Supporting Information (SI)

Structure of aryl *O*-demethylase offers molecular insight into a catalytic tyrosine-dependent mechanism

Amanda C. Kohler[1,2], Matthew J.L. Mills[1,2], Paul D. Adams[1,3,4], Blake A. Simmons[1,3], Kenneth L. Sale[1,2*]

[1]Joint BioEnergy Institute, Emeryville, CA, 94608, USA; [2]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, CA, 94550, USA; [3]Biosciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720 USA; [4]Department of Bioengineering, University of California Berkeley, Berkeley, CA, 94720, USA; *Correspondence should be addressed to Kenneth Sale, KLSale@lbl.gov

TABLE OF SI CONTENTS

Appendix 1 I Geometry Optimization of Substrate-bound LigM

To provide more compelling evidence for the direct methyl transfer reaction mechanism suggested by the docking results, a cluster model was built including the vanillate/3MGA and H_4folate substrates and the LigM residues involved in their binding. Additional models were built including explicit water molecules. The residues included, and their charge and pK_a are given in Table S2**.**

For each residue, the entire side chain and the peptide bond (capped with hydrogen) atoms were included in a cluster model and the geometry was optimized with *ab initio* quantum chemistry methods as described in the SI Materials and Methods. The final geometry agrees well with the direct vanillate methoxy methyl to H_4folate-N_5 methyl transfer mechanism hypothesized based on the electron density and docking results. The active site geometry is shown in Fig. S4*A*, and the corresponding values of important internal coordinates are given for all optimized cluster models in Fig. S4*B*. The H_4folate-vanillate, $N_5 \ldots C_V$, distance in the final geometry is 3.1 Å and the $N_5 \ldots O_M$-C_V angle is 146°, while the Tyr247 to vanillate $O_H H \ldots O_M$ hydrogen bond distance (O to O) is 2.7 Å and the angle is 151°. This geometry is near-exactly preserved when water is added (Fig. S4*B*). This geometry allows for the simultaneous transfer of the methyl group of vanillate to the N_5 atom of H_4folate with proton transfer from Tyr247 O_H to vanillate's O_M, keeping the atoms at appropriate reaction distances and maintaining close to straight lines for transfer.

Met61 is the only residue in the proposed vanillate-binding site that adopts two conformations in the crystal structure; therefore, binding site geometry optimization was performed for both conformations (Met61A, Met61B) independently. The orientation of the reacting methyl group of vanillate is slightly changed when Met61 is in its B conformation, relative to the A conformation, although it is still positioned for direct transfer to N_5 of H_4folate. Again the addition of water has essentially no effect on the cluster geometry. The geometry observed for the LigM-3MGA cluster is very similar to LigM-vanillate for Met61A. The only significant geometry deviation is observed for 3MGA with Met61B, where the methyl transfer distance increases from 3.3 to 4.2 Å, with a concomitant 4° decrease in the O_M-C_V…N_5 angle. These results suggest that Met61 orientation is important for catalysis, however much more detailed calculations that evaluate the free energy with proper sampling of the phase space are needed to draw conclusions about this. In all cases, the O_HH…O_M hydrogen bond is kept intact. It can be concluded from the similarity of the results that the mechanism is preserved throughout, and thus the discussion of the interatomic interactions in the active site will be focused on the LigM-vanillate cluster without water, with the remaining clusters discussed in terms of their difference to that system.

Appendix 2 I Quantum Chemical Topology

It is typical in discussions of enzyme structure to highlight atom-atom interactions (other than covalent bonds) believed to be of importance to enzyme-substrate binding or the reaction mechanism (see Fig. 4, Fig. 5, and Fig. 6 of the main text, for example). Throughout this note, such interactions are written in the format XH…Y, which denotes an interaction between the covalently bound pair of atoms X (the *donor*) and H with the atom Y (the *acceptor*). For example, the hydrogen bonds found between water molecules would be written OH…O. For interactions without an intermediate H atom, the simplified notation X…Y is used. Normally the existence of interatomic interactions is determined by visual inspection of an enzyme-substrate complex geometry, or by the application of tools that search for sets of atoms with geometric criteria commensurate with the presence of an interaction. For example hydrogen bonds are identified between atoms XH…Y when $r_{\mathrm{min}} < r_{\mathrm{XY}} < r_{\mathrm{max}}$ and $\theta_{\mathrm{min}} < \theta_{\mathrm{XHY}} < \theta_{\mathrm{max}}$, where r and θ are distances and angles respectively. There is, however, no physical basis for the required maximum and minimum values in these equations and thus there exists no rigorous basis for this kind of method.

Given that the geometry optimization of the cluster model with quantum chemistry methods provided a wavefunction for the system, it was possible to apply an alternative, quantitative approach based in quantum mechanics for locating atom-atom interactions present in the complex. To this end, a more tightly converged B3LYP/6-31G(d) wavefunction was obtained for the optimized cluster model geometries for analysis (SI Materials and Methods).

The theory of Atoms in Molecules (1, 2) (AIM) provides a parameter-free route to determination of the atomic connections present in a chemical system described by a wavefunction. The topology of the electron density (a scalar field computed from the wavefunction) is determined, and topological properties are associated with quantitative definitions of empirical chemical concepts. In particular, locating a bond critical point (BCP) (1) indicates the presence of an interatomic interaction between the pair of atoms whose nuclei are connected by the unique gradient paths originating at that BCP (together called an atomic interaction line, or AIL). Given the system wavefunction it is therefore possible to analyze the topology of the electron density of the LigM-vanillate cluster model and locate BCPs positioned between atoms of interest, namely those of the substrates and their local enzyme environment with substrates bound. The BCPs and AILs of the pertinent interactions between LigM, H_4folate, and vanillate are highlighted in Fig. S4*C* and *D* for the LigM-vanillate cluster model.

Locating interactions is not solely sufficient to allow comparison between them. For this purpose, several properties that can be evaluated at BCPs have been suggested. The simplest scheme involves specification of the value of the electron density ($\rho(\mathbf{r}_{\mathrm{cp}})$) and its Laplacian ($\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$) at the BCP of an atomic interaction (3, 4). The electron density is accumulated between interacting atoms, and larger values have been shown to correlate with higher bond orders. The Laplacian measures the local concentration of electron density and can be used to classify interactions as closed-shell or shared(5). Beyond this, a frequently cited set of criteria for defining interactions as hydrogen bonds is widely used in classification, although the eight necessary criteria listed in Koch, *et al.* include atomic and surface properties, which were not feasible to compute for wavefunctions as large as those determined in this work (6). The use of only the topological, electron density and Laplacian conditions described in Koch, *et al.* (rather than all 8 provided) for locating hydrogen bonds is widespread, and therefore, herein only the electron density and Laplacian values are reported and the computation of what is (or is not) a hydrogen bond is not attempted. This allows all interactions to be treated on an equal footing, and to be compared directly.

For the LigM-vanillate system, the enzyme-substrate interactions that correspond to the binding of H_4folate and vanillate in the LigM active site are of interest, as well as the interactions between atoms involved directly in the catalytic mechanism. Images of the complex, highlighting the appropriate regions of the system, are provided in Fig. S4*C* and *D*. The AIM properties are collected in Table S3 for LigM-H_4folate interactions and Table S4*A* for LigM-vanillate interactions. All measured values of $\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$ are positive, indicating that all of the interactions are closed-shell in nature. As expected, there is a strong linear correlation between the values of $\rho(\mathbf{r}_{\mathrm{cp}})$ and $\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$ measured at the critical points ($R^2 = 0.99$ for the entire set of LigM-vanillate/H_4folate interactions), and thus the relative strength of the interactions can be discussed in terms of electron density ($\rho(\mathbf{r}_{\mathrm{cp}})$) alone.

Appendix 3 I LigM – H_4folate Interactions

The H_4folate substrate forms five intramolecular interactions, all of which are found in or between the glutamate and p-aminobenzoate (PABA) moieties, save for the long (4.7 Å) $C_{5'}H\ldots O_4$ interaction, which has an exceedingly low $\rho(\mathbf{r}_{\mathrm{cp}})$ value (Table S3). The lack of intramolecular interactions around the bridging amine suggests that the kinked conformation of H_4folate observed when its bound to LigM is a result of the LigM environment rather than any innate tendency of H_4folate to exist in this conformation. As the intramolecular interactions are likely also present regardless of whether H_4folate is bound to enzyme or free in solution, they are not discussed further.

In order to bind H_4folate, LigM forms 34 intermolecular interactions with H_4folate. Specifically, the following residues form one or more interactions with H_4folate that vary in strength: Gln57 (1 interaction), Met61 (1), Gln93 (3), Gly107 (1), Ile108 (5), Phe110 (1), Val120 (6), Arg163 (1), Gln165 (1), Phe188 (3), Phe189 (3), His206 (2), Glu215 (1), and Trp256 (5). Several of the cases where only one interaction is formed with a particular residue feature long XY distances and small $\rho(\mathbf{r}_{\mathrm{cp}})$ values. These are the interactions between H_4folate and Gly107 (r_{XY} = 4.5 Å, $\rho(\mathbf{r}_{\mathrm{cp}})$ = 0.003 au), Phe110 (3.4, 0.009) and Arg163 (4.3, 0.002), which are weak and thus these residues are likely of limited utility in the direct binding of H_4folate. Met61 also forms a long interaction (4.7 Å) with H_4folate, although it is stronger (0.012 au). The remaining interactions can be roughly divided into those that bind the pteridine moiety, those that bind PABA, and those that bind the glutamate tail.

Gln57 and Gln165 both form short (2.8 and 3.2 Å) $N_{E2}H…O_4$ interactions (0.037 and 0.011 au, respectively) with the pteridine ring. N_2 is engaged in five interactions, of which two are long and weak (those with Phe110 and Arg163 as discussed previously). Atom N_2 interacts with Glu215 through a $N_2H…O_{E1}$ interaction (3.0, 0.026) and its remaining pair of interactions (both of type CH…N) are with Val120 (both 3.7 and 0.005). Val120 forms a further similar interaction with N_1 (3.4, 0.006), as well as a stronger $C_BH…HN_1$ interaction (3.6, 0.011). Additionally, the backbone oxygen of Val120 accepts NH…O interactions from N_1 and N_8 (3.3, 0.009 and 2.9, 0.029, respectively). Interactions with Ile108 constitute the remainder of the LigM-pteridine interactions. Two $C_{G1}H…N$ interactions are formed with N_1 and N_8 (3.4, 0.007 and 3.8, 0.006, respectively). The three other interactions of Ile108 are with the pteridine-PABA linkage involve CH donors, and are generally long (4.0 Å) and weak (0.004 to 0.005) and are therefore not predicted to be critical for binding.

The PABA moiety is bound by Gln93, Phe189, His206 and Trp256. Trp256 interacts with both the PABA and pteridine moieties, and four of its interactions with H_4folate have relatively long (3.5 to 3.8 Å) distances. Each of these four has a carbon atom acceptor and all but one has a CH donor atom. These facts are indicative of weak interactions, and consequently, none of the four observed interactions exhibit a significant value of $\rho(\mathbf{r}_{\mathrm{cp}})$ (the range is from 0.004 to 0.009 au). The remaining Trp256-H_4folate interaction, $C_{3'}H…C_{D1}$ is shorter at 3.3 Å with $\rho(\mathbf{r}_{\mathrm{cp}})$ = 0.013 au. His206 accepts two $CH…N_{E2}$ interactions from the PABA aryl ring atoms $C_{5'}$ and $C_{6'}$ (3.3, 0.007 and 3.1, 0.016, respectively). Gln93 has long $C_{2'}H…N_{E2}$ (3.9, 0.003) and $C_{2'}H…O_{E1}$ (3.8, 0.002) interactions and a short (2.7, 0.045) $N_{E2}H…O_{X1}$ interaction with the glutamate moiety. Phe188 assists Phe189 in binding the glutamate, together forming five CH…O interactions as donors, of which two $C_{D2}H…O_{E2}$ interactions (one each from Phe188 and Phe189) are notably shorter and stronger (both 3.1, 0.022). O_{E2} of H_4folate also accepts a $C_BH…O$ interaction from Phe188 (3.3, 0.013). Each of these residues also forms a $C_{E2}H…O_{11}$ interaction of moderate strength (3.21, 0.016 and 3.5, 0.010). Finally, Phe189 forms a much weaker (3.7, 0.007) $C_ZH…C_{6'}$ interaction with the PABA aryl ring.

Appendix 4 I LigM – Vanillate Interactions

Interactions between Arg122 and the carboxylate group of vanillate are to be expected based on the close proximity of oppositely charged groups, and Tyr29 is well placed to form an interaction with the carboxylate. The three highest values of $\rho(\mathbf{r}_{\text{cp}})$ are found for the binding interactions of Arg122 and Tyr29 with vanillate, amongst which the $O_HH...O_1$ is strongest (0.062 versus two of 0.040 and 0.041 au – $N_EH...O_2$ and $N_{H2}H...O_1$). These are also the shortest interactions located (2.6 and two 2.8 Å). This indicates that the motion of the substrate carboxylate group is tightly restrained around the binding geometry shown (Fig. S4*D*). Following these interactions in strength are four of the five mechanistic interactions illustrated in Fig. S4*C*. Short interactions are seen between Gln57-H_4folate ($N_{E2}H...O_4$) and Tyr247-vanillate ($O_HH...O_M$; in the reaction mechanism the proton of this interaction is transferred), which are similar in strength (0.037 and 0.036 au) and length (2.8 and 2.7 Å respectively) to the NH...O Arg122-vanillate interactions. The intramolecular vanillate ($O_3H...O_M$) and Gln57-Tyr247 ($N_{E2}H...O$) interactions are approximately half as strong (0.020 and 0.024 au). The former of these (between the vanillate hydroxyl group and the O of the reacting methoxy) could promote the upward positioning of the transferring methyl group, pointing it towards N_5 of H_4folate. A similar argument for intramolecular preorganization in substrates has been made before(7). The fifth interaction, $C_VH...N_5$, formed between the transferring methyl group and H_4folate is relatively strong at 0.015 au compared to the remainder, although a single interaction (Pro248 $C_AH...O_3$) is only of slightly higher strength. Given the longer length of the $C_VH...N_5$ interaction (3.1 Å compared to ≤ 3.0 Å for the remainder of the interactions in Fig. S4*C*) and the fact that interactions with CH donors are weaker than those with OH or NH donors, the low value of $\rho(\mathbf{r}_{\text{cp}})$ for this interaction is not surprising.

The remainder of the interactions exhibit values of $\rho(\mathbf{r}_{\text{cp}})$ in the range $0.003 < \rho(\mathbf{r}_{\text{cp}}) < 0.015$ au corresponding to strengths between 0.15 and ≈0.00 relative to the strength of the $O_HH...O_1$ interaction. The strongest of these (with $\rho(\mathbf{r}_{\text{cp}}) > 0.009$ au) are all CH...O interactions, five with vanillate as acceptor (with Pro248, Tyr29, Tyr31, Met61 and Thr251) and two as donor (with Tyr31 and Pro248), with lengths ranging from 3.1 to 3.4 Å. The remaining 14 interactions are a variety of types, but all have distances in the range $3.4 < r_{\text{XH...Y}} < 4.0$ and are significantly weaker, the strength of all 14 of these interactions is less than 10% the strength of the strongest interaction and 10 are less than 5%. These interactions are less able to preserve geometry in the active site and may be transient as the substrate moves. In conclusion, the carboxylate-enzyme interactions (Fig. S4*D*) and the mechanistic interactions (Fig. S4*C*) are predicted to be majorly responsible for the observed binding orientation of vanillate, with CH...O interactions constituting a secondary binding factor, and the remainder of the interactions having relatively little effect. Fig. 5 and Fig. 6 reflect this result, with the most important interactions highlighted.

The final point of discussion for LigM-vanillate is the difference in the interactions when the Met61 conformation is changed (observations when water is introduced to the cluster or the substrate is changed to 3MGA follow). The set of interactions observed for LigM-vanillate (no water) is conserved across Met61A and Met61B, except for the $C_VH...N_{10}$ interaction. Two additional vanillate interactions are observed in Met61B,

$C_vH...C_{E3}$ (3.93) with Trp256 and $C_VH...HC_6$ (r_{HH} = 2.19, r_{CC} = 3.78) with H_4folate. Both of these interactions fall into the final, least important set of interatomic interactions observed in the cluster. Thus, vanillate binding is not significantly affected by the conformation of Met61 and Met61B will not be discussed for the remaining systems.

Appendix 5 I Inclusion of Explicit Water Molecules

The inclusion of five water molecules does not disrupt the interactions between vanillate and LigM, as these waters reside near to but not within the binding cavity. The only observed difference is that the vanillate-Met61 $C_{O2}H...C_E$ interaction is replaced by a $C_{O2}H...HC_E$ interaction, i.e. a CH...C interaction is replaced by an H...H interaction. A number of interactions with water are observed, however these are not catalytically important nor do they influence vanillate or H_4folate binding. Vanillate forms an $O_3H...O_W$ interaction and a further three $C_VH...O_W$ interactions with three of the five water molecules. These result in secondary interactions being observed: Asn250 interacts with one water molecule through $N_{D2}H...O_W$ and $O_WH...O_{D1}$ interactions, and that water molecule interacts in turn with vanillate through $O_3H...O_W$ and $C_VH...O_W$ interactions. A second water molecule interacts with His206 through $C_{E1}H...O_W$ and with Gln165 through an $O_W...N_{E2}$ interaction and in turn interacts with vanillate through a $C_VH...O_W$ interaction and with H_4folate through $O_WH...O_4$ and $C_{5'}H...O_W$ interactions. The third and final water to form secondary interactions forms a $C_{E3}H...O_W$ interaction with Trp256 and an $O_WH...O_{G1}$ with Thr251, and forms a $C_VH...O_W$ interaction with vanillate and an $N_{10}H...O_W$ with H_4folate. In a real system, these interactions will all be transient as water diffuses in and out of the active site. The two water molecules not discussed form no secondary interactions with vanillate or H_4folate, only water-water and water-LigM interactions.

Appendix 6 I LigM – 3MGA Interactions

Replacing vanillate with 3MGA has a minimal effect on the observed interactions. The only structural difference between vanillate and 3MGA is the replacement of a hydrogen atom at C_{M1} with a hydroxyl group (atom label O_2). This results in the absence of a C_{M1}-H group, removing the $C_{M1}H...O$ and $C_{M1}H...HC_B$ interactions with Pro248, and the $C_{M1}H...HC_Z$ interaction with Phe393. Of these, two are weak interactions and one is within the third strongest set; the CH...O interactions. A fourth interaction, $C_{G2}H...C_{M1}$ with Thr251 is absent in the 3MGA cluster but is in the weakest group of interactions located in the vanillate cluster. The effect of these four lost interactions is minimal. The introduced hydroxyl group bonded to C_{M1} engages in five new interactions. It accepts a $C_{G2}H...O_2$ interaction from Thr251 (3.3) and a $C_BH...O_2$ from Pro248 (3.4), and engages in an $O...O_2$ interaction with Pro248 (2.6). It also acts as a donor in an $OH...C_Z$ interaction with Phe393 (3.1) and forms an $OH...HC_G$ interaction with Pro258 (2.3, 3.8). The first four of these interactions can be considered direct replacements for those lost by the addition of the hydroxyl group, and the interactions are collected in Table S4*B*. The values of $\rho(\mathbf{r}_{cp})$ for each interaction are also shown. From the data it is clear that the interactions between 3MGA and LigM are sufficiently strong to preserve the binding position. This suggests that the 3MGA substrate will bind and react in the same manner as vanillate inside the enzyme active site.

Appendix 7 I SI Materials and Methods

Protein Expression, Purification, and Mutagenesis. Full-length *Sphingomonas paucimobilis* LigM (1413 bp) was codon optimized for expression in *Escherichia coli*, synthesized (GenScript), and cloned into a modified pCDF-Duet-1 vector (Novagen-EMD Millipore) between the *Sal*I and *Not*I restriction sites. All LigM vanillate-binding site and catalytic point mutants (Y29A, Y31A, H60A, M61A, V62A, R122A, R122A/R147A, Y247F) were generated by site-directed PCR mutagenesis of pCDF-Duet-1-LigM. For protein expression, pCDF-Duet-1-LigM, wild-type and point mutants, were co-transformed with chaperone plasmid pG-KJE8 (Clontech Laboratories, Inc.) into *Escherichia coli* BL21(DE3) cells (New England Biolabs, Inc.). Cells were grown in Miller's LB media at 37°C to an OD_{600} of 0.4-0.5 and chaperone expression was induced with the addition of 0.5 mg ml^{-1} L-arabinose and 5 μg ml^{-1} tetracycline for 1 hour at 37°C. LigM protein expression was then induced with 1 mM IPTG for 16 hours at 19°C. Cells were harvested by centrifugation, resuspended in 50 mM Tris, pH 8.0, 250 mM NaCl, and 5 mM imidazole, 0.1 mM PMSF, and lysed via homogenization. Cleared lysate was incubated with 5 mM ATP for 30 minutes on ice before being passed over Ni-NTA resin (Qiagen). The protein was eluted from the resin in 50 mM Tris, pH 8.0, 250 mM NaCl, 500 mM imidazole and was dialyzed against 25 mM Tris, pH 8.0, 150 mM NaCl at 4°C for 16 hours. The N-terminal histidine tag was cleaved using 6xHis-HRV 3C Protease (Sino Biological, Inc.), leaving behind a four-residue (GVPD) scar on the N-terminus of LigM. The resulting material was applied to Ni-NTA resin to separate the cleaved tag and protease from LigM, and as a final purification step, LigM protein was purified by size-exclusion chromatography using a Superdex 200 HiLoad 16/600 column (GE Healthcare) into a final buffer of 25 mM Tris, pH 8.0, 150 mM NaCl for enzymatic assays. For crystallization trials, LigM protein was purified into a final buffer of 25 mM HEPES, pH 7.5, 25 mM NaCl, 2 mM DTT. For vanillate-binding site and catalytic analysis only, LigM, wild-type and mutants, were produced and purified as described above, with the exception that the N-terminal histidine tag was not removed. All purified protein was flash-frozen and stored at -80°C.

Crystallization, X-ray Data Collection, and Structure Determination. Wild-type LigM, residues 1-471, was crystallized in a sitting-drop format at a concentration of 50 mg ml^{-1} at ambient temperature (298 K). LigM crystallized readily under a range of conditions containing 12-25% PEG3350, but grew resolutely as needle clusters despite N-terminal truncation of the crystallization construct, altered purification conditions, and optimization of crystal growth conditions (protein concentration, additive screening, crystal seeding). Optimization of purification conditions led to an improvement in needle thickness, producing rod-like clusters that could be broken by hand into single crystals prior to data collection. However, extensive screening of these crystal fragments was required as many led to unproductive diffraction data sets, plagued by varying degrees of smeared or twinned reflections and predicted pseudo-translational symmetry. Of 45 crystal fragments screened, 19 of which were heavy atom soaked and 26 were native, a single, fruitful heavy atom derivative dataset and corresponding native dataset were obtained. High-resolution, native datasets were collected from crystals grown in 0.2 M ammonium tartrate dibasic and 20% PEG3350. Crystals used for anomalous data collection grew in 0.2 M potassium iodide and 20% PEG3350, appearing in 9-12 days, and heavy atom soaks were performed with hexatantalum tetradecabromide (Jena Bioscience GmbH) for

48 hours at ambient temperature. All crystals were cryoprotected with their corresponding mother liquor supplemented with 20% glycerol prior to flash-cooling in liquid nitrogen for data collection. Anomalous and native X-ray data were collected at 100 K on the Berkeley Center for Structural Biology beamline 8.2.2 of the Advanced Light Source at Lawerence Berkeley National Laboratory.

The structure of LigM was phased by the multiple-wavelength anomalous dispersion (MAD) method using anomalous datasets collected at the tantalum peak wavelength (1.25 Å), and a remote wavelength dataset was collected at 1.54 Å. A high-resolution, native dataset was collected at 1.00 Å. Diffraction data were indexed, integrated, and scaled in HKL2000 (8), and MAD phasing, density modification, and initial model building were performed using the program phenix.autosol in the Phenix suite (9, 10). The initial model was 71% complete (339 residues built out of 475), and was rebuilt to a completeness of 97% (460 residues) in phenix.autobuild (11). The subsequent model was used as a search model in molecular replacement (phaser.phenix) to phase a high-resolution, native structure of LigM. Further model building was performed in phenix.autobuild, and additional manual model building and refinement was carried out in Coot and phenix.refine, respectively (9, 12). The tncs correction was set to on during refinement, and TLS refinement was used in the last stages of refinement to generate a final model (13, 14). The asymmetric unit contained four molecules of LigM, which were arranged in two pairs related by translational non-crystallographic symmetry. This asymmetric unit arrangement is likely crystallographic and is not representative of a biological assembly as LigM is a functional monomer and formation of a LigM higher-order assembly has not been documented (15). The most complete chain (Chain A) in the asymmetric unit spanned residues Thr3-Ala467. Chain A, like the others, did not display discernable electron density for the extreme N- and C-termini as these regions were likely highly disordered, and thus, were not modeled into the structure. Additionally, no electron density was discernable for Chain A residues Gly207-Gly210, which reside on a disordered loop, and these residues were also not modeled into the final structure. The refinement statistics for the final model of LigM are given in Table S1. Structure validation was performed in the Phenix suite using MolProbity (16). All structure images were generated in CCP4MG(17) with the exception of the image in Figure S2*C*, which was generated in Chimera (18).

Enzymatic Assays. An HPLC-based activity assay was utilized for an in-depth analysis of the LigM-catalyzed, H_4folate-dependent *O*-demethylation of vanillate, which produces PCA and 5-methyl-H_4folate (15). To determine the optimal pH and temperature reaction conditions for LigM-catalyzed *O*-demethylation of vanillate, 76 nM LigM was incubated with 1 mM vanillate (Sigma-Aldrich Co., ≥97% purity) and 5 mM H_4folate (Caymen Chemical Co., ≥95% purity) in 100 mM buffering agent for 10 minutes at temperatures ranging from 10-80°C, in 10°C steps. Buffering agents included: 100 mM citric acid, pH 4.0; 100 mM citric acid, pH 6.0; 100 mM Tris, pH 8.0; 100 mM CHES, pH 10.0. The reactions were quenched with the addition of methanol to a final concentration of 25%. LigM activity was assessed by HPLC analysis of the quenched reaction mixture to quantify the amount of PCA product formed. Reactions were run on a ZORBAX Eclipse Plus C18 HPLC column (Agilent Technologies) in a mobile phase of 25% acetonitrile and 1% acetic acid with a flow rate of 1 mL/min (15). PCA and vanillate were detected spectrophotometrically at a wavelength of 254 nm and had retention times of 3 and 4.1 minutes, respectively. Under the assay conditions, H_4folate displayed multiple elution peaks (2.1, 2.3, and 2.6 min) that were clearly discernable from PCA and vanillate. No measurable peak was detected for LigM. Reactions were run in duplicate and non-enzyme control reactions were subtracted from the data before analysis. PCA concentration was determined by converting the area under the PCA peak to a corresponding concentration value based on PCA standards. PCA standards were run from 0.01-1 mM and displayed a linear relationship between PCA concentration and peak area.

LigM wild-type, vanillate binding site, and catalytic mutants were assayed using the HPLC-based assay described above with the following changes. Reactions contained 5 mM vanillate, 5 mM H_4folate, 76 nM LigM (wild-type or mutant), and 100 mM Tris, pH 8.0, and were run for 2 hours at 30°C. Assays were run in triplicate and a non-enzyme control reaction was subtracted from the data prior to analysis. For presentation purposes, PCA production for the mutants was set relative to that of WT-LigM.

The pH-rate profile for LigM-catalyzed vanillate demethylation was determined following a protocol similar to that of Reinhardt, *et al.* (19) where H_4folate was held at a saturating concentration while the vanillate concentration was varied at pH 6-10. For each reaction, 76 nM of LigM was incubated with a fixed concentration of H_4folate (5 mM) and a varied vanillate concentration (0.1-5 mM) in 100mM buffering agent. Buffering agents were: 100 mM citric acid, pH 6.0; 100 mM Tris, pH 7.0; 100 mM Tris, pH 8.0; 100 mM Tris, pH 9.0; 100 mM CHES, pH 10.0. Reactions were incubated at 30°C for 10 minutes, and reactions were quenched by the addition of methanol to a final concentration of 25%. The amount of PCA produced was quantified using the HPLC-based method described above. Reactions were run in duplicate and non-enzyme control reactions were subtracted from the data before analysis. The observed reaction rate (the catalytic turnover number, $k_{cat,\ obs}$) was determined by fitting the plot of velocity vs. vanillate concentration generated for each pH with the Michaelis-Menten equation (1) (19, 20):

$$(1)\ v = V_{max}[A]/(K_M + [A])$$

where v is the measured velocity, V_{max} is maximal velocity, [A] is the concentration of vanillate, and K_M is the Michaelis-Menten constant for vanillate. The apparent k_{cat} under these reaction conditions ($k_{cat,\ obs}$) was calculated by replacing the V_{max} term in equation 1 with the equivalent term $k_{cat,obs}[E]_t$, in which $[E]_t$ is the total concentration of enzyme active sites. A Dixon-Webb plot was generated by plotting $\log(k_{cat,\ obs}/K_{M,\ obs})$ vs. pH (Fig. S6) (19-21).

Bi-Substrate Kinetics. For steady-state kinetic analysis of LigM-catalyzed *O*-demethylation of vanillate, assays were performed under conditions that were identified as optimal for activity (30°C, pH 8.0) and that yielded linear, initial velocity profiles with respect to LigM concentration and reaction time. A series of end point assays were run with respect to each substrate (vanillate or H_4folate), where the concentration of one substrate was varied while the concentration of the other substrate was held at different fixed concentrations. With respect to vanillate, the vanillate concentration was varied from 0.1-5 mM while H_4folate was fixed at 0.5, 0.75, 1.5, 5, 10, 15 mM, and for H_4folate, the concentration of H_4folate was varied between 0.5-15 mM, while vanillate was held at 0.1, 0.25, 0.5, 1.5, 5 mM. All reactions contained 76 nM LigM and 100 mM Tris, pH 8.0

in a 500µL total volume, and were incubated at 30°C for 10 minutes. LigM activity was quenched by the addition of methanol to a final concentration of 25%, and PCA production was quantified by HPLC analysis as described above. Assays were run in duplicate or triplicate and the non-enzyme control reactions were subtracted from the sample reactions prior to data analysis. To distinguish between steady-state sequential and steady-state ping-pong mechanisms, data sets were fit with the Michaelis-Menten equation and linearized into Lineweaver-Burk plots for mechanism visualization by means of plot intersection pattern, and the non-linearized, velocity vs. substrate concentration data (Figure S1*A* and *B*) was fit globally with non-linear least-squares regression rate equations for sequential (2) and ping-pong (3) steady-state bireactant mechanisms using PRISM 7.0a (GraphPad Software, Inc., La Jolla, CA) (22-26).

(2) $v = V_{max}[A][B]/(K_{Ma}[B]+K_{Mb}[A]+K_{ia}K_{Mb}+[A][B])$
(3) $v = V_{max}[A][B]/(K_{Ma}[B]+K_{Mb}[A]+[A][B])$

where v is the measured velocity, V_{max} is maximal velocity, [A] and [B] are the concentrations of substrates A and B, K_{Ma} and K_{Mb} are the Michaelis-Menten constants for substrates A and B, and K_{ia} is the dissociation constant for substrate A. Substrate A is the varied substrate and substrate B is the varied, fixed substrate. The catalytic turnover number, k_{cat}, was calculated by replacing the V_{max} term in equation 2 with the equivalent term $k_{cat}[E]_t$, in which $[E]_t$ is the total concentration of enzyme active sites.

Inhibition Kinetics. Product inhibition experiments were performed with LigM product 5-methyl-H_4folate (Sigma-Aldrich Co., ≥88% purity). Inhibition of LigM activity by 5-methyl-H_4folate was determined by varying one substrate at varied, fixed concentrations of 5-methyl-H_4folate and a fixed, saturating concentration of the other substrate. Reactions were run for 10 minutes at 30°C and quenched with 25% methanol. LigM inhibition in the presence of saturating H_4folate (7 mM) was examined by varying vanillate (0.1-1.5 mM) at 5-methyl-H_4folate concentrations of 0, 0.1, 0.3, 0.5, 1.5, 5, and 7 mM. LigM inhibition in the presence of saturating vanillate (4 mM) was investigated by varying H_4folate (0.1-2 mM) at fixed concentrations of 5-methyl- H_4folate (0, 0.005, 0.02, 0.05, 0.1, 0.2 mM). All reactions contained 76 nM LigM and 100 mM Tris, pH 8.0 in a 500 µL total volume. Assays were run in duplicate and the non-enzyme control reactions were subtracted from the sample reactions prior to data analysis. Data sets were linearized and visualized as Lineweaver-Burk plots and the nature of intersection and location of the intersection point guided determination of the type of inhibition caused by the presence of 5-methyl-H_4folate (Fig. 2*C* and *D*). Since the varied vanillate, constant saturating H_4folate plot intersected to the left of the y-axis, 5-methyl-H_4folate was determined to display a non-competitive inhibition with respect to vanillate and thus, the non-linear velocity vs. vanillate concentration data (Fig. S1*C*) was globally fit with a non-competitive inhibition non-linear least-squares regression rate equation (4) in PRISM 7.0a (GraphPad Software, Inc., La Jolla, CA) (21, 23). The varied H_4folate, constant saturating vanillate plot intersected on the y-axis, indicating that 5-methyl-H_4folate acted as a competitive inhibitor with respect to H_4folate, and therefore, the non-linear velocity vs. H_4folate concentration data set (Fig. S2*D*) was fit globally with a competitive inhibition non-linear least-squares regression rate equation (5) in PRISM 7.0a (GraphPad Software, Inc., La Jolla, CA) (21, 23).

(4) $v = (V_{max}[A]/(1+[I]/K_i))/(K_{Ma}+[A])$

(5) $v = V_{max}[B]/((K_{Mb} + K_{Mb}[I]/K_i) + [B]$

where v is measured velocity, V_{max} is maximal velocity, [A] is the concentration of vanillate, [B] is the concentration of H_4folate, K_{Ma} is the Michaelis-Menten constant for vanillate, K_{Mb} is the Michaelis-Menten constant for H_4folate, [I] is the concentration of 5-methyl-H_4folate, and K_i is the inhibition constant for 5-methyl-H_4folate.

Placement of H_4folate and Vanillate Docking. Initial LigM enzyme coordinates were taken from Chain A of the reported crystal structure. Substrate structures were obtained from ChemSpider (27). The coordinates of the H_4folate substrate in the enzyme structure were estimated by sequence and structure alignment (using MultiSeq (28)) to the T-protein of the glycine cleavage system from *Thermotoga maritima* (PDB: 1WOO), which has a similar H_4folate binding domain to LigM. Visualization and image generation was performed with VMD (29, 30). The docking study was performed with AutoDock Vina (31). Input files were prepared with the AutoDockTools(32) program. The raw coordinate data was augmented with polar hydrogen atoms, and the vanillate and 3MGA substrates were prepared with all bonds involving exactly one aryl carbon and a second heteroatom made rotatable. All reported docking calculations were repeated 100 times, with the 'exhaustiveness' parameter (which controls the number of independent runs in a single repeat, each with the ligand mode randomly reset) equal to 12, the maximum number of output modes equal to 20 and the maximum affinity of the output modes limited to 20 kcal mol^{-1} above the mode with the most negative affinity.

An initial 'blind' docking procedure (33) was performed using a grid covering the entire LigM/H_4folate receptor of size (76.6, 70.2, 58.1) Å, placed at the spatial center of the enzyme, **r** = (46.8, 33.5, 30.2). The majority of the discovered vanillate modes (orientations of the ligand relative to the receptor) were in the region of the N_5 atom of H_4folate and this fact, combined with the knowledge that this is where methylation occurs in the product, suggested focusing docking in this region. A grid of size (22.0, 20.0, 26.0) Å was centered on the H_4folate-N_5 atom, at **r** = (37.6, 30.6, 28.8). The AutoDock Vina output revealed three potential binding sites. The best mode in each of the three potential binding sites was selected for further scrutiny. A box of dimension 12.0 x 12.0 x 12.0 Å was centered on each site (using the coordinates of the carboxylate C atom of the ligand from the previous calculation) and of the three optimal binding modes, two appeared to lie too far from N_5 ($r_{\mathrm{C_V \ldots N_5}}$ = 15.3 and 19.7 Å) to be directly involved in the reaction. These modes had affinities (where negative values of affinity imply stronger binding) of -5.7 and -5.4 kcal mol^{-1}. The optimal mode for the third region, centered at **r** = (41.0, 28.5, 23.2), was positioned so as to suggest a direct transfer of the vanillate methyl group to the H_4folate N_5 atom ($r_{\mathrm{C_V \ldots N_5}}$ = 3.5 Å), with protonation of the vanillate oxygen atom by Tyr247 ($r_{\mathrm{O_H \ldots O_M}}$ = 2.6 Å). The affinity of this mode was the strongest at -6.2 kcal mol^{-1}.

Quantum Chemistry Calculations. Initial cluster model heteroatom coordinates were extracted from the geometries produced by the docking procedure (i.e. the crystallographic LigM coordinates, aligned H_4folate coordinates and optimal AutoDock vanillate coordinates) for both Met61 conformations (Met61A, B). Met61 was the only residue in the predicted binding site that was found to exist in multiple conformations in the crystal structure. To examine the involvement of solvent on H_4folate- and vanillate-binding, an additional hydrated cluster was created for each by adding five water molecules at positions near the active site based on water positions present in the crystallographic data. See Table S2 for the complete set of included residues, their charges and computed pK_as. The addition of hydrogen atoms was carried out with Avogadro (34, 35), with the appropriate side chain protonation states assigned using residue-specific pK_a calculations performed with PROPKA 3.1 (36, 37). The pH for protonation was set to 8, as experimental data suggested that the enzyme reached optimal activity at this pH value (Fig. 1*B*). The total charge of the system includes +2 from the amino acid side chains, -2 from H_4folate and -1 from the substrate, for a total charge of -1 au. The spin multiplicity was set to 1. The total number of nuclei in the LigM-vanillate cluster model was 571, and the hydrated vanillate cluster model contained 586 nuclei.

Geometry optimization of the LigM-vanillate systems was performed with TeraChem 1.9 (38, 39) using 3 NVIDIA Tesla K80 dual-GPU accelerators and 2 Intel Xeon E5-2620 v3 CPUs. The protocol used was as follows. Initial calculations were performed at HF/6-31G(d) (40, 41). First, all heteroatom positions were fixed and all Avogadro-generated hydrogen positions were relaxed. Subsequently, the atoms of the vanillate substrate and H_4folate cofactor were released in order to optimize their binding positions in the fixed field of the enzyme. Following this, the side chain constraints on His60, Met61, Arg122 and Tyr247 were released to relax these active site residues and the substrate together. Then, the side chains of Tyr29 and Tyr31, Arg147, Pro248, Asn250, Thr251, Trp256, Pro258 and Phe393, which together surround vanillate, were released and geometry optimized. The side chains of the remaining residues (Gln57/93, Gly107/121, Ile108, Phe110, Val120, Arg163, Gln165, Phe188/189, His206 and Glu215) lining the H_4folate binding cavity were then released and all residues and substrates were geometry optimized with the backbone fixed. On completion, all side chain and substrate atoms were fixed at their final positions and the backbone atom coordinates were optimized to remove any high-energy interactions in the chain, which can be problematic for converging higher-level calculations.

The final geometry was then subjected to geometry optimization at B3LYP(42-45)/6-31G(d) with the DFT-D3 dispersion correction (46) (using the Becke-Johnson damping function (47)), with the backbone fixed. In order to initially converge the self-consistent field (SCF) procedure, the hybrid ADIIS+DIIS (48) scheme was required, and full double precision arithmetic was employed throughout. The final geometries (with and without water) were recombined with the crystal structure without introducing any positional clashes. The final non-hydrated geometry was used to create an initial LigM-3MGA cluster model (572 nuclei) by mutation in Avogadro, which was then optimized at B3LYP/6-31G(d) (with DFT-D3). The LigM-substrate wavefunctions needed for topological analysis were obtained by more tightly converging a single point calculation at the final level of theory above. Two-electron integrals less than 1.0×10^{-11} au were neglected and the wavefunction convergence threshold (in terms of the largest component of the DIIS error vector) was set at 1.0×10^{-8}.

Finally, the Atoms in Molecules analysis of the LigM-substrate wavefunctions was completed with the professional version of AIMAll (49) using 16 CPUs. The analysis was limited to determination of the critical points and connections between bond and nuclear attractor critical points of the molecular graph, i.e. of all existing gradient paths only the atomic interaction lines were determined. Obtaining the full molecular graph was prohibitively expensive for the LigM-vanillate 6-31G(d) wavefunction (of 9196 primitive Gaussians). Conversion of the TeraChem .tcfchk wavefunction output into the required ProAIM .wfn format was achieved with a Perl script, and topological images were generated with Blender (50) via the RhoRix add-on.

Appendix 8 I SI References

1. Bader RFW (1994) *Atoms in Molecules: A Quantum Theory* (Clarendon Press) p 458.
2. Popelier PLA (2000) *Atoms in Molecules: An Introduction* (Prentice Hall) p 188.
3. Boyd RJ & Choi SC (1985) A bond-length-bond-order relationship for intermolecular interactions based on the topological properties of molecular charge distributions. *Chemical Physics Letters* 120(1):80-85.
4. Boyd RJ & Choi SC (1986) Hydrogen bonding between nitriles and hydrogen halides and the topological properties of molecular charge distributions. *Chemical Physics Letters* 129(1):62-65.
5. Carroll MT & Bader RFW (1988) An analysis of the hydrogen bond in BASE-HF complexes using the theory of atoms in molecules. *Molecular Physics* 65(3):695-722.
6. Koch U & Popelier PLA (1995) Characterization of C-H-O Hydrogen Bonds on the Basis of the Charge Density. *The Journal of Physical Chemistry* 99(24):9747-9754.
7. Popelier PLA & Bader RFW (1992) The Existence of an Intramolecular C-H-O Hydrogen-Bond in Creatine and Carbamoyl Sarcosine. *Chem Phys Lett* 189(6):542-548.
8. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307-326.
9. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
10. Terwilliger TC*, et al.* (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65(Pt 6):582-601.
11. Terwilliger TC*, et al.* (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(Pt 1):61-69.
12. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
13. Winn MD, Murshudov GN, & Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 374:300-321.
14. Painter J & Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62(Pt 4):439-450.
15. Abe T, Masai E, Miyauchi K, Katayama Y, & Fukuda M (2005) A tetrahydrofolate-dependent O-demethylase, LigM, is crucial for catabolism of vanillate and syringate in Sphingomonas paucimobilis SYK-6. *J Bacteriol* 187(6):2030-2037.

16. Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12-21.
17. McNicholas S, Potterton E, Wilson KS, & Noble ME (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):386-394.
18. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.
19. Reinhardt LA, Thoden JB, Peters GS, Holden HM, & Cleland WW (2013) pH-rate profiles support a general base mechanism for galactokinase (Lactococcus lactis). *FEBS Lett* 587(17):2876-2881.
20. Allen KG, Stewart JA, Johnson PE, & Wettlaufer DG (1978) Identification of the functional ionic groups of papain by pH/rate profile analysis. *Eur J Biochem* 87(3):575-582.
21. Segel IH (1993) *Enzyme kinetics : behavior and analysis of rapid equilibrium and steady-state enzyme systems* (Wiley, New York) Wiley Classics Library Ed pp xxii, 957 p.
22. Bartling CM & Raetz CR (2008) Steady-state kinetics and mechanism of LpxD, the N-acyltransferase of lipid A biosynthesis. *Biochemistry* 47(19):5290-5302.
23. Cleland WW (1963) The kinetics of enzyme-catalyzed reactions with two or more substrates or products. III. Prediction of initial velocity and inhibition patterns by inspection. *Biochim Biophys Acta* 67:188-196.
24. Emptage RP, Pemble CWt, York JD, Raetz CR, & Zhou P (2013) Mechanistic characterization of the tetraacyldisaccharide-1-phosphate 4'-kinase LpxK involved in lipid A biosynthesis. *Biochemistry* 52(13):2280-2290.
25. Falk MD*, et al.* (2014) Enzyme Kinetics and Distinct Modulation of the Protein Kinase N Family of Kinases by Lipid Activators and Small Molecule Inhibitors. *Biosci Rep* 34(2):93-106.
26. Zhang G, Huang, R., Qi, w., Wang, Y., Su, R., He, Z. (2016) Engineering peptide-based biomimetic enzymes for enhances catalysis. *RSC Advances* 6:40828-40834.
27. http://www.chemspider.com**.**
28. Roberts E, Eargle J, Wright D, & Luthey-Schulten Z (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7(1):1-11.
29. Humphrey W, Dalke A, & Schulten K (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14(1):33-38.
30. Stone J (1998) An Efficient Library for Parallel Ray Tracing and Animation. (University of Missouri-Rolla).
31. Trott O & Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31(2):455-461.
32. Morris GM*, et al.* (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 30(16):2785-2791.
33. Hetenyi C & van der Spoel D (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci* 11(7):1729-1737.
34. Hanwell MD*, et al.* (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* 4(1):1-17.

35. Lokanath NK, Kuroishi C, Okazaki N, & Kunishima N (2004) Purification, crystallization and preliminary crystallographic analysis of the glycine-cleavage system component T-protein from Pyrococcus horikoshii OT3. *Acta Crystallogr D Biol Crystallogr* 60(Pt 8):1450-1452.
36. Søndergaard CR, Olsson MHM, Rostkowski M, & Jensen JH (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* 7(7):2284-2295.
37. Olsson MHM, Søndergaard CR, Rostkowski M, & Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* 7(2):525-537.
38. Ufimtsev IS & Martinez TJ (2009) Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *Journal of Chemical Theory and Computation* 5(10):2619-2628.
39. Kästner J*, et al.* (2009) DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *The Journal of Physical Chemistry A* 113(43):11856-11865.
40. Hariharan PC & Pople JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretica chimica acta* 28(3):213-222.
41. Francl MM*, et al.* (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics* 77(7):3654-3665.
42. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* 98(7):5648-5652.
43. Lee C, Yang W, & Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* 37(2):785-789.
44. Stephens PJ, Devlin FJ, Chabalowski CF, & Frisch MJ (1994) Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* 98(45):11623-11627.
45. Vosko SH, Wilk L, & Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* 58(8):1200-1211.
46. Grimme S, Antony J, Ehrlich S, & Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 132(15):154104.
47. Grimme S, Ehrlich S, & Goerigk L (2011) Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 32(7):1456-1465.
48. Hu X & Yang W (2010) Accelerating self-consistent field convergence with the augmented Roothaan–Hall energy function. *The Journal of Chemical Physics* 132(5):054109.
49. Keith TA (2016) AIMAll (TK Gristmill Software, Overland Park, KS, USA), 16.01.09.
50. Anonymous (2016) Blender), 2.77a.
51. Lee HH, Kim DJ, Ahn HJ, Ha JY, & Suh SW (2004) Crystal structure of T-protein of the glycine cleavage system. Cofactor binding, insights into H-protein

recognition, and molecular basis for understanding nonketotic hyperglycinemia. *J Biol Chem* 279(48):50514-50523.

52. Olsson MH, Sondergaard CR, Rostkowski M, & Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* 7(2):525-537.
53. Sondergaard CR, Olsson MH, Rostkowski M, & Jensen JH (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput* 7(7):2284-2295.
54. Tipton KF & Dixon HB (1979) Effects of pH on enzymes. *Methods Enzymol* 63:183-234.
55. Milstein S, Kapatos, G., Levine, R.A., Shane, B. (2002) *Chemistry and Biology of Pteridines and Folates: Proceedings of the 12th International Symposium on Pteridines and Folates, National Institutes of Health, Bethesda, Maryland, June 17-22, 2001* (Springer US).

Appendix 9 I SI Figures

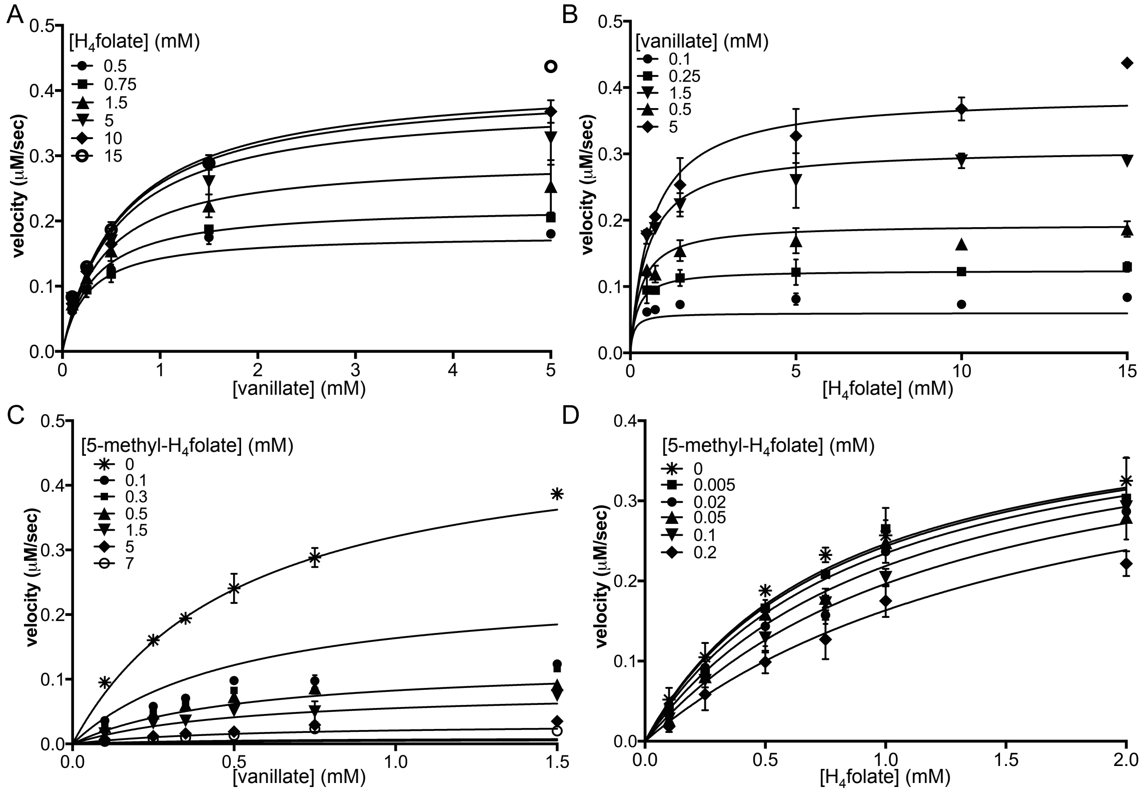


Fig. S1. Kinetic Data: velocity vs. substrate concentration. Kinetic data was collected and plotted as non-linear velocity (μM sec^{-1}) vs. substrate concentration (mM) curves and kinetic parameters were determined by globally fitting each family of curves with the appropriate non-linear least-squares rate equation using *PRISM 7.0a* (GraphPad, Inc.). (*A*) The curves of velocity vs. vanillate concentration at different fixed H$_4$folate concentrations (mM) and (*B*) the curves of velocity vs. H$_4$folate concentration at different fixed vanillate concentrations (mM) were fit with ordered sequential rate equations. For product inhibition experiments, 5-methyl-H$_4$folate was used as a product inhibitor at different fixed concentrations (*C*) while H$_4$folate was held at a saturating concentration and the vanillate concentration was varied, and (*D*) while vanillate was held at a saturating concentration and H$_4$folate was varied. The family of curves in (*C*) was fit with a non-competitive rate equation and the family of curves in (*D*) was fit with a competitive rate equation. Data shown is mean with S.E.M. Lineweaver-Burk plots of these data are depicted in Fig. 2.

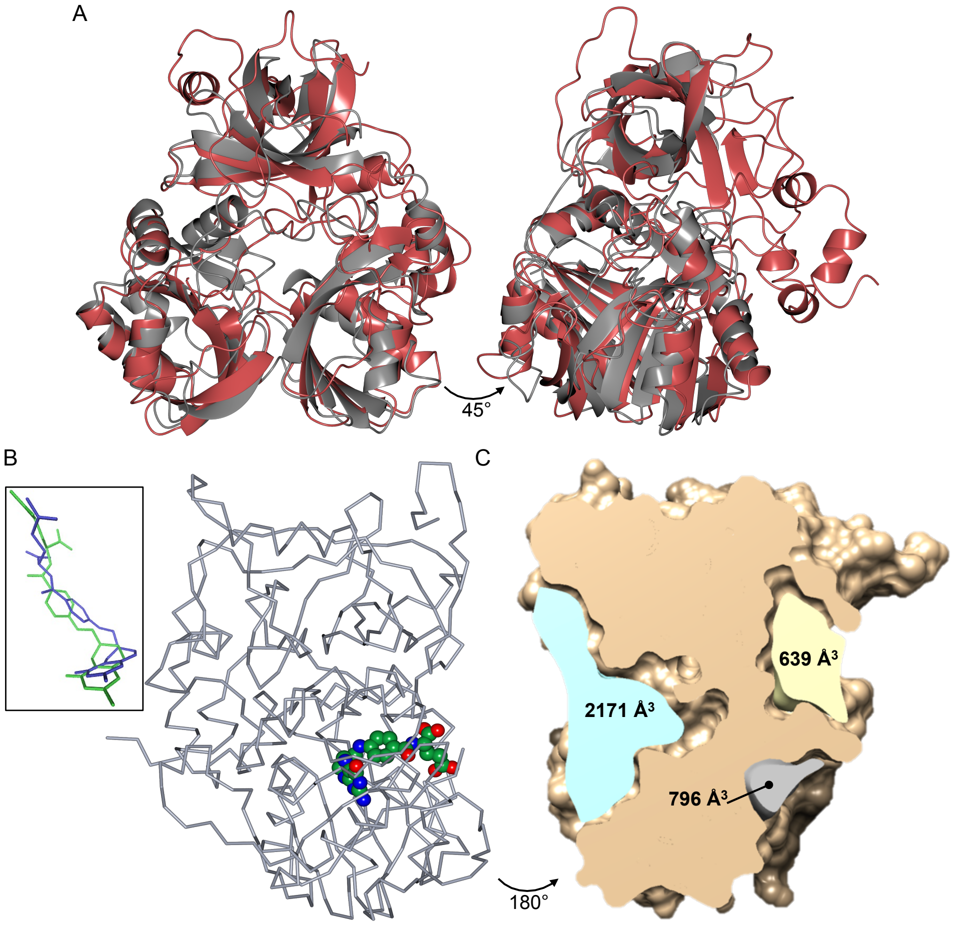


Fig. S2. Homolog alignments and analysis of LigM cavity space. (*A*) LigM contains unique structural features beyond the canonical folate-binding domain that it shares with its structural homologs. LigM (red) is shown aligned to its closest structural homolog (*Tm*_T, PDB: 1WOS) (gray) (51). The proteins align well through their respective tripartite folate-binding domains; however, LigM has novel structural elements, largely composed of sequence insertions in connecting loop regions of the folate-binding domain. This novel fold resides predominantly on one face of LigM (as indicated in Fig. 4*A*). (*B*) LigM's H_4folate-binding site was identified through structural homolog alignments and the geometry of H_4folate and the surrounding LigM residues were optimized with DFT calculations. Inset depicts the superposition of the conformation of H_4folate pre-geometry optimization (blue) and post-optimization (green). The RMSD between the conformations is 1.69 Å. As observed in homologous structures, H_4folate is predicted to bind to LigM in a kinked conformation with its pteridine ring bent at a near 90° angle towards Domain 1 and 2 of the canonical folate-binding fold. The structure of LigM is depicted in gray and H_4folate is colored according to element (green, carbon; red, oxygen; blue, nitrogen). (*C*) To provide initial insight into potential locations of LigM's second substrate-binding site, a cavity search of the LigM structure was performed in Chimera (18). Three major cavities were identified and their respective volumes are indicated. The largest cavity (blue) corresponds to LigM's H_4folate-binding site.

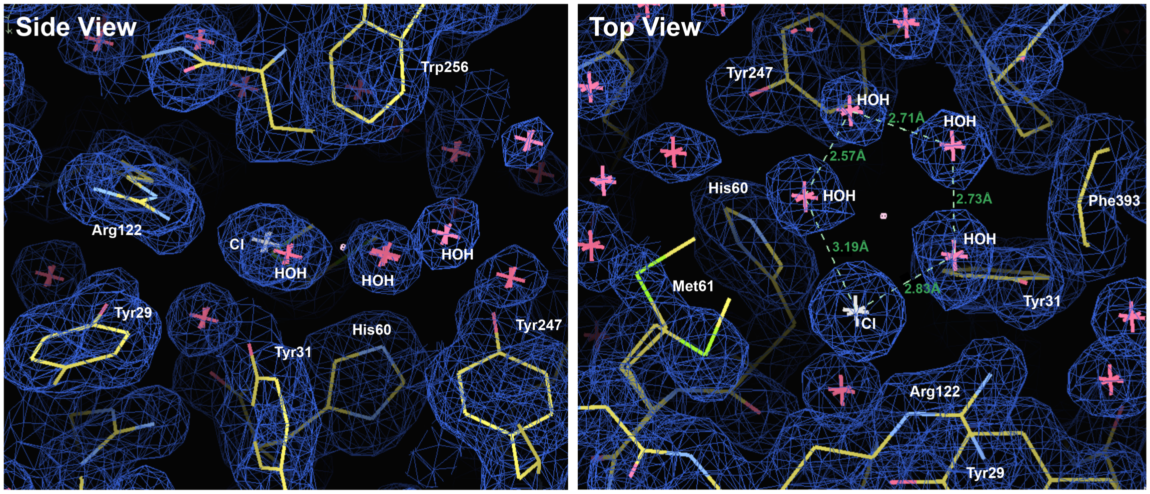
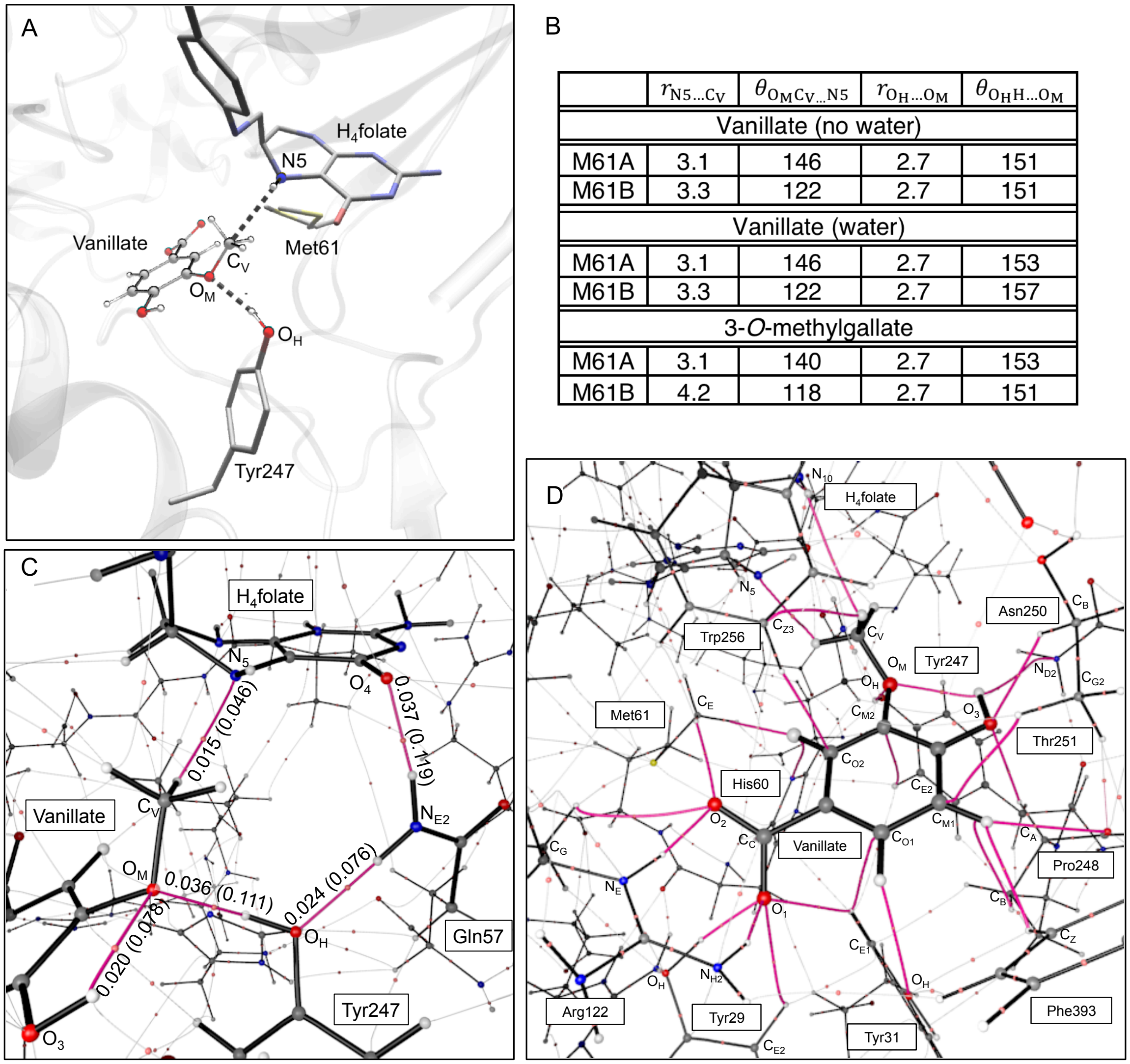
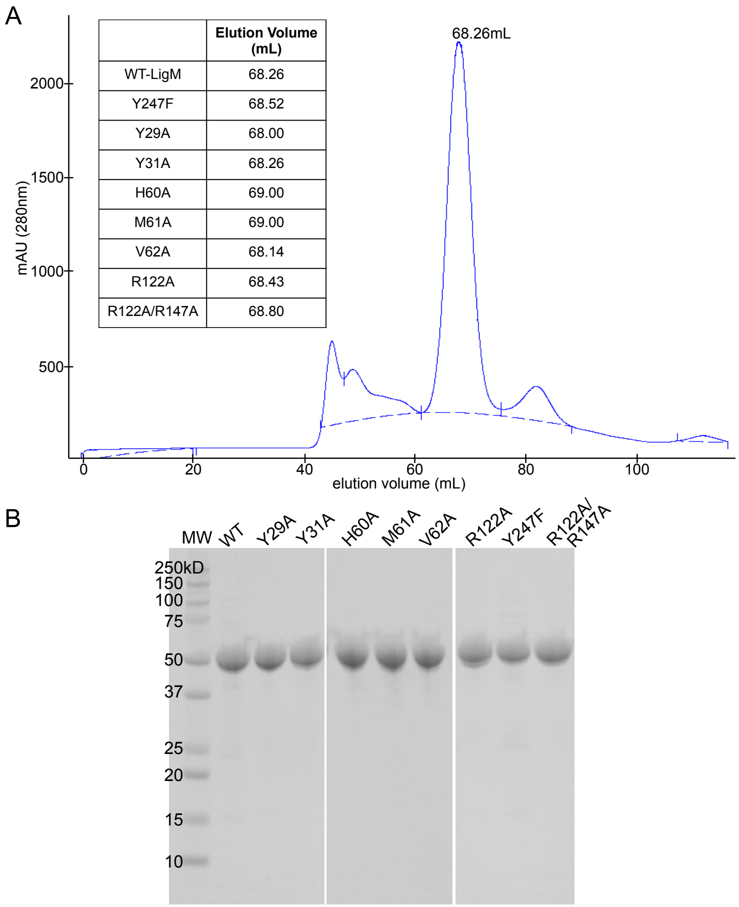


Fig. S3. Vanillate-binding site identification from electron density. Inspection of the LigM electron density map identified four unusually ordered water molecules and a chloride ion in an internal cavity of LigM, roughly 5.3 Å from the predicted H_4folate-binding site. These molecules sit on relatively the same plane, as viewed from the side, and viewed from the top, they appear to be arranged in a pentagon-like formation. The water molecules are spaced roughly 2.7 Å from each other and the two waters located nearest to the chloride are approximately 3 Å from the chloride ion. The water molecules are labeled HOH, the chloride ion is labeled Cl, and the surrounding LigM residues are labeled: Tyr29, Tyr31, His60, Met61, Arg122, Tyr247, Trp256, and Phe393. The measured distances between molecules are shown in green next to the distance, represented as a dashed line, to which they correspond. The $2F_o - F_c$ model-phased electron density map is illustrated in blue with the final refined model of LigM and is contoured at 1σ. Further evidence suggesting the importance of this cavity came from a residue-specific pK_a analysis using PROPKA 3.1 (52, 53), which indicated that out of the 461 residues resolved in the LigM structure, only two, His60 and Arg122, were calculated to exhibit dramatically altered pK_a's due to their surrounding electronic environment. Such dramatic perturbations of pK_a, especially involving two residues located within the same solvent-accessible cavity, suggested that the environment of this cavity could play a role in substrate interaction and/or catalysis.



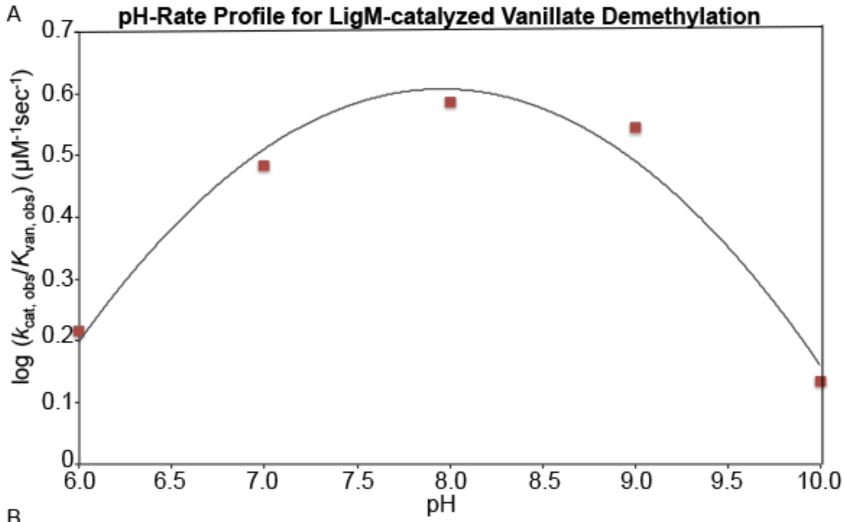
	$r_{\mathrm{N5...Cv}}$	$\theta_{\mathrm{O_MCv...N5}}$	$r_{\mathrm{O_H...O_M}}$	$\theta_{\mathrm{O_HH...O_M}}$
Vanillate (no water)				
M61A	3.1	146	2.7	151
M61B	3.3	122	2.7	151
Vanillate (water)				
M61A	3.1	146	2.7	153
M61B	3.3	122	2.7	157
3-*O*-methylgallate				
M61A	3.1	140	2.7	153
M61B	4.2	118	2.7	151

Fig. S4. Identification of LigM's aryl-binding site. (*A*) Optimized active site cluster model geometry for LigM-vanillate (no water) superimposed on the entire enzyme structure (crystallographic coordinates). Atoms directly involved in the reaction are highlighted, and those forming $\mathrm{O_MH}$ and $\mathrm{C_VN_5}$ bonds are connected with dotted lines. Both experimentally observed Met61 orientations are shown. (*B*) All the geometric properties relevant to the reaction mechanism for all clusters, i.e. those with vanillate (with and without water) and those with 3MGA, are shown. Distances are in Angstroms (Å) and angles are in degrees. (*C*) Interactions identified by Atoms in Molecules calculations that are directly involved in the catalytic action of LigM are shown in pink. Only nuclear attractor and bond critical points are shown, along with their connecting atomic interaction lines. Each atomic interaction line is marked with its value of $\rho(\mathbf{r}_{\mathrm{cp}})$ followed by the value of $\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$ in parentheses. (*D*) Interactions identified by Atoms in Molecules calculations between the vanillate substrate and the enzyme residues in the binding site of LigM. Vanillate is drawn with thick bonds and the surrounding protein residues are displayed with thin bonds and labeled accordingly.



	Elution Volume (mL)
WT-LigM	68.26
Y247F	68.52
Y29A	68.00
Y31A	68.26
H60A	69.00
M61A	69.00
V62A	68.14
R122A	68.43
R122A/R147A	68.80

Fig. S5. Solution behavior of WT-LigM and vanillate-binding site/catalytic site mutants. (*A*) WT-LigM and vanillate-binding site and catalytic site mutants, Y29A, Y31A, H60A, M61A, V62A, R122A, R122A/R147, and Y247F, were purified first by affinity chromatography and finally by size exclusion chromatography on a Superdex 200 HiLoad 16/600 column (GE Healthcare). The WT-LigM size exclusion elution profile is shown and the peak at 68.26 mL corresponds to nearly pure WT-LigM running as a dimer. Protein from this peak was used for all subsequent experiments. The elution volumes for the mutants are displayed in the inset chart. All mutants were soluble, behaving similarly to WT-LigM in solution as determined by the essentially identical size exclusion elution volume for WT-LigM and the mutants. Size exclusion chromatography provides a measure of relative shape and size, thus similarity in elution volume is indicative of similar overall structure and solution behavior (monomer, dimer, etc.) when comparing WT-LigM to the single point mutants. (*B*) To assess the purity of WT-LigM and the mutants, the protein from each major size exclusion peak (~68-69 mL) was concentrated and 20 µg was visualized by SDS-PAGE (8-16% acrylamide tris/glycine gels) with Coomassie staining. The LigM construct plus 12xHistidine tag ran at a molecular weight (MW) of approximately 50 kD.



B

pH	$k_{cat, obs}$ (sec^{-1})†	$K_{vanillate, obs}$ (μM)†
6.0	0.93 +/ 0.04	0.57 +/- 0.07
7.0	1.72 +/- 0.05	0.56 +/- 0.05
8.0	1.76 +/- 0.06	0.45 +/- 0.05
9.0	2.36 +/- 0.11	0.67 +/- 0.10
10.0	1.41 +/- 0.4	1.03 +/- 0.07

†Mean +/- S.E.M.

Fig. S6. Effect of pH on LigM-catalyzed *O*-demethylation of vanillate. (*A*) The Dixon-Webb pH-rate profile of log ($k_{cat, obs}$/$K_{van, obs}$) vs. pH indicates that LigM achieves maximal catalytic efficiency ($k_{cat, obs}$/$K_{van, obs}$) between pH 7-9 with a peak at pH 8. Using the method described by Tipton and Dixon (54), pK_a's for two different ionizing groups were identified with values of 5.6 and 10.3. These pK_a values are hypothesized, based on structural, biochemical, and computational analyses, to correspond to the pK_a's of N_5 of H_4folate (reported pK_a=5.1 (55)) and Tyr247, respectively. Thus, at pH 8, N_5-H_4folate would be deprotonated while Tyr247 would be protonated, enabling proper reaction progression. The importance of Tyr247 for LigM-catalyzed vanillate *O*-demethylation was confirmed via site-directed mutagenesis (Y247F) (Fig. 5). (*B*) The observed kinetic constants at each pH are displayed as their mean with S.E.M. These apparent constants were determined by holding the concentration of H_4folate at a saturating concentration and varying the concentration of vanillate.

Appendix 10 | SI Tables

Table S1. Data collection and refinement statistics for LigM (PDB code: 5TL4)

	Native[a]	Ta_6Br_{12}-soak[a]	
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	
Cell dimensions			
a, *b*, *c* (Å)	112.84, 126.18, 155.65	112.84, 126.58, 154.89	
α, β, γ (°)	90, 90, 90	90, 90, 90	
		Peak (Ta)	*Remote*
Wavelength	1.0	1.25	1.54
Resolution (Å)[b,c]	49.01 - 1.75 (1.81 - 1.75)	50.0 - 2.2 (2.28 - 2.2)	50.0 - 2.71 (2.81 - 2.7)
R_{meas} (%)	14.3 (0.0[d])	19.0 (65.4)	23.2 (53.8)
$I/\sigma(I)$	14.33 (1.23)	18.11 (2.78)	17.85 (2.8)
$CC_{1/2}$	0.99 (0.67)	0.99 (0.89)	0.99 (0.92)
Completeness (%)	99.8 (97.8)	99.56 (96.9)	99.42 (95.5)
Redundancy	7.1 (5.3)	7.2 (6.0)	7.1 (5.8)
Refinement			
Resolution (Å)	49.01 - 1.75		
No. reflections	222,187 (21,801)		
R_{work} / R_{free} (%)	16.94 / 21.21		
No. atoms	17,087		
Protein	14,399		
Ligand/ion	34		
Water	2,654		
B-factors	22.07		
Protein	20.03		
Ligand/ion	30.94		
Water	33.04		
R.m.s deviations			
Bond lengths (Å)	0.006		
Bond angles (°)	0.79		
Molprobity Analysis[e]			
Ramachandran Plot			
Favored (%)	98		
Allowed (%)	1.8		
Outliers (%)	0		
Rotamer outliers (%)	0.067		
Clash score	3.73		

[a] One crystal used for native dataset and one crystal used for Ta_6Br_{12} datasets. [b] Values in parentheses are for highest-resolution shell. [c] The resolution limit according to an $I/\sigma(I)$ of 2 is 1.81 Å for the native dataset. [d] Data was processed in HKL2000 (scalepack) where R_{meas} > 100% is reported as 0. [e] S.I. Reference 16.

Table S2. Summary of LigM cluster model residues. Residues retained in the cluster model of LigM, their pK_a values if ionizable (as computed by PROPKA 3.1 (52, 53)) and contributions to the total charge of the system. These cluster model residues were used in subsequent LigM-H_4folate, LigM-vanillate, and LigM-3MGA calculations.

LigM Residue	pK_a	Charge	LigM Residue	pK_a	Charge
Tyr29	9.34	0	Arg163	14.31	+1
Tyr31	12.16	0	Gln165	-	-
Gln57	-	-	Phe188	-	-
His60	2.81	0	Phe189	-	-
Met61	-	-	His206	4.17	0
Gln93	-	-	Glu215	6.34	-1
Gly107	-	-	Tyr247	12.68	0
Ile108	-	-	Pro248	-	-
Phe110	-	-	Asn250	-	-
Val120	-	-	Thr251	-	-
Gly121	-	-	Trp256	-	-
Arg122	15.37	+1	Pro258	-	-
Arg147	11.99	+1	Phe393	-	-

Table S3. **Summary of LigM-H_4folate cluster interactions.** X-Y interatomic distances and AIM properties (the electron density, $\rho(\mathbf{r}_{\mathrm{cp}})$, and Laplacian, $\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$), measured at the bond critical point of each listed interaction XH…Y between LigM and H_4folate in the LigM-vanillate cluster model.

Donor	X	Y	Acceptor	r_{XY} (Å)	$\rho(\mathbf{r}_{\mathrm{cp}})$	$\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$
Intramolecular Interactions						
H_4folate	N	O_{X1}	H_4folate	2.50	0.0430	0.1426
H_4folate	N-H	H-$C_{2'}$	H_4folate	2.84 (1.85)	0.0154	0.0657
H_4folate	C_B	O_{11}	H_4folate	3.20	0.0114	0.0402
H_4folate	C_G	O_{X2}	H_4folate	3.16	0.0098	0.0346
H_4folate	$C_{5'}$	O_4	H_4folate	4.69	0.0005	0.0025
Intermolecular Interactions						
Gln93	N_{E2}	O_{X1}	H_4folate	2.69	0.0448	0.1452
Gln57	N_{E2}	O_4	H_4folate	2.80	0.0370	0.1189
H_4folate	N_8	O	Val120	2.88	0.0295	0.0991
H_4folate	N_2	O_{E1}	Glu215	2.96	0.0256	0.0833
Phe188	C_{D2}	O_{E2}	H_4folate	3.09	0.0220	0.0678
Phe189	C_{D2}	O_{E2}	H_4folate	3.09	0.0214	0.0709
H_4folate	$C_{6'}$	N_{E2}	His206	3.12	0.0161	0.0518
Phe188	C_{E2}	O_{11}	H_4folate	3.21	0.0157	0.0560
Phe188	C_B	O_{E2}	H_4folate	3.33	0.0131	0.0399
H_4folate	$C_{3'}$	C_{D1}	Trp256	3.32	0.0130	0.0466
Met61	C_E	N_8	H_4folate	3.56	0.0120	0.0346
Val120	C_B-H	H-N_1	H_4folate	3.58 (1.94)	0.0110	0.0424
Gln165	N_{E2}	O_4	H_4folate	3.15	0.0109	0.0388
Phe189	C_{E2}	O_{11}	H_4folate	3.46	0.0101	0.0341
H_4folate	C_6	C_{H2}	Trp256	3.51	0.0092	0.0280
Phe110	C_{E1}-H	H-N_2	H_4folate	3.44 (2.07)	0.0091	0.0389
H_4folate	N_1	O	Val120	3.30	0.0087	0.0331
H_4folate	$C_{5'}$	N_{E2}	His206	3.31	0.0075	0.0261
Ile108	C_{G1}	N_1	H_4folate	3.41	0.0073	0.0237
Phe189	C_Z	$C_{6'}$	H_4folate	3.69	0.0066	0.0217
Val120	C_{G1}	N_1	H_4folate	3.37	0.0058	0.0207
Ile108	C_{G1}	N_8	H_4folate	3.78	0.0056	0.0195
H_4folate	N_{10}	C_{D2}	Trp256	3.85	0.0053	0.0168
Val120	C_B	N_2	H_4folate	3.65	0.0052	0.0177
Val120	C_{G1}	N_2	H_4folate	3.65	0.0052	0.0176
H_4folate	C_7-H	H-C_{H2}	Trp256	3.62 (2.50)	0.0051	0.0169
Ile108	C_{G1}-H	H-C_7	H_4folate	4.03	0.0049	0.0184
Ile108	C_{D1}-H	H-C_9	H_4folate	4.06	0.0044	0.0163
H_4folate	C_9	C_{Z2}	Trp256	3.86	0.0040	0.0115
H_4folate	C_9	C_{G1}	Ile108	3.95	0.0037	0.0135
H_4folate	$C_{2'}$	N_{E2}	Gln93	3.92	0.0028	0.0106
Gly107	C_A-H	H-C_7	H_4folate	4.47 (2.73)	0.0025	0.0080
H_4folate	$C_{2'}$	O_{E1}	Gln93	3.82	0.0023	0.0101
Arg163	C_D	N_2	H_4folate	4.33	0.0017	0.0062

Table S4. Summary of LigM-aryl cluster interactions. (*A*) X-Y interatomic distances and AIM properties (the electron density, $\rho(\mathbf{r}_{\mathrm{cp}})$, and Laplacian, $\nabla^2\rho(\mathbf{r}_{\mathrm{cp}})$), measured at the bond critical point of each listed interaction XH…Y (from Fig. S4*C* and *D*) in the LigM-vanillate cluster model. (*B*) Acceptor/donor groups and values of the electron density (au) for interactions present only in the LigM-3MGA cluster. The corresponding (absent) interactions from the LigM-vanillate cluster are marked with superscript numbers.

Donor	X	Y	Acceptor	$r_{\mathbf{XY}}$ (Å)	$\rho(\mathbf{r}_{\mathbf{cp}})$	$\nabla^2\rho(\mathbf{r}_{\mathbf{cp}})$
A) LigM-Vanillate Cluster Interactions						
Tyr29	O	O_1	Vanillate	2.60	0.0618	0.1715
Arg122	N_{H2}	O_1	Vanillate	2.75	0.0412	0.1278
Arg122	N_E	O_2	Vanillate	2.81	0.0396	0.1171
Tyr247	O_H	O_M	Vanillate	2.72	0.0359	0.1105
Gln57	N_{E2}	O_H	Tyr247	3.00	0.0241	0.0762
Vanillate	O_3	O_M	Vanillate	2.69	0.0203	0.0784
Pro248	C_A	O_3	Vanillate	3.23	0.0152	0.0484
Vanillate	C_V	N_5	H_4folate	3.10	0.0146	0.0461
Tyr29	C_{E2}	O_1	Vanillate	3.07	0.0114	0.0448
Tyr31	C_{E1}	O_1	Vanillate	3.36	0.0103	0.0355
^{1}Vanillate	C_{M1}	O	Pro248	3.42	0.0101	0.0339
Met61	C_E	O_2	Vanillate	3.37	0.0099	0.0344
Thr251	C_B	O_3	Vanillate	3.43	0.0092	0.0309
Vanillate	C_{O1}	O_H	Tyr31	3.36	0.0090	0.0326
Vanillate	C_{O2}	C_E	Met61	3.76	0.0080	0.0310
Tyr31	C_{E1}	C_{O1}	Vanillate	3.50	0.0079	0.0274
Trp256	C_{Z3}	C_{O2}	Vanillate	3.79	0.0066	0.0207
Arg122	C_G	O_2	Vanillate	3.48	0.0062	0.0226
Pro258	C_G	C_{O1}	Vanillate	3.40	0.0059	0.0186
^{2}Pro248	C_B-H	H-C_{M1}	Vanillate	3.41	0.0057	0.0214
Pro258	C_B	O_2	Vanillate	3.69	0.0050	0.0200
^{3}Vanillate	C_{M1}-H	H-C_Z	Phe393	3.90	0.0043	0.0169
^{4}Thr251	C_{G2}	C_{M1}	Vanillate	3.74	0.0041	0.0134
Vanillate	C_{O2}	N_{E2}	His60	3.59	0.0039	0.0137
Vanillate	C_V	C_{Z3}	Trp256	3.91	0.0037	0.0115
Vanillate	O_3	N_{D2}	Asn250	3.55	0.0036	0.0155
Tyr247	C_{E2}	C_{M2}	Vanillate	3.53	0.0033	0.0109
Vanillate	C_V	N_{10}	H_4folate	3.74	0.0027	0.0101
His60	C_{D2}	C_C	Vanillate	4.22	0.0014	0.0053
B) LigM-3MGA Differing Cluster Interactions						
^{3}3MGA	O_2	C_Z	Phe393	-	0.0139	-
^{1}Pro248	O	O_2	3MGA	-	0.0185	-
^{4}Thr251	C_{G2}	O_2	3MGA	-	0.0090	-
^{2}Pro248	C_B	O_2	3MGA	-	0.0073	-
Pro258	C_G-H	H-O_2	3MGA	-	0.0053	-