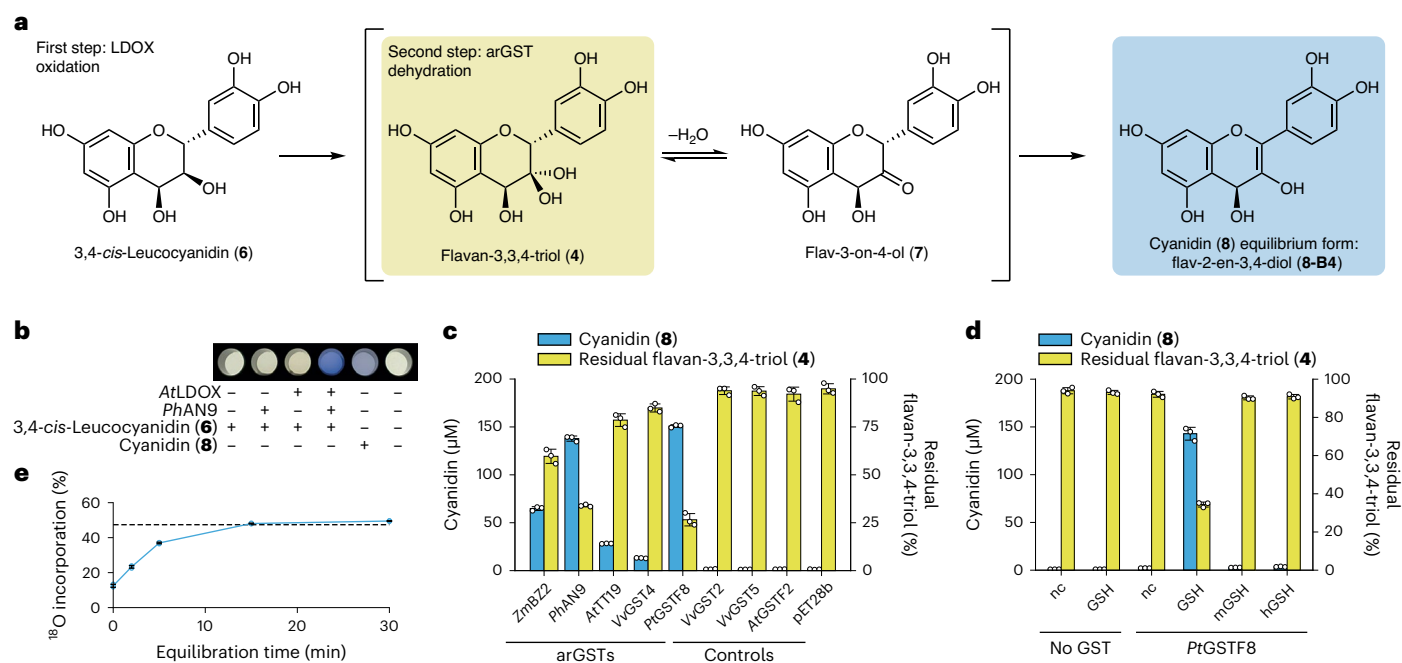
The penultimate step in the anthocyanin biosynthetic pathway (Extended Data Fig. 1) from 3,4-*cis*-leucoanthocyanidins (**2**) to anthocyanidins (**3**) is thought to be catalysed by the

Fe/α-ketoglutarate-dependent dioxygenase leucoanthocyanidin dioxygenase (LDOX), also called anthocyanidin synthase (ANS)<sup>8,9</sup>. Based on a protein crystal structure<sup>10</sup>, biochemical analysis<sup>11–14</sup> and feeding of isotopically labelled precursors in plants<sup>15</sup>, LDOX is proposed to catalyse an α-face C-3 hydroxylation of leucoanthocyanidins (**2**). Notably, however, the unstable product flavan-3,3,4-triol (**4**) is converted mainly into the by-products epidihydroquercetin (**9**), dihydroquercetin (**10**) and quercetin (**11**)<sup>13</sup>, while the transformation into the coloured anthocyanidin has only been observed as a minor side reaction *in vitro*<sup>11–13</sup> or *in vivo* by engineered microorganisms<sup>6,16</sup>.

Another debated aspect of the anthocyanin pathway is the molecular function of anthocyanin-related glutathione transferases (arGSTs)<sup>17,18</sup>. Loss-of-function mutations or downregulation of genes encoding arGSTs result in strongly decreased accumulation of anthocyanins (**1**) in a wide range of plants<sup>19–28</sup>, including reduced fruit colouration in cultivated *Fragaria vesca* (strawberry) through transient knock-down of reduced anthocyanins in petioles (*RAP*)<sup>20</sup> or white

<sup>1</sup>Competence Center for Biocatalysis, Zurich University of Applied Sciences, Wädenswil, Switzerland. <sup>2</sup>Department of Biotechnology, Delft University of Technology, Delft, Netherlands. <sup>3</sup>Department of Biochemistry, University of Zurich, Zurich, Switzerland. <sup>4</sup>Center for Cooperative Research in Biosciences, Basque Research and Technology Alliance, Derio, Spain. <sup>5</sup>Ikerbasque, Basque Foundation for Science, Bilbao, Spain. <sup>6</sup>Lantana Bio, Toulouse, France.

✉e-mail: [rebecca.buller@zhaw.ch](mailto:rebecca.buller@zhaw.ch)



**Fig. 1** arGSTs catalyse formation of cyanidin. **a**, Proposed enzymatic reactions from 3,4-*cis*-leucocyanidin (6) to flav-2-en-3,4-diol (8-B4), the 4-hydrated form within the reaction network of cyanidin in solution (Extended Data Fig. 4). **b**, Blue colour formation in coupled enzymatic reactions from 3,4-*cis*-leucocyanidin (6) using *AtLDOX* and *PhAN9*. **c**, Cyanidin (8) formation from ultrafiltered flavan-3,3,4-triol (4) with clarified *E. coli* lysates expressing GSTs of varying plant origins. **d**, Effect of GSH and selected cofactor analogues (mGSH and hGSH) at 20  $\mu$ M on

the conversion of ultrafiltered flavan-3,3,4-triol (4) to cyanidin (8) employing purified apo-*PtGSTF8*. nc, no cofactor; mGSH, *S*-methylglutathione; hGSH, *S*-hexylglutathione. **e**, Percentage of <sup>18</sup>O incorporation into cyanidin (8) as a function of *PtGSTF8* addition at varying time points after dilution of ultrafiltered flavan-3,3,4-triol (4) into H<sub>2</sub><sup>18</sup>O containing buffer. The dashed line marks the H<sub>2</sub><sup>18</sup>O content of the reaction buffer. Data in **c–e** are the mean values  $\pm$  s.d. of three independent replicates.

*Prunus persica* (peach) flowers caused by an arGST encoding regulator involved in anthocyanin transport (*Riant*) gene with a 2 bp insertion leading to a premature stop codon<sup>21</sup>. In some cases the phenotype is linked to mislocalization of the pigments to the cytoplasm<sup>29</sup> or vesicle-like structures<sup>19,30</sup>. While arGSTs were shown to catalyse the nucleophilic transfer of the tripeptide glutathione (GSH) onto the model substrate 1-chloro-2,4-dinitrobenzene (CDNB) (5) *in vitro*, such GSH conjugates with anthocyanins (1) were never observed *in vitro* or in planta<sup>19,31,32</sup>. In contrast, arGSTs bind to various flavonoids *in vitro*<sup>19,26,31</sup> and localize on membranes in tissues with high anthocyanin (1) production in planta<sup>19,30</sup>. This evidence has inspired the ligandin model for vacuolar sequestration of anthocyanins (1), in which arGSTs are believed to protect the labile anthocyanidins (3) from degradation and/or guide them from their biosynthetic origin on the cytoplasmic surface of the endoplasmic reticulum to transporters located on the tonoplast membrane<sup>18,19,31</sup>.

Yet, the results of several plant studies on the involvement of arGSTs in anthocyanin and oligomeric proanthocyanidin accumulation cannot be fully explained by the ligandin model. For example, when introduced via microparticle bombardment, cDNA encoding the arGST from *Zea mays* (*ZmBZ2*) can complement mutations of the corresponding genes encoding arGST (*PhAN9*) in *Petunia* spp., and vice versa. Notably, the pigmented aleurone (*Zea mays*) or corolla (*Petunia*) spots—indicative of functional complementation—have a dark purple centre and a halo of paler but distinctly pigmented cells<sup>28</sup> suggesting that the result of arGSTs *ZmBZ2* and *PhAN9* activity goes beyond intracellular transport. Additionally, epicatechin (19), a starter unit for proanthocyanidins<sup>33</sup>, is produced cytosolically from cyanidin (8-AH<sup>+</sup>) by the enzyme anthocyanidin reductase (ANR) (Extended Data Fig. 1). Notably, in *Arabidopsis thaliana* the mutation of arGST *tt19* completely eliminates production of epicatechin<sup>34</sup>—an unexpected outcome in the frame of the ligandin model as the cytosolic availability

of cyanidin (8-AH<sup>+</sup>), the precursor molecule of epicatechin (19), should be unaffected by mutating a transport protein.

In this Article, we show that arGSTs have a catalytic role in heterologous anthocyanin biosynthesis. Using *PtGSTF8* from poplar as a model enzyme, we use biochemical, structural and computational methods to demonstrate that arGSTs employ a GSH-dependent Brønsted base mechanism for conversion of the LDOX product flavan-3,3,4-triol (4) or flavan-3-on-4-ol (7) into flav-2-en-3,4-diol (8-B4), an equilibrium form of cyanidin. In *Saccharomyces cerevisiae* microbial cell factories constructed for anthocyanin biosynthesis, inclusion of selected arGSTs boosts the cyanidin-3-*O*-glucoside (13) titre by 36.5-fold. In addition, we show that heterologous production of pelargonidin-3-*O*-glucoside (14) and delphinidin-3-*O*-glucoside (15) can be similarly enhanced.

## Results

### Characterization of *AtLDOX* product

To elucidate the late steps of anthocyanidin biosynthesis, we set out to more closely investigate the LDOX oxidation product of 3,4-*cis*-leucocyanidin (6). This compound was only recently characterized in *in vitro* reactions with LDOX from *Vitis vinifera* (*VvLDOX*) and proposed as flavan-3,3,4-triol (4) based on HPLC–MS analysis and the observed degradation products<sup>13</sup>. In our hands, reactions performed with purified LDOX from *A. thaliana* (*AtLDOX*) led to an enzymatic product with an equivalent MS spectrum to the literature report (Extended Data Fig. 2d). As observed previously<sup>13</sup>, the putative flavan-3,3,4-triol (4) was not stable during purification, and confirmatory time-course experiments showed almost complete degradation of the intermediate (4) within 24 h (Extended Data Fig. 3a). Thus, to gather additional evidence about the identity of the LDOX oxidation product, we performed enzymatic reactions in buffer containing H<sub>2</sub><sup>18</sup>O, exploiting the expected equilibrium between the gem-diol flavan-3,3,4-triol (4) and the corresponding ketone flavan-3-on-4-ol (7) (Fig. 1a).

In aqueous media, incorporation of the isotopic label from bulk solvent into flavan-3,3,4-triol (**4**) was anticipated to lead to distinct  $m/z$  and  $^{18}\text{O}$  incorporation patterns of the observed adducts and flavonoid-typical C-ring cleavage fragments. The *At*LDOX oxidation product was analysed via ESI-MS in positive and negative modes, leading to spectra in accordance with a geminal diol at C3 (Extended Data Fig. 2), further confirming flavan-3,3,4-triol (**4**) as the oxidation product of LDOX from 3,4-*cis*-leucocyanidin (**6**). As expected, the reactions with *At*LDOX also resulted in formation of epidihydroquercetin (**9**), dihydroquercetin (**10**) and quercetin (**11**), the known by-products of the investigated reaction<sup>12,13</sup> (Extended Data Figs. 1a,c and 3b).

### arGSTs catalyse formation of cyanidin

arGSTs are essential for the accumulation of anthocyanins (**1**) in plants<sup>19–28</sup>. The observed inconsistencies between available in planta, in vitro and in vivo (cell factory) data, however, led us to hypothesize that arGSTs might not (only) have the literature-proposed transport function but could play a central role in the biosynthetic step from leucoanthocyanidins (**2**) to anthocyanidins (**3**). Thus, unlike previous in vitro biochemical studies or metabolic engineering efforts, which had excluded arGSTs, we set out to confirm our hypothesis by performing coupled enzymatic reactions with 3,4-*cis*-leucocyanidin (**6**) combining clarified *Escherichia coli* cell lysates expressing *At*LDOX and the well-studied arGST *PhAN9* from petunia. The reaction in the presence of both enzymes resulted in the formation of a partially soluble blue pigment with the same  $\lambda_{\text{max}}$  as authentic cyanidin (**8**) in reaction buffer (Fig. 1b and Extended Data Fig. 3e). Reconstitution experiments with different buffer components suggested that the blue pigment was an iron co-pigment of the quinoidal base anion form of cyanidin (**8-AM**) (Extended Data Fig. 4a,b). HPLC-MS analysis after solubilization with acidified methanol confirmed an almost stoichiometric conversion of 3,4-*cis*-leucocyanidin (**6**) into cyanidin (**8**), while the formation of the side products epidihydroquercetin (**9**), and quercetin (**11**) was reduced (Extended Data Fig. 3c,d).

In the early steps of the anthocyanin pathway, chalcone isomerase-like proteins modulate the product specificity of chalcone synthase through protein–protein interactions<sup>35</sup>. To determine whether arGSTs have such a modulating activity on LDOX, we ran two-step reactions, in which we separated *At*LDOX from the flavan-3,3,4-triol (**4**) product through ultrafiltration before addition of *PhAN9*-clarified cell lysate. Again, this experiment resulted in the conversion of flavan-3,3,4-triol (**4**) to cyanidin (**8**), suggesting a catalytic rather than a modulating role of the arGSTs in the further processing of the semi-stable LDOX product. Importantly, we established that ultrafiltered flavan-3,3,4-triol (**4**) was converted to cyanidin (**8**) by additional well-characterized arGSTs from maize (*ZmBZ2*), arabidopsis (*AtTT19*) and grapevine (*VvGST4*), and from the structure-elucidated phi-class orthologue from poplar (*PtGSTF8*), while application of GSTs unrelated to anthocyanin biosynthesis (*VvGST2*, *VvGST5*, *AtGSTF2*) did not result in cyanidin (**8**) formation (Fig. 1c). To exclude false negatives, all investigated GSTs were confirmed to be solubly and functionally produced by SDS-PAGE analysis and an activity test with the generic GST substrate CDNB (**5**) (Extended Data Fig. 3f,g).

### Biochemical characterization of *PtGSTF8*

To characterize the enzymatic function of arGSTs in the anthocyanin pathway in more detail, we decided to use the structure-elucidated *PtGSTF8* as a representative model enzyme in our biochemical assays. After expression of *PtGSTF8* in *E. coli*, we reduced the reported disulfide bond between the active site cysteine C13 and GSH in the clarified cell lysate using dithiothreitol (DTT) and removed the reduced GSH with an extended on-column washing step during purification. Using these enzyme preparations with a defined GSH concentration (20  $\mu\text{M}$ ) in the established two-step in vitro reactions, we found that the extent of conversion of the ultrafiltered flavan-3,3,4-triol (**4**) depended on

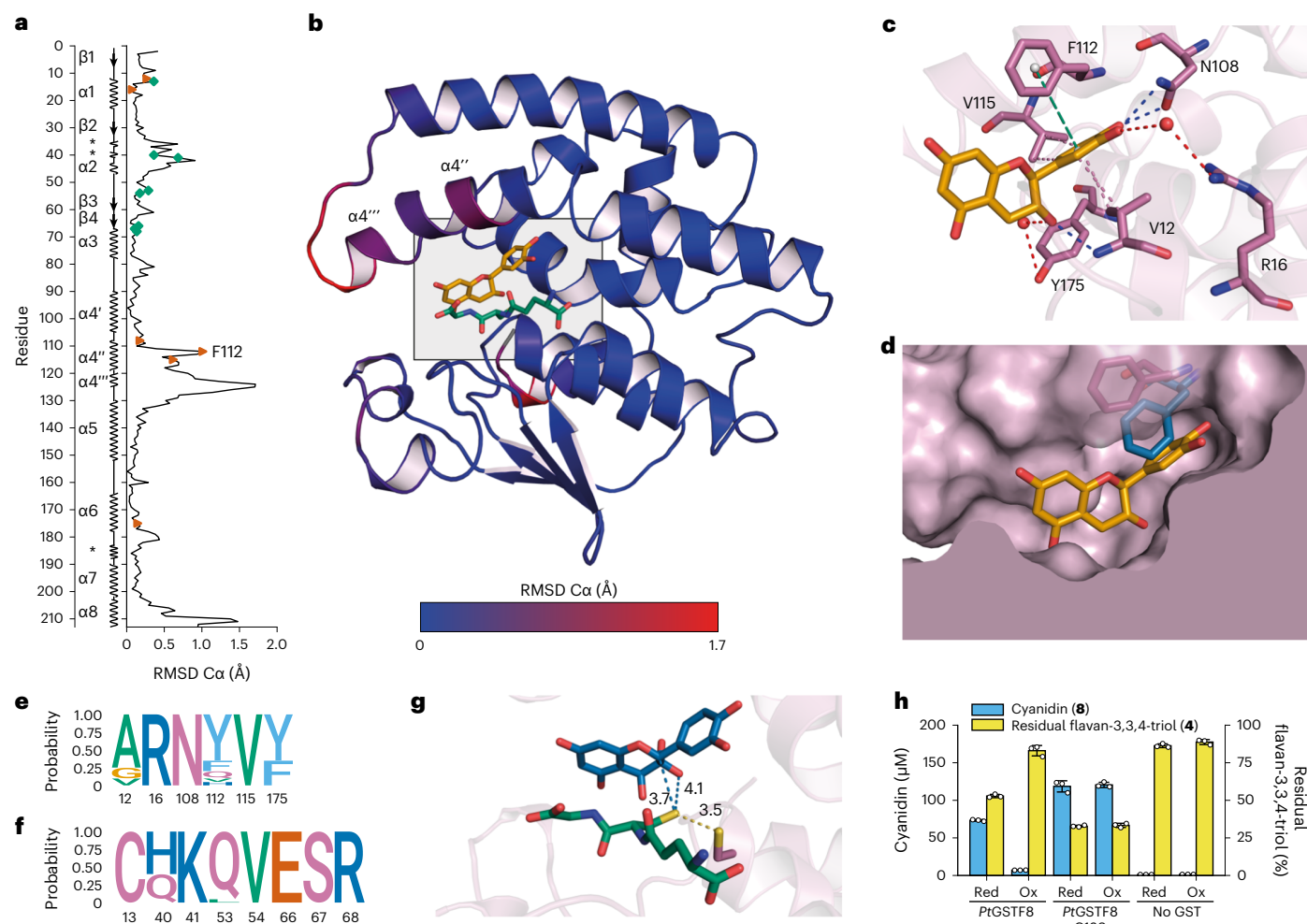
the applied concentration of the purified *PtGSTF8*, confirming that this enzyme is required and sufficient for the observed conversion to cyanidin (Extended Data Fig. 3h). Recording of Michaelis–Menten kinetics of *PtGSTF8* with flavan-3,3,4-triol (**4**) revealed an apparent catalytic rate constant ( $k_{\text{cat}}$ ) of 7.1  $\text{min}^{-1}$  and a Michaelis constant ( $K_{\text{m}}$ ) of  $147 \pm 8 \mu\text{M}$  (Extended Data Fig. 5) in line with catalytic parameters described for enzymes active in the secondary metabolism<sup>36</sup>. In addition, our experiments revealed that the enzymatic reaction strictly required the addition of catalytic amounts of GSH (Extended Data Fig. 3i). In contrast, addition of *S*-substituted methyl- and hexyl-glutathione did not result in the formation of cyanidin (**8**) (Fig. 1d), implying an involvement of the thiol group of GSH in catalysis.

To shed more light on the identity of the substrate of arGSTs, we performed labelling experiments with  $\text{H}_2^{18}\text{O}$ , exploiting the same principle of  $^{18}\text{O}$  incorporation as described earlier. In detail, we diluted ultrafiltered flavan-3,3,4-triol (**4**) into  $\text{H}_2^{18}\text{O}$ -labelled reaction buffer to initiate  $^{18}\text{O}$  incorporation at C-3 of the flavonoid structure. By adding purified *PtGSTF8* at varying time points, we expected to halt the incorporation of the isotopic label into the final product cyanidin (**8**). As hypothesized, the  $^{18}\text{O}$  content in cyanidin (**8**) increased as a function of equilibration time and the expected level (47.4%) for stochastic incorporation of a single  $^{18}\text{O}$  label was reached after 15 min of pretreatment (Fig. 1e). These results confirmed that the enzymatic reaction catalysed by *PtGSTF8* terminated the hydration equilibrium between flavan-3,3,4-triol (**4**) and flavan-3-on-4-ol (**7**), leading us to propose a GST-dependent dehydration of flavan-3,3,4-triol (**4**) or a tautomerization of flavan-3-on-4-ol (**7**) to form flav-2-en-3,4-diol (**8-B4**), an equilibrium form of anthocyanins (**1**) in solution (Fig. 1a and Extended Data Figs. 4a and 6a)<sup>37,38</sup>.

### (–)-Catechin bound structure of *PtGSTF8*

To better understand the catalytic machinery of arGSTs, we determined the crystal structure of *PtGSTF8* as a ternary complex with GSH and the substrate analogue (–)-catechin (**12**) (Extended Data Fig. 1d) to a resolution of 1.09 Å (Fig. 2b, Extended Data Fig. 7a and Extended Data Table 1). The structure comprises the typical GST fold with an N-terminal thioredoxin-like domain ( $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ ) and an all-helical C-terminal domain ( $\alpha_4\alpha_5\alpha_6\alpha_7\alpha_8$ ). The active-site cavity, which is located between the two domains, comprises a glutathione binding site (G-site), formed mainly by the N-terminal domain, and a hydrophobic substrate-binding site (H-site), which is formed by the C-terminal domain. We found the structure to be very similar to the published binary complex of *PtGSTF8* with GSH (PDB 5F07)<sup>39</sup> at an overall root-mean-square deviation (RMSD) ( $\text{C}\alpha$ ) of 0.16 Å (Fig. 2a,b). However, presumably to accommodate (–)-catechin (**12**), the active site is substantially enlarged by a shift of the  $\alpha_4''$  and  $\alpha_4'''$  helices by up to 1.71 Å (Fig. 2a,b) and an upwards movement of the side chain of phenylalanine F112 (Fig. 2d and Extended Data Fig. 7c). Within the H-site, the substrate analogue (**12**) interacts through hydrogen bonds to N108 and the backbone of V12, a  $\pi$ – $\pi$  interaction with F112, water bridges to R16 and Y175, and hydrophobic interactions with V12 and V115 (Fig. 2c). Interestingly, the amino acid composition of the substrate-binding H-site and the G-site, which binds to GSH through hydrogen bonds and salt bridges (Extended Data Fig. 7b)<sup>39</sup>, is strikingly uniform amongst arGSTs (Fig. 2e,f), making the existence of similar substrate–enzyme interactions in other arGSTs plausible.

To shed light on the interactions of the enzyme with the natural substrates, we docked flavan-3,3,4-triol (**4**) and flavan-3-on-4-ol (**7**) into the active pocket of the crystal structure of *PtGSTF8* obtained with (–)-catechin (**12**) (Extended Data Fig. 7d,e). The best binding poses for both substrates had a highly similar conformation to (–)-catechin (**12**), with an RMSD of matched atoms of 0.51 Å for flavan-3,3,4-triol (**4**) and 0.54 Å for flavan-3-on-4-ol (**7**). The docking analyses revealed that the positioning of the GSH-thiolate in relation to C-2 and O-3 of the docked substrates could enable an acid–base mechanism of dehydration or



**Fig. 2 | Structural basis of dehydratase activity of PtGSTF8.** **a**, RMSD (C $\alpha$ ) between PtGSTF8 as a ternary complex with bound (-)-catechin (**12**) and GSH, and as a binary complex without **12** (PDB 5F07). Green diamonds, G-site; red triangles, H-site; asterisk,  $3_{10}$  helix. **b**, Overall structure of PtGSTF8 as ternary complex with GSH (green) and (-)-catechin (**12**) (orange). The ribbon is coloured according to RMSD (C $\alpha$ ) to the structure of PtGSTF8 as binary complex. The active pocket is highlighted with a grey square. **c**, H-site interactions to (-)-catechin (**12**) (orange). Dashes are coloured according to the interaction type: blue, hydrogen bonds; red, water bridges; green,  $\pi$ - $\pi$  interactions; magenta, hydrophobic interactions. **d**, Surface representation of H-site of ternary complex

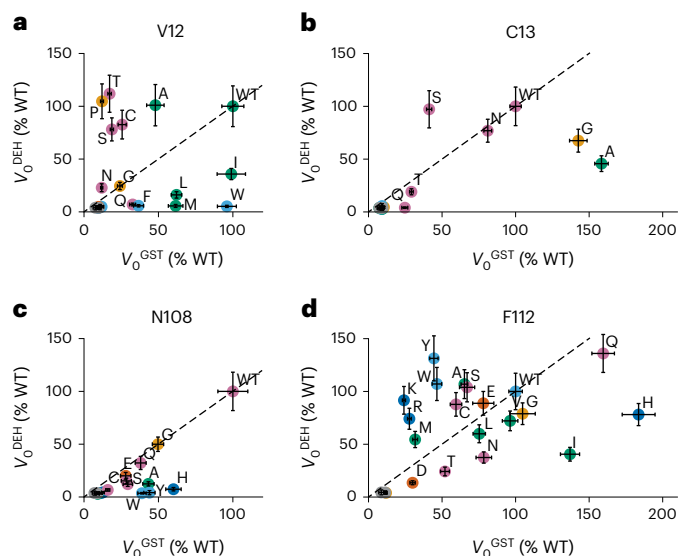
with GSH and (-)-catechin (**12**) (orange). F112 is shown in the conformations from the ternary complex with **12** (magenta) and the binary complex (blue) without **12**. The surface of F112 is shown transparently. **e, f**, Sequence logo highlighting the probability of amino acid occurrence within the H-site (**e**) and G-site (**f**) of 24 published arGSTFs. **g**, Docking of flavan-3,3,4-triol (**4**) (blue) into the active site of PtGSTF8 and distances (Å) from the GSH (green) thiolate to C-2 and O-3 of **4** and distance of the C-13 thiolate (purple) to GSH (green). **h**, Dehydratase activity of PtGSTF8 wild type and C13S variants towards ultrafiltered flavan-3,3,4-triol (**4**) before and after oxidation with GSSG. Data in **h** are mean values  $\pm$  s.d. of three independent replicates. Red, reduction; Ox, oxidation.

tautomerization, in which the GSH cofactor acts as a proton-relay system (Fig. 2g and Extended Data Figs. 6a and 7f).

### Structural basis of arGST activity

Guided by the structural information derived from the ternary complex of GSH, (-)-catechin (**12**) and PtGSTF8, we selected a set of amino acids in PtGSTF8 for mutagenesis studies as a probe for the positions' importance in catalysis. Amino acids V12 (putative oxyanion hole), F112 (active site architecture), N108 (flavonoid binding) and C13 (GSH stabilization) were fully randomized by site saturation mutagenesis (NNK codons), and the created PtGSTF8 variants were tested for conversion of flavan-3,3,4-triol (**4**) to cyanidin (**8**), and for their GSH-transferase activity towards CDNB (**5**) as a control of catalytic competence. For all four tested residues, we found that depending on the substitution introduced, the two enzymatic activities were impacted to different extents (Fig. 3). For positions V12 and F112, many of the tested variants exhibited wild-type-level dehydratase activity (Fig. 3a,d). Unsurprisingly, we

found that the putative oxyanion hole, formed by the NH backbone of V12, could be provided by other amino acid residues. In addition, the observed side-chain flexibility of V12 and F112 indicated that these residues are mainly involved in the overall shaping of the active site and help to position the flavonoid substrate in accordance with the presented structural evidence. Interestingly, the observation that other amino acids than V12 and F112 can principally support flavonoid binding is also reflected in the more relaxed amino acid conservation in native arGSTFs at these two positions (Fig. 2e). Probing position N108, we found that incorporation of any other amino acid resulted in a substantially reduced formation of cyanidin (**8**) suggesting that the hydrogen-bond interaction of the asparagine to the flavonoid substrate is critical for dehydratase activity, while GSH-transferase activity was comparatively less affected for N108 variants comprising an alanine, histidine, tyrosine or tryptophan (Fig. 3c). C13 is optimally placed to form a hydrogen bond with the deprotonated cofactor GSH (Fig. 2g and Extended Data Fig. 7f). In line with this hypothesis, the activity-conferring substitution pattern



**Fig. 3 | Mutational study of *PtGSTF8*.** **a–d**, Initial reaction velocities of the dehydratase activity ( $V_0^{\text{DEH}}$ ) towards ultrafiltered flavan-3,3,4-triol (**4**) and GSH-transferase activity ( $V_0^{\text{GST}}$ ) towards CDNB (**5**) of *PtGSTF8* variants at V12 (**a**), C13 (**b**), N108 (**c**) and F112 (**d**) measured in colorimetric 96-well-plate assays. Data points are coloured according to amino acid properties: green, hydrophobic aliphatic; light blue, hydrophobic aromatic; purple, polar; dark blue, basic; red, acidic; yellow, special. The dashed bisector line marks where an equivalent effect on both tested enzymatic activities would be expected. Data are mean values  $\pm$  s.d. of three independent replicates. Individual data points can be found in Supplementary Table 7. WT, wild type.

of C13, which is highly conserved in arGSTs (Fig. 2f), was found to be much less permissive than for V12 and F112 (Fig. 3b). While a substitution with the polar serine at position 13 retained wild-type dehydratase activity, presumably due to the alcohol's ability to assume cysteine's function in stabilizing deprotonated GS<sup>−</sup>, the only other *PtGSTF8* variants to keep an—albeit reduced—flavan-3,3,4-triol (**4**) dehydration activity contained a glycine, alanine or asparagine mutation (Fig. 3b). These variants equally preserved their GSH-transferase activity, suggesting the importance of additional factors in the modulation of the negative log of GSH's acid dissociation constant ( $\text{p}K_{\text{a}}$ ), including the ability to optimally arrange the well-described electron-sharing network which generates a positively charged electrostatic field in the G-site of GSTs<sup>40,41</sup>.

To assess the general relevance of the investigated positions in the context of alternative arGSTs, we introduced three selected substitutions (A/V12M, C13S and N108H) into the two well-studied arGSTs *PhAN9* and *VuGST4*. In this experiment we observed similar trends on dehydratase and GST-transferase activity (Extended Data Fig. 7g–i) as observed for *PtGSTF8*, suggesting that the proposed catalytic mechanism and active site architecture are conserved between arGSTs from different plant origins.

To further probe the catalytic mechanism of arGSTs, we conducted quantum mechanics/molecular mechanics (QM/MM) calculations based on the ternary complex of GSH, (−)-catechin (**12**) and *PtGSTF8*. Extensive experimental and computational studies have shown that the binding of GSH to GSTs decreases the  $\text{p}K_{\text{a}}$  of the GSH thiol group from around 9 to about 6 in the active site through hydrogen bonding and salt bridges and through the positive electrostatic potential of the G-site<sup>41,42</sup>. Consequently, the GSH deprotonation step was considered non-rate-determining and thermodynamically favoured, leading us to use the GSH-thiolate as the starting point to compute the reaction mechanism for the conversion of flav-3-on-4-ol (**7**) into cyanidin (**8–B4**) in the wild-type enzyme and the C13S variant (Extended Data Fig. 8).

The analysis of the optimized structures of the Michaelis–Menten complex indicate that the polar residues at position 13 (Cys or Ser) stabilize the deprotonated GS<sup>−</sup>, which in turn acts as the catalytic Brønsted base and abstracts the proton at position C2 of substrate **7** (Extended Data Fig. 8 and Supplementary Figs. 1 and Fig. 2). Notably, the hydroxyl group at the C4 position of the tetrahydropyran ring of substrate **7** also stabilizes the GSH-thiolate through an additional hydrogen bond (Extended Data Fig. 8). As hypothesized from the available structural data, the NH backbone of V12 serves as an oxyanion hole for the enolate formed upon deprotonation.

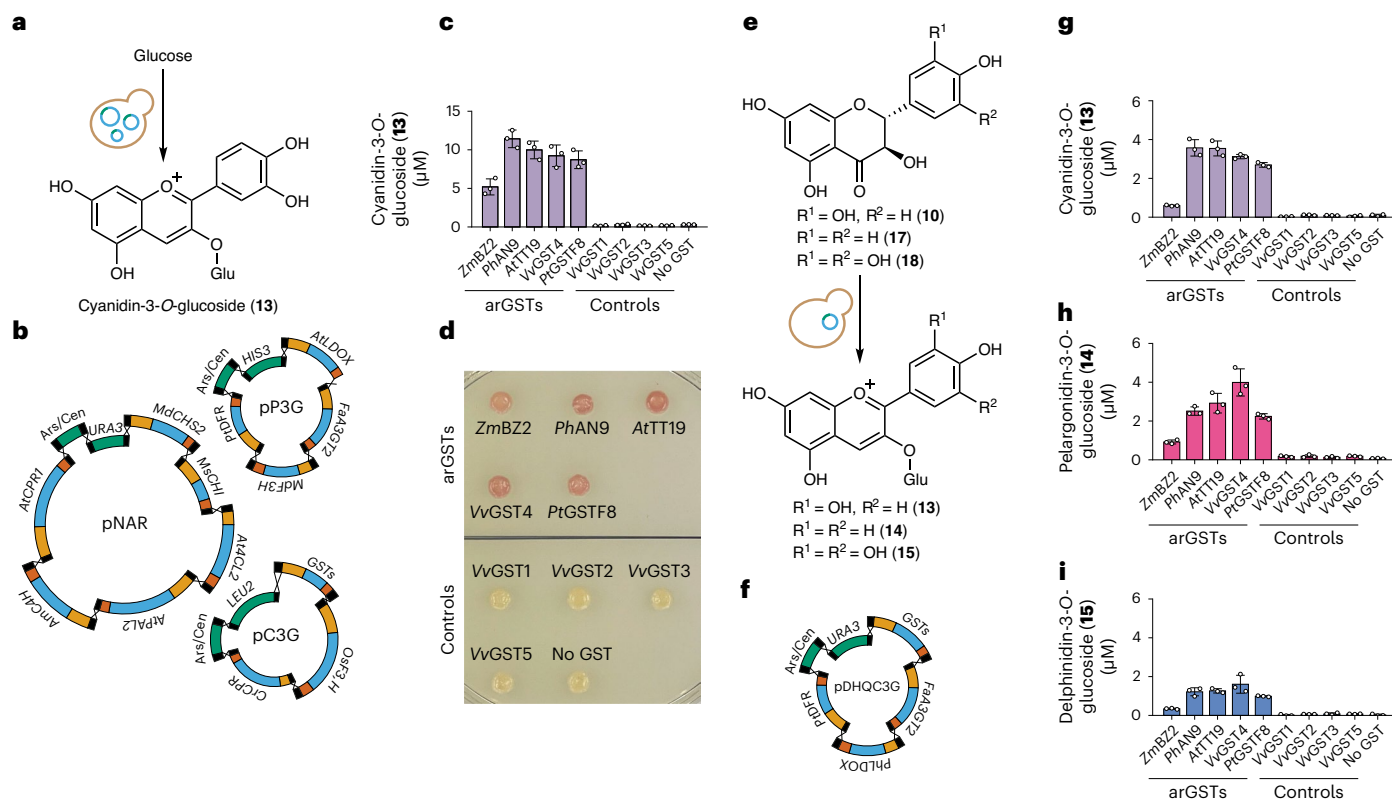
Informed by the structural data of *PtGSTF8*, the accompanying mutagenesis experiments and the QM/MM studies, we investigated the impact of a disulfide bond between C13 and GSH on dehydratase activity. To this end, we attempted to form the mixed disulfide bond with purified apo-*PtGSTF8* and its C13S variant using oxidized glutathione (GSSG). The expected increase of  $m/z$  was only observed in the matrix-assisted laser desorption ionization (MALDI) spectra of the wild-type enzyme, in line with an S-glutathionylation at C13 (Extended Data Fig. 9). When assayed with ultrafiltered flavan-3,3,4-triol (**4**), oxidation of C13 in the wild-type enzyme resulted in a drastically lower formation of cyanidin (**8**), while we observed no effect for the C13S variant (Fig. 2h), giving further evidence of the importance of free GSH for catalysis.

### Anthocyanin production in yeast

To highlight the practical relevance of our pathway elucidation, we set out to show that the catalytic activity of arGSTs is also required for anthocyanin biosynthesis *in vivo* in a eukaryotic cell and thus is key for production of anthocyanins (**1**) in microbial cell factories. To this end, we engineered *S. cerevisiae* to produce cyanidin-3-*O*-glucoside (**13**), the first stable compound in the anthocyanin pathway, using a three-plasmid strategy (Fig. 4a,b). For the known anthocyanin pathway, we selected genes encoding enzymes previously found to be efficient in *S. cerevisiae*<sup>6</sup> and combined them with a set of nine representative plant-derived GSTs. In this *in vivo* system, all tested arGSTs boosted the titre of the end-product cyanidin-3-*O*-glucoside (**13**) by a factor of up to 36.5, while the control GSTs unrelated to anthocyanin biosynthesis were comparable to the control without GST (Fig. 4c,d). To further explore the substrate scope of arGSTs, additional yeast systems were engineered using a one-plasmid strategy containing genes encoding dihydroflavonol-4-reductase (DFR), LDOX, a range of arGST and non-arGSTs (as negative controls) and anthocyanidin-3-*O*-glucosyl transferase (A3GT) (Fig. 4e,f). Feeding of selected dihydroflavonols (that is, dihydroquercetin (**10**), dihydrokaempferol (**17**) and dihydromyricetin (**18**)) led to the formation of the corresponding coloured and stable anthocyanidin-3-*O*-glucosides, namely cyanidin-3-*O*-glucoside (**13**), pelargonidin-3-*O*-glucoside (**14**) and delphinidin-3-*O*-glucoside (**15**), underscoring the arGSTs' biosynthetic relevance in the formation of additional natural pigments. As observed in the case of the whole biosynthetic pathway, yeast strains engineered to incorporate GSTs unrelated to anthocyanin biosynthesis did not produce any of the anthocyanidin-3-*O*-glucosides (Fig. 4g–i).

### Discussion

This work demonstrates that arGSTs have a catalytic role in the biosynthetic pathway to anthocyanins (**1**). Based on the here-presented biochemical, structural and computational evidence, we propose that arGSTs utilize a GSH-dependent Brønsted base mechanism for conversion of the semi-stable LDOX product flavan-3,3,4-triol (**4**) or flavan-3-on-4-ol (**7**) into flav-2-en-3,4-diol (**8–B4**), the 4-hydration species of cyanidin (**8**) (Extended Data Fig. 6a). Similar mechanisms in which the GSH-thiolate acts as a Brønsted base catalyst have been suggested for the well-studied isomerization of ketosteroids by GST A3-3<sup>43</sup> (Extended Data Fig. 6b), and for the tautomerization of (R)-2-hydroxymenthofuran by GST A1-1<sup>44</sup> (Extended Data Fig. 6c).



**Fig. 4 | Anthocyanin production in engineered *S. cerevisiae*.** **a**, Biosynthesis of cyanidin-3-*O*-glucoside (**13**) from glucose in yeast. **b**, Stepwise assembly of cyanidin-3-*O*-glucoside (**13**) pathway (Supplementary Table 1) on three plasmids using transcription-associated recombination. Blue, open reading frames; orange, promoters; red, terminators; black, linkers; green, plasmid backbone elements. **c**, Cyanidin-3-*O*-glucoside (**13**) product titres in liquid cultures of engineered *S. cerevisiae* strains grown on glucose and expressing the anthocyanin biosynthetic pathway with representative anthocyanin-related and control GSTs. **d**, Colour formation on SC-ULH agar plates of engineered *S. cerevisiae* strains expressing the anthocyanin biosynthetic pathway with representative anthocyanin-related and control GSTs. **e**, Biosynthesis of cyanidin-3-*O*-glucoside (**13**), pelargonidin-3-*O*-glucoside (**14**) and delphinidin-3-*O*-glucoside (**15**) from the respective dihydroflavonols (dihydroquercetin (**10**), dihydrokaempferol (**17**) and dihydromyricetin (**18**)). **f**, Assembly of the partial anthocyanin (**1**) pathway on a single plasmid using transcription-associated

recombination. Blue, open reading frames; orange, promoters; red, terminators; black, linkers; green, plasmid backbone elements. **g–i**, Cyanidin-3-*O*-glucoside (**13**) (**g**), pelargonidin-3-*O*-glucoside (**14**) (**h**) and delphinidin-3-*O*-glucoside (**15**) (**i**) product titres of engineered *S. cerevisiae* strains expressing the partial anthocyanin biosynthetic pathway with representative anthocyanin-related and control GSTs. The cultures were grown in media containing 300  $\mu$ M of the respective dihydroflavonol. Data in **c**, **g**, **h** and **i** are mean values  $\pm$  s.d. of three independent replicates. *A3GT* encodes a cyanidin-3-*O*-glucosyl transferase; *CHI* encodes a chalcone isomerase; *CHS* encodes a chalcone synthase; *CHH* encodes a cinnamate-4-hydroxylase; *CPR* encodes a cytochrome P450 reductase; *DFR* encodes a dihydroflavonol-4-reductase; *F3H* encodes a flavanone-3-hydroxylase; *F3'H* encodes a flavonoid-3'-hydroxylase; *GST* encodes a glutathione transferase; *LDOX* encodes a leucoanthocyanidin dioxygenase; *PAL* encodes a phenylalanine ammonia lyase; *4CL* encodes a 4-coumarate-CoA ligase.

The main phenotype of loss-of-function mutations in arGSTs is the decreased content of anthocyanins (**1**)<sup>19–28</sup> and proanthocyanidins<sup>27</sup>, which can be well explained by an incomplete biosynthesis due to the absence of an essential enzyme. Furthermore, hitherto puzzling observations, such as the halo formation in complementation experiments carried out with arGST-deficient maize and Arabidopsis plants<sup>28</sup>, and the requirement for arGSTs in proanthocyanidin formation in planta<sup>27</sup>, can be rationalized with the catalytic role of arGSTs in producing anthocyanidin precursors. In Arabidopsis, arGST loss-of-function mutations were additionally linked with an increased content of flavonols<sup>19,45</sup>, most probably due to a second oxidation of accumulating flavan-3,3,4-triol (**4**), which is catalysed by *AtLDOX* in vitro<sup>11–13</sup>. Notably, anthocyanin biosynthesis is believed to occur on the surface of the endoplasmic reticulum in membrane-associated enzyme complexes, so-called metabolons<sup>46</sup>. The here-proposed catalytic involvement of arGSTs in the biosynthesis would profit from such association with metabolons and is indirectly observed through the enzymes' localization to membranes in planta<sup>30</sup>. Further evidence for this hypothesis comes from recent yeast two-hybrid

studies performed with enzymes from sweet peony, where the arGST *PsGSTF3* was found to interact with *PsDFR*, the dihydroflavonol reductase catalysing the formation of 3,4-*cis*-leucocyanidin (**6**)<sup>47</sup>. Considering our results, the currently accepted role of arGSTs in transport of anthocyanins (**1**) in planta will have to be reconsidered as less plausible.

The elucidation of arGSTs' role as biosynthetic enzymes in the anthocyanin pathway has a direct implication for production of these important pigments in microbial cell factories, as highlighted by a 36.5-fold improvement of the cyanidin-3-*O*-glucoside (**13**) titre through the coexpression of arGSTs in our prototype *S. cerevisiae* production strains. Accordingly, we show that the inclusion of arGSTs in our production strains substantially boosts formation of pelargonidin-3-*O*-glucoside (**14**) and delphinidin-3-*O*-glucoside (**15**). In conclusion, the present work provides crucial biosynthetic insights for a more sustainable production of designer anthocyanins (**1**) and is anticipated to drive the construction of powerful, industrial anthocyanin-producing strains, further fuelling the ongoing biologization of the food and nutraceutical sectors.

## Methods

### Materials

Chemicals were ordered from Sigma, ABCR, Carl Roth, Lucerna Chem, Plantmetachem or Formedium, and were used without further purification. Oligonucleotides were ordered from Microsynth. Q5 High-Fidelity polymerase, T4 DNA ligase and restriction enzymes were ordered from New England BioLabs. Clonal genes and gene fragments were ordered from Twist Bioscience. An Excel table containing additional information (company name, catalogue number) of all employed commercial reagents is available as Supplementary Data.

### Synthesis of 3,4-*cis*-leucocyanidin (6)

Dihydroquercetin (**10**) at 5 mg ml<sup>-1</sup> was reduced with NaBH<sub>4</sub> at 5 mg ml<sup>-1</sup> in ethanol to form 3,4-*trans*-leucocyanidin (**14**) under vigorous stirring for 90 min at 23 °C. Next, 10 vol of 1% acetic acid was added, and the solution was stirred vigorously for 3 h at 40 °C to isomerize 3,4-*trans*-leucocyanidin (**14**) to 3,4-*cis*-leucocyanidin (**6**). After adding 3-morpholino-2-hydroxypropanesulfonic acid (MOPSO) to a final concentration of 10 mM and setting the pH to 6.2 using 5 N NaOH, the product was extracted four times with 1 vol of ethyl acetate. The extract was rotary evaporated at 30 °C to a volume of 2 ml, acidified to pH 4 with formic acid, and 200 µl aliquots were purified using preparative HPLC (Agilent Technologies, I260 Infinity) over a Luna 5 µm phenyl-hexyl 100 Å (250 × 10 mm) column and a 15 min linear gradient from 0% acetonitrile in H<sub>2</sub>O to 16% acetonitrile in H<sub>2</sub>O at 4 ml min<sup>-1</sup> flow. After pooling the fractions containing pure 3,4-*cis*-leucocyanidin (**6**), MOPSO was added to a final concentration of 10 mM, the pH was set to 6.2 using 5 N NaOH and the pooled fractions were rotary-evaporated to a concentration of 7 mM of 3,4-*cis*-leucocyanidin (**6**), quantified by HPLC-UV (280 nm) with an external calibration curve with 3,4-*trans*-leucocyanidin (**14**). Finally, 3,4-*cis*-leucocyanidin (**6**) was aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C until further use.

### Plasmids, subcloning and creation of NNK libraries

Genes, codon optimized for expression in *S. cerevisiae*, were obtained from Twist Bioscience as DNA strands or subcloned into pET-28b(+) or pTwist Amp MC (Supplementary Tables 2 and 3). Coding sequences of all genes used can be found in Supplementary Table 6.

Basic parts of yeast expression cassettes, which were flanked by linkers for assembly of multi-expression plasmid by homologous recombination (adapted from a method described by Zhao et al.<sup>48</sup>), were obtained from Twist Bioscience in a pTwist Amp backbone (Supplementary Table 2). Additional expression cassettes were constructed by restriction-enzyme-based subcloning according to the manufacturer's instructions, performed in *E. coli* NEB Turbo (New England BioLabs) as outlined in Supplementary Table 2.

Point mutations and NNK site saturation libraries of *PtGSTF8* were constructed by overlap extension PCR as previously reported<sup>49</sup> using primers T7 fw, T7 term and:

*PtGSTF8*(V12X): V12C13\_rv and V12X\_fw, *PtGSTF8*(C13X): V12C13\_rv and C13X\_fw, *PtGSTF8*(C13S): V12C13\_rv and C13S\_fw, *PtGSTF8*(N108X): N108\_rv, N108X\_fw, *PtGSTF8*(F112X): F112\_rv, F112X\_fw (Supplementary Table 4) using pANT35 as template.

*PhAN9*(A12M): *PhAN9*\_A12M\_fw, *PhAN9*\_A12M\_rv, *PhAN9*(C13S): *PhAN9*\_C13S\_fw, *PhAN9*\_C13S\_rv, *PhAN9*(N108H): *PhAN9*\_N108H\_fw, *PhAN9*\_N108H\_rv (Supplementary Table 4) using pANT15 as template.

*VuGST4*(A12M): *VuGST4*\_A12M\_fw, *VuGST4*\_A12M\_rv, *VuGST4*(C13S): *VuGST4*\_C13S\_fw, *VuGST4*\_C13S\_rv, *VuGST4*(N108H): *VuGST4*\_N108H\_fw, *VuGST4*\_N108H\_rv (Supplementary Table 4) using pANT20 as template.

### Production and purification of AtLDOX with C-terminal Strep-tag

For expression of AtLDOX with a C-terminal Strep-tag, *E. coli* BL21(DE3) containing plasmid pANT43 was inoculated from an overnight

preculture at a 1:100 ratio in four 2 l baffled Erlenmeyer flasks filled with 400 ml TB medium containing 50 µg ml<sup>-1</sup> kanamycin sulfate and incubated at 37 °C, 140 rpm using a 5 cm shaking diameter for 2.75 h. Enzyme expression was induced by the addition of 100 µM isopropyl β-D-1-thiogalactopyranoside, the cultures were incubated for another 3.5 h at 30 °C, 140 rpm using a 5 cm shaking diameter, and the final OD<sub>600</sub> (optical density at 600 nm) was determined using a CO8000 cell density meter (WPA biowave). Cells were harvested by centrifugation at 4,400g at 4 °C for 15 min and the medium was discarded. Cell pellets were resuspended in LDOX lysis buffer (20 mM potassium phosphate, 200 mM NaCl, pH 7.4, 0.01 mg ml<sup>-1</sup> DNaseI) to an OD<sub>600</sub> of 140 and lysed by two rounds of sonication for 2 min at 50% amplitude with 2 s intervals using a Sonopuls (Bandelin) sonicator. Cell debris was removed by centrifugation at 4,400g at 4 °C for 30 min. The clarified cell lysate was either directly submitted to further protein purification steps or flash-frozen in liquid nitrogen after addition of 10% glycerol and stored at -80 °C for use in biocatalytic reactions.

For purification, a 5 ml StrepTrap HP (GE Healthcare) was equilibrated with at least five column volumes of lysis buffer. After filtering the clarified cell lysate through a 0.45 µm filter, it was loaded onto the column at a flow rate of 3 ml min<sup>-1</sup>. The column was washed with 15 column volumes of PPS buffer (20 mM potassium phosphate, 200 mM NaCl, pH 7.4) at a flow rate of 5 ml min<sup>-1</sup>. Next, the protein was eluted using LDOX elution buffer (20 mM potassium phosphate, 200 mM NaCl, 2.5 mM *d*-desthiobiotin, pH 7.4) and the fractions were combined according to the recorded absorption 280 nm (A<sub>280nm</sub>). Finally, the buffer was exchanged to PPS buffer using three serially connected 5 ml HiTrap Desalting Columns (GE Healthcare). The concentration of purified AtLDOX was measured with a NanoDrop spectrophotometer (Thermo Fisher Scientific) at A<sub>280nm</sub> using an extinction coefficient estimated with ProtParam<sup>50</sup>. After supplementation with 10% glycerol, aliquots of purified protein were flash-frozen in liquid nitrogen and stored at -80 °C until further use.

### Small-scale production of GSTs for reactions with clarified cell lysates

For small-scale expression of GST variants, *E. coli* containing pET-28b(+) based plasmids (pANT13-16, pANT18, pANT20-21, pANT35, pANT95, pJK1-3, pJK7-9 or site saturation variants of pANT35) were inoculated from an overnight preculture at a ratio of 1:100 in 96-deep-well plates equipped with a CR1296 sandwich cover (EnzyScreen) in 0.5 ml TB medium containing 50 µg ml<sup>-1</sup> kanamycin sulfate per well and incubated at 37 °C, 300 rpm using a 5 cm shaking diameter for 2.5 h. Enzyme expression was induced by adding 100 µM isopropyl β-D-1-thiogalactopyranoside, and the cultures were incubated for another 20.5 h at 20 °C, 300 rpm, 5 cm shaking diameter. Cells were harvested by centrifugation at 4,400g at 4 °C for 15 min, and the medium was discarded. Cell pellets were resuspended in 0.3 ml small-scale lysis buffer (20 mM potassium phosphate, 200 mM NaCl, pH 7.4, 1 mg ml<sup>-1</sup> lysozyme, 0.75 mg ml<sup>-1</sup> polymyxin B, 0.01 mg ml<sup>-1</sup> DNase I) per well and incubated at 30 °C, 300 rpm, 5 cm shaking diameter for 30 min. Cell debris was removed by centrifugation at 4,400g at 4 °C for 30 min, and the clarified cell lysate was directly used for biocatalytic reactions. GST expression levels were analysed via SDS-PAGE. Gel pictures were captured using UV Lab (4.1.0) software.

### Large-scale production and purification of GSTs

For large-scale expression of *PtGSTF8* and *PtGSTF8*(C13S), *E. coli* containing plasmid pANT35 or pANT62 was inoculated from an overnight culture at a ratio of 1:100 in a 2 l baffled Erlenmeyer flask filled with 400 ml Zym-5052 medium<sup>51</sup> containing 50 µg ml<sup>-1</sup> kanamycin sulfate per flask, incubated at 20 °C, 140 rpm, 5 cm shaking diameter for 24 h, and the final OD<sub>600</sub> was determined using a CO8000 cell density meter (WPA biowave). Cells were harvested by centrifugation at 4,400g at 4 °C for 15 min and the medium was discarded. Cell pellets were

resuspended in reducing lysis buffer (20 mM potassium phosphate, 200 mM NaCl, pH 7.4, 10 mM DTT, 0.01 mg ml<sup>-1</sup> DNase I) to an OD<sub>600</sub> of 140 and lysed by two rounds of sonication for 2 min at 50% amplitude with 2 s intervals using a Sonopuls (Bandelin) sonicator. Cell debris was removed by centrifugation at 4400g at 4 °C for 30 min and the clarified cell lysate was used for protein purification.

For purification, a 5 ml GSTrap FF (GE Healthcare) was equilibrated with at least five column volumes of reducing lysis buffer. After filtering the clarified cell lysate through a 0.45 µm filter, the crude protein solution was loaded onto the column at a flow rate of 3 ml min<sup>-1</sup>. The disulfide bond between *Pt*GSTF8 residue C13 and GSH was reduced on-column with 20 column volumes of rPPS buffer (20 mM potassium phosphate, 200 mM NaCl, 10 mM DTT, pH 7.4), at a flow rate of 5 ml min<sup>-1</sup> and the column was then washed with 40 column volumes of PPS buffer at a flow rate of 5 ml min<sup>-1</sup> to remove free GSH. Next, the apo-enzyme was eluted using GST elution buffer (50 mM Tris, 10 mM methylglutathione, pH 8) before relevant fractions were combined according to their A<sub>280nm</sub>. Finally, the buffer was exchanged to PPS buffer, the concentration measured and the apo-enzyme flash-frozen as described for *At*LDOX.

### Oxidation of disulfide bonds of *Pt*GSTF8

Disulfide bonds between C13 of purified apo-*Pt*GSTF8 variants and GSH were formed by incubating 20 µM *Pt*GSTF8 with 0 mM or 1 mM GSSG in the presence of 200 µM GSH in 10 mM MOPSO, pH 6.45 for 2 h at 30 °C, 800 rpm using a 3 mm shaking diameter. Then, 10 µl aliquots of these reactions were directly added to 90 µl ultrafiltered flavan-3,3,4-triol (**4**) to initiate the second step of the two-step in vitro assays with 3,4-*cis*-leucocyanidin (**6**) as described below.

For analysis of disulfide bond formation by MALDI, the samples were desalted and concentrated using C18 Ziptip pipette tips (Millipore). Next, 1 µl of the samples were mixed with 1 µl of sinapinic acid (Sigma-Aldrich) matrix solution and spotted onto a ground steel MALDI target and allowed to dry. Mass spectra were measured in linear positive mode, using an AutoflexSpeed MALDI mass spectrometer (Bruker Daltonics). The instrument was equipped with Nd:YAG laser, emitting at 355 nm. Data collection was carried out using Compass for flexSeries (1.4).

### In vitro GST assay with CDNB (**5**)

Glutathione transferase assays with CDNB (**5**) were performed in 50 mM Tris, pH 7.5 with 5% clarified cell lysate, 1 mM CDNB and 1 mM GSH. All components were mixed, and reactions were performed in 100 µl volume in 96-well microtitre plates (PS, F-Bottom, clear, Greiner). Plates were incubated for 15 min in a Spark 20M (Tecan) plate reader at 30 °C and A<sub>340nm</sub> was measured every 26 s with a 5 s shaking interval between measurements at 180 rpm using a 3 mm shaking diameter. Data collection was carried out using the software Sparkcontrol (2.3). Initial reaction velocities were calculated from the slope of the A<sub>340nm</sub> data from 1 min to 3 min using  $\epsilon = 9,600 \text{ cm}^{-1}\text{M}^{-1}$  for the concentration calculation of the conjugation product.

### In vitro assays with 3,4-*cis*-leucocyanidin (**6**)

Standard reaction conditions for in vitro one-step and two-step biotransformations of 3,4-*cis*-leucocyanidin (**6**) were 100 mM MOPSO, pH 6.2, 20 mM sodium ascorbate, 1 mM  $\alpha$ -ketoglutaric acid, 0.4 mM ammonium iron(II) sulfate, 200 µM 3,4-*cis*-leucocyanidin (**6**) and 20 µM GSH. When using clarified cell lysates, each lysate was added to 5% of the reaction volume. When using purified protein, *At*LDOX was used at a concentration of 1 µM and *Pt*GSTF8 variants were used at a concentration of 2 µM. For the experiment shown in Fig. 1d, the GSH cofactor was replaced with the indicated GSH analogues at the concentration noted on the x-axis label. For the experiment shown in Extended Data Fig. 3a,b, extractions were carried out at the time points indicated on the x-axis label. For the experiment shown in Extended Data Fig. 3h, the

concentration of *Pt*GSTF8 was adapted as indicated on the x-axis label. For the experiment shown in Extended Data Fig. 3i, the concentration of GSH was adapted as shown on the x-axis label.

For one-step reactions, all components were mixed in 96-well microtitre plates in 100 µl volume (PS, F-Bottom, clear, Greiner) and reactions were performed for 30 min in a Spark 20 M (Tecan) plate reader at 30 °C. A<sub>580nm</sub> was measured every 26 s with a 5 s shaking interval between measurements at 180 rpm using a 3 mm shaking diameter.

For two-step reactions, an initial reaction was performed in 90% of the final reaction volume with all components present except for GSH and GST. The reaction was carried out for 15 min in 2 ml microcentrifuge tubes at 30 °C, 1,200 rpm using a 3 mm shaking diameter. Following the first reaction step, the *At*LDOX protein was removed by filtering the reaction through an Amicon Ultra-4 (10 kDa cut-off) (Merck Millipore) and distributing 90 µl aliquots in 96-well microtitre plates (PS, F-Bottom, clear, Greiner). The second reaction was initiated through the addition of the remaining 10% (10 µl) reaction volume containing GSH and GST and the same plate reader method as used for one-step reactions was applied to follow the reaction progress. Initial reaction velocities were calculated from the slope of the A<sub>580nm</sub> data of the first minute using a calibration curve with a commercial cyanidin (**8**) reference compound.

For the Michaelis–Menten kinetics, the first step was adapted to 700 µM 3,4-*cis*-leucocyanidin (**6**) and 2 µM *At*LDOX. The second step after filtering was performed with 0.23 µM *Pt*GSTF8 and varying dilutions of the first step reaction in buffer to cover substrate concentrations from 389 µM to 6.5 µM. The same plate reader method as used for one-step reactions was applied to follow the reaction progress by measuring A<sub>580nm</sub> every 12 s. Exact flavan-3,3,4-triol (**4**) concentration was evaluated by a full conversion control reaction containing 5 µM *Pt*GSTF8 in the second step and quantification of the product cyanidin (**8**) by HPLC–MS. Initial reaction velocities were calculated from the slope of the A<sub>580nm</sub> data of the first minute using a calibration curve with a commercial cyanidin (**8**) reference compound.

For analysis of flavonoids by HPLC–MS, reaction volumes were split after the plate reader incubations, and a 40 µl aliquot was extracted with 80 µl of 75% methanol in H<sub>2</sub>O, 1 mM EDTA, 0.3% hydrochloric acid for quantification of cyanidin (**8**) concentration while a second 40 µl aliquot was extracted with 80 µl of 75% acetonitrile in H<sub>2</sub>O, 1 mM EDTA for quantification of all other flavonoids. After thoroughly mixing the extract by vortexing, the precipitated protein was removed by centrifugation (16,000g, 5 min, 4 °C) before the supernatants were analysed by HPLC–MS.

Ultraviolet–visible spectra were captured using a Lambda 465 (PerkinElmer) spectrometer. For ultraviolet–visible measurements, one-step in vitro reactions with the standard reaction conditions as defined above were performed in a total volume of 1.2 ml in 2 ml microcentrifuge tubes incubated at 30 °C for 15 min at 1,200 rpm using a 3 mm shaking diameter.

### <sup>18</sup>O incorporation into cyanidin (**8**)

Flavan-3,3,4-triol (**4**) was first formed in a reaction with 100 mM MOPSO, pH 6.2, 20 mM sodium ascorbate, 1 mM  $\alpha$ -ketoglutaric acid, 0.4 mM ammonium iron(II) sulfate, 500 µM 3,4-*cis*-leucocyanidin (**6**) and 2.5 µM *At*LDOX, which was incubated at 30 °C for 15 min at 1,200 rpm using a 3 mm shaking diameter. Next, the *At*LDOX protein was removed using an Amicon Ultra-4 device (10 kDa cut-off) (Merck Millipore) and the eluent was diluted with 1.375 vol of the same reaction buffer containing 82% H<sub>2</sub><sup>18</sup>O and incubated at 30 °C, 1,200 rpm using a 3 mm shaking diameter to achieve <sup>18</sup>O equilibration. At varying time points (0, 2, 5, 15, 30 min), purified *Pt*GSTF8 (8 µM) and GSH (20 µM) were added to an aliquot of the equilibration reaction and the resulting mixtures were incubated at 30 °C for 30 min at 1200 rpm using a 3 mm shaking diameter before extraction with 2 vol of 75% methanol in H<sub>2</sub>O, 1 mM EDTA, 0.3% hydrochloric acid. The ratio of <sup>18</sup>O incorporation



was measured by HPLC–MS and calculated as the ratio of peak areas of selected ion monitoring of the  $[M]^+$  ions. The expected  $^{18}\text{O}$  incorporation ratio for stochastic incorporation of 47.4% was calculated from the  $\text{H}_2^{18}\text{O}$  content before addition of *PtGSTF8*.

### HPLC–MS analysis of flavonoids

HPLC–MS analyses of samples from *in vitro* reactions or *S. cerevisiae* cultures were analysed with either an Agilent Technologies 1260 Infinity LC System coupled to a 1260 DAD (G4212B) and a LC/MSD XT (G6135B) for quantification of flavonoids or an Agilent Technologies 1290 Infinity coupled to a 1290 DAD (G1316C) and a 6540 UHD Accurate-Mass Q-TOF LC/MS (G6540B) for recording high-resolution mass spectra. For HPLC–MS analyses, data collection was carried out with OpenLAB CDS ChemStation Edition (C.01.07 SR4), while for data analysis the OpenLAB CDS (2.4 Update\_06) was used. For high-resolution HPLC–MS runs, data collection was done via MassHunter Workstation Software (B.08.00) and data analysis relied on MassHunter Workstation Software Qualitative Analysis (10.0). Flavonoids were separated with an InfinityLab Poroshell 120, EC-C18 (2.1 × 50 mm, 2.7 μm, Agilent) column at 30 °C. Solvents used were A: 5% acetonitrile in water, 0.2% formic acid; and B: acetonitrile, 0.2% formic acid with a gradient profile of: 0–3 min 0–40% B; 3–3.5 min 40–95% B; 3.5–4.5 min 95% B, 4.5–4.6 min 95–0% B; 4.6–6 min 0% B at 0.8 ml min<sup>-1</sup> flow. Flavonoids were quantified using peak areas of  $A_{280\text{nm}}$ ,  $A_{500\text{nm}}$  or selective ion monitoring in positive mode with external calibration curves prepared in the same solvent as the samples. Epidihydroquercetin (**9**) was quantified using a calibration curve of dihydroquercetin (**10**) assuming a similar ionization, and its retention time was verified using dihydroquercetin (**10**) epimerized as previously reported<sup>13</sup>.

### Yeast strain construction

A high-efficiency lithium acetate transformation protocol<sup>52</sup> was used for integration of *MET15* and assembly of multi-expression-plasmids in *S. cerevisiae* strains. The native *MET15* ORF was amplified from genomic DNA of *S. cerevisiae* S288C using primers *MET15\_fw* and *MET15\_rv* and was integrated into yeast strain BY4741 to restore prototrophy for methionine. Next, the multi-expression plasmids were assembled in yeast using transcription-associated recombination as previously described<sup>53</sup>. The construction of all yeast strains is shown in Supplementary Table 5.

### In vivo anthocyanin (1) production in *S. cerevisiae*

For production of flavonoids, engineered *S. cerevisiae* strains C3G1-C3G10 and DHQC3G1–DHQC3G10 (Supplementary Table 5) were restreaked on SC-ULH plates (1.47 g l<sup>-1</sup> Synthetic Complete (Kaiser) Drop Out: Leu, His, Ura, 6.7 g l<sup>-1</sup> yeast nitrogen base (without amino acids), 20 g l<sup>-1</sup> D-(+)-glucose, 16 g l<sup>-1</sup> agar) or SC-U plates (SC-ULH, 380 mg l<sup>-1</sup> leucine, 76 mg l<sup>-1</sup> histidine), respectively. Colonies were used to inoculate 0.25 ml SC-ULH medium (1.47 g l<sup>-1</sup> Synthetic Complete (Kaiser) Drop Out: Leu, His, Ura, 6.7 g l<sup>-1</sup> yeast nitrogen base (without amino acids), 20 g l<sup>-1</sup> D-(+)-glucose), pH set to 5.8 with hydrochloric acid) or SC-U medium (SC-ULH, 380 mg l<sup>-1</sup> leucine, 76 mg l<sup>-1</sup> histidine), respectively, in 96-deep-well plates equipped with a CR1296 sandwich cover (EnzyScreen). The medium to produce flavonoids from dihydroflavonols with the DHQC3G1–DHQC3G10 strains contained 300 μM of the precursor (dihydrokaempferol (**17**), dihydroquercetin (**10**) or dihydromyricetin (**18**)). After incubation for 5 days (C3G strains) or 3 days (DHQC3G strains) at 30 °C, 300 rpm, 5 cm shaking diameter, 0.5 ml methanol containing 0.3% hydrochloric acid was added for extraction of flavonoids. After incubation for 15 min at 30 °C, 1,000 rpm using a 3 mm shaking diameter, cell debris was removed by centrifugation (4,400g at 4 °C for 30 min) and the supernatants were analysed by HPLC–MS at two different dilutions (undiluted and four times diluted with 67% methanol, 0.2% hydrochloric acid in H<sub>2</sub>O).

### Multiple sequence alignments and sequence logos

A multiple protein sequence alignment of known arGSTs of the phi class (*PhAN9*<sup>28</sup> (O24261), *ArTT19*<sup>27</sup> (Q9FE46), *VvGST4*<sup>30</sup> (Q56AY1), *PfGST1*<sup>54</sup> (BIB5E6), *CsGST3*<sup>55</sup> (I2FHU9), *DcGSTF2*<sup>56</sup> (I4DUE3), *PpRiant*<sup>21</sup> (MSWTIS), *LcGST4*<sup>57</sup> (AOA0U4AVK2), *FvRAP*<sup>20</sup> (UPI0002C2E8B9), *MdGST1*<sup>25</sup> (G3LX80), *AcGST1*<sup>26</sup> (AOA2R6R622), *IbGSTF4*<sup>58</sup> (AOA8E4DFU3), *CsGSTa*<sup>59</sup> (AOA4P8P9K5), *PcGSTF12*<sup>60</sup> (AOA6G5X592), *GtGST1*<sup>24</sup> (NCBI accession number: BCD52748.1), *LhGST*<sup>23</sup> (AOA5B9G8S8), *EpBract1*<sup>61</sup> (NCBI accession number: QTX16320.1), *GhGSTF12*<sup>22</sup> (AOA5D2YA6), *RsGST1*<sup>62</sup> (AOA6JOK8F1), *PtGSTF8*<sup>39</sup> (NCBI accession number, 5F07\_A) – Uniprot accession numbers unless stated otherwise) was constructed using Clustal Omega 1.2.2<sup>63</sup> with standard settings. Sequence logos were generated from the multiple sequence alignment using the R package ggseqlogo<sup>64</sup>.

### Crystallization of *PtGSTF8* in presence of (±)-catechin and GSH

Purified *PtGSTF8* in 30 mM Tris, pH 8, 200 mM NaCl, 1 mM EDTA was concentrated to 12.4 mg ml<sup>-1</sup> using an Amicon Ultra-4 device (10 kDa cut-off) (Merck Millipore) and supplemented with 1 mM GSH and 25 mM (±)-catechin immediately before crystallization. Crystallization was performed using the sitting drop method at 4 °C in an Intelli-Plate 96-3 LVR (Hampton Research). A drop of 150 nl enzyme/cofactor/substrate solution was mixed with 150 nl precipitant (100 mM MES, pH 6.4, 200 mM MgCl<sub>2</sub>, 17.9% PEG200) and equilibrated against 50 μl reservoir solution (100 mM MES, pH 6.4, 200 mM MgCl<sub>2</sub>, 17.9% PEG200). Crystals were observed after 5 days of crystallization and picked after 21 days of total incubation. Finally, 2 μl 30% glycerol (in reservoir solution) was added to the drop before picking the crystal using 0.1 mm litholoops and flash-freezing it in liquid nitrogen.

Diffraction data were collected at the Swiss Light Source (Paul Scherrer Institut, Villigen AG) on beamline X06DA-PXIII at a temperature of 100 K and 0.999995 Å wavelength using the serial synchrotron crystallography software suite. The data were integrated by XDS<sup>65</sup> run through autoproc<sup>66</sup>. A dataset with 1.09 Å resolution was obtained. PDB 5F07 was used in molecular replacement to determine the starting phases for refinement with anisotropic B-factors and hydrogen atoms using MOLREP<sup>67</sup>. Refinement and manual model building were performed with REFMAC<sup>68</sup> and coot<sup>69</sup>, respectively, using the ccp4i2<sup>70</sup> interface. The final model contains all protein residues except Met1, as well as GSH and (–)-catechin. The MolProbity<sup>71</sup> validation of this model gave 97.16% Ramachandran favoured residues and 0% outliers with a Molprobity score of 1.16.

### Analysis of *PtGSTF8* crystal structures

$F_0 - F_c$  composite omit maps were calculated with the CCP4 7.1.018: comit<sup>72</sup> (v.0.1.0) function using the fast method. Molecular docking was performed with AutoDock Vina 1.2.0<sup>73</sup> using default settings, except for increasing the exhaustiveness to 32. Figure 2b with colour-coded RMSD between crystal structures was generated using the ColorByRMSD Pymol script (<https://pymolwiki.org/index.php/ColorByRMSD>, accessed on 16 June 2022). The interactions of ligands to the enzyme were analysed using PLIP<sup>74</sup> with standard settings. The RMSD of matched atoms between ligands was calculated using LigRMSD<sup>75</sup> with standard settings.

### QM/MM calculations

Enzyme–substrate complexes were generated through molecular docking using the GOLD software<sup>76</sup> (ChemScore fitness function) on the crystal structure of *PtGSTF8* preserving crystallographic Na<sup>+</sup> ions and flav-3-on-4-ol coordinates optimized at the B3LYP/6-31G(d) quantum mechanical level (see below for details). The docking cavity was centred on the sulfur atom of the GSH's cysteine and allowed to extend in a spherical surrounding region with a 15 Å radius. The number of GA (genetic algorithm) runs was set to 30.

The top-ranking binding pose was then prepared for classical MD relaxation with the AMBER 20<sup>77</sup> suite using the ff14SB<sup>78</sup> force field for the protein and gaff2<sup>79</sup> for GSH and flav-3-on-4-ol. Enzyme–substrate complexes were immersed in a water box with an 8 Å buffer of OPC3<sup>80</sup> water. A two-stage geometry optimization approach was implemented. The first stage minimizes only the positions of solvent molecules and ions, and the second stage is an unrestrained minimization of all the atoms in the simulation cell. The systems were then heated for 100 ps by incrementing the temperature from 0 to 300 K under a constant pressure of 1 atm and periodic boundary conditions with a timestep of 1 fs. Harmonic restraints of 50.0 kcal mol<sup>-1</sup> Å<sup>-2</sup> were applied to the solutes, including the crystallographic ions, to allow solvent relaxation while preserving the maximum fidelity to the docking pose, and the Andersen temperature coupling scheme<sup>81,82</sup> was used to control and equalize the temperature. The last frame from the MD trajectory was then extracted and excess solvent removed leaving a 6 Å thick water shell around the enzyme–substrate complexes.

Full geometry optimizations, relaxed scans and transition structure searches were carried out with Gaussian 16<sup>83</sup> with the ONIOM<sup>84</sup> QM/MM method using the B3LYP hybrid functional<sup>85</sup> and 6-31 G(d) basis set with ultrafine integration grids for the QM part and the ff14SB<sup>78</sup>, gaff2<sup>79</sup> and TIP3P<sup>86</sup> force fields for the MM part (protein, substrate and water, respectively) with electrostatic embedding. Only water molecules were kept frozen, allowing the enzyme to adapt along the reaction path. The definition of the QM and MM layers is depicted in Extended Data Fig. 8. The possibility of different conformations was considered for all structures. All stationary points were characterized by a frequency analysis performed at the same level used in the geometry optimizations from which thermal corrections were obtained at 298.15 K. Single-point energies were calculated on the optimized geometries using the ωB97X-D hybrid functional<sup>87</sup> and the 6-311 + G(2d,p) basis set with ultrafine integration grids for the QM part and the ff14SB, gaff2 and TIP3P force fields for the MM part (protein, substrate and water, respectively) with electrostatic embedding. The lowest-energy conformer for each calculated stationary (available in PDB format as Supplementary Figs. 1 and 2) was considered in the discussion; computed geometries and energies can be accessed through the Zenodo repository (<https://doi.org/10.5281/zenodo.8069429>). ONIOM energies, entropies, enthalpies, Gibbs free energies and lowest frequencies of the calculated structures are summarized in Supplementary Table 8.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Nucleotide sequences of codon-optimized genes can be found in Supplementary Table 6. The diffraction images were deposited to Integrated Resource for Reproducibility in Macromolecular Crystallography<sup>88</sup> (<http://proteindiffraction.org/>) and can be accessed via PDB 8AGQ. Crystallographic coordinates of the ternary complex of PtGSTF8 have been deposited in the PDB as 8AGQ. The binary PtGSTF8 crystal structure used in molecular replacement experiments can be accessed via PDB 5F07. Computed geometries and energies can be accessed through the Zenodo repository (<https://doi.org/10.5281/zenodo.8069429>). Source data are provided with this paper.

### References

- Harborne, J. B. & Grayer, R. J. in *The Flavonoids: Advances in Research Since 1980s* (ed. Harborne, J.) 1–20 (Springer, 1988).
- Mannino, G., Gentile, C., Ertani, A., Serio, G. & Berteà, C. M. Anthocyanins: biosynthesis, distribution, ecological role, and use of biostimulants to increase their content in plant foods—a review. *Agriculture* **11**, 212 (2021).
- Pojer, E., Mattivi, F., Johnson, D. & Stockley, C. S. The case for anthocyanin consumption to promote human health: a review. *Compr. Rev. Food Sci. Food Saf.* **12**, 483–508 (2013).
- Belwal, T. et al. Anthocyanins, multi-functional natural products of industrial relevance: recent biotechnological advances. *Biotechnol. Adv.* **43**, 107600 (2020).
- Yang, S., Mi, L., Wu, J., Liao, X. & Xu, Z. Strategy for anthocyanins production: from efficient green extraction to novel microbial biosynthesis. *Crit. Rev. Food Sci. Nutr.* <https://doi.org/10.1080/10408398.2022.2067117> (2022).
- Eichenberger, M., Hansson, A., Fischer, D., Dürr, L. & Naesby, M. De novo biosynthesis of anthocyanins in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **18**, 103 (2018).
- Zha, J., Wu, X., Gong, G. & Koffas, M. A. G. Pathway enzyme engineering for flavonoid production in recombinant microbes. *Metab. Eng. Commun.* **9**, e00104 (2019).
- Preuß, A. et al. *Arabidopsis thaliana* expresses a second functional flavonol synthase. *FEBS Lett.* **583**, 1981–1986 (2009).
- Chaves-Silva, S. et al. Understanding the genetic regulation of anthocyanin biosynthesis in plants—tools for breeding purple varieties of fruits and vegetables. *Phytochemistry* **153**, 11–27 (2018).
- Wilmouth, R. C. et al. Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. *Structure* **10**, 93–103 (2002).
- Turnbull, J. J. et al. Mechanistic studies on three 2-oxoglutarate-dependent oxygenases of flavonoid biosynthesis. *J. Biol. Chem.* **279**, 1206–1216 (2004).
- Turnbull, J. J. et al. The C-4 stereochemistry of leucocyanidin substrates for anthocyanidin synthase affects product selectivity. *Bioorg. Med. Chem. Lett.* **13**, 3853–3857 (2003).
- Zhang, J., Trossat-Magnin, C., Bathany, K., Delrot, S. & Chaudière, J. Oxidative transformation of leucocyanidin by anthocyanidin synthase from *Vitis vinifera* leads only to quercetin. *J. Agric. Food Chem.* **67**, 3595–3604 (2019).
- Welford, R. W. D., Turnbull, J. J., Claridge, T. D. W., Schofield, C. J. & Prescott, A. G. Evidence for oxidation at C-3 of the flavonoid C-ring during anthocyanin biosynthesis. *Chem. Commun.* **18**, 1828–1829 (2001).
- Heller, W., Britsch, L., Forkmann, G. & Grisebach, H. Leucoanthocyanidins as intermediates in anthocyanidin biosynthesis in flowers of *Matthiola incana* R. Br. *Planta* **163**, 191–196 (1985).
- Yan, Y., Chemler, J., Huang, L., Martens, S. & Koffas, M. A. G. Metabolic engineering of anthocyanin biosynthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* **71**, 3617–3623 (2005).
- Zhao, J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* **20**, 576–585 (2015).
- Chanoca, A. et al. Anthocyanin vacuolar inclusions form by a microautophagy mechanism. *Plant Cell* **27**, 2545–2559 (2015).
- Sun, Y., Li, H. & Huang, J.-R. *Arabidopsis* TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Mol. Plant* **5**, 387–400 (2012).
- Luo, H. et al. Reduced anthocyanins in petioles codes for a GST anthocyanin transporter that is essential for the foliage and fruit coloration in strawberry. *J. Exp. Bot.* **69**, 2595–2608 (2018).
- Cheng, J. et al. A small indel mutation in an anthocyanin transporter causes variegated colouration of peach flowers. *J. Exp. Bot.* **66**, 7227–7239 (2015).
- Shao, D. et al. GhGSTF12, a glutathione S-transferase gene, is essential for anthocyanin accumulation in cotton (*Gossypium hirsutum* L.). *Plant Sci.* **305**, 110827 (2021).
- Cao, Y. et al. LhGST is an anthocyanin-related glutathione S-transferase gene in Asiatic hybrid lilies (*Lilium* spp.). *Plant Cell Rep.* **40**, 85–95 (2021).
- Tasaki, K. et al. Molecular characterization of an anthocyanin-related glutathione S-transferase gene in Japanese gentian with the CRISPR/Cas9 system. *BMC Plant Biol.* **20**, 370 (2020).

25. Jiang, S. et al. MdGSTF6, activated by MdMYB1, plays an essential role in anthocyanin accumulation in apple. *Hortic. Res.* **6**, 40 (2019).
26. Liu, Y. et al. Molecular cloning and functional characterization of AcGST1, an anthocyanin-related glutathione S-transferase gene in kiwifruit (*Actinidia chinensis*). *Plant Mol. Biol.* **100**, 451–465 (2019).
27. Kitamura, S., Shikazono, N. & Tanaka, A. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J.* **37**, 104–114 (2004).
28. Alfenito, M. R. et al. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *Plant Cell* **10**, 1135–1149 (1998).
29. Marrs, K. A., Alfenito, M. R., Lloyd, A. M. & Walbot, V. A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene *Bronze-2*. *Nature* **375**, 397–400 (1995).
30. Gomez, C. et al. In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. *Plant J.* **67**, 960–970 (2011).
31. Mueller, L. A., Goodman, C. D., Silady, R. A. & Walbot, V. AN9, a petunia glutathione S-transferase required for anthocyanin sequestration, is a flavonoid-binding protein. *Plant Physiol.* **123**, 1561–1570 (2000).
32. Li, X. et al. The *Arabidopsis* *tt19-4* mutant differentially accumulates proanthocyanidin and anthocyanin through a 3' amino acid substitution in glutathione S-transferase. *Plant Cell Environ.* **34**, 374–388 (2011).
33. Lu, N., Jun, J. H., Liu, C. & Dixon, R. A. The flexibility of proanthocyanidin biosynthesis in plants. *Plant Physiol.* **190**, 202–205 (2022).
34. Kitamura, S. et al. Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* flavonoid accumulation mutants. *Plant J.* **62**, 549–559 (2010).
35. Waki, T. et al. A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat. Commun.* **11**, 870 (2020).
36. Bar-Even, A. et al. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).
37. Brouillard, R. & Lang, J. The hemiacetal–*cis*-chalcone equilibrium of malvin, a natural anthocyanin. *Can. J. Chem.* **68**, 755–761 (1990).
38. Forino, M., Gambuti, A., Luciano, P. & Moio, L. Malvidin-3-O-glucoside chemical behavior in the wine pH range. *J. Agric. Food Chem.* **67**, 1222–1229 (2019).
39. Pégeot, H. et al. Structural plasticity among glutathione transferase phi members: natural combination of catalytic residues confers dual biochemical activities. *FEBS J.* **284**, 2442–2463 (2017).
40. Kubo, Y., Graminski, G. F. & Armstrong, R. N. Spectroscopic and kinetic evidence for the thiolate anion of glutathione at the active site of glutathione S-transferase. *Biochemistry* **28**, 3562–3568 (1989).
41. Perperopoulou, F., Pouliou, F. & Labrou, N. E. Recent advances in protein engineering and biotechnological applications of glutathione transferases. *Crit. Rev. Biotechnol.* **38**, 511–528 (2018).
42. Deponte, M. Glutathione catalysis and the reaction mechanisms of glutathione-dependent enzymes. *Biochim. Biophys. Acta* **1830**, 3217–3266 (2013).
43. Mannervik, B., Ismail, A., Lindström, H., Sjödin, B. & Ing, N. H. Glutathione transferases as efficient ketosteroid isomerases. *Front. Mol. Biosci.* **8**, 765970 (2021).
44. Khojasteh-Bakht, S. C., Nelson, S. D. & Atkins, W. M. Glutathione S-transferase catalyzes the isomerization of (R)-2-hydroxymenthofuran to mint lactones. *Arch. Biochem. Biophys.* **370**, 59–65 (1999).
45. Jiang, N. et al. Synergy between the anthocyanin and RDR6/SGS3/DCL4 siRNA pathways expose hidden features of *Arabidopsis* carbon metabolism. *Nat. Commun.* **11**, 2456 (2020).
46. Nakayama, T., Takahashi, S. & Waki, T. Formation of flavonoid metabolites: functional significance of protein–protein interactions and impact on flavonoid chemodiversity. *Front. Plant Sci.* **10**, 821 (2019).
47. Han, L., Zhou, L., Zou, H., Yuan, M. & Wang, Y. PsGSTF3, an anthocyanin-related glutathione S-transferase gene, is essential for petal coloration in tree peony. *Int. J. Mol. Sci.* **23**, 1423 (2022).
48. Yuan, Y., Andersen, E. & Zhao, H. Flexible and versatile strategy for the construction of large biochemical pathways. *ACS Synth. Biol.* **5**, 46–52 (2016).
49. Eichenberger, M. et al. Asymmetric cation-olefin monocyclization by engineered squalene–hopene cyclases. *Angew. Chem. Int. Ed.* **60**, 26080–26086 (2021).
50. Gasteiger, E. et al. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, 2005); <https://doi.org/10.1385/1-59259-890-0:571>
51. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
52. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
53. Eichenberger, M. et al. Metabolic engineering of *Saccharomyces cerevisiae* for de novo production of dihydrochalcones with known antioxidant, antidiabetic, and sweet tasting properties. *Metab. Eng.* **39**, 80–89 (2017).
54. Yamazaki, M. et al. Differential gene expression profiles of red and green forms of *Perilla frutescens* leading to comprehensive identification of anthocyanin biosynthetic genes. *FEBS J.* **275**, 3494–3502 (2008).
55. Kitamura, S., Akita, Y., Ishizaka, H., Narumi, I. & Tanaka, A. Molecular characterization of an anthocyanin-related glutathione S-transferase gene in cyclamen. *J. Plant Physiol.* **169**, 636–642 (2012).
56. Larsen, E. S., Alfenito, M. R., Briggs, W. R. & Walbot, V. A carnation anthocyanin mutant is complemented by the glutathione S-transferases encoded by maize *Bz2* and petunia *An9*. *Plant Cell Rep.* **21**, 900–904 (2003).
57. Hu, B. et al. *LcGST4* is an anthocyanin-related glutathione S-transferase gene in *Litchi chinensis* Sonn. *Plant Cell Rep.* **35**, 831–843 (2016).
58. Kou, M. et al. A novel glutathione S-transferase gene from sweet potato, *IbGSTF4*, is involved in anthocyanin sequestration. *Plant Physiol. Biochem.* **135**, 395–403 (2019).
59. Liu, Y. et al. Three *Camellia sinensis* glutathione S-transferases are involved in the storage of anthocyanins, flavonols, and proanthocyanidins. *Planta* **250**, 1163–1175 (2019).
60. Zhang, Z. et al. Transcriptomic and metabolomic analysis provides insights into anthocyanin and procyanidin accumulation in pear. *BMC Plant Biol.* **20**, 129 (2020).
61. Vilperte, V., Boehm, R. & Debener, T. A highly mutable GST is essential for bract colouration in *Euphorbia pulcherrima* Willd. Ex Klotsch. *BMC Genomics* **22**, 208 (2021).
62. Lai, B. et al. Identification and functional characterization of *RsgST1*, an anthocyanin-related glutathione S-transferase gene in radish. *J. Plant Physiol.* **263**, 153468 (2021).
63. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
64. Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).
65. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).

66. Vonnrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D* **67**, 293–302 (2011).
67. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
68. Murshudov, G. N. et al. REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
69. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
70. Potterton, L. et al. CCP 4 i 2: the new graphical user interface to the CCP 4 program suite. *Acta Crystallogr. D* **74**, 68–84 (2018).
71. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
72. Bhat, T. N. Calculation of an OMIT map. *J. Appl. Crystallogr.* **21**, 279–281 (1988).
73. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).
74. Adasme, M. F. et al. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Res.* **49**, W530–W534 (2021).
75. Velázquez-Libera, J. L., Durán-Verdugo, F., Valdés-Jiménez, A., Núñez-Vivanco, G. & Caballero, J. LigRMSD: a web server for automatic structure matching and RMSD calculations among identical and similar compounds in protein–ligand docking. *Bioinformatics* **36**, 2912–2914 (2020).
76. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).
77. Case, D. A. et al. Amber 2022 (Univ. California, 2022).
78. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
79. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
80. Izadi, S. & Onufriev, A. V. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **145**, 074501 (2016).
81. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72**, 2384–2393 (1980).
82. Andrea, T. A., Swope, W. C. & Andersen, H. C. The role of long-ranged forces in determining the structure and properties of liquid water. *J. Chem. Phys.* **79**, 4576–4584 (1983).
83. Frisch, M. J. et al. Gaussian 16 Revision C.01 (Gaussian, 2016).
84. Vreven, T. et al. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* **2**, 815–826 (2006).
85. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **98**, 11623–11627 (1994).
86. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
87. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615 (2008).
88. Grabowski, M. et al. A public database of macromolecular diffraction experiments. *Acta Crystallogr. D* **72**, 1181–1193 (2016).

## Acknowledgements

This work was created as part of NCCR Catalysis, a National Centre of Competence in Research funded by the Swiss National Science Foundation (grant number 180544 to R.M.B.) and by AEI (Spain) (grant number RTI2018-099592-B-C22 to G.J.-O.). F.P. thanks MINECO for a Juan de la Cierva Incorporación fellowship (IJC2020-045506-I). We thank I. Krosiakova for acquiring the MALDI data, S. Kern for support in interpreting MS data, M. Niklaus for assistance with sequence logos and D. Patsch for assistance with Python.

## Author contributions

M.E., M.N. and R.M.B. initiated and designed the project. M.E., T.S. and R.M.B. designed the experiments. M.E., S.H., T.S. and J.K. carried out the experiments and M.E., S.H., T.S., J.K. and R.M.B. analysed the data. S.H., T.S. and P.R.E.M. performed the structure elucidation, F.P. and G.J.-O. carried out the QM/MM calculations. M.E. and R.M.B. wrote the paper with feedback from all authors. R.M.B. supervised the project.

## Competing interests

M.N. is employed by Lantana Bio. The other authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41929-023-01018-y>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41929-023-01018-y>.

**Correspondence and requests for materials** should be addressed to Rebecca M. Buller.

**Peer review information** *Nature Catalysis* thanks Mattheos Koffas, Bengt Mannervik, Boas Pucker, Dingguo Xu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

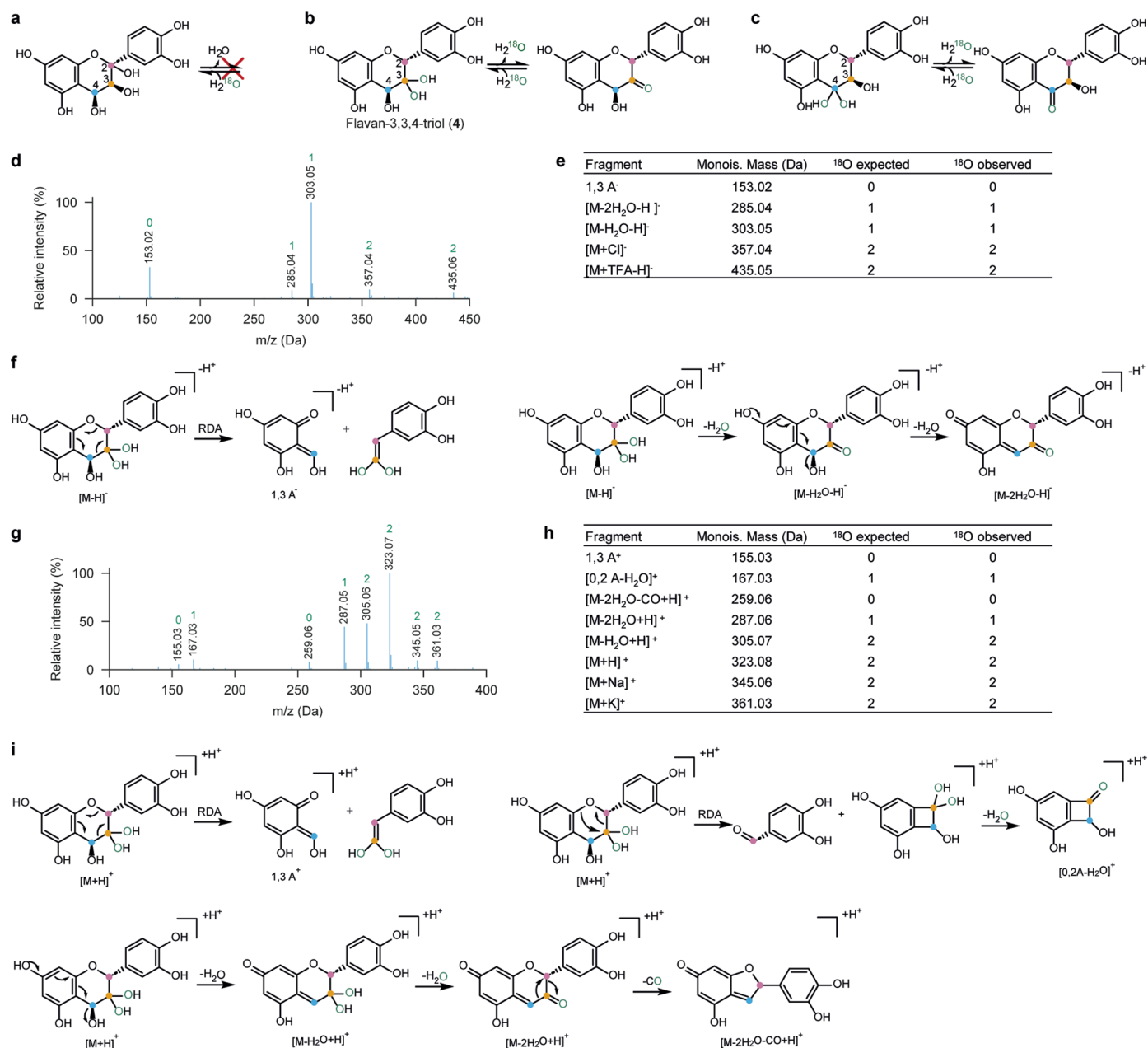
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

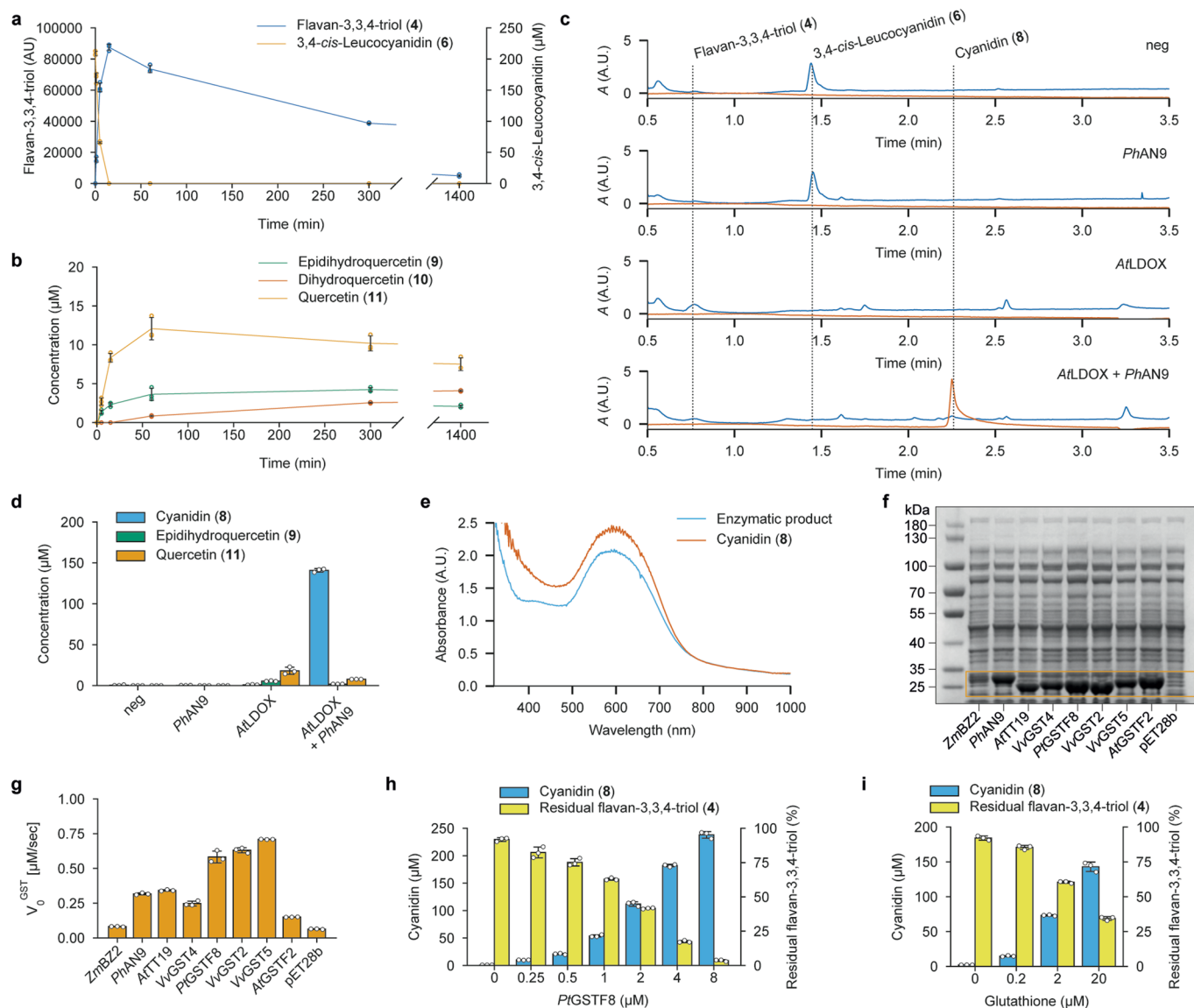
© The Author(s) 2023





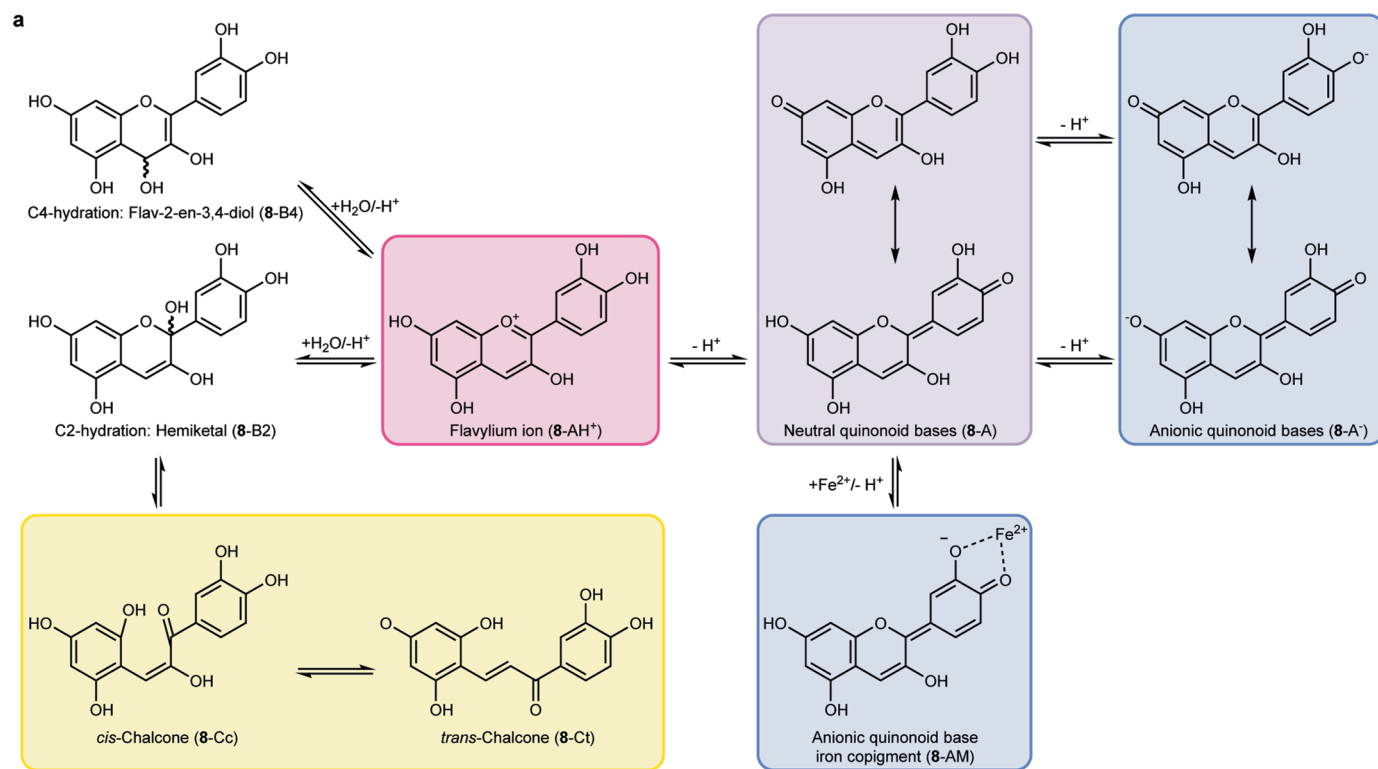
**Extended Data Fig. 2 | Mass spectrometric analysis of the LDOX oxidation product flavan-3,3,4-triol (4).** **a,b,c** Mechanistically, the AtLDOX catalysed hydroxylation of 3,4-*cis*-leucocyanidin (6) may result in hydroxylation products at C<sub>2</sub> (**a**), C<sub>3</sub> (**b**) or C<sub>4</sub> (**c**). **a**, Hydroxylation of C<sub>2</sub> would preclude the incorporation of an O<sup>18</sup> label as no equilibrium between a ketone and a gem-diol form exists for this hydroxylation product. **b,c**, Hydroxylation at either C<sub>3</sub> (**b**) or C<sub>4</sub> (**c**) allows the incorporation of the heavy oxygen isotope at the corresponding positions. **d,g**, Negative (**d**) and positive (**g**) ion ESI-MS spectrum of the AtLDOX oxidation product from 3,4-*cis*-leucocyanidin (6). The green labels show the observed <sup>18</sup>O incorporations into the fragments when reactions are performed with H<sub>2</sub><sup>18</sup>O.

The observed incorporation of up to two O<sup>18</sup> labels rules out formation of the C<sub>2</sub> hydroxylated product (**a**). **e,h**, Observed adducts and fragments in negative (**e**) and positive (**h**) ion ESI-MS, their calculated monoisotopic masses, as well as the expected and observed number of <sup>18</sup>O incorporations. **f,i**, Proposed fragmentation pathways of flavan-3,3,4-triol (4) for fragments observed in negative (**f**) and positive (**i**) ion ESI-MS. For fragment [1,3 A<sup>-</sup>] (**d,e,f**) and fragment [1,3 A<sup>+</sup>] (**g,h,i**) comprising C<sub>4</sub>, the incorporation of an O<sup>18</sup> label is not observed, while fragment [0,2 A-H<sub>2</sub>O]<sup>+</sup> (**g,h,i**), including C<sub>3</sub>, contains an O<sup>18</sup> label supporting formation of flavan-3,3,4-triol (4).



**Extended Data Fig. 3 | arGSTs catalyse the conversion of flavan-3,3,4-triol (4) into cyanidin (8).** **a**, Time course of formation of flavan-3,3,4-triol (4) from 3,4-*cis*-leucocyanidin (6) using purified *AtLDOX*. **b**, Timecourse of formation of side products from 3,4-*cis*-leucocyanidin (6) using purified *AtLDOX*. **c**, HPLC-UV traces of coupled reactions using clarified *E. coli* cell lysates expressing *PhAN9* and *AtLDOX* with 3,4-*cis*-leucocyanidin (6). Negative control reactions were performed with buffer in place of the purified enzymes (neg). Blue: 280 nm, Red: 500 nm. **d**, Quantification of reaction products of coupled reactions using clarified *E. coli* cell lysates expressing *PhAN9* and *AtLDOX* with 200  $\mu\text{M}$  3,4-*cis*-leucocyanidin (6). Negative control reactions were performed with buffer

in place of the purified enzymes (neg). **e**, Comparison of UV-Vis spectra of the enzymatic product and authentic cyanidin (8) in reaction buffer. **f**, SDS Page of clarified *E. coli* lysates expressing different GSTs. **g**, GSH-transferase activity of *E. coli* lysates expressing different GSTs towards the model substrate CDNB (5). **h**, Conversion of ultrafiltered flavan-3,3,4-triol (4) into cyanidin (8) using purified PtGSTF8 at different concentrations. **i**, Conversion of ultrafiltered flavan-3,3,4-triol (4) into cyanidin (8) using 2  $\mu\text{M}$  purified PtGSTF8 and different concentrations of GSH. Data of panels **a**, **b**, **d**, **g**, **h**, **i** are mean values  $\pm$  standard deviation of three independent replicates.

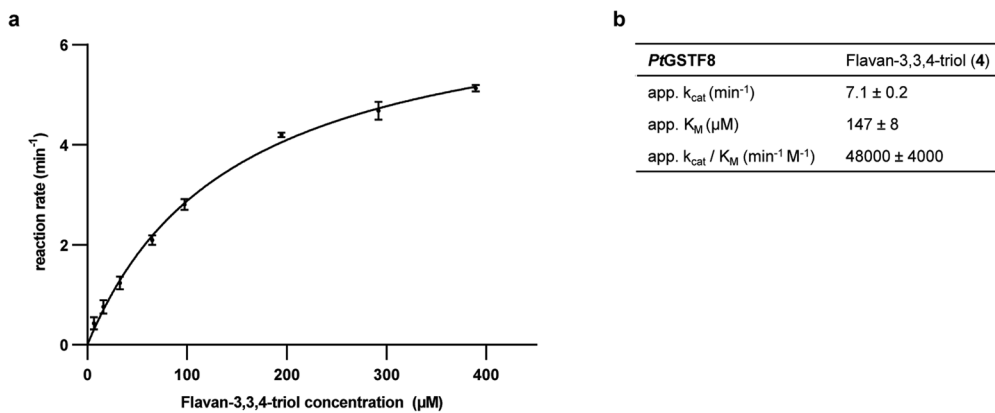


**b**

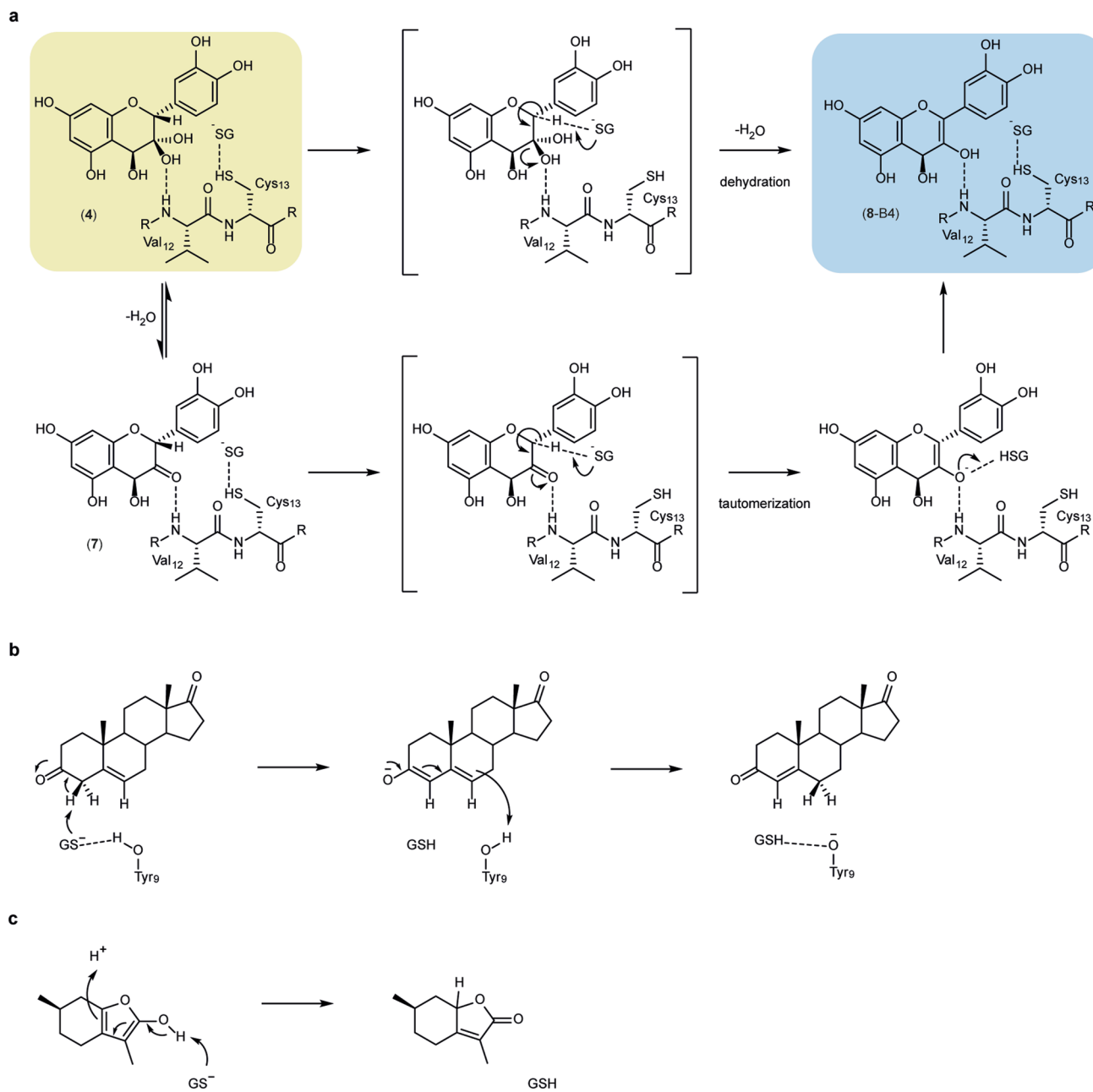
Buffer	MOPSO pH 6.2	Sodium phosphate pH 6.2	100 mM buffer	0.4 mM ammonium iron(II) sulfate	20 mM sodium ascorbate	1 mM $\alpha$ -ketoglutarate	2mM EDTA
	+	+	+	+	+	+	+
	-	-	+	+	+	-	-
	-	-	+	+	-	-	+
	-	-	+	+	-	-	+
	-	-	-	+	-	+	-

**Extended Data Fig. 4 | Reaction network of cyanidin (8).** **a**, Equilibrium reaction network of cyanidin (8) in solution. The boxes show the colours of the different species. **b**, Colour expression of authentic cyanidin (8) in reaction buffers with different combinations of co-factors and EDTA, showing the formation of the blue iron co-pigment of the anionic quinonoid base (8-AM).

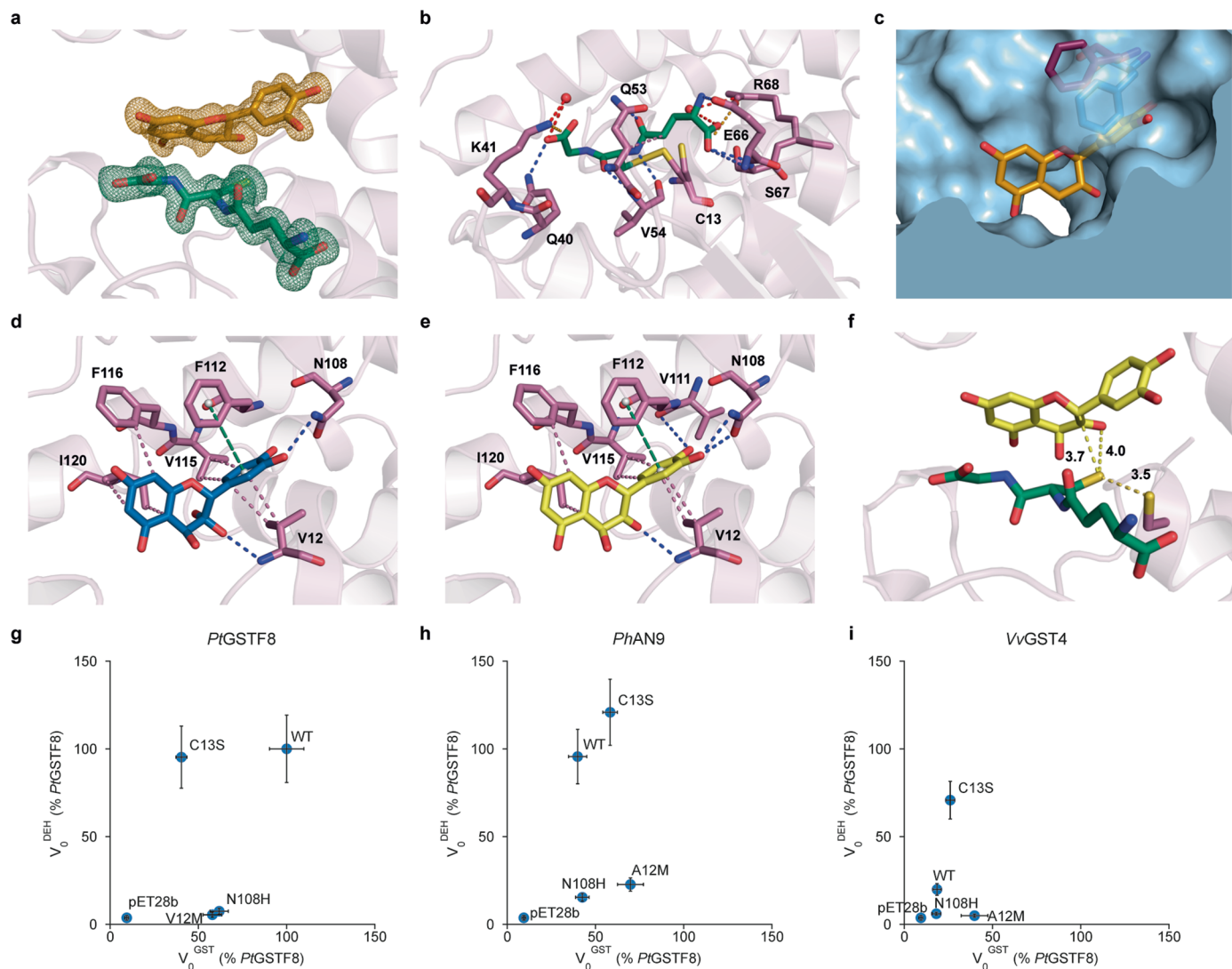




**Extended Data Fig. 5 | Michaelis–Menten-Kinetics of *Pr*GSTF8 with flavan-3,3,4-triol (**4**).** **a**, Michaelis–Menten plot of measured reaction rate at different substrate concentration. **b**, Table of apparent Michaelis–Menten parameters. Data of panels **a**, **b** are mean values  $\pm$  standard deviation of three independent replicates.

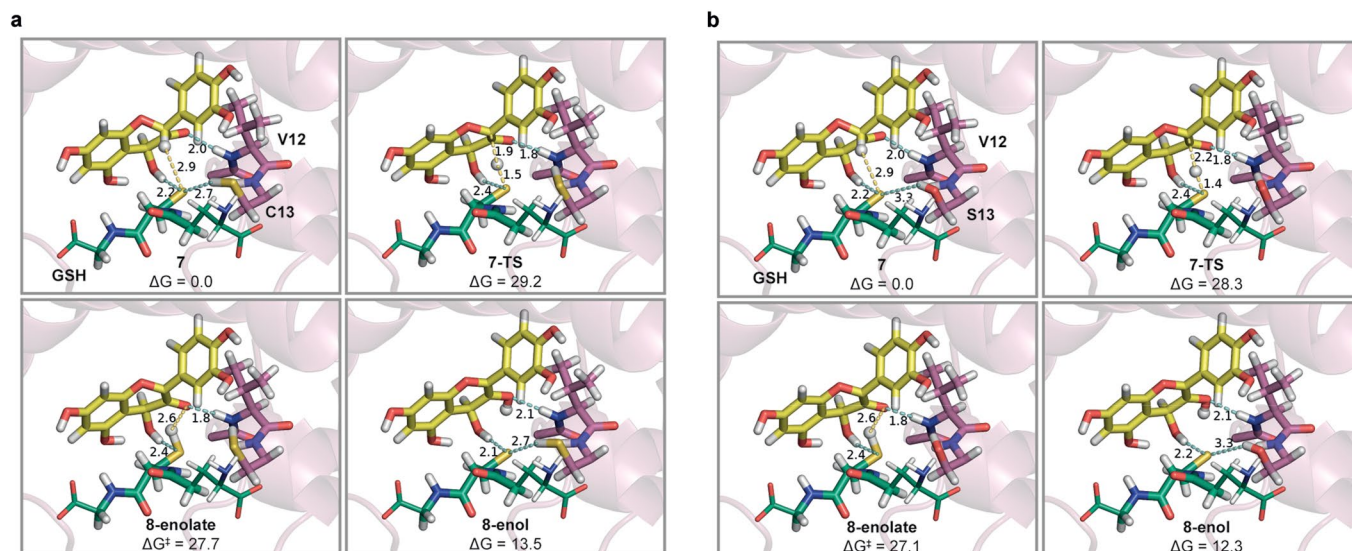


**Extended Data Fig. 6 | Proposed Brønsted base mechanism of GST catalysed reactions. a,** Proposed mechanisms for the dehydration of flavan-3,3,4-triol (**4**) to flav-2-en-3,4-diol (**8-B4**) by arGSTs. **b,** Proposed mechanism for the isomerization of 5-androsten-3,17-dione by GST A3-3<sup>43</sup>. **c,** Proposed mechanism for tautomerization of (R)-2-hydroxymenthofuran by GST A1-1<sup>44</sup>.



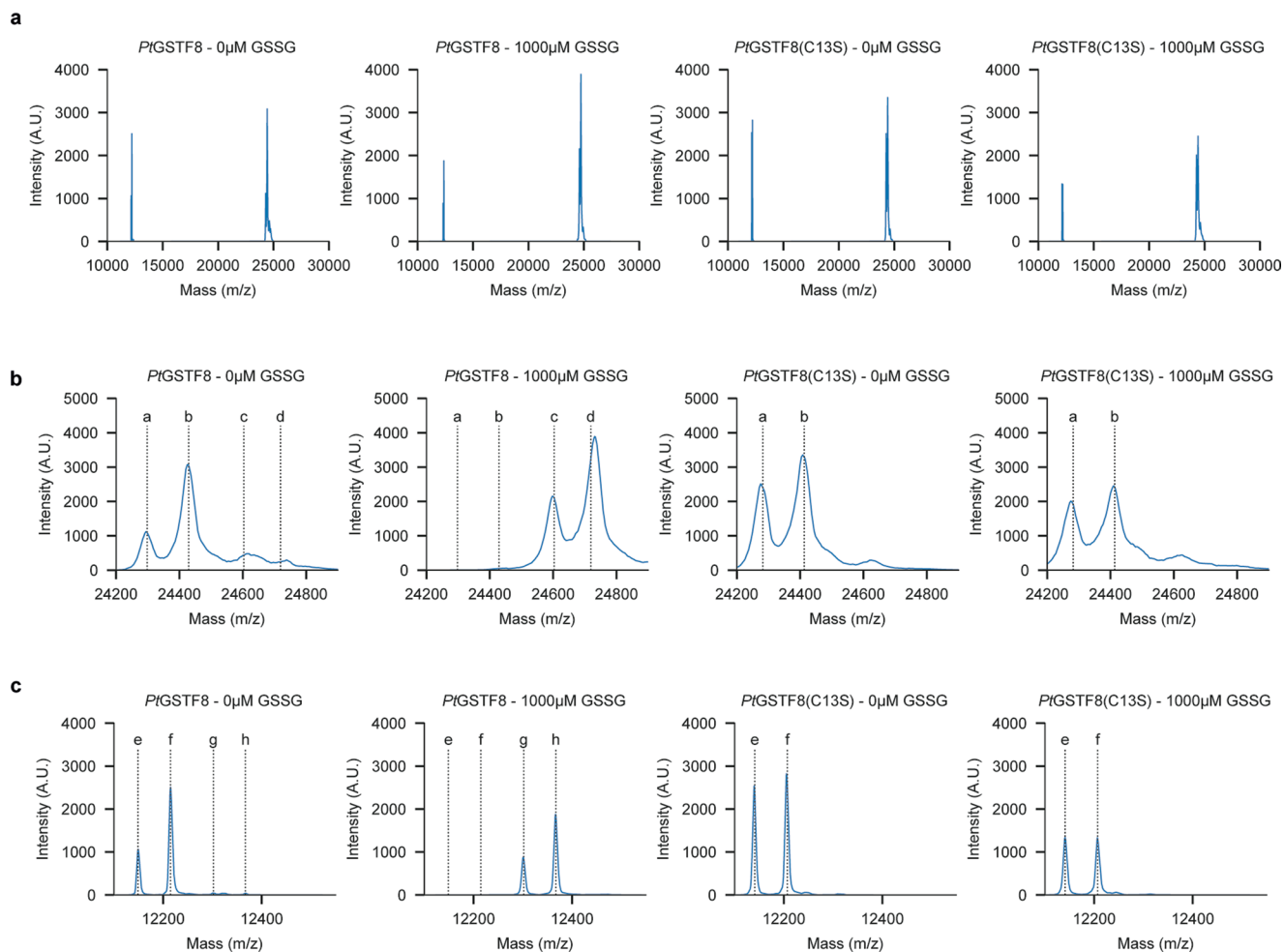
**Extended Data Fig. 7 | Structural basis of *PrGSTF8* dehydratase activity. a,**  $F_0 - F_c$  composite omit maps of glutathione (green) and (-)-catechin (**12**) (orange) in the ternary complex rendered at  $3\sigma$ . **b,** G-site interactions to GSH. Blue: hydrogen bond, red: water bridge, green:  $\pi$ - $\pi$  interaction, magenta: hydrophobic interaction. **c,** Surface representation of H-site of *PrGSTF8* crystallized without (-)-catechin (**12**) (PDB:5F07). F112 is shown in the conformations in crystal structures obtained without (blue) and with (magenta) (-)-catechin (**12**) and its surface is rendered transparently. (-)-catechin (**12**) is superimposed from the ternary complex. **d,e,** H-site interactions to docked flavan-3,3,4-triol (**4**) (**d**,

blue) and flavan-3-on-4-ol (**7**) (**e**, yellow). Dashes are coloured according to the interaction types: Blue: hydrogen bond, red: water bridge, green:  $\pi$ - $\pi$  interaction, magenta: hydrophobic interaction. **f,** Docking of flavan-3-on-4-ol (**7**) (yellow) into the active site of *PrGSTF8* and distances (in Å) from the GSH (green) thiolate to C-2 and O-3 of **7** as well as distance of the C13 thiolate (purple) to GSH (green) **7**. **g,h,i,** Dehydratase activity towards ultrafiltered flavan-3,3,4-triol (**4**) and GST-transferase activity towards CDNB (**5**) of variants of *PrGSTF8* (**g**), *PhAN9* (**h**), *VvGST4* (**i**) measured in colorimetric 96-well-plate assays. Data of panels **g,h,i** are mean values  $\pm$  standard deviation of three independent replicates.



**Extended Data Fig. 8 | QM/MM calculations of the *Pt*GSTF8 wild-type and C13S reaction pathways. **a,b****, Representation of the stationary points calculated with ONIOM( $\omega$ B97X-D/6-311 + G(2d,p):ff14SB,gaff2,TIP3P)// ONIOM(B3LYP/6-31 G(d):ff14SB,gaff2,TIP3P) for the *Pt*GSTF8-catalysed deprotonation-protonation reaction: flavan-3-on-4-ol (**7**), transition state for deprotonation (**7-TS**), deprotonation product (**8-enolate**) and cyanidin (**8-B4**) bound to the wild-type enzyme (**a**) and the C13S variant (**b**). The calculated activation energies ( $\Delta G^\ddagger$  in kcal mol<sup>-1</sup>) for the abstraction the C2 proton of **7** to produce the cyanidin enolate (**8-enolate**) were found to be almost identical for the wild-type and the C13S variant, in agreement with their similar dehydratase

activity measured experimentally. Following the rate determining formation of **8-enolate**, this intermediate shows only marginal stabilization over the corresponding transition state and evolves to **8-B4** in a barrierless step. The overall reaction is endergonic for both variants. However, unbinding of the product was calculated to be -7 kcal mol<sup>-1</sup> more favorable than the reactant, ensuring catalytic turnover. Substrate is shown in yellow, GS<sup>-</sup> in green and residues C/S13, V12 and the backbone of A11 in magenta. The protein backbone is shown as a semi-transparent magenta ribbon. Atoms represented in thick sticks were included in the high layer (QM level), while the rest were treated at the MM level. Water molecules are omitted for clarity.



**Extended Data Fig. 9 | MALDI analysis of oxidation of *PtGSTF8* variants.** **a**, Full MALDI mass spectrum of *PtGSTF8* wild-type and C13S variant after oxidation with 0 or 1 mM GSSG. **b**, Zoom into the area of the single charged ions in the MALDI mass spectra. Expected  $m/z$  for [E-Met+H]<sup>+</sup> (a), [E + H]<sup>+</sup> (b), [E-Met+GSH + H]<sup>+</sup> (c),

and [E + GSH + H]<sup>+</sup> (d) are indicated. **c**, Zoom into the area of the double charged ions in the MALDI mass spectra. Expected  $m/z$  for [E-Met+2H]<sup>2+</sup> (e), [E + 2H]<sup>2+</sup> (f), [E-Met+GSH + 2H]<sup>2+</sup> (g), and [E + GSH + 2H]<sup>2+</sup> (h) are indicated.

## Extended Data Table 1 | X-ray crystallographic data collection and refinement statistics

	<i>Pt</i> GSTF8 with GSH and (-)-catechin
<b>Data collection</b>	
Space group	C 1 2 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	89.95, 55.45, 54.79
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 113.55, 90.00
No. unique reflections	101092 (5007)*
Resolution (Å)	38.47-1.09 (1.11-1.09)
$R_{\text{meas}}$	0.055 (0.667)
$R_{\text{pim}}$	0.021 (0.267)
CC1/2	0.999 (0.817)
Mean(I) / sd(I)	15.7 (2.3)
Completeness (%)	99.2 (97.6)
Redundancy	6.4 (6.0)
Wilson B-factor (Å <sup>2</sup> )	9.9
<b>Refinement</b>	
Resolution (Å)	38.47-1.09
No. reflections all / free	100738 / 4917
$R_{\text{work}} / R_{\text{free}}$	0.133 / 0.149
No. atoms	
Protein	1911
Ligand/ion	41 / 5
Water	312
<i>B</i> -factors	
Protein	16.3
Ligand/ion	13.8 / 26.78
Water	31.55
R.m.s. deviations	
Bond lengths (Å)	0.0110
Bond angles (°)	1.727
Clashscore	3
Ramachandran favoured (%)	97
Ramachandran outliers	0

\*Values in parentheses are for highest-resolution shell.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                                       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

HPLC-MS: OpenLAB CDS ChemStation Edition (C.01.07 SR4)  
 High-resolution HPLC-MS: MassHunter Workstation Software (B.08.00)  
 Spectrometry: Sparkcontrol (2.3), UV Lab (4.1.0)  
 MALDI: Compass for flexSeries (1.4)  
 X-ray: Serial synchrotron crystallography (SSX) software suite (CY+ version git-7dabea6a7cf35)  
 QM/MM studies: GOLD 2020.3.0 (CCDC), Gaussian 16 (Gaussian, Inc) and Amber 20 (UCSF).

#### Data analysis

HPLC-MS: OpenLab CDS (2.4 Update\_06)  
 High-resolution HPLC-MS: MassHunter Workstation Software Qualitative Analysis (10.0)  
 Data processing and plotting: Microsoft Excel 365 (2108), Python (3.8.8), Pymol (2.5.2), Chemdraw (19.0.0.22), Adobe Illustrator 2022 (26.3.1), Affinity Photo (1.10.5)  
 Bioinformatics: R (4.1.2), Clustal Omega (1.2.2), Autodock Vina (1.2.0), PLIP (2.2.0), LigRMSD (1.0), ProtParam (Expasy, no version number available)  
 Structure elucidation: REFMAC (5.8.0352) and Coot (0.9.8.7) within ccp4i2 interface (v8.0.002), autoPROC (1.0.5), XDS (Version January 10, 2022), MolREP (11.9.02), MolProbity (4.5.1), CCP4 7.1.018: comit (0.1.0)  
 QM/MM studies: Python 3.8, Pymol 2.4 and Microsoft Excel 365.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Nucleotide sequences of codon-optimized genes can be found in the Supplementary Table 6. The diffraction images were deposited to Integrated Resource for Reproducibility in Macromolecular Crystallography88 (<http://proteindiffraction.org/>) and can be accessed using Protein Data Bank entry number 8AGQ. Crystallographic coordinates of the ternary complex of PtGSTF8 have been deposited in the Protein Data Bank as 8AGQ. The binary PtGSTF8 crystal structure used in molecular replacement experiments can be accessed via PDB ID: 5F07. Computed geometries and energies can be accessed through the Zenodo repository (<https://doi.org/10.5281/zenodo.8069429>). All source data are provided with this manuscript.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="n.a."/>
Population characteristics	<input type="text" value="n.a."/>
Recruitment	<input type="text" value="n.a."/>
Ethics oversight	<input type="text" value="n.a."/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="No sample size calculation was used. In pre-experiments, the experimental data was found to be reproducible and no outliers were observed, therefore 3 independent replicates were performed."/>
Data exclusions	<input type="text" value="No data were excluded from the analyses."/>
Replication	<input type="text" value="At least three independent replicates were performed and reported for each experiment. Most experiments were performed at least twice with almost identical results, showing a good reproducibility."/>
Randomization	<input type="text" value="This study does not involve human/animal subjects and group allocation, therefore randomization was not necessary. Nevertheless, appropriate control experiment were performed."/>
Blinding	<input type="text" value="This study does not involve human/animal subjects and group allocation, therefore blinding was not necessary. Nevertheless, appropriate control experiment were performed."/>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.



## Materials & experimental systems

- | n/a                                 | Included in the study                                  |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |

## Methods

- | n/a                                 | Included in the study                           |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |