Title
Structural basis of stereospecificity in the bacterial enzymatic cleavage of β-aryl ether bonds in lignin

Permalink
https://escholarship.org/uc/item/7pt1595b

Journal
Journal of Biological Chemistry, 291(10)

ISSN
0021-9258

Authors
Helmich, KE
Pereira, JH
Gall, DL
et al.

Publication Date
2016-03-04

DOI
10.1074/jbc.M115.694307

Peer reviewed
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 β-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 β-ethersases, 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 β-aryl ether bonds account for 50–70% of all interunit linkages in lignin, understanding the mechanism of enzymatic β-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 β-aryl ether (termed simply β-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 β-ether-linked phenolic lignin model substrate, guaiacylglycerol-β-guaiacyl ether (GGE; Fig. 1), three enzymatic reactions composing the β-ether degradation pathway were identified in Sphingobium sp. strain SYK-6 (7, 9, 10). Following oxidation of the α-hydroxyl group in GGE by a Ca-dehydrogenase, stereospecific glutathione (GSH)-dependent cleavage of the β-ether linkage in β-(3'-methoxyphenoxy)-γ-hydroxypropiovanillone (MPHPV) is catalyzed by the glutathione S-transferase (GST) enzymes LigE and LigF, forming β-glutathionyl-γ-hydroxypropiovanillone (GS-HPV) and guaiacol. LigE catalyzes stereospecific cleavage of (βR)-MPHPV to (βS)-GS-HPV, whereas LigF catalyzes the cleavage of (βS)-MPHPV to (βR)-GS-HPV. Finally, GSH-dependent and stereospecific elimination of GSH from (βS)-GS-HPV is catalyzed by the GST lyase LigG, generating glutathione disulfide (GSSG) and the

5 The abbreviations used are: GGE, guaiacylglycerol-β-guaiacyl ether; MPHPV, β-(3'-methoxyphenoxy)-γ-hydroxypropiovanillone; GS-HPV, β-glutathionyl-γ-hydroxypropiovanillone; CPD, cysteine protease domain; FPHPV, β-(1'-formyl-3'-methoxyphenyl)-γ-hydroxypropioveratrole; F-FPHPV, fluoro-(1'-formyl-3'-methoxyphenyl)-γ-hydroxypropioveratrole.
achiral derivative γ-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 β-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 β-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 β-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 cholerae 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...
of interest and CPD. The pVP80K_LigFΔ242 vector was prepared as described previously (9). Insert and vector backbone PCR products were mixed 1:1 and immediately trans-ferred into the laboratory strain E. coli TOP10, 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 pVP80K_LigF vector was purified from the supernatant by immobilized nickel affinity resin (GE Healthcare). After protein binding and washing twice with HF buffer, inositol hexakisphosphate was added to a final concentration of 200 μM. Note that the inositol hexakisphosphate 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

---

TABLE 1
LigE and LigF statistics
Summary of crystal parameters, data collection, and refinement statistics. Values in parentheses are for the highest resolution shell.

<table>
<thead>
<tr>
<th>Crystal parameters</th>
<th>LigFΔ242-GSH</th>
<th>LigFΔ255</th>
<th>LigFΔ255-GSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P6322</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td>a, b, c (Å)</td>
<td>123.71, 123.71, 66.42</td>
<td>122.55, 97.15, 131.38</td>
</tr>
<tr>
<td>β</td>
<td></td>
<td>106.65</td>
<td>81.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data collection statistics</th>
<th>LigFΔ242-GSH</th>
<th>LigFΔ255</th>
<th>LigFΔ255-GSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.97857</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.00–2.07</td>
<td>50–1.90</td>
<td>50–2.6 (2.65–2.60)</td>
</tr>
<tr>
<td>No. of reflections (measured/unique)</td>
<td>326,246/18,884</td>
<td>575,459/114,153</td>
<td>160,219/42,163</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (98.5)</td>
<td>99.2 (99.9)</td>
<td>97.7 (93.8)</td>
</tr>
<tr>
<td>Rmerge (a)</td>
<td>0.076 (0.637)</td>
<td>0.137 (0.63)</td>
<td>0.135 (0.61)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>17.3 (12.9)</td>
<td>5.0 (4.9)</td>
<td>3.8 (3.6)</td>
</tr>
<tr>
<td>Mean fobs(β)</td>
<td>9.1 (3.8)</td>
<td>9.0 (1.6)</td>
<td>6.1 (2.5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement and model statistics</th>
<th>LigFΔ242-GSH</th>
<th>LigFΔ255</th>
<th>LigFΔ255-GSH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>40.49–2.07</td>
<td>48–1.90</td>
<td>48–2.6 (2.65–2.60)</td>
</tr>
<tr>
<td>No. of reflections (work/test)</td>
<td>17,278/964</td>
<td>114,138/1,999</td>
<td>42,158/2,000</td>
</tr>
<tr>
<td>Rmerge (b)</td>
<td>0.161 (0.182)</td>
<td>0.227 (0.289)</td>
<td>0.222 (0.266)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.214 (0.263)</td>
<td>0.271 (0.350)</td>
<td>0.267 (0.290)</td>
</tr>
<tr>
<td>Root mean square deviation bonds (Å)</td>
<td>0.008</td>
<td>0.004</td>
<td>0.005</td>
</tr>
<tr>
<td>Root mean square deviation angles (degrees)</td>
<td>1.022</td>
<td>0.956</td>
<td>0.776</td>
</tr>
<tr>
<td>B factor (protein/solvent) (Å²)</td>
<td>39.58/43.93</td>
<td>29.13/37.54</td>
<td>34.22/27.38</td>
</tr>
<tr>
<td>B factor (GSH) (Å²)</td>
<td>26</td>
<td>146, 148, 149, 146</td>
<td></td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1,983</td>
<td>9,405</td>
<td>8,491</td>
</tr>
<tr>
<td>No. of waters</td>
<td>229</td>
<td>1,159</td>
<td>165</td>
</tr>
<tr>
<td>Auxiliary molecules (real space correlation coefficient (CC))</td>
<td>1 glutathione (0.97), 1 Tris (0.95), 1 PEG (0.95)</td>
<td>4 glutathione (1 per chain), A = 0.71, B = 0.58, C = 0.58, D = 0.61</td>
<td></td>
</tr>
</tbody>
</table>

| Ramachandran plot | 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 |

---

(a) Rmerge = \(\frac{\sum |I(h)| - \langle |F_{hk,\text{cal}}|\rangle}{\sum |I(h)|}\), where I(h) is the intensity of an individual measurement of the reflection, and \(\langle |F_{hk,\text{cal}}|\rangle\) is the mean intensity of the reflection.

(b) Rmerge = \(\frac{\sum |F_{hk,\text{obs}}| - \sum |F_{hk,\text{calc}}|}{\sum |F_{hk,\text{obs}}|}\), where \(F_{hk,\text{obs}}\) and \(F_{hk,\text{calc}}\) are the observed and calculated structure factor amplitudes, respectively.

(R) Rfree was calculated as Rmerge using 5.0% of randomly selected unique reflections that were omitted from the structure refinement.
ml. Final size exclusion purification was performed on a HiLoad™ 26/60 Supradex™ 200 preparation grade column. Enzyme Kinetic Assays—In vitro /H9252 -etherase assays with LigE, LigF, LigF/H9004_242, and LigF/H9004_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/H9004_242), 3.9 /H9262 (LigF/H9004_242-S13A), 11.2 /H9262 (LigE with (βS)-F-FPHPV), or 12.0 /H9262 (LigF with (βR)-F-FPHPV). Enantiopure preparations of (βR)-FPHPV and (βS)-FPHPV were obtained from chiral chromatographic separation of the parent racemate as described previously (9). Similarly, chiral chromatography was used for the separation of (βS)-F-FPHPV and (βR)-F-FPHPV, with (βS)-F-FPHPV being used as a substrate in the LigE assays. \( \text{Enzymatic substrates FPHPV and F-FPHPV are described in the supplemental material.} \)

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 μl were added to a 2-ml vial; 2) 875 μl 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 μl of 100 mM GSH was added (100 mM GSH stock solution was prepared by adding GSH to 25 mM Tris (pH X)); 4) 50 μl of 20× concentrated enzyme was added; 5) 150-μl 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 μl 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 C18-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.

**FIGURE 2. LigE and LigF structures.** A, schematic representation of LigF, including the N-terminal thioredoxin domain (blue), the C-terminal α-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 α-helical domain (brown), and the short linker (gray). Bound GSH is shown as yellow spheres. D, schematic representation of the dimer of LigF with the proposed binding site (Fig. 5B) circled.
Crystallization—LigE was concentrated to 9 mg ml⁻¹ and dialyzed against 20 mM HEPES, pH 7.4, and 50 mM NaCl. LigF was dialyzed in 10 mM HEPES buffer, pH 7.5, containing 50 mM sodium chloride, 0.5 mM tris(2-carboxyethyl)phosphine, and 1 mM GSH, and concentrated to 18.5 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 M ammonium citrate, 0.1 M MES, pH 5.5, 20% PEG 3,350, and 5% isopropyl alcohol; LigF, 25% polyethylene glycol monomethyl ether 2000, 0.25 M trimethylamine N-oxide, and 0.1 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 µl of protein solution and 0.2 µl of reservoir solution. LigF crystals were obtained in <24 h with drops containing a mixture of 1 µl of protein solution, 0.8 µl of reservoir solution, and 0.2 µl of seed crystals (pulverized LigFΔ242 crystals in 0.2 M magnesium formate, 30% polyethylene glycol 3350, and 1 mM GSH).
and otherwise default parameters, with the search space limited to a $10 \times 10 \times 10\text{Å}$ 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 OpenBabel (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). LigEΔ255 and the LigEΔ255-GSH complex crystallized in the space group C2 with four molecules in the asymmetric unit with electron density for the bound GSH molecule. LigFΔ242-GSH crystallized in the space group P6$_3$2$_2$ 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.
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 α-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 β-strands and three α-helices following the topology β1α1β2α2β3β4α3. The loop between β1 and α1 is longer in LigE than in LigF and occupies the space between the thioredoxin domain and the α-helical domain, whereas in LigF, this loop is moved away from the domain interface toward the surface of the thioredoxin domain. The loop between β2 and α2 is longer in LigF than in LigE, but both interact with the α-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 α-helices, respectively. The root mean square deviation between the C-α 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 Å² 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 χ value of 2.4 and 1.4 for LigF and LigE, respectively (Fig. 3).

The LigF dimer forms via interactions between helices α3 and α4, 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 α3 (the final helix of the thioredoxin domain) and α4 (the first helix in the C-terminal domain) (13, 15, 55, 56). Variability in the arrangement of secondary structural elements away from the α3/α4 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 α3/α4 canonical 4-helix bundle dimer structure (58–65). In the LigE...
Sphingobium LigE and LigF Crystal Structures

FIGURE 6. Substrate binding sites in LigF and LigE. A, model of ternary complex LigΔ242-GSH-(βS)-MPHPV. 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)-MPHPV (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 α4 of one monomer is interdigitated between α4 and α7 of the other monomer, and the entire dimer interface is contained within the α-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.

The enzymatic active sites of these GST family members are often located in a cleft between the thioredoxin domain and the α-helical domain. Both the LigE and LigF enzymes contain the βαα motif required for anchoring GSH in the active site. In LigF, Glu-65 and Ser-66 located in the turn connecting β4 and α3 recognize the γ-glutamyl moiety of GSH as part of the βαα motif. 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 β4 and α3, hydrogen-bond with the amino and carboxylate groups, respectively, of the γ-glutamyl residue of the GSH molecule.

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 LigΔ242-GSH-(βS)-MPHPV complex model (Fig. 6A) from the LigΔ242-GSH structure and a molecular model of (βS)-MPHPV. 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).

Enzymatic Analysis and Mutagenesis—To analyze the enzymatic activities of the GSH-dependent β-etherase enzymes, FPHPV degradation rates were measured by the accumulation of vanillin, a monoaromatic product of FPHPV cleavage. Whereas β-etherase catalysis with MPHPV 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 β-etherase catalysis for LigE variant LigE-S21A and two LigF variants, LigFΔ242 and LigFΔ242-S13A.

We found that LigE catalysis resulted in stereospecific (βR)-FPHPV cleavage, whereas LigF selectively degraded the (βS)-FPHPV enantiomer, as is consistent with previous reports.
The inhibitory to predicted disordered region in the C terminus may actually be
2) of the wild type and LigF
respectively; Fig. 5) and because the specific activities of the
LigF Ser-13 hydroxyls to the GSH thiol (4.1 and 5.4 Å, respec-
tively) were 3 orders of magnitude more than LigF
S21A and LigF
S13A have pH optima at pH 8.0 (Fig. 8). The activity
albeit at a much lower velocity compared with cleavage of
an SN2 catalytic mechanism would not involve deprotonation of
the β-carbon of the substrate. (BS)-F-FPHPV and (BR)-FPHPV (despite
their Cahn-Ingold-Prelog-derived R/S notations (66)) have the
same enantiomeric configuration with respect to the orienta-
tion of their β-ether bonds and differ only in replacement of the
hydrogen at the β-carbon in (BR)-FPHPV with a fluorine in
(BS)-F-FPHPV, and this fluorine is predicted to prohibit depro-
tonation. We found that LigE catalyzed conversion of (BS)-F-
FPHPV to vanillin and a glutathione-conjugated coproduct,
albeit at a much lower velocity compared with cleavage of (BR)-
FPHPV (Table 2), exactly as predicted based on the hypothesis that
an $S_{n}2$-type catalytic mechanism would not involve deprotonation of
the β-proton. Based on NMR analysis of the reaction products, we
conclude that the LigE-catalyzed β-ether cleavage of (BS)-F-
FPHPV resulted in formation of the expected glutathione-conju-
gated product, (BS)-F-GS-HVP. Although it is unclear why the
LigE-catalyzed reaction, converting FPHPV to vanillin and GS-
HVP. It is possible that these effects were even more pronounced in the
active site of LigF, because LigF showed no detectable activity with the (BR)-F-FPHPV enantiomer.
Table 2

LigE and LigF kinetic parameters

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>V_maxa (units mg⁻¹)</th>
<th>% WT activity with (βS)-F-HPHPVb</th>
<th>k_cat (s⁻¹)</th>
<th>K_m (μM)</th>
<th>k_cat/K_m (10⁻² s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LigE</td>
<td>(βR)-F-HPHPV</td>
<td>59.7 ± 1.2</td>
<td>-</td>
<td>31.9 ± 0.6</td>
<td>554 ± 16</td>
<td>57.6 ± 4.8</td>
</tr>
<tr>
<td>LigE-S21A</td>
<td>(βR)-MPHPV</td>
<td>-</td>
<td>13.5b</td>
<td>269 ± 1</td>
<td>118.4 ± 1.1</td>
<td></td>
</tr>
<tr>
<td>LigF</td>
<td>(βS)-F-HPHPV</td>
<td>63.8 ± 0.4</td>
<td>100</td>
<td>69.3 ± 4.9</td>
<td>108.7</td>
<td></td>
</tr>
<tr>
<td>LigF ΔS242</td>
<td>(βS)-F-HPHPV</td>
<td>1.5 ± 0.1a</td>
<td>2.3</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>LigF ΔS242-13A</td>
<td>(βS)-F-HPHPV</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>LigF</td>
<td>(ββ)-F-HPHPV</td>
<td>-</td>
<td>NDA</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

a 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.
b Where noted, independent assays using substrate (ββ)-MPHPV 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 *Sphingobium* 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 β-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 β-ether-linked lignin dimer molecules via an S_n,2 nucleophilic attack on the β-carbon of the substrate that is consistent with previous results showing inversion of the chiral center at the β-carbon (9). Because LigE catalyzed the conversion of (βS)-F-FPHPV to (βS)-F-GS-HVP, we conclude that the LigE mechanism is unlikely to involve formation of an enzyme-substrate adduct and does not involve Cβ 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 sulphydryl 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_n,2 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 stereocchemistry 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 β-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.
Sphingobium LigE and LigF Crystal Structures

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 crystallomemics, 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 BioEnergy 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 08SP1000817. 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


