



# Structural Basis of Stereospecificity in the Bacterial Enzymatic Cleavage of $\beta$ -Aryl Ether Bonds in Lignin<sup>\*[S]</sup>

Received for publication, September 24, 2015, and in revised form, December 2, 2015. Published, JBC Papers in Press, December 4, 2015, DOI 10.1074/jbc.M115.694307

Kate E. Helmich<sup>§§1</sup>, Jose Henrique Pereira<sup>¶¶1</sup>, Daniel L. Gall<sup>§§\*1</sup>, Richard A. Heins<sup>¶¶</sup>, Ryan P. McAndrew<sup>¶¶</sup>, Craig Bingman<sup>‡</sup>, Kai Deng<sup>¶¶</sup>, Keefe C. Holland<sup>¶¶</sup>, Daniel R. Noguera<sup>§§\*</sup>, Blake A. Simmons<sup>¶¶</sup>, Kenneth L. Sale<sup>¶¶</sup>, John Ralph<sup>§§</sup>, Timothy J. Donohue<sup>§§2</sup>, Paul D. Adams<sup>¶¶||3</sup>, and George N. Phillips Jr.<sup>|||4</sup>

From the <sup>‡</sup>Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706, the <sup>§</sup>United States Department of Energy Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, University of Wisconsin, Madison, Wisconsin 53726, the <sup>¶</sup>Joint BioEnergy Institute, Emeryville, California 94608, the <sup>||</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, the Departments of <sup>\*\*</sup>Civil and Environmental Engineering and <sup>§§</sup>Bacteriology, University of Wisconsin, Madison, Wisconsin 53706, the <sup>¶¶</sup>Biological and Engineering Sciences Center, Sandia National Laboratories, Livermore, California 94551, the <sup>¶¶</sup>Department of Bioengineering, University of California, Berkeley, California 94720, and the <sup>|||</sup>Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77251

Lignin is a combinatorial polymer comprising monoaromatic units that are linked via covalent bonds. Although lignin is a potential source of valuable aromatic chemicals, its recalcitrance to chemical or biological digestion presents major obstacles to both the production of second-generation biofuels and the generation of valuable coproducts from lignin's monoaromatic units. Degradation of lignin has been relatively well characterized in fungi, but it is less well understood in bacteria. A catabolic pathway for the enzymatic breakdown of aromatic oligomers linked via  $\beta$ -aryl ether bonds typically found in lignin has been reported in the bacterium *Sphingobium* sp. SYK-6. Here, we present x-ray crystal structures and biochemical characterization of the glutathione-dependent  $\beta$ -etherases, LigE and LigF, from this pathway. The crystal structures show that both enzymes belong to the canonical two-domain fold and glutathione binding site architecture of the glutathione *S*-transferase family. Mutagenesis of the conserved active site serine in both LigE and LigF shows that, whereas the enzymatic activity is reduced, this amino acid side chain is not absolutely essential for catalysis. The results include descriptions of cofactor binding sites, substrate binding sites, and catalytic mechanisms. Because  $\beta$ -aryl ether bonds account for 50–70% of all interunit linkages in lignin, understanding the mechanism of enzymatic  $\beta$ -aryl ether cleavage has significant potential for informing ongoing studies on the valorization of lignin.

The primary obstacle in the production of lignocellulosic biofuels is the release of sugars in high quantities at low cost from recalcitrant biomass feedstocks (1). Lignin is the prime source of this recalcitrance, and there has been renewed interest in the microbial enzymes capable of lignin degradation and catabolism of lignin-derived compounds (2, 3). Generally, white rot fungi secrete lignin peroxidases, versatile peroxidase, manganese peroxidases, and laccases that are involved in the initial degradation of lignin (4, 5), whereas bacteria are thought to play a role in further degradation of lignin-derived lower molecular weight compounds (6).

*Sphingobium* sp. strain SYK-6, one of the most well studied bacteria implicated in lignin-derived compound degradation, has the ability to grow on a wide variety of dimeric aromatic compounds representing the various units, with their characteristic interunit linkages, present in plant lignins (6, 7). The cleavage of  $\beta$ -aryl ether (termed simply  $\beta$ -ether hereafter) linkages is an essential step in any catabolic process for degradation of lignin-derived aromatic oligomers, because this bond type accounts for 50–70% of all interunit linkages in lignin polymers (8). Using a  $\beta$ -ether-linked phenolic lignin model substrate, guaiacylglycerol- $\beta$ -guaiacyl ether (GGE<sup>5</sup>; Fig. 1), three enzymatic reactions composing the  $\beta$ -ether degradation pathway were identified in *Sphingobium* sp. strain SYK-6 (7, 9, 10). Following oxidation of the  $\alpha$ -hydroxyl group in GGE by a  $\alpha$ -dehydrogenase, stereospecific glutathione (GSH)-dependent cleavage of the  $\beta$ -ether linkage in  $\beta$ -(3'-methoxyphenoxy)- $\gamma$ -hydroxypropiovanillone (MPPHV) is catalyzed by the glutathione *S*-transferase (GST) enzymes LigE and LigF, forming  $\beta$ -glutathionyl- $\gamma$ -hydroxypropiovanillone (GS-HPV) and guaiacol. LigE catalyzes stereospecific cleavage of ( $\beta$ R)-MPPHV to ( $\beta$ S)-GS-HPV, whereas LigF catalyzes the cleavage of ( $\beta$ S)-MPPHV to ( $\beta$ R)-GS-HPV. Finally, GSH-dependent and stereospecific elimination of GSH from ( $\beta$ S)-GS-HPV is catalyzed by the GST lyase LigG, generating glutathione disulfide (GSSG) and the

<sup>\*</sup> This work was supported in part by the United States Department of Energy (DOE) Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). This work was also supported by National Institutes of Health Protein Structure Initiative Grant GM098248 and National Institutes of Health Grant GM109456 (to G.N.P.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

[S] This article contains supplemental material.

The atomic coordinates and structure factors (codes 4XT0, 4YAM, and 4YAN) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> To whom correspondence may be addressed. E-mail: tdonohue@bact.wisc.edu.

<sup>3</sup> To whom correspondence may be addressed. E-mail: pdadams@lbl.gov.

<sup>4</sup> To whom correspondence may be addressed. E-mail: georgep@rice.edu.

<sup>5</sup> The abbreviations used are: GGE, guaiacylglycerol- $\beta$ -guaiacyl ether; MPPHV,  $\beta$ -(3'-methoxyphenoxy)- $\gamma$ -hydroxypropiovanillone; GS-HPV,  $\beta$ -glutathionyl- $\gamma$ -hydroxypropiovanillone; CPD, cysteine protease domain; FPPHV,  $\beta$ -(1'-formyl-3'-methoxyphenoxy)- $\gamma$ -hydroxypropioveratrone; F-FPPHV, fluoro-(1'-formyl-3'-methoxyphenoxy)- $\gamma$ -hydroxypropioveratrone.

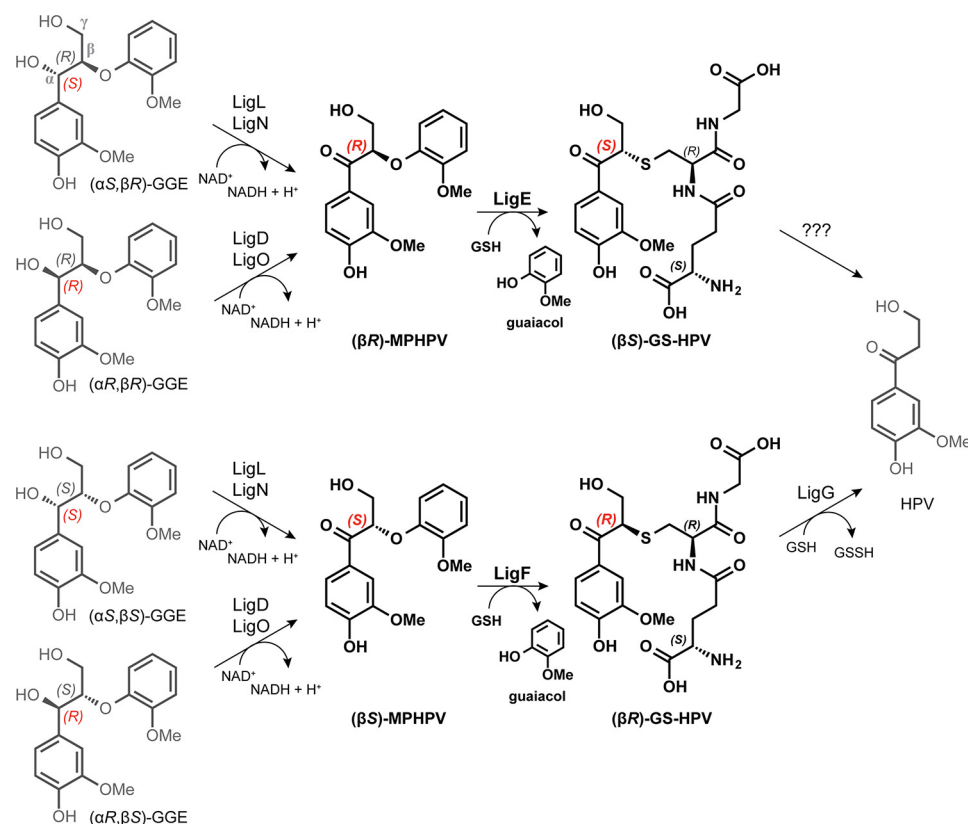


FIGURE 1. The *Sphingobium* sp. strain SYK-6  $\beta$ -etherase pathway. Chiral carbons at which stereospecific reactions occur are highlighted (red). Stereospecific reactions for ( $\alpha$ S, $\beta$ R)-GGE and ( $\alpha$ S, $\beta$ S)-GGE oxidation (by LigL and LigN) and for ( $\alpha$ R, $\beta$ R)-GGE and ( $\alpha$ R, $\beta$ S)-GGE oxidation (by LigD and LigO), the GSH-dependent stereospecific cleavage reactions of ( $\beta$ R)-MPPHV (by LigE) and ( $\beta$ S)-MPPHV (by LigF), and the stereospecific lyase reaction of LigG with ( $\beta$ S)-GS-HPV are shown.

achiral derivative  $\gamma$ -hydroxypropiovanillone, which ultimately serves as the growth substrate for strain SYK-6 (6, 10) (Fig. 1). The fate of the corresponding (*R*)-stereoisomer of GS-HPV is not presently understood. Recently, it has been reported that these GST family member enzymes have the ability to work with lignin-derived materials *in vitro* (11, 12).

GST superfamily members are multifunctional enzymes often involved in cellular detoxification processes via GSH conjugation (13). However, some bacterial GSTs are implicated in basal metabolism and supply bacterial cells with carbon (14). GSTs with >40% sequence identity are traditionally considered to be in the same class, whereas proteins of different classes have typically <25% protein sequence identity (15). However, these classifications are also based on a number of other considerations, including structure, function, and biochemical properties (15). Although there are seven classes of GSTs in mammals (Alpha, Mu, Pi, Sigma, Theta, Omega, and Zeta), there is an ever-increasing number of non-mammalian classes, including Beta, Chi, Delta, Epsilon, Lambda, Phi, and Tau, as well as a number of more recently defined novel classes (15–17). Previous studies have suggested that the  $\beta$ -etherase enzymes LigE and LigF might be classified in the fungal GST-FuA class of GSTs based on sequence phylogeny (18).

Because plant lignins are racemic polymers, complementary stereospecificities of the multiple enzymes in the  $\beta$ -ether degradation pathway are required by bacteria to oxidize and cleave the various stereoisomers that are present in lignin polymers

(19–22). Here, we describe three protein crystal structures and provide the corresponding biochemical data for the LigE and LigF enzymes involved in the  $\beta$ -ether cleavage step of the *Sphingobium* sp. strain SYK-6 degradation pathway. The modest structural homology of these two enzymes highlights the fitness adaptation afforded in this and probably other microbial catabolic pathways that can degrade lignin-derived materials, required for enzymatic degradation of such racemic products. This work provides new insights into the structure-function relationships and biochemistry of this pathway, expanding our knowledge of the bacterial catabolism of lignin-derived compounds. Because lignin is the most abundant aromatic polymer in nature, this study informs broader lignin valorization efforts that will ultimately enable the development of efficient pathways for the conversion of lignin into renewable aromatics with applications in advanced biofuels and chemicals (23).

## Experimental Procedures

**Gene Cloning**—LigE was synthesized and cloned into a custom vector (pCPD) assembled by GenScript (Piscataway, NJ). This vector combined the pVP16 backbone (provided by the Center for Eukaryotic Structural Genomics, Madison, WI) with the gene of interest and a C-terminal fusion protein tag containing the *Vibrio cholera* MARTX toxin cysteine protease domain (CPD) (24). During protein purification, the CPD tag can be activated by the addition of inositol hexakisphosphate, cleaving at a leucine positioned between the N terminus protein

TABLE 1

## LigE and LigF statistics

Summary of crystal parameters, data collection, and refinement statistics. Values in parentheses are for the highest resolution shell.

	LigFΔ242-GSH	LigEΔ255	LigEΔ255-GSH
<b>Crystal parameters</b>			
Space group	P6322	C2	C2
Unit cell parameters			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	123.71, 123.71, 66.42	122.55, 97.15, 131.38	121.00, 96.13, 126.16
$\beta$		106.65	81.52
<b>Data collection statistics</b>			
Wavelength (Å)	0.97857	0.999	1.000
Resolution range (Å)	50.00–2.07 (2.11–2.07)	50–1.90 (1.93–1.90)	50–2.6 (2.65–2.60)
No. of reflections (measured/unique)	326,246/18,884	575,459/114,153	160,219/42,163
Completeness (%)	99.8 (98.3)	99.2 (99.9)	97.7 (93.8)
$R_{\text{merge}}^a$	0.076 (0.637)	0.137 (0.63)	0.135 (0.61)
Redundancy	17.3 (12.9)	5.0 (4.9)	3.8 (3.6)
Mean $I/\sigma(I)$	9.1 (3.8)	9.0 (1.6)	6.1 (2.5)
<b>Refinement and model statistics</b>			
Resolution range (Å)	40.49–2.07 (2.17–2.07)	48–1.90 (1.95–1.90)	48–2.60 (2.65–2.60)
No. of reflections (work/test)	17,278/964	114,138/1,999	42,158/2,000
$R_{\text{cryst}}^b$	0.161 (0.182)	0.227 (0.289)	0.222 (0.266)
$R_{\text{free}}^c$	0.214 (0.263)	0.271 (0.350)	0.267 (0.290)
Root mean square deviation bonds (Å)	0.008	0.004	0.003
Root mean square deviation angles (degrees)	1.022	0.956	0.776
<i>B</i> factor (protein/solvent) (Å <sup>2</sup> )	39.58/43.93	29.13/37.54	34.22/27.38
<i>B</i> factor (GSH) (Å <sup>2</sup> )	26		146, 148, 149, 146
No. of protein atoms	1,983	9,405	8,491
No. of waters	229	1,159	165
Auxiliary molecules (real space correlation coefficient (CC))	1 glutathione (0.97), 1 Tris (0.95), 1 PEG (0.95)		4 glutathione (1 per chain, <i>A</i> = 0.71, <i>B</i> = 0.58, <i>C</i> = 0.58, <i>D</i> = 0.61)
<b>Ramachandran plot</b>			
Favorable region	98.4	95.8	94.4
Additional allowed region	1.6	3.0	4.3
Disallowed region	0	1.2	1.3
Protein Data Bank entry	4XT0	4YAM	4YAN

<sup>a</sup>  $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$ , where  $I_i(h)$  is the intensity of an individual measurement of the reflection, and  $\langle I(h) \rangle$  is the mean intensity of the reflection.<sup>b</sup>  $R_{\text{cryst}} = \sum_h |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_h |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes, respectively.<sup>c</sup>  $R_{\text{free}}$  was calculated as  $R_{\text{cryst}}$  using 5.0% of randomly selected unique reflections that were omitted from the structure refinement.

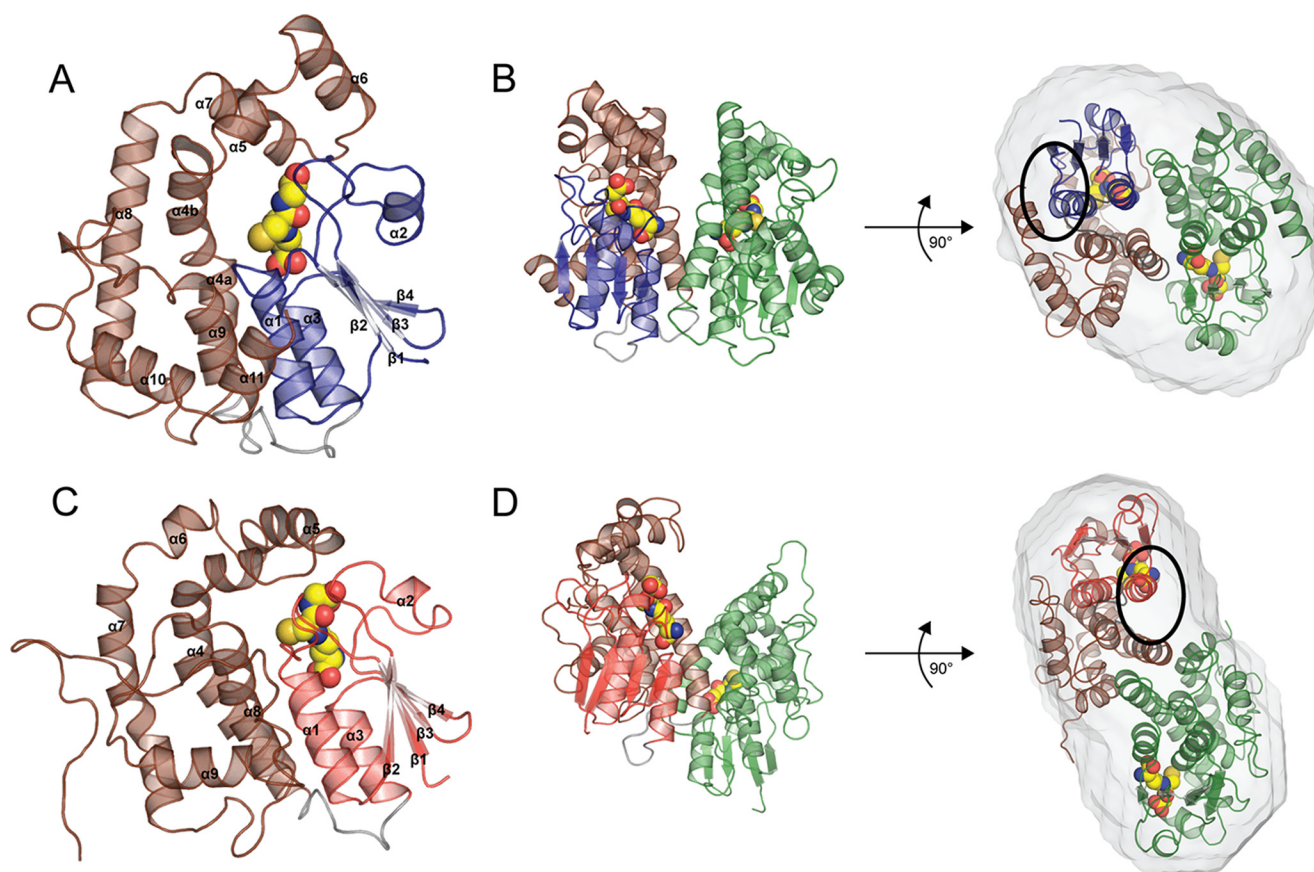
of interest and CPD. The pVP80K\_LigFΔ242 vector was prepared using polymerase incomplete primer extension as described previously using Phusion High-Fidelity PCR master mix with HF buffer (New England Biolabs Inc., Ipswich, MA), and primers from Integrated DNA Technologies (Coralville, IA) (25). The pVP80K vector was provided by the Center for Eukaryotic Structural Genomics (Madison, WI), and the pVP102KSSLigF vector containing full-length wild type LigF was prepared as described previously (9). Insert and vector backbone PCR products were mixed 1:1 and immediately transformed into *Escherichia coli* One Shot® TOP10 cells (Invitrogen). The pVP80K\_LigFΔ242 vector was purified from *E. coli* (One Shot® TOP10, 10 ml of LB with kanamycin, 18 h at 37 °C) using the QIAprep® spin miniprep kit (Qiagen, Germantown, MD) and transformed into the laboratory strain *E. coli* B834(DE3) Z-competent cells (Zymo Research, Orange, CA).

**Enzyme Expression and Purification**—NEB Express protein expression cells (New England Biolabs Inc., Ipswich, MA) containing pCPD-LigE were grown in autoinducing selenomethionine medium as described previously (26) and harvested via centrifugation. Harvested cells were resuspended in 30 ml of lysis buffer (50 mM HEPES buffer, pH 7.4, 150 mM NaCl, and 40 mM imidazole) and lysed by an Avestin EmulsiFlex-C3 homogenizer. The C-terminally His-tagged proteins were purified from the clarified supernatant using precharged nickel-IMAC resin (GE Healthcare). After protein binding and washing twice with lysis buffer, inositol hexakisphosphate was added to a final concentration of 200 μM. Note that the inositol hexakisphos-

phate was first diluted to 10 mM in lysis buffer to neutralize the acidic pH of the stock solution. After 1 h of incubation, the resin was washed with 1 ml of lysis buffer to elute the cleaved protein. Following buffer exchange into 20 mM Tris, pH 8, the LigE protein was further purified using a HiTrap Q HP anion exchange column. Fractions containing LigE, as confirmed by SDS-PAGE, were pooled and concentrated. Final protein cleanup was done using gel filtration on a Superdex 200 10/300 GL column (GE Healthcare).

Laboratory strain *E. coli* B834(DE3) Z-competent cells (Zymo Research, Orange, CA) containing the pVP80K\_LigFΔ242 plasmid were grown in autoinducing selenomethionine medium as described previously (26) and harvested via centrifugation. Harvested cells were resuspended in 20 ml of lysis buffer (20 mM sodium phosphate buffer, pH 7.5, 500 mM sodium chloride, 20% ethylene glycol) and lysed by sonication. The N-terminally His-tagged LigFΔ242 fusion protein was purified from the supernatant by immobilized nickel affinity chromatography using a HiTrap Q HP anion exchange column on an ÄKTA FPLC system (GE Healthcare, Piscataway, NJ). Fractions containing LigFΔ242, as determined by SDS-PAGE, were combined and dialyzed overnight at 4 °C. LigFΔ242 was cleaved from the fusion protein using tobacco etch virus protease (1 mg/100 mg of protein; provided by the Center for Eukaryotic Structural Genomics). Following cleavage, LigFΔ242 and the polyhistidine tag were separated using a HiTrap Q HP anion exchange column. Pooled fractions containing LigFΔ242, as confirmed by SDS-PAGE, were pooled and concentrated to 3





**FIGURE 2. LigE and LigF structures.** *A*, schematic representation of LigF, including the N-terminal thioredoxin domain (blue), the C-terminal  $\alpha$ -helical domain (brown), and the short linker (gray). Bound GSH is shown as yellow spheres. *B*, schematic representation of the LigF dimer with the proposed substrate binding site (Fig. 5A) circled. *C*, schematic representation of LigE, including the N-terminal thioredoxin domain (red), the C-terminal  $\alpha$ -helical domain (brown), and the short linker (gray). Bound GSH is shown as yellow spheres. *D*, schematic representation of the dimer of LigE with the proposed binding site (Fig. 5B) circled.

ml. Final size exclusion purification was performed on a HiLoad™ 26/60 Supradex™ 200 preparation grade column.

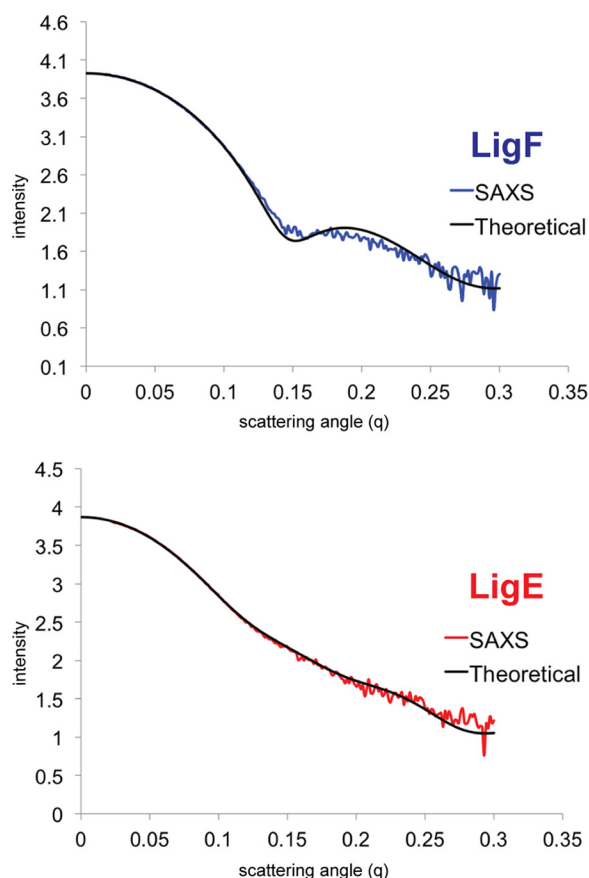
**Enzyme Kinetic Assays**—*In vitro*  $\beta$ -etherase assays with LigE, LigF, LigF $\Delta$ 242, and LigF $\Delta$ 242-S13A were conducted in an aqueous assay buffer (25 mM Tris, 2.5% DMSO, 5 mM GSH, pH 7.0–10.0) at 30 °C with an initial substrate concentration of 1.5 mM and enzyme concentrations of either 160 nM (LigE), 170 nM (LigF), 180 nM (LigF $\Delta$ 242), 3.9  $\mu$ M (LigF $\Delta$ 242-S13A), 11.2  $\mu$ M (LigE with ( $\beta$ S)-F-FPHPV), or 12.0  $\mu$ M (LigF with ( $\beta$ R)-F-FPHPV). Enantiopure preparations of ( $\beta$ R)-F-FPHPV and ( $\beta$ S)-F-FPHPV were obtained from chiral chromatographic separation of the parent racemate as described previously (9). Similarly, chiral chromatography was used for the separation of ( $\beta$ S)-F-FPHPV and ( $\beta$ R)-F-FPHPV, with ( $\beta$ S)-F-FPHPV being used as a substrate in the LigE assays. Synthesis and purification details for enzymatic substrates FPHPV and F-FPHPV are described in the supplemental material.

Michaelis-Menten curves were generated by measuring the enzymatic specific activities over a range of initial substrate concentrations (1.50, 1.25, 1.00, 0.75, 0.50, and 0.25 mM) obtained from serial dilution of a 1.5 mM substrate buffer made immediately prior to conducting the assays. The 1-ml assays were conducted in triplicate and were managed as follows: 1) the substrate was dissolved in DMSO at 60 mM, and 25  $\mu$ l were added to a 2-ml vial; 2) 875  $\mu$ l of 25.7 mM Tris, pH *X*, was added

(where *X* is higher than the intended pH of the assay to account for the acidic effect of GSH (e.g. pH *X* = 11.5 drops to pH 8.0 after the addition of 5 mM GSH); 3) 50  $\mu$ l of 100 mM GSH was added (100 mM GSH stock solution was prepared by adding GSH to 25 mM Tris (pH *X*)); 4) 50  $\mu$ l of 20 $\times$  concentrated enzyme was added; 5) 150- $\mu$ l samples were collected after 0, 6, 12, 18, 24, and 30 s of incubation, and enzymatic activity was abolished by pipetting each sample into 5  $\mu$ l of 5 M phosphoric acid; and 6) the remaining reaction volume was used to measure the pH of the mixture with pH paper.

Each sample was then subjected to  $C_{18}$ -reversed phase HPLC using a Beckman 125NM solvent delivery module equipped with a Beckman 168 UV detector. Samples and external standards were quantified by UV absorption at 280 nm. The HPLC mobile phase was a mixture of aqueous buffer (5 mM formic acid in 95:5 water/acetonitrile) and methanol at a flow rate of 1.0 ml/min. The ratio of buffers was adjusted as follows: 0–6 min, 30% methanol; 6–15 min, gradient from 30 to 80% methanol; 15–25 min, 80% methanol; 25–26 min, gradient from 80 to 30% methanol; 26–33 min, 30% methanol. Vanillin concentrations were quantified for each time point, and a linear regression was generated over the 30-s assay period in order to calculate the specific activity of each reaction. Averages of the triplicate assays were reported.

## Sphingobium LigE and LigF Crystal Structures



**FIGURE 3. Theoretical and experimental small angle x-ray scattering scatter curves.** The scattering angle ( $q$ ) versus the intensity of the scattering plots shows the experimentally observed data and the theoretical scattering determined using CRY SOL from the x-ray structures of the dimers. LigF is shown at the top in blue, and LigE is shown at the bottom in red.

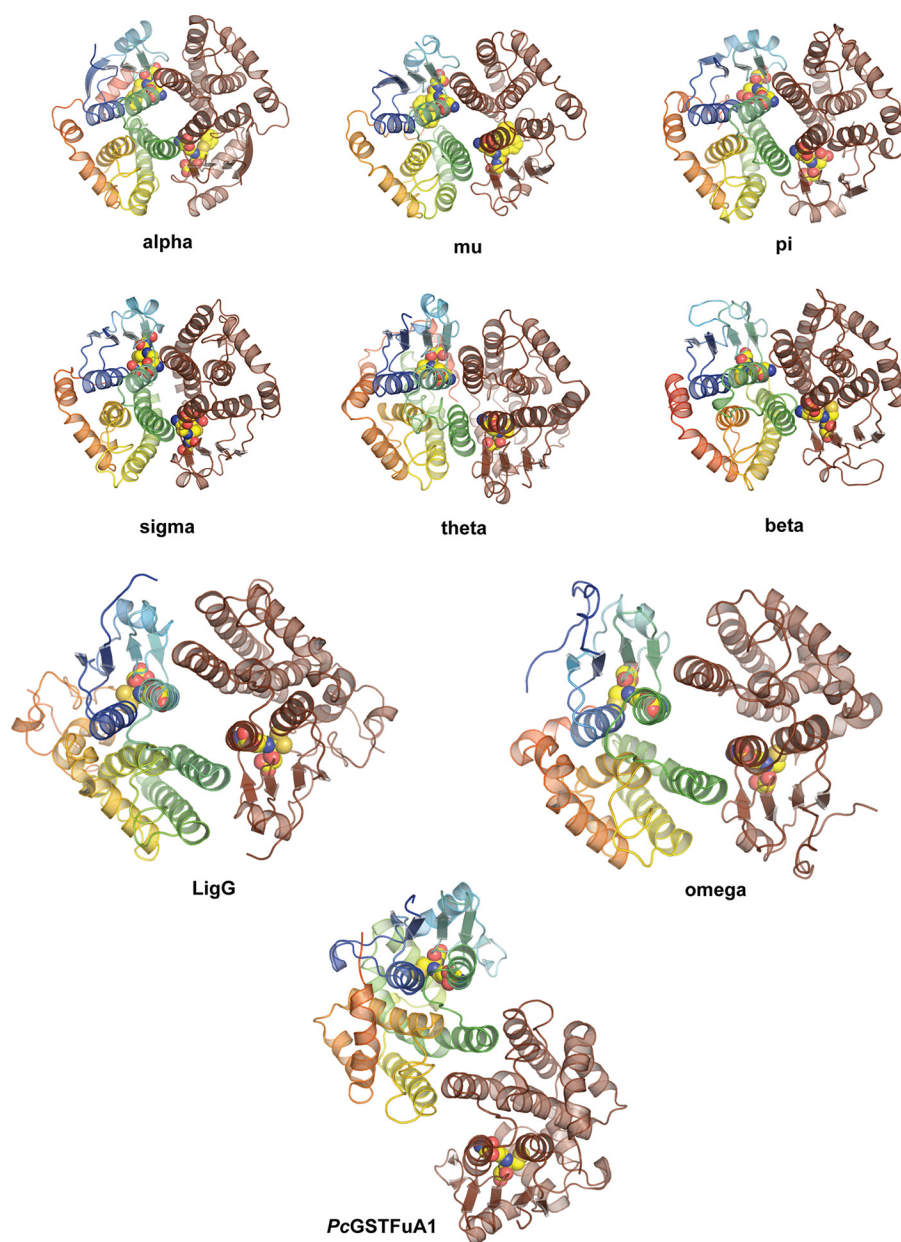
**Crystallization**—LigE was concentrated to  $9 \text{ mg ml}^{-1}$  and dialyzed against  $20 \text{ mM}$  HEPES, pH 7.4, and  $50 \text{ mM}$  NaCl. LigF was dialyzed in  $10 \text{ mM}$  HEPES buffer, pH 7.5, containing  $50 \text{ mM}$  sodium chloride,  $0.5 \text{ mM}$  tris(2-carboxyethyl)phosphine, and  $1 \text{ mM}$  GSH, and concentrated to  $18.5 \text{ mg ml}^{-1}$ . LigE and LigF proteins were screened using the sparse matrix method (27) with a Phoenix Robot (Art Robbins Instruments, Sunnyvale, CA) and a Mosquito dispenser (TTP LabTech, Melbourn, UK) utilizing the following crystallization screens: Berkeley Screen (Lawrence Berkeley National Laboratory), Crystal Screen, SaltRx, PEG/Ion, Index and PEGRx (Hampton Research, Aliso Viejo, CA), and JSCG-plus HT-96 and PACT premier HT-96 (Molecular Dimensions, Altamonte Springs, FL). The optimum conditions for crystallization of the different pathway proteins were found as follows: LigE,  $0.1 \text{ M}$  ammonium citrate,  $0.1 \text{ M}$  MES, pH 5.5, 20% PEG 3,350, and 5% isopropyl alcohol; LigF, 25% polyethylene glycol monomethyl ether 2000,  $0.25 \text{ M}$  trimethylamine *N*-oxide, and  $0.1 \text{ M}$  Tris, pH 8.5. LigE crystals were obtained after 2–7 days by the sitting drop vapor diffusion method with the drops consisting of a mixture of  $0.2 \mu\text{l}$  of protein solution and  $0.2 \mu\text{l}$  of reservoir solution. LigF crystals were obtained in  $<24 \text{ h}$  with drops containing a mixture of  $1 \mu\text{l}$  of protein solution,  $0.8 \mu\text{l}$  of reservoir solution, and  $0.2 \mu\text{l}$  of seed crystals (pulverized LigF $\Delta$ 242 crystals in  $0.2 \text{ M}$  magnesium formate, 30% polyethylene glycol 3350, and  $1 \text{ mM}$  GSH).

**X-ray Data Collection and Structure Determination**—The LigE crystals were placed in a reservoir solution containing 10–20% (v/v) glycerol and then flash-cooled in liquid nitrogen. The x-ray data sets for LigE were collected at the Berkeley Center for Structural Biology beamlines 8.2.1 and 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory. LigF crystals were cryoprotected with a reservoir solution containing 30% polyethylene glycol monomethyl ether 2000 and  $1 \text{ mM}$  GSH. X-ray diffraction data were collected at Life Sciences Collaborative Access Team Sector 21 with x-ray wavelength  $0.9793$  at the Advanced Photon Source at Argonne National Laboratory. Data sets were indexed and scaled using HKL2000 (28). The LigF crystal structure was determined by molecular replacement using the program PHASER (29) within the Phenix suite (30) with the coordinates of a LigF homologue (Lig37), whose sequence was identified from a metagenomic analysis of a rice-straw-enriched compost microbial community (Berkeley, CA) (31, 32). The crystal structure of LigE was solved using selenomethionine-labeled protein by the single-wavelength anomalous dispersion method (33) with the *phenix.autosol* (34) and *phenix.autobuild* (35) programs. Structure refinement was performed using the *phenix.refine* program (36). Manual rebuilding using COOT (37) and the addition of water molecules allowed construction of the final models. Root mean square deviation differences from ideal geometries for bond lengths, angles, and dihedrals were calculated with Phenix (30). The overall stereochemical quality of all final models was assessed using the program MOLPROBITY (38), and all figures were generated in PyMOL (39). Structures were observed and analyzed using a stereoscopic television display (40).

**Small Angle X-ray Scattering**—LigE and LigF were dialyzed for 15 h at  $4^\circ\text{C}$  into buffer containing  $10 \text{ mM}$  HEPES, pH 7.5,  $50 \text{ mM}$  sodium chloride,  $1 \text{ mM}$  GSH, and  $0.5 \text{ mM}$  tris(2-carboxyethyl)phosphine. Prior to data collection, samples were filtered through a  $0.2\text{-}\mu\text{m}$  syringe filter and diluted to the working concentrations. After dilution, samples were clarified via centrifugation. The buffer blank was also syringe-filtered and clarified by centrifugation. Small angle scattering data were collected on a Bruker NANOSTAR x-ray generator located at the National Magnetic Resonance Facility at the University of Wisconsin (Madison, WI). Three data collections of 1 h each were taken for each sample and buffer. Data were merged and indexed using the Bruker NANOSTAR small angle x-ray scattering system software (Bruker AXS, Madison, WI). The scattering intensity was obtained by subtracting the scattering of the buffer blank from the sample scattering using the PRIMUS software (41). All SAXS data were processed using GNOM, integrated in the PRIMUS software, to obtain the pair distance distribution function (42). The GNOM output was used with DAMMIF to calculate 10 *ab initio* dummy atom models (43). Models were averaged using DAMAVER and aligned to x-ray crystal structures using SUPCOMB (44, 45). Theoretical scattering curves for the x-ray crystal structure of LigE and a model of the dimer of LigF were calculated using CRY SOL (46).

**Molecular Docking**—Docking of MHPV to the LigF $\Delta$ 242-GSH structure was performed using the SwissDock server (47, 48). Docking was performed using the “Accurate” parameter





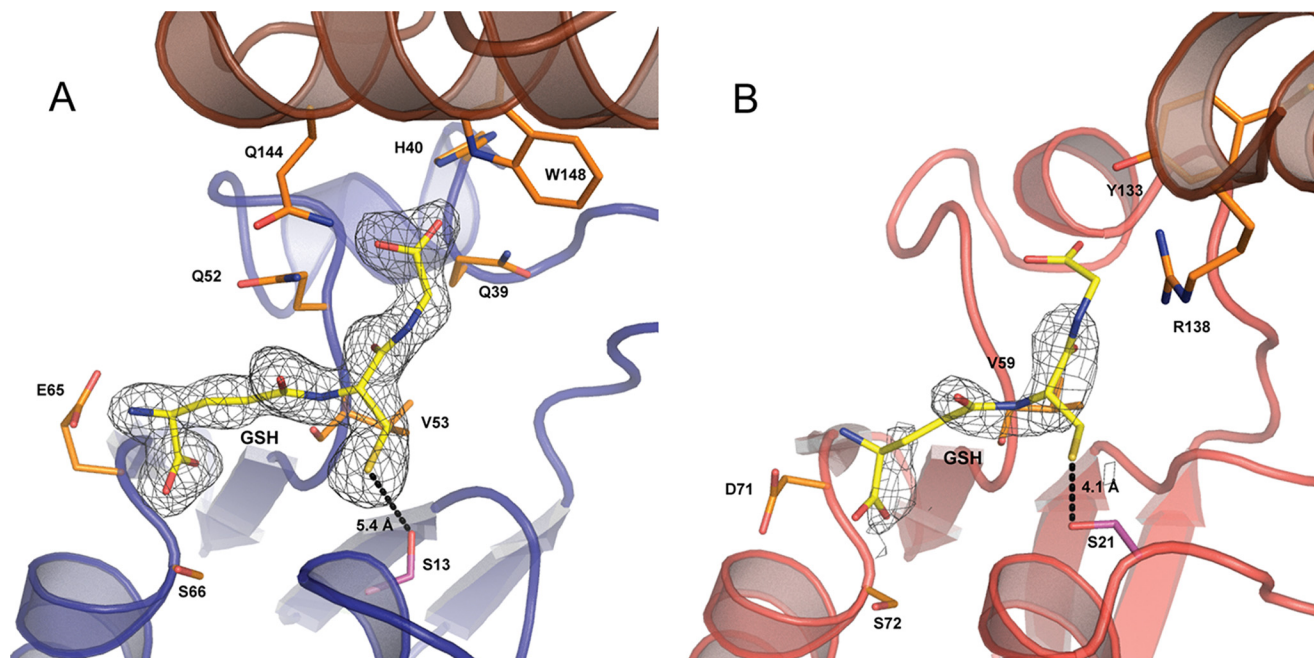
**FIGURE 4. Representative cytosolic GST dimer forms.** Representatives from several GST classes are shown, in which one molecule of the dimer is shown in brown, the second molecule of the dimer is shown in rainbow colors (N terminus in blue to C terminus in red), and the bound glutathione or glutathione analog is shown as yellow spheres. The Alpha (Protein Data Bank entry 1GUH; human GST A1-1), Mu (2GST; rat), Pi (2GSR; pGST P1-1 from pig), Sigma (1GSQ; squid), Theta (1LJR; human hGST T2-2), Beta (2PMT; bacterial GST from *P. mirabilis*), Omega (3LFL; human GST Omega-1), and LigG (4G10; *Spingobium* sp. SYK-6) dimers show variations on the  $\alpha 3/\alpha 4$  canonical four-helix bundle dimer structure, whereas the GSTFuA structure from *P. chrysosporium* shows a non-canonical dimer formed via interaction between  $\alpha 4$  and the C-terminal domain of the second molecule of the dimer.

and otherwise default parameters, with the search space limited to a  $10 \times 10 \times 10 \text{ \AA}$  region around the GSH binding. Both the protein and the MPHPV ligand were rigid during docking. The structure of MPHPV was built in ChemDraw (49), converted to three-dimensional coordinates using OpenBable (50). Docking results were visualized and screened using the UCSF Chimera molecular modeling system (51).

## Results

**Structural Analysis**—Attempts to solve the structure of full-length wild-type LigE (282 residues) and LigF (254 residues) were unsuccessful, but C-terminal truncation constructs of

both proteins were generated, successfully crystallized, and used for structural analysis. Truncations of LigE and LigF were designed based on homology models generated by I-TASSER Online and disorder predictions generated using PONDR (52, 53). LigE $\Delta 255$  and the LigE $\Delta 255$ -GSH complex crystallized in the space group C2 with four molecules in the asymmetric unit with electron density for the bound GSH molecule. LigF $\Delta 242$ -GSH crystallized in the space group P6<sub>3</sub>22 with one molecule in the asymmetric unit. Well defined electron density corresponding to the GSH molecule is also visible in the structure. Data collection, refinement, and model statistics for LigE and LigF are summarized in Table 1.



**FIGURE 5. Glutathione binding sites in LigF and LigE.** A, the GSH binding site in LigF is located in a cleft between the thioredoxin and  $\alpha$ -helical domains. Density for the bound GSH (yellow sticks) is shown in gray contoured to  $1.0\sigma$  (CC = 0.97). Residues interacting with the  $\gamma$ -glutamyl (Gln-52 and Val-53), and glycine (Gln-144, His-40, Trp-148, and Gln-39) residues of the bound GSH are shown as orange sticks. The distance between the GSH sulfur and the active site serine 13 (purple sticks) is 5.4 Å. B, the GSH binding in LigE is located on a surface-exposed face between the thioredoxin and  $\alpha$ -helical domains. Density for the bound GSH (yellow sticks) from chain A of the model is shown in gray contoured to  $1.0\sigma$  (CC = 0.71). Residues interacting with the  $\gamma$ -glutamyl (Asp-71 and Ser-72), cysteinyl (Val-59), and glycine (Arg-138 and Tyr-133) of the GSH are shown as orange sticks. The distance between the GSH sulfur and the active site serine 21 (purple sticks) is 4.1 Å.

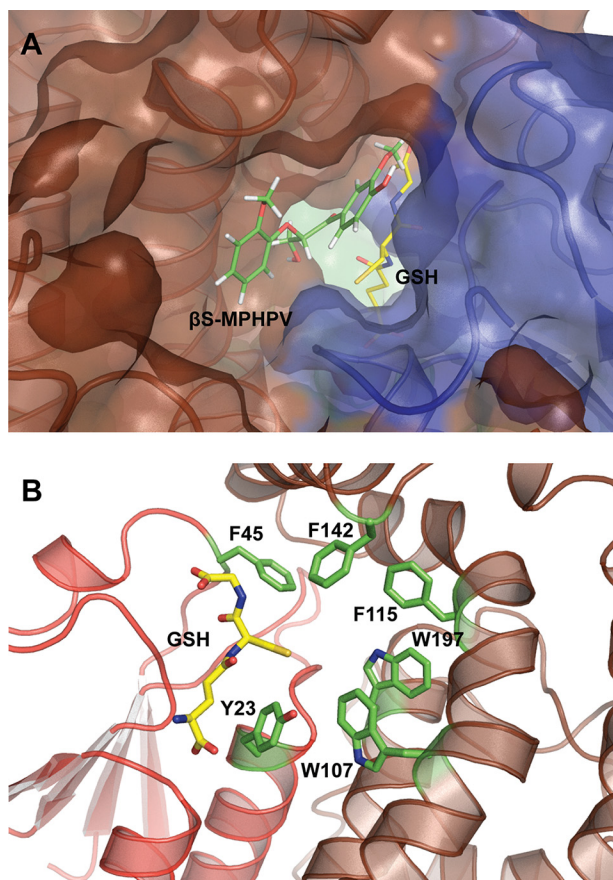
Consistent with their classification as GST enzymes, LigE and LigF each adopt the canonical GST domain fold with an N-terminal thioredoxin domain (residues 1–82 and 1–76, respectively) and a C-terminal  $\alpha$ -helical domain (residues 93–255 and 93–242, respectively) connected by a short linker (residues 83–92 and 77–92, respectively) (Fig. 2). In both LigE and LigF, the thioredoxin domain comprises four  $\beta$ -strands and three  $\alpha$ -helices following the topology  $\beta 1\alpha 1\beta 2\alpha 2\beta 3\beta 4\alpha 3$ . The loop between  $\beta 1$  and  $\alpha 1$  is longer in LigE than in LigF and occupies the space between the thioredoxin domain and the  $\alpha$ -helical domain, whereas in LigF, this loop is moved away from the domain interface toward the surface of the thioredoxin domain. The loop between  $\beta 2$  and  $\alpha 2$  is longer in LigF than in LigE, but both interact with the  $\alpha$ -helical domain on the protein face opposite the linker (Fig. 2). The C-terminal domains of both LigE and LigF are composed of six and eight  $\alpha$ -helices, respectively. The root mean square deviation between the C- $\alpha$  locations of monomers of LigE and LigF is 4.42 Å, indicating that, although they catalyze very similar reactions, the enzymes display significant structural differences.

Biochemical and small angle x-ray scattering data suggest that both LigE and LigF exist as dimers in solution, and these dimers, related by 2-fold symmetry, can be seen in the respective crystal structures. The dimer interface accounts for 1,066 and 1,092 Å<sup>2</sup> of buried surface area in LigE and LigF, respectively (PISA European Bioinformatics Institute) (54). The overall dimeric shapes of both LigF and LigE were confirmed using small angle x-ray scattering on both the truncated and full-length proteins. The protein envelopes determined by *ab initio* modeling align well with the crystal structures of both proteins

(Fig. 2). The theoretical scattering curves predicted from the x-ray structures match well with the experimentally determined scattering curves with a  $\chi$  value of 2.4 and 1.4 for LigF and LigE, respectively (Fig. 3).

The LigF dimer forms via interactions between helices  $\alpha 3$  and  $\alpha 4$ , in the thioredoxin and C-terminal domains, respectively, of each monomer, forming a four-helix bundle. The dimer interface is largely polar, lacking the traditional lock-and-key motif or hydrophobic surface common in other GST dimers, specifically the Alpha, Pi, and Mu classes (14, 15). The LigF dimer more closely matches those of the Beta or Theta class, which, like LigF, lack a hydrophobic lock-and-key motif, and there is no open V-shape to the dimer interface (14). Although the arrangement and characterization of the dimer forms in GST structures differ within and between classes, most are canonically anchored through contacts between  $\alpha 3$  (or the final helix of the thioredoxin domain) and  $\alpha 4$  (the first helix in the C-terminal domain) (13, 15, 55, 56). Variability in the arrangement of secondary structural elements away from the  $\alpha 3/\alpha 4$  four-helix bundle changes the total buried surface area of the various GST dimers as well as changing the architecture of the enzyme in the vicinity of the active site (57). Representative structures demonstrating the variability of dimer packing in GSTs are shown in Fig. 4. The Alpha (Protein Data Bank entry 1GUH, human GST A1-1), Mu (2GST, rat), Pi (2GSR, pGST P1-1 from pig), Sigma (1GSQ, squid), Theta (1LJR, human hGST T2-2), Beta (2PMT, bacterial GST from *Proteus mirabilis*), Omega (3LFL, human GST Omega-1), and LigG (4G10, *Sphingobium* sp. SYK-6) dimers show variations on the  $\alpha 3/\alpha 4$  canonical 4-helix bundle dimer structure (58–65). In the LigE





**FIGURE 6. Substrate binding sites in LigF and LigE.** A, model of ternary complex LigFΔ242-GSH-(βS)-MPPHV. Schematic representations are shown of the N-terminal thioredoxin domain (blue) and the C-terminal α-helical domain (brown) with the circled region from Fig. 2B detailed in a transparent surface rendering. The bound glutathione (yellow) and docked (βS)-MPPHV (green) are shown as sticks. B, proposed substrate binding surface in LigE. Schematic representations are shown for the LigE dimer, and the circled region from Fig. 2D is detailed, showing the hydrophobic aromatic substrate binding pocket formed by Phe-45, Phe-142, Phe-115, Trp-197, Trp-107, and Tyr-23 as green sticks.

dimer, helix  $\alpha 4$  of one monomer is interdigitated between  $\alpha 4$  and  $\alpha 7$  of the other monomer, and the entire dimer interface is contained within the  $\alpha$ -helical domain. The dimer is anchored by a hydrophobic lock-and-key motif in which Phe-101 of each monomer is in a hydrophobic pocket formed in the second monomer. This motif is seen in several GST classes, including Alpha, Mu, and Phi, which display the more typical four-helix bundle dimer mode rather than the elongated dimer of LigE (57). This elongated atypical dimer form in a GST was first described for GSTFuA1 from *Phanerochaete chrysosporium* (Fig. 4) in the GSTFuA class of GSTs, of which LigE is a member (13, 18). An additional  $\beta$ -hairpin motif between  $\alpha 2$  and  $\beta 3$  in GST5118 hinders the formation of the regular  $\alpha 3/\alpha 4$  GST dimer; however, this  $\beta$ -hairpin is not present in LigE (13). An extended loop between  $\alpha 5$  and  $\alpha 6$  in the C-terminal domain, which protrudes above the normal  $\alpha 3/\alpha 4$  packing site, may be responsible for the alternate dimer formation in LigE. LigG, a known GST lyase in the *Sphingobium*  $\beta$ -ether degradation pathway, is an Omega class GST with the canonical  $\alpha 3/\alpha 4$  GST dimer with a wider opening in the C-terminal domain, allowing for a pro-

posed substrate binding site on the same face as that in LigE (65).

The enzymatic active sites of these GST family members are often located in a cleft between the thioredoxin domain and the  $\alpha$ -helical domain (15). Both the LigE and LigF enzymes contain the  $\beta\beta\alpha$  motif required for anchoring GSH in the active site (56). In LigF, Glu-65 and Ser-66 located in the turn connecting  $\beta 4$  and  $\alpha 3$ , recognize the  $\gamma$ -glutamyl moiety of GSH as part of the  $\beta\beta\alpha$  motif (Fig. 5A). Additionally, Gln-52 and the backbone of Val-53 interact with the cysteinyl moiety, whereas Gln-144, His-40, Tyr-148, and Gln-39 anchor the glycine residue of the active site GSH molecule. In LigE, Asp-71 and Ser-72, both located in the turn between  $\beta 4$  and  $\alpha 3$ , hydrogen-bond with the amino and carboxylate groups, respectively, of the  $\gamma$ -glutamyl residue of the GSH molecule (Fig. 5B). Additionally, the backbone of Val-59 interacts with the cysteinyl moiety, whereas weak hydrogen bonds are formed between the GSH glycine and Arg-138 and Tyr-133.

Due to the occlusion of one face of the GSH binding pocket in LigF, we propose that the substrate binding site is located on the opposite face of the LigF monomer from the dimer interface (Fig. 2B, black circle). In the absence of a substrate-bound structure, SwissDock (47, 48) was used to generate a LigFΔ242-GSH-(βS)-MPPHV complex model (Fig. 6A) from the LigFΔ242-GSH structure and a molecular model of (βS)-MPPHV. The model supports our assignment of the substrate binding site. However, in LigE, this side of the GSH binding pocket is blocked by a number of loops, whereas the face of the GSH binding site shared with the dimer has been opened, due to the dimer rearrangement (Fig. 2D, black circle). Based on the binding site of GSH in LigE, we propose a potential location for the native substrate-binding site at the highly hydrophobic region consisting of residues Tyr<sup>23</sup>, Phe<sup>45</sup>, Trp<sup>107</sup>, Phe<sup>115</sup>, Phe<sup>142</sup>, and Trp<sup>197</sup> (Fig. 6B). The aromatic rings of these hydrophobic residues are probably important in stacking interactions with the aromatic compounds from low molecular weight lignin derivative compounds.

The LigEΔ255-GSH and LigFΔ242-GSH structures revealed LigE Ser-21 (Fig. 5B) and LigF Ser-13 (Fig. 5A) as potential catalytic residues, based on their proximities to the thiol of the bound GSH. To further investigate the roles of LigE Ser-21 and LigF Ser-13 in  $\beta$ -etherase catalysis, variants LigE-S21A and LigFΔ242-S13A, in which serine residues were replaced with alanine, were expressed, purified, and tested for activity in the  $\beta$ -etherase assays.

**Enzymatic Analysis and Mutagenesis**—To analyze the enzymatic activities of the GSH-dependent  $\beta$ -etherase enzymes, FPHPV degradation rates were measured by the accumulation of vanillin, a monoaromatic product of FPHPV cleavage (Figs. 7 and 8). Whereas  $\beta$ -etherase catalysis with MPPHV results in the release of guaiacol (Fig. 1), vanillin is more easily detected by UV absorption, thus improving the sensitivity of the assays. In addition to LigE and LigF, we tested the rates of  $\beta$ -etherase catalysis for LigE variant LigE-S21A and two LigF variants, LigFΔ242 and LigFΔ242-S13A.

We found that LigE catalysis resulted in stereospecific ( $\beta R$ )-FPHPV cleavage, whereas LigF selectively degraded the ( $\beta S$ )-FPHPV enantiomer, as is consistent with previous reports (7,



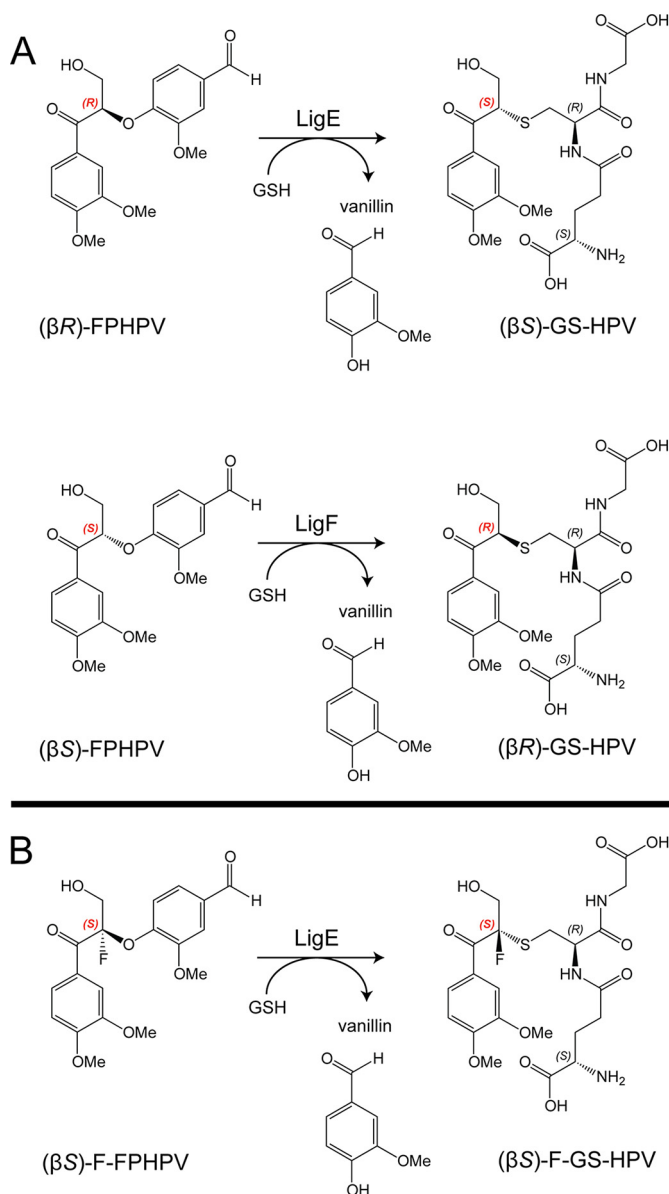


FIGURE 7. A, structure of an MPHPV analog substrate, FPHPV, that was used in the LigE- and LigF-catalyzed reactions, converting FPHPV to vanillin and GS-HPV. B, LigE-catalyzed  $\beta$ -ether elimination reaction with fluorinated model substrate (*S*)-F-FPHPV, resulting in formation of vanillin and (*S*)-F-GS-HPV.

9). The effect of pH on  $\beta$ -etherase activities was determined for each enzyme, revealing that LigE, LigF, LigF $\Delta$ 242, and LigF $\Delta$ 242-S13A have pH optima at pH 8.0 (Fig. 8). The activity of LigE was relatively unaffected by pH, whereas the activity of LigF $\Delta$ 242 and LigF $\Delta$ 242-S13A was significantly reduced above pH 8.0. The truncated LigF $\Delta$ 242 exhibited higher rates of catalysis than full-length LigF at all pH values, indicating that the predicted disordered region in the C terminus may actually be inhibitory to  $\beta$ -etherase activity. The specific activities of LigE-S21A and LigF $\Delta$ 242-S13A were 14% and <5% (Fig. 8 and Table 2) of the wild type and LigF $\Delta$ 242, respectively, consistent with the structure-based predictions that these serine residues are involved in catalysis. Given the proximities of LigE Ser-21 and LigF Ser-13 hydroxyls to the GSH thiol (4.1 and 5.4 Å, respectively; Fig. 5) and because the specific activities of the  $\beta$ -etherases did not steadily increase as a function of increasing pH (Fig.

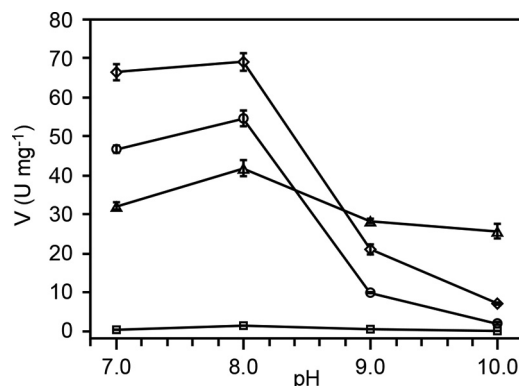


FIGURE 8. **LigE and LigF pH rate profile.** The effect of pH on  $\beta$ -etherase activities was determined for each enzyme, revealing that LigE (triangles), LigF (circles), LigF $\Delta$ 242 (diamonds), and LigF $\Delta$ 242-S13A (squares) have pH optima at pH 8.0. Plotted as a function of pH (x axis) are the specific enzymatic activities (y axis) of  $\beta$ -etherases with either (*R*)-FPHPV (LigE) or (*S*)-FPHPV (LigF, LigF $\Delta$ 242, and LigF $\Delta$ 242-S13A) as the assay substrate (1.5 mM initial concentration). Error bars, S.D. of triplicate measurements.

8), it is unlikely that these serine residues activate the GSH thiol for nucleophilic attack; rather, they act in GSH binding or thiol orientation or serve a different catalytic purpose.

Because LigF Ser-13 was the only potential acid-base catalyst revealed in the active site of the LigF $\Delta$ 242-GSH structure (Fig. 5A), we hypothesized that an  $S_N2$ -type nucleophilic attack mechanism is responsible for catalysis in LigF. The LigF $\Delta$ 242-GSH-(*S*)-MPHPV complex model, generated using SwissDock (47, 48), revealed that the GSH thiolate is in the appropriate orientation for an  $S_N2$  attack relative to the substrate  $\beta$ -carbon.

The LigE $\Delta$ 255-GSH structure revealed several potential catalytic residues in the active site, leaving open the possibility that the LigE  $\beta$ -etherase mechanism involves additional acid-base reactions. A substrate analog model compound, (*S*)-fluoro-(1'-formyl-3'-methoxyphenoxy)- $\gamma$ -hydroxypropioveratrone [(*S*)-F-FPHPV] (Fig. 7), was used to test the possibility of a non- $S_N2$  mechanism that would involve the deprotonation of the  $\beta$ -carbon of the substrate. (*S*)-F-FPHPV and (*R*)-FPHPV (despite their Cahn-Ingold-Prelog-derived *R/S* notations (66)) have the same enantiomeric configuration with respect to the orientation of their  $\beta$ -ether bonds and differ only in replacement of the hydrogen at the  $\beta$ -carbon in (*R*)-FPHPV with a fluorine in (*S*)-F-FPHPV, and this fluorine is predicted to prohibit deprotonation. We found that LigE catalyzed conversion of (*S*)-F-FPHPV to vanillin and a glutathione-conjugated coproduct, albeit at a much lower velocity compared with cleavage of (*R*)-FPHPV (Table 2), exactly as predicted based on the hypothesis that an  $S_N2$  catalytic mechanism would not involve deprotonation of the  $\beta$ -proton. Based on NMR analysis of the reaction products, we conclude that the LigE-catalyzed  $\beta$ -ether cleavage of (*S*)-F-FPHPV resulted in formation of the expected glutathione-conjugated product, (*S*)-F-GS-HPV. Although it is unclear why the reaction with (*S*)-F-FPHPV was some 3 orders of magnitude slower than LigE-catalyzed cleavage of (*R*)-FPHPV (Table 2), we hypothesize that the fluorine atom affects the  $\beta$ -ether bond angle and inhibits the approach of the thiolate ion for  $S_N2$  elimination. It is possible that these effects were even more pronounced in the active site of LigF, because LigF showed no detectable activity with the (*R*)-F-FPHPV enantiomer.

TABLE 2

LigE and LigF kinetic parameters

Kinetic parameters, determined from Michaelis-Menten curves for GSH-dependent  $\beta$ -etherases LigE, LigF, and their variants with substrates ( $\beta$ R)-MHPHV, ( $\beta$ R)-FPHPV, ( $\beta$ S)-F-FPHPV, and ( $\beta$ S)-F-FPHPV at pH 8.0. NDA, no detectable activity.

Enzyme	Substrate	$V_{\max}^a$	% WT activity with ( $\beta$ S)-F-FPHPV <sup>b</sup>	$k_{\text{cat}}$	$K_m$	$k_{\text{cat}}/K_m$
		units $\text{mg}^{-1}$	%	$\text{s}^{-1}$	$\mu\text{M}$	$\text{mM}^{-1} \text{s}^{-1}$
LigE	( $\beta$ R)-F-FPHPV	$59.7 \pm 1.2$		$31.9 \pm 0.6$	$554 \pm 16$	$57.6 \pm 4.8$
LigE-S21A	( $\beta$ R)-MHPHV		13.5 <sup>b</sup>			
LigF	( $\beta$ S)-F-FPHPV	$63.8 \pm 0.4$	100	$31.9 \pm 0.1$	$269 \pm 1$	$118.4 \pm 1.1$
LigF $\Delta$ 242	( $\beta$ S)-F-FPHPV	$69.3 \pm 4.9^a$	108.7			
LigF $\Delta$ 242-S13A	( $\beta$ S)-F-FPHPV	$1.5 \pm 0.1^a$	2.3			
LigE	( $\beta$ S)-F-F-FPHPV	$0.02^a$				
LigF	( $\beta$ R)-F-F-FPHPV	NDA				

<sup>a</sup> Where noted (i.e. in the absence of Michaelis-Menten curves), activity is reported as the velocity from assays in which the initial substrate concentration was 1.5 mM.

<sup>b</sup> Where noted, independent assays using substrate ( $\beta$ R)-MHPHV and either LigE or LigE-S21A as catalysts indicated that the LigE  $V_{\max}$  was approximately 7-fold greater than that for LigE-S21A.

## Discussion

The biocatalytic breakdown of lignin-derived compounds represents a potential source of aromatic products that would be valuable for the chemical, food, and pharmaceutical industries (2). In contrast to known fungal systems, the bacterium *Spingobium* sp. strain SYK-6 possesses an enzymatic route to the breakdown of lignin-derived components that is stereospecific and independent of chemical mediators and requires common cellular cofactors, such as pyridine nucleotides and glutathione. These combined structural and biochemical studies of the  $\beta$ -aryl ether cleavage pathway enzymes provide insights into the features important for substrate and cofactor binding and catalysis. We propose that both LigE and LigF cleave  $\beta$ -ether-linked lignin dimer molecules via an  $S_N2$  nucleophilic attack on the  $\beta$ -carbon of the substrate that is consistent with previous results showing inversion of the chiral center at the  $\beta$ -carbon (9). Because LigE catalyzed the conversion of ( $\beta$ S)-F-F-FPHPV to ( $\beta$ S)-F-GS-HVP, we conclude that the LigE mechanism is unlikely to involve formation of an enzyme-substrate adduct and does not involve C $\beta$  deprotonation or substrate enolization.

Although the sequences and x-ray crystal structures show a conserved serine in the active site of both LigE and LigF (serine 21 and 13, respectively) near the thiol of the bound glutathione (4.1 and 5.4 Å, respectively; Fig. 5), the serine is not essential for catalysis. In both LigE and LigF, mutation of the active site serine greatly reduced, but did not abolish, the enzymatic activity and did not shift the pH optimum, indicating that it may play a role other than deprotonation of the GSH thiol or perturbation of the apparent  $pK_a$  of the bound glutathione. A conserved catalytic serine is a characteristic of the Theta class, Zeta class, and some bacterial GSTs (15), but there is evidence of GSTs from the bacteria *P. mirabilis*, *Ochrubactrum anthropi*, and *E. coli* in which this active site serine is not critical for catalytic activity (67–69). Based on the data presented here and support from previous studies, it is clear that although the active site serine is not responsible for the direct activation of the thiolate anion by deprotonation or perturbation of the  $pK_a$  of the bound glutathione, it may be active in binding GSH in the active site, orienting the sulfhydryl group of GSH in the catalytic step, or stabilization of the transition state. Because GSH-dependent cleavage of these molecules does not occur readily *in vitro* in the absence of enzyme, it may be that the enzyme is able to stabilize the thiolate anion via a network of interactions

within the active site or that the binding of the substrates in the optimal orientation and distance for the  $S_N2$  attack is sufficient for catalysis.

The structures of the LigE and LigF enzymes also highlight the nature of stereospecific control that is key to this pathway. These enzymes possess dramatically different structural arrangements within the monomers and different dimer interfaces, reflected in very different dimer shapes. As a result, the substrate binding surfaces of the two enzymes are on opposite faces of the thioredoxin domain and glutathione binding site. This observation means that if a substrate with the wrong stereochemistry were to bind, it would not be in the correct orientation with respect to the glutathione for catalysis, hence introducing stereospecificity. Due to the completely different geometry of the active site, there is no simple set of mutations that would switch substrate specificity or make each individual enzyme more promiscuous.

Based on structural properties, LigE is most similar to the fungal GSTFuA class (13), suggesting that the enzymes in this class are present in both prokaryotes and fungi. Other representatives in this class are from saprotrophic fungi, suggesting a functional connection among the members of the class (18). Although it has been suggested that LigF also belongs in the GSTFuA class (13), the dimer interface present in the structure is inconsistent with other members of the class. Based on our data, LigF is best placed in a new structural class closely related to GSTFuAs or as a fungal Ure2p-like GST based on structural similarities and function in saprotrophic organisms, although it does not strictly fit the class (70). Assignments to different GST family classes, combined with the structural and biochemical information presented here, suggest that LigE and LigF evolved to cleave unique stereoisomers of the aromatic dimers that are predicted to be found in plant lignins.

The detailed structural and biochemical characterization of LigE and LigF in this study and other members of the  $\beta$ -aryl etherase pathway reveal important new aspects of the enzyme mechanism and the determinants of substrate stereospecificity. Future enzyme engineering studies informed by these results may focus on optimizing the pathway for catalysis of specific lignin-derived compounds, formed as the byproducts of industrial biomass processing, into suitable products for use as, or precursors of, advanced biofuels and renewable chemicals.

**Author Contributions**—K. E. H. designed experiments, produced protein, solved structure, analyzed LigF structure, performed all small angle x-ray scattering experiments, co-wrote the initial draft of manuscript, designed and compiled figures, and edited the manuscript; J. H. P. designed experiments, produced protein, solved structure, analyzed LigE structures, co-wrote initial draft of manuscript, and edited. D. L. G. designed experiments, synthesized substrates, cloned genes, expressed protein, performed enzymatic assays, co-wrote the initial draft of the manuscript, and edited the manuscript; R. A. H. designed experiments, cloned genes, expressed protein, performed enzymatic assays, and edited the manuscript; R. P. M. performed crystallographic data collection; C. B. designed experiments, assisted in construct design and crystallography, and coordinated x-ray data collection; K. D. synthesized substrates; K. C. H. produced protein; D. R. N. designed experiments, led enzymology on LigF, provided substrates, and edited the manuscript; B. A. S. provided research direction, contributed lignin-specific and ligninolytic enzyme expertise, and edited the manuscript; K. L. S. provided research direction, contributed lignin-specific and ligninolytic enzyme expertise, and edited the manuscript; J. R. provided research direction, contributed lignin-specific expertise, and edited the manuscript; T. J. D. provided research direction, contributed microbiological specific expertise, and edited the manuscript; P. D. A. provided research direction, contributed crystallographic and structural biology expertise, and edited the manuscript; G. N. P. provided research direction, contributed crystallographic and structural enzymology expertise, and edited the manuscript.

**Acknowledgments**—We are grateful to the staff of the Berkeley Center for Structural Biology at the Advanced Light Source of Lawrence Berkeley National Laboratory. The Berkeley Center for Structural Biology is supported in part by NIGMS, National Institutes of Health (NIH). The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the United States Department of Energy under Contract DE-AC02-05CH11231. The Joint Bio-Energy Institute is supported by the United States Department of Energy, Office of Science, Office of Biological and Environmental Research, through Grant DE-AC02-05CH11231. We thank Dr. Brian G. Fox and Dr. Christopher M. Bianchetti for helpful discussion and the NIH-funded Center for Eukaryotic Structural Genomics for general access to computers and equipment for the structural work. Use of the Advanced Photon Source was supported by the United States Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Use of Life Sciences Collaborative Access Team Sector 21 was supported by the Michigan Economic Development Corporation and by Michigan Technology Tri-Corridor Grant 085P1000817. Use of GM/CA@APS Sector 23 was supported by Federal funds NCI, NIH, Grant ACB-12002 and NIGMS, NIH, Grant AGM-12006. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIGMS, NIH, Grant P41GM103399. Small angle x-ray scattering studies were supported by funds from NIH Shared Instrumentation Grant S10RR027000 and the University of Wisconsin (Madison, WI).

## References

- Simmons, B. A., Loque, D., and Blanch, H. W. (2008) Next-generation biomass feedstocks for biofuel production. *Genome Biol.* **9**, 242
- Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Rahmanpour, R. (2011) Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* **28**, 1883–1896
- Bugg, T. D. H., Ahmad, M., Hardiman, E. M., and Singh, R. (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Curr. Opin. Biotechnol.* **22**, 394–400
- Leonowicz, A., Cho, N. S., Luterek, J., Wilkolazka, A., Wojtas-Wasilewska, M., Matuszewska, A., Hofrichter, M., Wesenberg, D., and Rogalski, J. (2001) Fungal laccase: properties and activity on lignin. *J. Basic Microbiol.* **41**, 185–227
- Martínez, A. T., Speranza, M., Ruiz-Dueñas, F. J., Ferreira, P., Camarero, S., Guillén, F., Martínez, M. J., Gutiérrez, A., and del Río, J. C. (2005) Biodegradation of lignocelluloses: microbial chemical, and enzymatic aspects of the fungal attack of lignin. *Int. Microbiol.* **8**, 195–204
- Masai, E., Katayama, Y., and Fukuda, M. (2007) Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. *Biosci. Biotechnol. Biochem.* **71**, 1–15
- Masai, E., Ichimura, A., Sato, Y., Miyauchi, K., Katayama, Y., and Fukuda, M. (2003) Roles of the enantioselective glutathione *S*-transferases in cleavage of  $\beta$ -aryl ether. *J. Bacteriol.* **185**, 1768–1775
- Adler, E. (1957) Newer views of lignin formation. *Tappi* **40**, 294–301
- Gall, D. L., Kim, H., Lu, F., Donohue, T. J., Noguera, D. R., and Ralph, J. (2014) Stereochemical features of glutathione-dependent enzymes in the *Sphingobium* sp. strain SYK-6  $\beta$ -aryl etherase pathway. *J. Biol. Chem.* **289**, 8656–8667
- Sato, Y., Moriuchi, H., Hishiyama, S., Otsuka, Y., Oshima, K., Kasai, D., Nakamura, M., Ohara, S., Katayama, Y., Fukuda, M., and Masai, E. (2009) Identification of three alcohol dehydrogenase genes involved in the stereospecific catabolism of arylglycerol- $\beta$ -aryl ether by *Sphingobium* sp. strain SYK-6. *Appl. Environ. Microbiol.* **75**, 5195–5201
- Picart, P., Sevenich, M., Dominguez de Maria, P., and Schallmeyer, A. (2015) Exploring glutathione lyases as biocatalysts: paving the way for enzymatic lignin depolymerization and future stereoselective applications. *Green Chem.* **17**, 4931–4940
- Reiter, J., Strittmatter, H., Wiemann, L. O., Schieder, D., and Sieber, V. (2013) Enzymatic cleavage of lignin  $\beta$ -O-4 aryl ether bonds via net internal hydrogen transfer. *Green Chem.* **15**, 1373–1381
- Mathieu, Y., Prosper, P., Buée, M., Dumarçay, S., Favier, F., Gelhaye, E., Gérardin, P., Harvengt, L., Jacquot, J. P., Lamant, T., Meux, E., Mathiot, S., Didierjean, C., and Morel, M. (2012) Characterization of a *Phanerochaete chrysosporium* glutathione transferase reveals a novel structural and functional class with ligandin properties. *J. Biol. Chem.* **287**, 39001–39011
- Allocati, N., Federici, L., Masulli, M., and Di Ilio, C. (2009) Glutathione transferases in bacteria. *FEBS J.* **276**, 58–75
- Sheehan, D., Meade, G., Foley, V. M., and Dowd, C. A. (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.* **360**, 1–16
- Hayes, J. D., Flanagan, J. U., and Jowsey, I. R. (2005) Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* **45**, 51–88
- Wiktelius, E., and Stenberg, G. (2007) Novel class of glutathione transferases from cyanobacteria exhibit high catalytic activities towards naturally occurring isothiocyanates. *Biochem. J.* **406**, 115–123
- Mathieu, Y., Prosper, P., Favier, F., Harvengt, L., Didierjean, C., Jacquot, J. P., Morel-Rouhier, M., and Gelhaye, E. (2013) Diversification of fungal specific class A glutathione transferases in saprotrophic fungi. *PLoS One* **8**, e80298
- Akiyama, T., Magara, K., Matsumoto, Y., Meshitsuka, G., Ishizu, A., and Lundquist, K. (2000) Proof of the presence of racemic forms of arylglycerol- $\beta$ -aryl ether structure in lignin: studies on the stereo structure of lignin by ozonation. *J. Wood Sci.* **46**, 414–415
- Ralph, J., Peng, J., Lu, F., Hatfield, R. D., and Helm, R. F. (1999) Are lignins optically active? *J. Agric. Food Chem.* **47**, 2991–2996
- Sugimoto, T., Akiyama, T., Matsumoto, Y., and Meshitsuka, G. (2002) The erythro/threo ratio of  $\beta$ -O-4 structures as an important structural characteristic of lignin: Part 2. Changes in erythro/threo (E/T) ratio of  $\beta$ -O-4 structures during delignification reactions. *Holzforschung* **56**, 416–421
- Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P. F., Marita, J. M., Hatfield, R. D., Ralph, S. A., Christensen, J. H., and Boerjan, W. (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochem. Rev.* **3**, 29–60



23. Ragauskas, A. J., Beckham, G. T., Biddy, M. J., Chandra, R., Chen, F., Davis, M. F., Davison, B. H., Dixon, R. A., Gilna, P., Keller, M., Langan, P., Naskar, A. K., Saddler, J. N., Tschaplinski, T. J., Tuskan, G. A., and Wyman, C. E. (2014) Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843
24. Shen, A., Lupardus, P. J., Morell, M., Ponder, E. L., Sadaghiani, A. M., Garcia, K. C., and Bogoy, M. (2009) Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS One* **4**, e8119
25. Klock, H. E., Koesema, E. J., Knuth, M. W., and Lesley, S. A. (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins* **71**, 982–994
26. Sreenath, H. K., Bingman, C. A., Buchan, B. W., Seder, K. D., Burns, B. T., Geetha, H. V., Jeon, W. B., Vojtik, F. C., Aceti, D. J., Frederick, R. O., Phillips, G. N., Jr., and Fox, B. G. (2005) Protocols for production of selenomethionine-labeled proteins in 2-L polyethylene terephthalate bottles using auto-induction medium. *Protein Expr. Purif.* **40**, 256–267
27. Jancarik, J., and Kim, S. H. (1991) Sparse-matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411
28. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
29. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
30. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
31. Reddy, A. P., Simmons, C. W., D'haeseleer, P., Khudyakov, J., Burd, H., Hadi, M., Simmons, B. A., Singer, S. W., Thelen, M. P., and Vanderghenst, J. S. (2013) Discovery of microorganisms and enzymes involved in high-solids decomposition of rice straw using metagenomic analyses. *PLoS One* **8**, e77985
32. Simmons, C. W., Reddy, A. P., D'haeseleer, P., Khudyakov, J., Billis, K., Pati, A., Simmons, B. A., Singer, S. W., Thelen, M. P., and Vanderghenst, J. S. (2014) Metatranscriptomic analysis of lignocellulolytic microbial communities involved in high-solids decomposition of rice straw. *Biotechnol. Biofuels* **7**, 495
33. Hendrickson, W. A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58
34. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 582–601
35. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69
36. Afonine, P. V., Grosse-Kunstleve, R. W., Urzhumtsev, A., and Adams, P. D. (2009) Automatic multiple-zone rigid-body refinement with a large convergence radius. *J. Appl. Crystallogr.* **42**, 607–615
37. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
38. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383
39. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, version 1.5.0.1, Schroedinger, LLC, New York
40. Yennamalli, R., Arangarasan, R., Bryden, A., Gleicher, M., and Phillips, G. N. (2014) Using a commodity high-definition television for collaborative structural biology. *J. Appl. Crystallogr.* **47**, 1153–1157
41. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282
42. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
43. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
44. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41
45. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864
46. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRY SOL: a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773
47. Grosdidier, A., Zoete, V., and Michielin, O. (2011) Fast docking using the CHARMM force field with EADock DSS. *J. Comput. Chem.* **32**, 2149–2159
48. Grosdidier, A., Zoete, V., and Michielin, O. (2011) SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **39**, W270–W277
49. Mills, N. (2006) ChemDraw ultra 10.0. *J. Am. Chem. Soc.* **128**, 13649–13650
50. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: an open chemical toolbox. *J. Cheminform.* **3**, 33
51. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
52. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **42**, 38–48
53. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738
54. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
55. Oakley, A. (2011) Glutathione transferases: a structural perspective. *Drug Metab. Rev.* **43**, 138–151
56. Armstrong, R. N. (1997) Structure, catalytic mechanism, and evolution of the glutathione transferases. *Chem. Res. Toxicol.* **10**, 2–18
57. Dirr, H., Reinemer, P., and Huber, R. (1994) X-ray crystal structures of cytosolic glutathione S-transferases: implications for protein architecture, substrate recognition and catalytic function. *Eur. J. Biochem.* **220**, 645–661
58. Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., and Board, P. G. (1993) Structure determination and refinement of human-Alpha class glutathione transferase A1-1, and a comparison with the Mu and Pi class enzymes. *J. Mol. Biol.* **232**, 192–212
59. Ji, X., Johnson, W. W., Sesay, M. A., Dickert, L., Prasad, S. M., Ammon, H. L., Armstrong, R. N., and Gilliland, G. L. (1994) Structure and function of the xenobiotic substrate-binding site of a glutathione S-transferase as revealed by X-ray crystallographic analysis of product complexes with the diastereomers of 9-(S-glutathionyl)-10-hydroxy-9,10-dihydrophenanthrene. *Biochemistry* **33**, 1043–1052
60. Dirr, H., Reinemer, P., and Huber, R. (1994) Refined crystal structure of porcine class Pi glutathione S-transferase (pGST P1-1) at 2.1 Å resolution. *J. Mol. Biol.* **243**, 72–92
61. Ji, X., von Rosenvinge, E. C., Johnson, W. W., Tomarev, S. I., Piatigorsky, J., Armstrong, R. N., and Gilliland, G. L. (1995) 3-dimensional structure, catalytic properties, and evolution of a Sigma-class glutathione transferase from squid, a progenitor of the lens S-crystallins of cephalopods. *Biochemistry* **34**, 5317–5328
62. Rossjohn, J., McKinstry, W. J., Oakley, A. J., Verger, D., Flanagan, J., Chelvanayagam, G., Tan, K. L., Board, P. G., and Parker, M. W. (1998) Human Theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site. *Structure* **6**, 309–322

63. Rossjohn, J., Polekhina, G., Feil, S. C., Allocati, N., Masulli, M., Di Ilio, C., and Parker, M. W. (1998) A mixed disulfide bond in bacterial glutathione transferase: functional and evolutionary implications. *Structure* **6**, 721–734
64. Zhou, H., Brock, J., Casarotto, M. G., Oakley, A. J., and Board, P. G. (2011) Novel folding and stability defects cause a deficiency of human glutathione transferase Omega 1. *J. Biol. Chem.* **286**, 4271–4279
65. Meux, E., Prosper, P., Masai, E., Mulliert, G., Dumarçay, S., Morel, M., Didierjean, C., Gelhaye, E., and Favier, F. (2012) Sphingobium sp SYK-6 LigG involved in lignin degradation is structurally and biochemically related to the glutathione transferase Omega class. *FEBS Lett.* **586**, 3944–3950
66. Cahn, R. S., Ingold, C., and Prelog, V. (1966) Specification of molecular chirality. *Angew. Chem. Int. Ed.* **5**, 385–415
67. Casalone, E., Allocati, N., Ceccarelli, I., Masulli, M., Rossjohn, J., Parker, M. W., and Di Ilio, C. (1998) Site-directed mutagenesis of the *Proteus mirabilis* glutathione transferase B1-1 G-site. *FEBS Lett.* **423**, 122–124
68. Federici, L., Masulli, M., Bonivento, D., Di Matteo, A., Gianni, S., Favalaro, B., Di Ilio, C., and Allocati, N. (2007) Role of Ser<sup>11</sup> in the stabilization of the structure of *Ochrobactrum anthropi* glutathione transferase. *Biochem. J.* **403**, 267–274
69. Nishida, M., Harada, S., Noguchi, S., Satow, Y., Inoue, H., and Takahashi, K. (1998) Three-dimensional structure of *Escherichia coli* glutathione S-transferase complexed with glutathione sulfonate: catalytic roles of Cys<sup>10</sup> and His<sup>106</sup>. *J. Mol. Biol.* **281**, 135–147
70. Thuillier, A., Ngadin, A. A., Thion, C., Billard, P., Jacquot, J.-P., Gelhaye, E., and Morel, M. (2011) Functional diversification of fungal glutathione transferases from the ure2p class. *Int. J. Evol. Biol.* **2011**, 938308