Title

Molecular characterization of novel pyridoxal-5′-phosphate-dependent enzymes from the human microbiome

Journal

Protein Science, 23(8)

ISSN

0961-8368

Authors

Fleischman, Nicholas M
Das, Debanu
Kumar, Abhinav
et al.

Publication Date

2014-08-01

DOI

10.1002/pro.2493

Peer reviewed

THE PROTEIN SOCIETY

Molecular characterization of novel pyridoxal-5′-phosphate-dependent enzymes from the human microbiome

Nicholas M. Fleischman,[1] Debanu Das,[2,3] Abhinav Kumar,[2,3] Qingping Xu,[2,3] Hsiu-Ju Chiu,[2,3] Lukasz Jaroszewski,[2,4,5] Mark W. Knuth,[2,6] Heath E. Klock,[2,6] Mitchell D. Miller,[2,3] Marc-André Elsliger,[2,7] Adam Godzik,[2,4,5] Scott A. Lesley,[2,6,7] Ashley M. Deacon,[2,3] Ian A. Wilson,[2,7]* and Michael D. Toney[1]

[1]Department of Chemistry, University of California, Davis, California 95616
[2]Joint Center for Structural Genomics, http://www.jcsg.org
[3]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025
[4]Center for Research in Biological Systems, University of California, San Diego, La Jolla, California 92093
[5]Program on Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, California 92037
[6]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121
[7]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037

Received 2 April 2014; Accepted 27 May 2014
DOI: 10.1002/pro.2493
Published online 29 May 2014 proteinscience.org

Abstract: Pyridoxal-5′-phosphate or PLP, the active form of vitamin B6, is a highly versatile cofactor that participates in a large number of mechanistically diverse enzymatic reactions in basic metabolism. PLP-dependent enzymes account for ~1.5% of most prokaryotic genomes and are estimated to be involved in ~4% of all catalytic reactions, making this an important class of enzymes. Here, we structurally and functionally characterize three novel PLP-dependent enzymes from bacteria in the human microbiome: two are from *Eubacterium rectale,* a dominant, nonpathogenic, fecal, Gram-positive bacteria, and the third is from *Porphyromonas gingivalis,* which plays a major role in human periodontal disease. All adopt the Type I PLP-dependent enzyme fold and

Abbreviations: AAT, aspartate aminotransferase; APS, Advanced Photon Source; ArAT, aromatic amino acid aminotransferase; DHDH, d-2-hydroxyisocaproate dehydrogenase; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GDH, glutamate dehydrogenase; HEPES, 4-(2-hydroxyethyl)−1-piperazineethanesulfonic acid; IPTG, isopropyl β-d-1-thiogalactopyranoside; JCSG, Joint Center for Structural Genomics; KAT, kynurenine aminotransferase; LB, Luria-Bertani broth; LDH, lactate dehydrogenase; MAD, multiwavelength anomalous diffraction; MDH, malate dehydrogenase; MSE, selenomethionine; NADH, nicotinamide adenine dinucleotide; OAS, O-acetyl-l-serine; OPA, o-phthaldialdehyde; pArAt, Paracoccus denitrificans aromatic amino acid aminotransferase; PCR, polymerase chain reaction; PEG, polyethylene glycol; PHP, 3-phosphohydroxypyruvate; PIPE, Polymerase Incomplete Primer Extension; PLP, pyridoxal-5′-phosphate; PLPDE, PLP-dependent enzyme; PMP, pyridoxamine-5′-phosphate; PP, phenylpyruvate; PS, phosphoserine; PSAT, phosphoserine aminotransferase; PSI, Protein Structure Initiative; SAM, Stanford Automated Mounting; SDS-PAGE; sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEC, size-exclusion chromatography; SSM, Secondary Structure Matching; SSRL, Stanford Synchrotron Radiation Lightsource; TAT, tyrosine aminotransferase; TCEP, tris(2-carboxyethyl)phosphine; TEA, triethylamine; TEV, Tobacco Etch Virus.

Additional Supporting Information may be found in the online version of this article.

Conflict of interest: The authors declare no conflict of interest in this work.

Grant sponsor: Protein Structure Initiative; Grant number: U54 GM094586; Grant sponsor: NIGMS Administrative Supplement for Functional Studies; Grant number: 3R01GM054779-13S1.

*Correspondence to: Ian A. Wilson, Joint Center for Structural Genomics, The Scripps Research Institute, BCC206, 10550 N. Torrey Pines Road, La Jolla, CA 92037. E-mail: wilson@scripps.edu (or) M. D. Toney, Department of Chemistry, University of California, Davis, CA 95616. E-mail: mdtoney@ucdavis.edu

structure-guided biochemical analysis enabled functional assignments as tryptophan, aromatic, and probable phosphoserine aminotransferases.

Keywords: human microbiome; PLP-dependent enzymes; crystal structure; biochemical characterization; structural genomics; Protein Structure Initiative

Introduction

To contribute to structure-based functional annotation of uncharacterized proteins, the Protein Structure Initiative (PSI) in PSI-1 and PSI-2 focused on structural coverage of under-represented sectors of protein sequence space.[1] For example, the PSI has to date determined structures of ~70 novel pyridoxal-5′-phosphate (PLP, the active form of vitamin B6)-dependent enzymes (PLPDEs), where novelty is defined as <30% sequence identity to any structure in the PDB at the outset of the study. Prior to structure determination, these proteins were either annotated as proteins of unknown function or as members of the very large and diverse aminotransferase superfamily. Forty of these PLP-dependent structures have been determined by the Joint Center for Structural Genomics (JCSG[2]; see TOPSAN[3] page: http://www.topsan.org/Groups/PLP-dependent_aminotransferase).

PLPDEs are classified into five different fold types.[4–7] Fold-Type I, as represented by aspartate aminotransferase, is by far the largest group of PLPDEs with eight known subclasses, three of which contain aminotransferases, two contain lyases, one contains decarboxylases, and two other smaller subclasses include serine hydroxymethyltransferase and 3-amino-5-hydroxybenzoic acid synthase.[6] These aminotransferases are obligatory dimers since the active site spans two subunits.

Despite structural similarities within fold types,[8] the catalytic activity of a particular PLPDE cannot simply be assigned based on its fold. All PLPDE fold types, except Type II, contain enzymes that catalyze more than one reaction. Together, PLPDEs catalyze up to 160 different functions by some estimates.[9] Here, we describe crystal structures of three novel PLPDEs from the human microbiome with PLP bound. These proteins were initially annotated as proteins of unknown function: Uniprot[10] accession code C4ZCJ5_EUBR3 (PDB id 3ele) and C4ZGW0_EUBR3 (PDB id 3f0h) from *Eubacterium rectale* (henceforth referred to by their Ordered Locus Names EUBREC_0560 and EUBREC_2651, respectively), a dominant nonpathogenic Gram-positive fecal bacteria; and Q7MUZ3_PORGI (PDB id 3g0t) from the periodontal disease strain *Porphyromonas gingivalis* W83 (henceforth referred to by its Ordered Locus Name PG_1327). Their crystal structures revealed that they adopt the PLPDE Type I fold. Through a supplement program for the biochemical characterization of novel proteins whose structures have been determined by the PSI, we investigated the biochemistry of these PLPDEs.

Results

Overall structure

The cloning, expression, purification, and crystallization of EUBREC_0560, EUBREC_2651, and PG_1327 were performed using standard JCSG protocols as described in "Materials and Methods." All three crystal structures were determined by the MAD method to 2.1 Å (EUBREC_0560), 1.7 Å (EUBREC_2651), and 1.75 Å (PG_1327) resolution (see "Materials and Methods"). The data collection, model, and refinement statistics are summarized in Supporting Information Table S1. All three proteins adopt the PLPDE Type I fold (the aspartate aminotransferase fold), consisting of a large domain composed of a seven-stranded β-sheet (N-terminus) and a small domain composed of a four-stranded β-sheet (C-terminus), flanked by ~16 α-helices.[6] The EUBREC_2651 (residues 1–357) and PG_1327 (residues 1–436) structures have endogenous PLP (derived during protein expression) bound in a Schiff base linkage to their respective active site lysine, while the EUBREC_0560 (residues 1–397) structure has PLP bound in the active site with no covalent linkage. The architecture of each enzyme depicted in Figure 1 reveals a triangular shape with the active site (Fig. 2) residing in a cavity on the concave side. The additional 40-residue segment in EUBREC_0560, as compared to EUBREC_2651, consists primarily of two tandem helices at the N-terminus that are proximal to the C-terminal domain. PG_1327 is the largest in size amongst these three PLPDEs and the additional residues primarily reside at the N-terminus where they form three helices that pack against the C-terminal domain.

Structure similarity searches were performed with SSM (Secondary Structure Matching) using the protein structure comparison service, PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm;[11] Table I). The listed SSM hits represent biochemically characterized aminotransferases with an average sequence identity of 23% and none >30%. The SSM results indicate that EUBREC_0560 is most similar to two aspartate aminotransferases, as well as a tyrosine and an alanine aminotransferase. EUBREC_2651 has greater similarity to enzymes that bind smaller substrates, including two alanine-glyoxylate

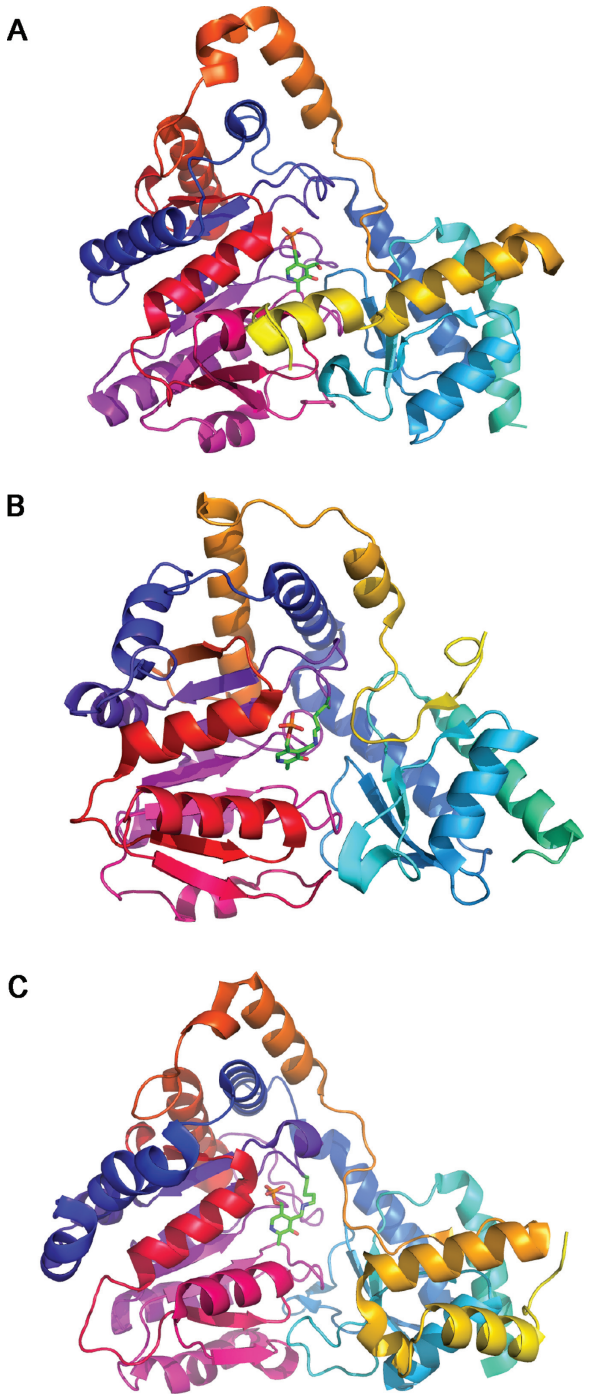

Figure 1. Crystal structures of (A) EUBREC_0560, (B) EUBREC_2651, and (C) PG_1327, colored from N- to C-terminus (yellow, red, magenta, blue, cyan, and green). PG_1327 is larger and the extra portion consists of helices at the N-terminus that are positioned near the C-terminal domain.

aminotransferases and a 2-aminoethylphosphonate aminotransferase. PG_1327 is most similar to an aspartate aminotransferase and a glutamine aminotransferase.

Pfam 27.0 (March 2013) clearly assigns both EUBREC_0560 and PG_1327 to Pfam Aminotransferase Classes I and II family (PF00155), a family of 52,439 proteins from 5588 species for which ~90 unique protein structures are known. EUBREC_2651 is assigned to Pfam Aminotransferase Class V family (PF00266), which includes 23,746 proteins from 5192 species for which ~35 unique structures are known.

A multiple structure superimposition of EUBREC_0560 and PG_1327 with 12 representative family members with annotated function and known structure (out of the 90 above) was performed with POSA[12] (Supporting Information Fig. S1). All share a common core of 228 residues that can be aligned with an average (for all residues in the common core for all pairs) RMSD of 2.75 Å based on Cα atoms. A structure-based multiple sequence alignment with the nearest structures (8 out of 12) from the multiple structure comparison performed using T-Coffee[13] (http://www.tcoffee.org/) revealed 9 identical and 9 other highly conserved residues (Supporting Information Fig. S2). These residues include the Schiff base lysine, an aspartate residue interacting with the pyridine nitrogen of PLP, and a conserved loop that interacts with the PLP phosphate moiety.

A multiple structure superimposition of EUBREC_2651 with 15 representative proteins from the Pfam Aminotransferase Class V (PF00266) family with annotated function and known structures shows that they share a common core of 265 residues with an average RMSD of 2.48 Å based on Cα atoms (Supporting Information Fig. S3). EUBREC_2651 has the highest overall structural similarity to alanine/glyoxalate and serine/pyruvate aminotransferases with a common structural core of 351 residues, and 2-aminoethylphosphonate aminotransferase with a common core of 342 residues. A structure-based multiple sequence alignment with EUBREC_2651 and eight of the nearest homologous structures (Supporting Information Fig. S4) shows 11 identical and 26 highly conserved residues, which, as for EUBREC_0560 and PG_1327, are mostly located in the active site.

Oligomerization and active site comparisons

All three structures are consistent with a dimer as the likely biological unit as calculated by PISA,[14] with the active sites at the interfaces of the subunits, as seen in other PLP-fold Type I enzymes. Analytical size-exclusion chromatography supports the assignment of a dimer as the biological unit (data not shown).

All three enzymes contain the three structural features identified by Singh *et al.*[15] as being strictly conserved in, and characteristic of, aminotransferases: (1) an aromatic residue which stacks with the PLP pyridine ring, (2) a D-X-[A,V,I] motif where Asp interacts with the pyridine nitrogen and Ala, Val, or Ile interacts with the *si* side of the PLP ring, and (3) either an Asn and Tyr, or Thr by itself, interaction with the 3-O′ of PLP. The only exception is EUBREC_0560 where a proline interacts with the *si* face of PLP as noted by Singh *et al.* and initially identified in kynurenine aminotransferase (KAT).

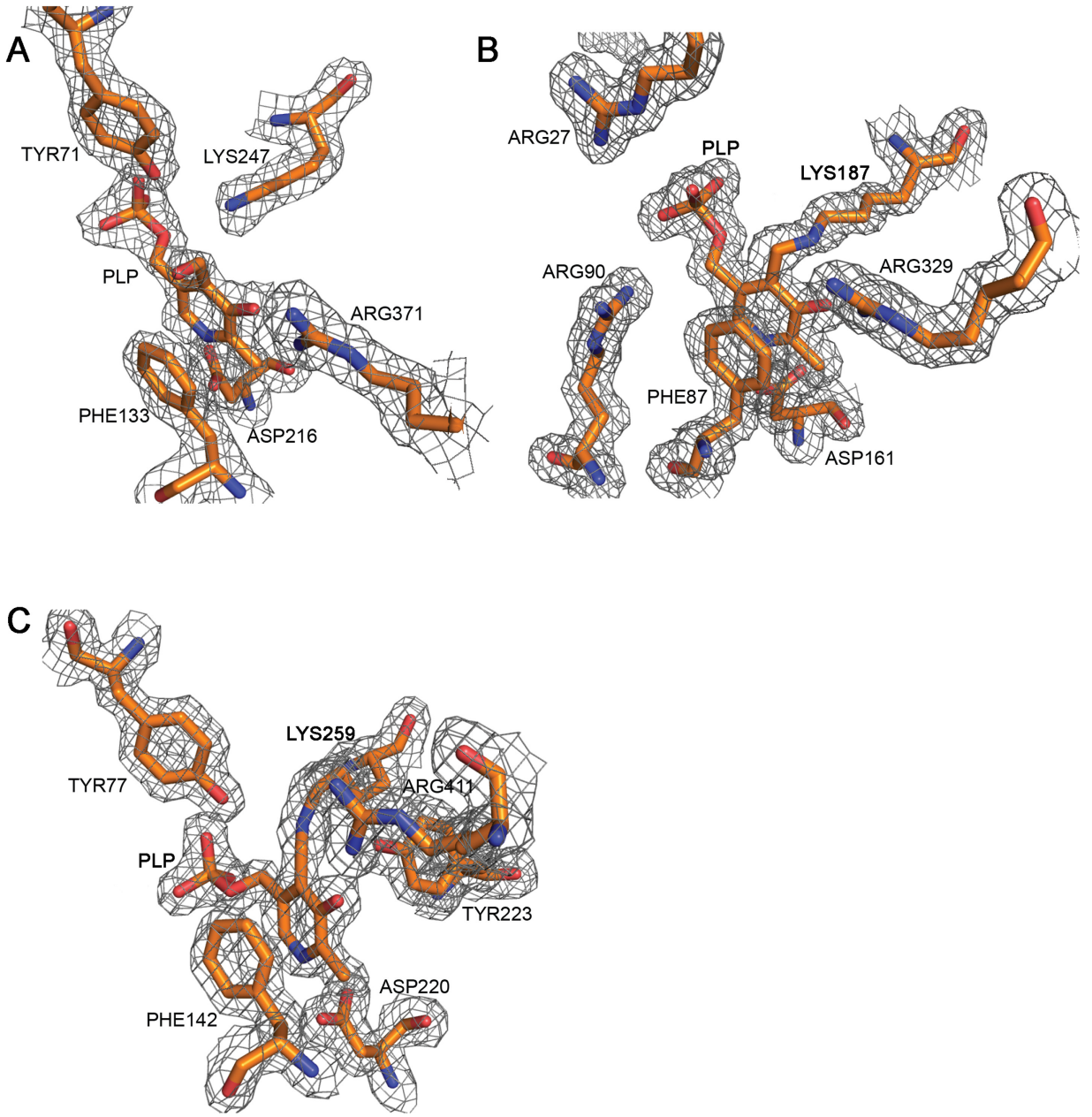

Figure 2. Active site architecture with the corresponding $2F_O - F_C$ electron density map contoured at 1σ for (A) EUBREC_0560, (B) EUBREC_2651, and (C) PG_1327 shows the quality of the electron density at the active site. The endogenous PLP co-factor bound to EUBREC_0560 (A) structure is not covalently linked to the active site, as opposed to the EUBREC_2651 (B) and PG_1327 (C) structures, where it forms a Schiff base linkage with the active site lysine.

A SPASM[16] search based on the 3D coordinates of user-defined active site residues was performed (Table II) to identify proteins with similar active site motifs but potentially different overall architectures. The active site residues chosen for the query were cofactor-binding residues, and hypothetical substrate binding residues as judged from the crystal structures. A multistructure overlay of each search query (i.e., the active site residues used in the search) and the associated SPASM hits are displayed in Figure 3.

The SPASM results show that the EUBREC_0560 active site is similar to aspartate and tyrosine aminotransferases (AAT and TAT, respectively), which are the two enzymes that were identified in the secondary-structure matching (SSM) search results. PG_1327 show no overlap with the top SSM matches. The only matches for EUBREC_2651 are two enzymes whose functions have not been confirmed, but are currently annotated as putative aspartate and serine aminotransferases.

Structure and function of EUBREC_0560

A qualitative transamination half-reaction activity assay indicates that EUBREC_0560 can deaminate

Table I. *Biochemically Characterized, Structurally Homologous Enzymes Identified with SSM and PDBeFold*

Query (PDB ID)	Name and species of hit	PDB ID	RMSD (Å)	Sequence identity	Q-score
EUBREC_0560	Human tyrosine aminotransferase	3dyd	1.6	22%	0.59
	Thermus thermophilus aspartate aminotransferase	1bjw	1.9	26%	0.58
	Pyrococcus furiosus alanine aminotransferase	1xi9	1.8	25%	0.58
	Thermotoga maritima aspartate aminotransferase	1o4s	1.9	29%	0.58
PG_1327	*Thermus thermophilus* aspartate aminotransferase	1bjw	1.7	18%	0.54
	Thermus thermophilus glutamine aminotransferase	1v2d	1.9	18%	0.51
EUBREC_2651	Yeast alanine-glyoxylate aminotransferase	2bkw	1.5	23%	0.67
	Salmonella typhimurium 2-aminoethylphosphonate transaminase	1m32	1.7	20%	0.66
	Human alanine-glyoxylate aminotransferase	3r9a	1.6	22%	0.66

13 of the 20 amino acids tested (Table III). Notably, EUBREC_0560 is particularly active towards the acidic residues aspartate and glutamate, and aromatic residues histidine, tryptophan, tyrosine, and phenylalanine. Of the nonproteinogenic amino acids tested, EUBREC_0560 has activity only towards L-cysteine sulfinate—a favorable reaction due to the irreversible release of SO_2. Note that the 330 nm absorbance peak in the UV–vis spectrum (characteristic of PMP formation) that is observed for EUBREC_0560 (as well as PG_1327 and EUBREC_2651) in the presence of cysteine during the qualitative half-reaction analysis, may be due to the production of a thiazolidine adduct that can form after the reaction of PLP with amino thiols.[17,18]

Kinetics assays were performed to identify the function of EUBREC_0560. In an assay for TAT activity using α-ketoglutarate and either tyrosine or phenylalanine as substrates, linear initial rates were observed for the first few seconds only (data not shown). This activity loss was not investigated further and complete kinetic profiles with these amino acids were not obtained. In an assay for AAT activity with 1 mM α-ketoglutarate and aspartate as substrates, no reaction above background was observed (data not shown). Steady-state kinetic parameters for EUBREC_0560 were successfully determined with tryptophan and α-ketoglutarate as substrates (Table IV). The micromolar binding constants (60 and 140 μ*M* for tryptophan and α-ketoglutarate, respectively) and catalytic efficiency on the order of 10^5 suggest that this substrate pair may be the natural ligands for this enzyme.

Unlike PG_1327 and EUBREC_2651, an internal aldimine linkage between Lys (position 247) and PLP is not present in the crystal structure of EUBREC_0560 (n.b. a 3.1 Å gap is found between the NZ atom and the C4′ atom of PLP), although the UV–vis spectral properties of the enzyme suggest the enzymes exists in the internal aldimine form (likely a hydrated carbinolamine or an enolimine,[19] based on the shift toward 340 nm) in the preparation used for functional studies (Fig. 4). It is possible that this disparity is due to X-ray induced structural damage during data collection, differences in the sample preparations used for crystallization and function studies, or the effects of the crystallization reagents (1*M* LiCl, 20% PEG 6000, and 0.1*M* HEPES pH 7.0).

In the EUBREC_0560 active site, PLP is sandwiched between Phe133 via a π-stacking interaction and Pro218 through a van der Waals interaction. Tyr219 and Asn182 hydrogen bond with O3′ of the pyridine ring, while Asp216 forms a salt bridge with N1. The phosphate moiety forms an ionic pair with Arg255 and accepts hydrogen bonds from Ser244, Ser246, and Tyr71 and the amides of Gly106 and Ala107. Lys247 donates a hydrogen bond to the oxygen atom in the phosphoester linkage.

Slow cooling molecular dynamics simulations with the tryptophan external aldimine modeled into the EUBREC_0560 active site (Fig. 5) shows that Tyr71, Ile41, and Phe133 can make van der Waals contacts with the side chain of the substrate. Arg18 forms a cation-π interaction with the electron-rich indole ring suggesting this residue may be important for the specificity towards tryptophan. Arg18 was fully modeled in only one of the four monomers present in the asymmetric unit, and was truncated at Cβ or Cδ in the remaining three molecules due to poorly defined electron density, indicating that the side chain is mobile in the absence of bound substrate. Arg371 forms an ion pair with the substrate carboxylate. As of March 2014, only one protein was annotated as a tryptophan aminotransferase (TrAT) in the PDB (3bwn, and 3bwo). This enzyme, from *Arabidopsis thaliana,* utilizes tryptophan as its preferred substrate.[20] EUBREC_0560 and TrAT share a sequence identity of 15.3%. TrAT has no residue in an analogous position to Arg18 of EUBREC_0560, which is located at the N-terminus of a 15-residue helix that is only present in EUBREC_0560. Instead, Arg350 of TrAT, which is located in a conserved loop present in both structures, may interact with tryptophan in a similar manner.

Structure and function of PG_1327

PG_1327 deaminated 8 of the 20 amino acids tested in qualitative half-reaction activity (Table III).

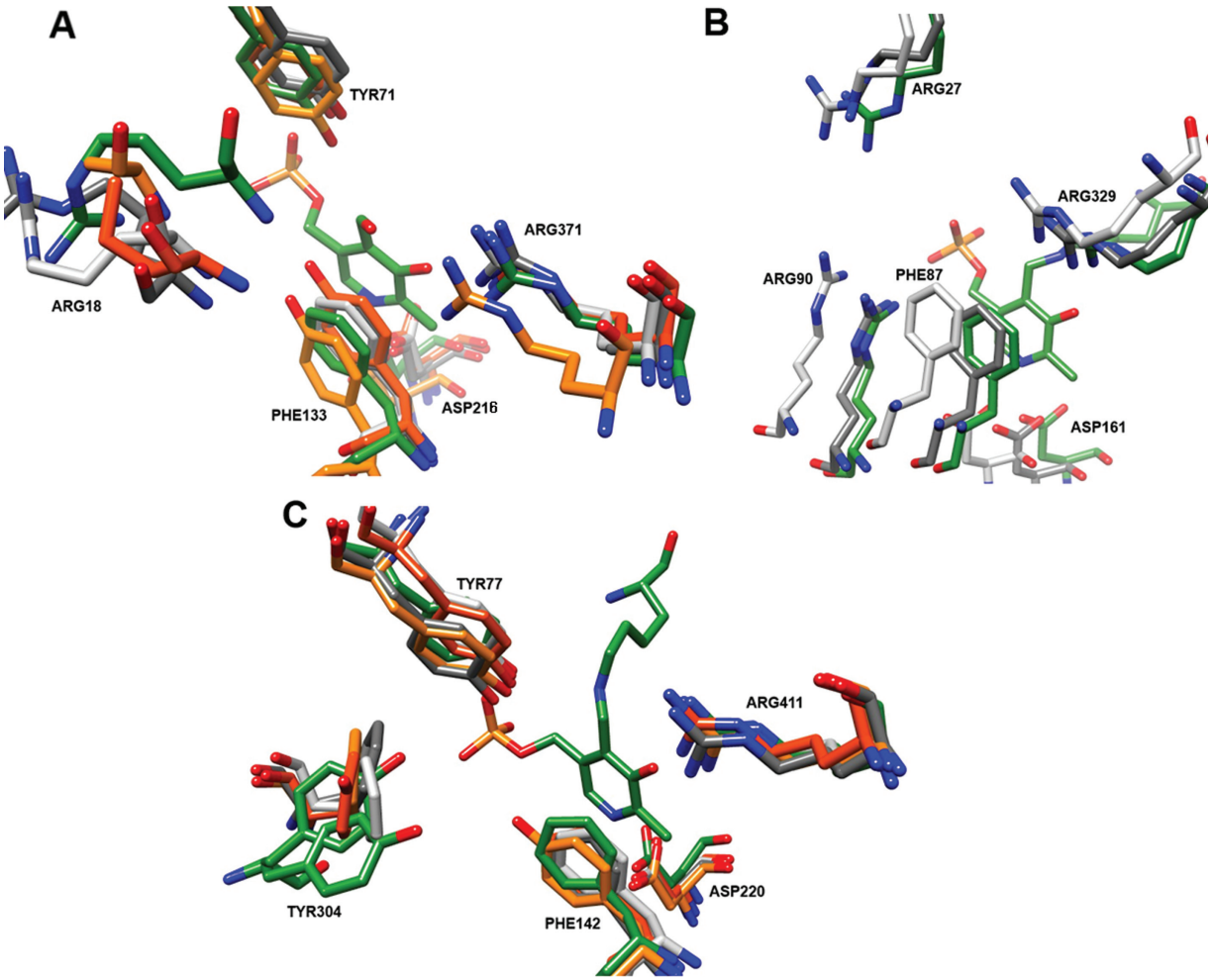

Figure 3. The active site residues that were used in SPASM searches of EUBREC_0560, EUBREC_2651, and PG_1327 are overlaid with their respective search results, as detailed in Table II, reveal the variation in active site structure of the homologs that have the most similar active site architectures. (A) EUBREC_0560 (green) compared with structures with PDB codes 1BW0, 1GDE, 1VP4, 2GB3, in light gray, dark gray, orange, and dark orange; (B) EUBREC_2651 (green) compared with 2DR1, 1IUG, in light gray, and dark gray; (C) PG_1327 (green) compared with 1W7L, 3B46, 1U08, 1YIZ, in light gray, dark gray, orange, and dark orange, respectively.

Similar to EUBREC_0560, PG_1327 has activity towards the acidic amino acids aspartate and glutamate, and the aromatic amino acids histidine, tyrosine and phenylalanine. Unlike EUBREC_0560, PG_1327 cannot deaminate asparagine, glutamine, alanine, serine, or tryptophan.

The SPASM and SSM structural similarity searches (Tables I and II) indicate PG_1327 may be a KAT. However, PG_1327 is unable to deaminate glutamine in the first half-reaction of transamination (Table III), suggesting that PG_1327 is not a KAT, as for *E. coli* KAT, and all studied mammalian KATs have glutamine aminotransferase activities.[21,22] Presteady-state kinetic parameters were determined (Table V) to define quantitatively the amino acid preference for the PG_1327 first half-reaction. Based on the k_{max}/K_{app} for the amino acids tested, PG_1327 has highest activity for phenylalanine (2.4×10^5 $M^{-1}s^{-1}$). The high K_{app} value of glutamate (25 m*M*) suggests it is not the natural amino group donor in reactions catalyzed by this enzyme. Presteady-state kinetic results show PG_1327 prefers large, aromatic substrates. Steady-state k_{cat}/K_M values of 3.3 and 3.5×10^4 M^{-1} s^{-1} (α-ketoglutarate and phenylalanine, respectively) (Table IV) support the assignment of PG_1327 as an aromatic amino acid aminotransferase (ArAT).

In each PG_1327 active site, PLP is covalently attached to Lys259 through a Schiff base linkage. Ala222 and Phe142 sandwich the pyridine ring from opposite faces. Asp220 forms a salt bridge with N1, and Tyr223 and Asn192 hydrogen bond with O3′. The phosphate group is anchored by hydrogen bonds from Tyr77, Ser258, Ser256 and the Gly112 and Ser113 amides. Arg267 forms an ionic bond with the phosphate moiety. Tyr77 is the only phosphate interacting residue that is not from the same subunit as the bound PLP cofactor, suggesting that the dimer is the active form.

PG_1327 was superimposed with the aromatic amino acid aminotransferase from *Paracoccus*

Table II. *SPASM Search Results Using a PDB Archive from July 2008*[a]

Query (active site residues used in search)	Name and species of hit	PDB ID
EUBREC_0560 (R18, Y71, F133, D216, R371)	*Trypanosoma cruzi* tyrosine aminotransferase	1BW0
	Pyrococcus horikoshii asparate aminotransferase	1GDE
	Thermotoga maritima hypothetical aminotransferase (function undefined)	1VP4
	Thermotogo maritima aspartate aminotransferase	2GB3
PG_1327 (Y77, F142, D220, Y304, R411)	*Escherichia coli* cystathionine beta-lyase	2FQ6
	Saccharomyces cerevisiae putative KAT	3B46
	Escherichia coli methionine aminotransferase	1U08
	Human KAT	1W7L
	Aedes aegypti KAT	1YIZ
EUBREC_2651 (R27, F87, R90, D161, R329)	*Pyrococcus horikoshii* hypothetical serine aminotransferase	2DR1
	Thermus thermophilus hypothetical aspartate aminotransferase	1IUG

[a] Results were filtered to display only hits below 1.5Å RMSD, 3Å Cα/Cα, and 3Å side-chain/side-chain cutoff. Residues were allowed substitutions as follows: Arg/His, Asp/Glu, and Phe/Trp/His/Tyr. Sequence directionality, gaps, and neighboring residues were not conserved. SPASM uses a modified version of the PDB database and this is the most recent version. SPASM was used because overall it provided the most robust results compared with other available choices.

denitrificans (pArAT)[23] (PDB id 1ay4) to compare active sites and identify a potential substrate recognition mechanism for aromatic substrates (Fig. 6). The sequences are 16.5% identical and share similar active site architectures. The aromatic substrate recognition residues of pArAT are Trp140, Tyr70, Ser296, Asn142, and Leu18 and overlay with Phe142, Tyr77, Ser308, Leu144, and Arg28 in PG_1327, respectively. The acidic substrate recognition residue of pArAT is primarily Arg292, which forms a salt bridge with the carboxylate of acidic substrates.

In one established mechanism for dual aromatic and acidic substrate specificity in ArATs, the arginine residue functions as a conformational switch in which it ion pairs with the carboxylate of acidic substrates or flips out to bulk solvent to accommodate aromatic substrates.[24] In PG_1327, Tyr304 corresponds to the position of Arg292, suggesting a different mechanism is employed for dual specificity, such as using hydrogen bonding networks that are seen in the *Pyrococcus horikoshii* ArAT.[25] Slow cooling molecular dynamics simulation of the phenylalanine external aldimine complexed with PG_1327 (Fig. 5) shows Phe142 π-stacking against the PLP pyridine ring in addition to van der Waals contacts with the substrate phenyl. Tyr304, Met114, Leu144, Val52, Leu80, Tyr77, and Ile27 surround the phenyl side

Table III. *Qualitative Transamination Half-Reaction Activity with Various Amino Acids*[a]

Substrate	EUBREC_2651	EUBREC_0560	PG_1327
2-Aminoethylphosphonate	–	–	–
Acetylserine	–	–	–
Alanine	+	+	–
Arginine	–	–	–
Asparagine	+	+	–
Aspartate	+	+	+
Cysteine	+	+	+
Glutamate	+	+	+
Glutamine	+	+	–
Glycine	–	–	–
Histidine	+	+	+
L-Cysteine sulfinate	+	+	+
Methionine	+	+	+
O-phospho-L-serine	+	–	–
Phenylalanine	+	+	+
Proline	–	–	–
Serine	+	+	–
Tryptophan	+	+	–
Tyrosine	+	+	+
Valine	–	–	–

[a] Enzyme (0.2 mg/mL) was incubated with 1 m*M* substrate for 10 min. Spectra of the enzyme before and after incubation were collected between 250 and 500 nm. Reactivity was assessed by changes in the cofactor absorbance spectrum consistent with conversion of enzyme bound PLP to PMP. The appearance of a peak at 330 nm (PMP) when incubated with amino acids suggests that they are aminotransferases, which produce this cofactor form as an obligatory intermediate.[5] A positive result indicates the appearance of a 330 nm peak.

Table IV. *Steady-State Kinetic Parameters*

Enzyme (PDB ID)	Substrate(s)	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M ($M^{-1}s^{-1}$)
EUBREC_0560	Tryptophan	21.6 ± 0.5	0.06 ± 0.01	3.6×10^5
	α-ketoglutarate		0.14 ± 0.01	1.5×10^5
PG_1327	Phenylalanine	46 ± 2	1.3 ± 0.1	3.5×10^4
	α-ketoglutarate		1.4 ± 0.2	3.3×10^4
EUBREC_2651	PS [a]	0.11 ± 0.01	2.7 ± 0.6	40.8
	Glutamate	4.9 ± 0.4	0.40 ± 0.08	1.2×10^4
	Oxaloacetate[b]		2.4 ± 0.4	2.0×10^3

One k_{cat} value is reported for each amino acid/ketoacid pair since the enzymes use two substrates and catalyze one overall reaction, except for the single substrate phosphate elimination catalyzed by EUBREC_2651.
[a] P_i elimination.
[b] The reaction of glutamate and oxaloacetate showed substrate inhibition by oxaloacetate. The K_I value for oxaloacetate is 13 ± 5 mM.

chain forming a hydrophobic pocket. Lys259 is in a position behind the substrate to accept the Cα proton in the step that leads to the carbanionic quinonoid intermediate in the PLP transamination mechanism.

Structure and function of EUBREC_2651

EUBREC_2651 has activity towards 14 of the 20 amino acids tested for half-reaction activity—making it the most promiscuous of the three enzymes. Notably, it is the only enzyme tested that produces coenzyme spectral changes in the presence of *O*-phospho-L-serine (PS) (Table III). Otherwise, the qualitative half-reaction activity profile of EUBREC_2651 is almost identical to EUBREC_0560.

SPASM and SSM searches reveal EUBREC_2651 has structural similarities to alanine, 2-aminoethylphosphonate (2-AEP), and aspartate aminotransferases. EUBREC_2651 was unable to deaminate 2-AEP and, thus, is likely not a 2-AEP aminotransferase. In stopped-flow experiments (Table V), alanine does not saturate the enzyme; therefore, EUBREC_2651 is not an alanine aminotransferase. The second order rate constant for the reaction of glutamate with the PLP enzyme in stopped-flow experiments is high, while that for PS is much lower. Nevertheless, the kinetic values observed for PS are consistent with steady-state values found with other known PSATs.[26–28]

Steady-state kinetic analysis of the AAT reverse reaction (Table IV) shows that EUBREC_2651 can utilize glutamate as an effective amino donor with a low micromolar K_M (k_{cat}/K_M is 1.2×10^4 and 2.0×10^3 for glutamate and oxaloacetate, respectively). For the forward reaction, aspartate was held at 25 mM. Under these conditions, α-ketoglutarate saturates the enzyme with an apparent K_M of 5 ± 1 μM and k_{cat} of 0.37 ± 0.02 s^{-1}. Likewise, when α-ketoglutarate was held at 100 μM, the K_M for aspartate is 15 ± 4 mM, and the k_{cat} is 0.47 ± 0.05 s^{-1}. The low micromolar α-ketoglutarate K_M, and the higher K_M values for oxaloacetate and aspartate suggest α-ketoglutarate is the natural amino group acceptor for this enzyme.

EUBREC_2651 was the only enzyme able to form PMP (Tables III and V) in the presence of phosphoserine (PS) and, hence, is likely to be a phosphoserine aminotransferase (PSAT). PSAT is an enzyme in the phosphorylated pathway of serine biosynthesis in which the glycolytic intermediate D−3-phosphoglycerate is transformed to L-serine in a three-step process. The second step is the transamination of glutamate and 3-phosphohydroxypyruvate (PHP) to α-ketoglutarate and PS.[29] To qualitatively test if EUBREC_2651 can catalyze transamination between α-ketoglutarate and PS, the enzyme was incubated with PS and α-ketoglutarate followed by derivatization with OPA as described in "Materials and Methods." The PS peak present initially disappears after 2.5 h of incubation with no concomitant formation of a glutamate peak (Fig. 4).

During attempts to assay for PSAT activity, EUBREC_2651 was observed to catalyze the β-elimination of phosphate from PS generating pyruvate and NH_3 (Fig. 4). The steady-state parameters of this side reaction were quantified (Table IV). Under the same conditions, EUBREC_2651 was also observed to eliminate acetate from *O*-acetyl-L-serine (OAS; $k_{obs} = 0.04$ s^{-1} at 5 mM OAS). Phosphate elimination has not been previously reported for PSAT, but is a known side reaction of other aminotransferases.[30]

The rapid formation of PMP from the reaction of EUBREC_2651 with PS was confirmed by HPLC (data not shown). Combined, these results are puzzling: PS reacts quickly to form enzyme-bound PMP but does not transaminate α-ketoglutarate to glutamate, while in the aspartate/α-ketoglutarate reaction, glutamate is formed. Others have reported steady-state kinetic assays for PS/α-ketoglutarate transamination that employed phosphoglycerate dehydrogenase and NADH. The presence of the coupling enzyme makes release of PHP, the product of the PS transamination half-reaction, irreversible. In these experiments, PHP was not trapped by a

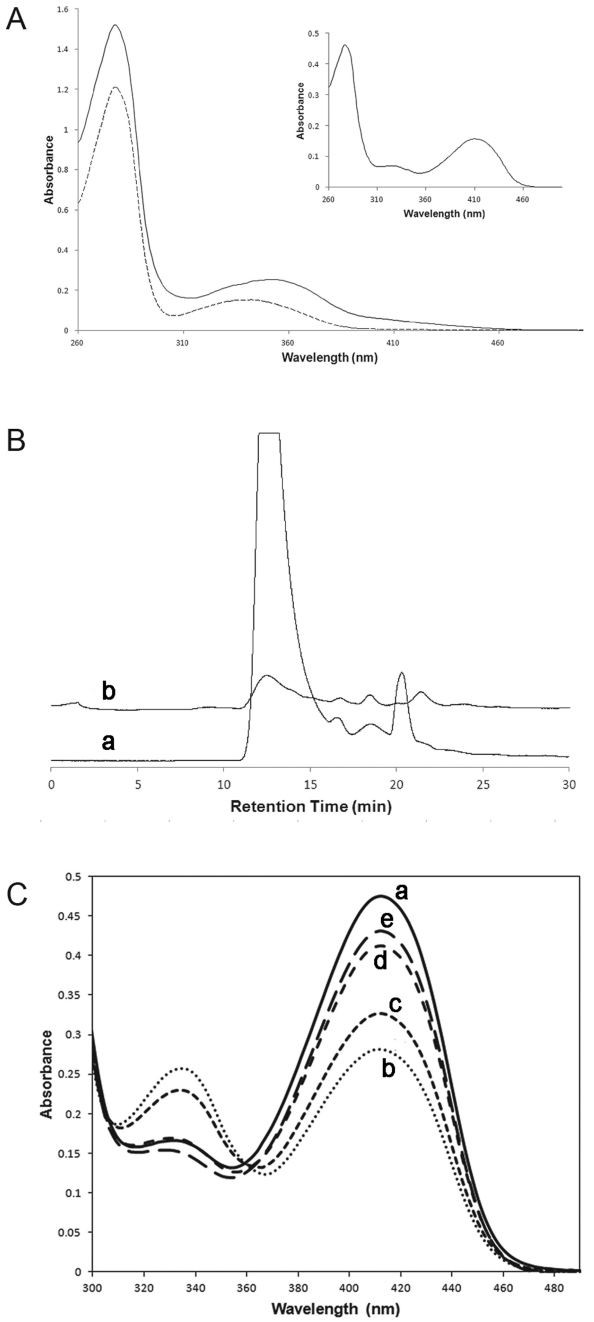

Figure 4. Spectral data of biochemical reactions. (A) UV–vis spectra of EUBREC_2651 (inset), PG_1327 (solid line), and EUBREC_0560 (dashed line). Each reaction contained 1 mg/mL enzyme in 50 m*M* pH 7.2 HEPES with 100 m*M* NaCl. (B) HPLC fluorescence chromatogram of EUBREC_2651 reaction with *O*-phospho-L-serine (PS). (a) Spectrum of control reaction after 2.5 h with no enzyme added. (b) Spectrum of reaction with EUBREC_2651 and PS after 2.5 h. (C) Spectra taken at various time intervals of EUBREC_2651 with and without PS. (a) Spectrum prior to incubation with PS. (b) taken immediately after addition of 1 m*M* PS. (c) Two minutes after addition of PS, (d) 10 min, and (e) 30 min. The reaction contained 100 m*M* pH 7.1 HEPES, 100 m*M* NaCl, 1 m*M* PS, and 3 mg/mL EUBREC_2651.

coupling enzyme or other means and is, therefore, stoichiometric with the PMP enzyme. We speculate that EUBREC_2651 is indeed a PSAT, but in all of our experiments, the PHP likely binds tightly to the PMP enzyme, preventing transamination of α-ketoglutarate to glutamate. This could explain the irreversible loss of phosphate in reactions containing α-ketoglutarate: the PMP-PHP complex equilibrates with the PLP-PS external aldimine intermediate, with the latter the branchpoint for the phosphate elimination reaction (Fig. 7). Unfortunately, PHP was not available for testing with glutamate in the physiological direction.

In the EUBREC_2651 active site, the PLP cofactor is covalently attached through a Schiff base linkage to Lys187. The pyridine ring is sandwiched between Val163 and Phe87. N1 forms an ionic pair with Asp161 while O3′ interacts with Thr136. The phosphate moiety accepts hydrogen bonds from Thr62, Ser60, Thr235, and the amide nitrogens of the peptide linkages between Ser60/Ser61 and Ser61/Thr62. Arg90, which is 4.9 Å away from the central phosphorous atom, may be involved in binding both substrate and the cofactor phosphate moiety. Simulated annealing with the PS external aldimine modeled into the EUBREC_2651 active site shows that the doubly negatively charged PS phosphate tightly interacts with Arg27, and to a lesser extent Arg90. Arg27 is involved in three ion-dipole pairs: one to the side chain of Gln231 and two to the carbonyls in the peptide linkages between Gln231/Thr232 and Thr232/Pro233, which fixes its position in the active site. Phe87 stacks against the PLP

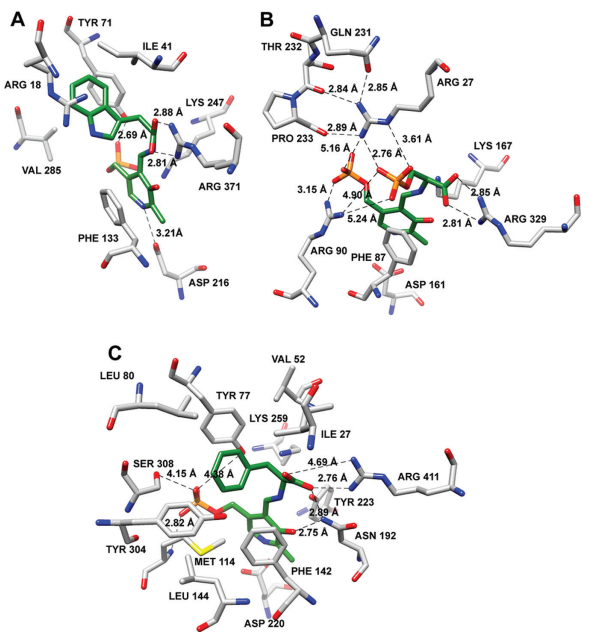

Figure 5. The active sites of (A) EUBREC_0560, (B) EUBREC_2651, and (C) PG_1327 after simulated annealing minimization with the tryptophan, PS, and phenylalanine external aldimine intermediates (green), respectively, which reveals details of the predicted protein-ligand interactions to understand how the different active site residues interact with their preferred ligands.

Table V. *Presteady-State Kinetic Parameters*[a]

Enzyme (PDB ID)	Substrate	k_{max} (s^{-1})	K_{app} (mM)	k_{max}/K_{app} ($M^{-1}s^{-1}$)
PG_1327	Phenylalanine	350 ± 20	1.5 ± 0.4	2.4×10^5
	Tyrosine	410 ± 51	5 ± 1	8.6×10^4
	Glutamate	207 ± 8	25 ± 2	8.2×10^3
	Valine	19 ± 2	7 ± 2	2.6×10^3
	Histidine	9 ± 1	17 ± 4	5.3×10^2
EUBREC_2651	Glutamate	315 ± 27	1.6 ± 0.4	2.0×10^5
	PS	0.80 ± 0.07	4 ± 1	186.0
	Alanine	NS[b]	NS	23.4

[a] Observed rate constants were plotted against substrate concentration. The data were fitted to: $k_{obs} = (k_{max}\ [S])/K_{app} + [S]$ to obtain the parameters presented here.
[b] NS, no saturation observed.

pyridine ring, and Arg329 binds the substrate α-carboxylate similar to the molecular dynamics results with the other two enzymes. The TT0402 protein from *Thermus thermophilus*[31] has been characterized biochemically and its active site is most similar to EUBREC_2651 (Table II, Fig. 3). TT0402 has a similar high affinity for glutamate ($k_{max}/K_{app} = 3.5 \times 10^4\ M^{-1}s^{-1}$), but unlike EUBREC_2651 displays no activity toward PS. TT0402 is annotated as a putative AAT, but similarly to our result, it is suggested that a different, acidic compound may be the substrate of this enzyme. The EUBREC_2651 crystal structure also revealed density (largely extended in shape within a triangulated region of $\sim 8 \times 8 \times 4$ Å^3) near PLP that was modeled as an unidentified ligand and could not be ascribed to any of the substrates tested.

Discussion

We present biochemical, structural, and computational evidence that support our functional assignment of PG_1327 as an aromatic amino acid aminotransferase, EUBREC_0560 as a tryptophan aminotransferase, and EUBREC_2651 as a PSAT.

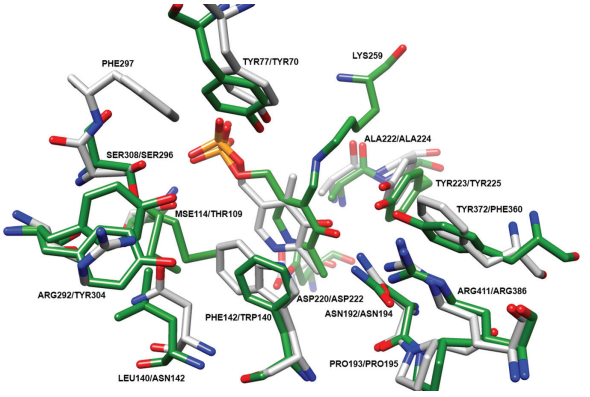

Figure 6. The active site of PG_1327 (green, first label) overlaid with *Paracoccus denitrificans* ArAT (gray, second label) shows the similarity of the active site architecture, which helps to support an aromatic aminotransferase activity for PG_1327.

EUBREC_2651 may utilize a different amino group donor than PS, but α-ketoglutarate is likely the natural amino group acceptor.

The wide range of activities in the qualitative transamination half-reaction studies (Table III) that we conducted is characteristic of only some aminotransferases,[21,32] which makes assignment of specific function more difficult, requiring further biochemical analysis. The same will likely be true for other uncharacterized aminotranferases. Our approach of using structure for initial functional inference followed by kinetic analyses and docking or molecular dynamics should be adaptable to future functional characterization of novel PLPDEs with known structures. In general, the flowchart presented in Supporting Information Figure S5 can be used as a general strategy for function classification of newly discovered PLP enzymes; however, more detailed investigations are needed to assign specific functions.

Others have used computational methods to successfully predict the specificity and/or function of uncharacterized enzymes. Raushel and coworkers[33] docked a library of high-energy intermediates into Tm0936, an enzyme with a known structure but unknown function from *T. maritima*. The computational results (i.e., docking score) corresponded well to catalytic rate constants observed during *in vitro* biochemical analysis. Jacobson and coworkers[34] used homology models of a group of dipeptide epimerases as targets for large-scale docking experiments, and confirmed their predictions *in vitro*. A similar *in vitro*/*in silico* approach could be attempted with uncharacterized PLPDEs. For example, after identifying likely aminotransferase activity through bioinformatics or biochemical (e.g., Table III) inference, a large library of external aldimine intermediates including proteinogenic and nonproteinogenic amino acids could be used for docking calculations. This would then be followed by *in vitro* experiments, similar to those in this work, for functional confirmation. A challenge for functional characterization of PLPDEs is the vast number of reactions they

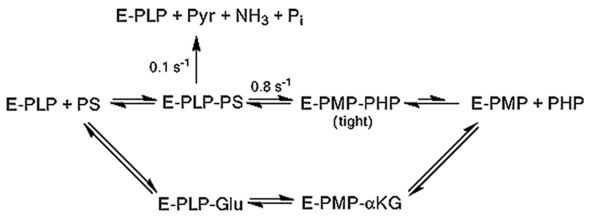

Figure 7. Reaction scheme for EUBREC_2651. This scheme explains the observations that EUBREC_2651 reacts quickly with PS to form E-PMP but does not react further in the presence of PHP to form glutamate from α-ketoglutarate. Elimination of phosphate from the external aldimine intermediate as a side reaction explains the spectral changes presented in Figure 4 and the formation of pyruvate from PS.

perform as a group. An approach streamlined by computation would be preferable to *in vitro* analyses on a large number of substrates.

Analysis of the genetic context of the genes encoding these proteins was performed using DOOR[35–37] (http://csbl.bmb.uga.edu/DOOR/). EUBREC_2651 was predicted to be located in a five-gene operon. The genes immediately upstream and downstream of EUBREC_2651 are a "hypothetical" glycerol-3-phosphate cytidyltransferase, and a "hypothetical" glycerophosphate transferase, respectively, two enzymes that are involved in teichoic acid biosynthesis in Gram-positive bacteria.[38,39] This operon analysis suggests EUBREC_2651 might play a yet unknown role in the metabolism of phosphorylated compounds involved in teichoic acid biosynthesis. PG_1327 was predicted to be in a two-gene operon with a "hypothetical" CoA ligase. EUBREC_0560 was not predicted to be in any operon that we could discern. Future studies on these proteins may help link our biochemical characterization and operon analysis with exact molecular functions and cellular roles for these proteins.

The Human Microbiome Project[40] aims to enable comprehensive characterization of the human microbiota and analysis of their role in human health and disease. Correlations have emerged between the distribution and composition of these microbial populations and human health and disease.[41–47] Most of the studies on the human microbiome have been at a metagenomics level, focusing on genome sequencing and organism distribution and relative abundancy, but some progress has also been made on understanding how individual proteins from these commensal bacteria modulate the microbiome and the host.[48] Several recent structural genomics approaches have started to focus on investigating individual proteins at the molecular level from bacteria found in the human microbiome.[49–56] Our results therefore contribute to the molecular characterization of individual human oral and gut microbiome proteins.

The three enzymes characterized here are aminotransferases present in human intestinal and oral bacteria. The structural and functional characterization of these novel PLPDEs paves the way for understanding their importance in symbiotic relationships of the human microbiome with the host in normal and disease states. Indeed, PLPDEs have been implicated in several diseases and have been explored in drug discovery.[57–61] In addition, treatment with probiotics (gut microbiome bacteria) has also been shown to be useful in several therapies and offers advantages due to the nature of their symbiotic existence with the human host. Thus this exciting new area of research may prove to be a new frontier for design and application of novel therapeutics.[62]

Materials and Methods

Protein expression and purification for structure determination

All three full-length proteins were produced using standard JCSG protocols.[2,63] Clones were generated using the Polymerase Incomplete Primer Extension (PIPE) cloning method.[64] The gene encoding EUBREC_0560 (residues 1–397) was amplified by polymerase chain reaction (PCR) from *Eubacterium rectale* genomic DNA using *PfuTurbo* DNA polymerase (Stratagene) and I-PIPE (Insert) primers (forward primer, 5′- ctgtacttccagggcATGGTAGTAAACG AAAGCATGTACCAGC-3′; reverse primer, 5′- aatt aagtcgcgttaTTTATTATATTTCTTGTATATCTTTTC-3′, target sequence in upper case) that included sequences for the predicted 5′ and 3′ ends. The gene encoding EUBREC_2651 (residues 1–357) was amplified by PCR from *E. rectale* genomic DNA using *PfuTurbo* DNA polymerase and I-PIPE (Insert) primers (forward primer, 5′-ctgtacttccagggcATGTTAAATTTT ACAGTAGGACCAG-3′; reverse primer, 5′-aattaagtc gcgttaTAAAAGATTTCTCTTTTGTAAATCT-3′, target sequence in upper case) that included sequences for the predicted 5′ and 3′ ends. The gene encoding PG_1327 (residues 1–436) was amplified by PCR from *Porphyromonas gingivalis* genomic DNA using *PfuTurbo* DNA polymerase and I-PIPE (Insert) primers (forward primer, 5′-ctgtacttccagggcATGAAT TTTCCAATCGATGAAAAACTG-3′; reverse primer, 5′-aattaagtcgcgttaTCCTTCGGCGTTCAGCATTTGGA GTCGC-3′, target sequence in upper case) that included sequences for the predicted 5' and 3' ends. The expression vector, pSpeedET, which encodes an amino-terminal tobacco etch virus (TEV) protease-cleavable expression and purification tag (MGSDKIHHHHHHENLYFQ/G), was PCR amplified with V-PIPE (Vector) primers (forward primer: 5′-taacgcgacttaattaactcgtttaaacggtctccagc-3′, reverse primer: 5′-gccctggaagtacaggttttcgtgatgatgatgatgatg-3′). V-PIPE and I-PIPE PCR products were mixed to anneal the amplified DNA fragments together. *Escherichia coli* GeneHogs (Invitrogen) competent cells

were transformed with the I-PIPE/V-PIPE mixture and dispensed on selective LB-agar plates. The cloning junctions were confirmed by DNA sequencing. Expression was performed in a selenomethionine-containing medium at 37°C. Selenomethionine was incorporated via inhibition of methionine biosynthesis, which does not require a methionine auxotrophic strain.[65] At the end of fermentation, lysozyme was added to the culture to a final concentration of 250 μg/mL, and the cells were harvested and frozen. After one freeze/thaw cycle, the cells were homogenized in lysis buffer [20 m*M* HEPES pH 8.0, 50 m*M* NaCl, 10 m*M* imidazole, 1 m*M* Tris(2-carboxyethyl)-phosphine-HCl (TCEP)] and passed through a Microfluidizer. The lysate was clarified by centrifugation at 32,500*g* for 30 min and loaded onto a nickel-chelating affinity column (GE Healthcare) pre-equilibrated with lysis buffer, the column washed with wash buffer [50 m*M* HEPES pH 8.0, 300 m*M* NaCl, 40 m*M* imidazole, 10% (v/v) glycerol, 1 m*M* TCEP], and the protein was eluted with elution buffer [20 m*M* HEPES pH 8.0, 300 m*M* imidazole, 10% (v/v) glycerol, 1 m*M* TCEP]. The eluate was buffer exchanged with TEV buffer (20 m*M* HEPES pH 8.0, 200 m*M* NaCl, 40 m*M* imidazole, 1 m*M* TCEP) using a PD-10 column (GE Healthcare), and incubated with 1 mg of TEV protease per 15 mg of eluted protein for 2 h at ambient temperature followed by overnight at 4°C. The protease-treated eluate was passed over nickel-chelating column (GE Healthcare) pre-equilibrated with HEPES crystallization buffer (20 m*M* HEPES pH 8.0, 200 m*M* NaCl, 40 m*M* imidazole, 1 m*M* TCEP) and the column was washed with the same buffer. Proteins were concentrated by centrifugal ultrafiltration (Millipore) for crystallization trials.

Crystallization

All three proteins were crystallized using the nanodroplet vapor diffusion method[66] using standard JCSG crystallization protocols.[2,63] Sitting drops composed of 200 nL protein solution mixed with 200 nL crystallization solution were equilibrated against a 50 μL reservoir at 277 K, with protein at 14.6, 20, and 20 mg/mL, for 16, 49, and 45 days prior to harvest, for EUBREC_0560, EUBREC_2651, and PG_1327, respectively. The crystallization reagent for EUBREC_0560 consisted of 1*M* LiCl, 20% PEG 6000, and 0.1*M* HEPES pH 7.0. Ethylene glycol was added to a final concentration of 10% (v/v) as a cryoprotectant. The crystallization reagent for EUBREC_2651 consisted of 20% PEG 6000, and 0.1*M* Bicine pH 9.0. Glycerol was added to a final concentration of 15% (v/v) as a cryoprotectant. The crystallization reagent for PG_1327 consisted of 0.2*M* NaF, 20% PEG 3350, pH 7.1. Ethylene glycol was added to a final concentration of 10% (v/v) as a cryoprotectant. Initial screening for diffraction was carried out using the Stanford Automated Mounting (SAM) system[67] and an X-ray microsource[68] at the Stanford Synchrotron Radiation Lightsource (SSRL, Menlo Park, CA). The diffraction data for EUBREC_0560, EUBREC_2651 and PG_1327 were indexed in space groups P1, $P6_122$, and $P2_12_12$, respectively.

X-ray data collection, structure determination, and refinement

For EUBREC_0560, MAD data were collected at SSRL, SLAC National Accelerator Laboratory, on beamline 9-2 at wavelengths corresponding to the high-energy remote (λ_1), inflection point (λ_2), and peak (λ_3) of a selenium MAD experiment using the BLU-ICE[69] data collection environment. The data sets were collected at 100 K using a MarMosaic 325 CCD detector (Rayonix). The MAD data were integrated and reduced using MOSFLM[70] and scaled with the program SCALA.[71,72] The heavy atom substructure was determined with SHELXD.[73] Phasing was performed with autoSHARP,[74] RESOLVE[75] was used for density modification and automatic model building.

For EUBREC_2651 and PG_1327, MAD data were collected at the APS, Argonne National Laboratory, on beamline 23-ID-B (GM/CA CAT) at wavelengths corresponding to the high-energy remote (λ_1) and inflection point (λ_2) of a selenium MAD experiment. The data sets were collected at 100 K using a MarMosaic 300 CCD detector (Rayonix). For EUBREC_2651, MAD data were integrated and reduced using XDS[76] and scaled with the program XSCALE. For PG_1327, MAD data were integrated and reduced using MOSFLM[70] and scaled with the program SCALA.[71,72] For EUBREC_2651 and PG_1327, the heavy atom sub-structure was determined with SHELXD.[73] Phasing was performed with autoSHARP[74] and SOLOMON[77] (implemented in autoSHARP) was used for density modification. ARP/wARP[78] was used for automatic model building.

For all three structures, model completion and crystallographic refinement were performed with the λ_1 data set using COOT[72] and REFMAC5.[79] The refinement protocol included experimental phase restraints in the form of Hendrickson–Lattman coefficients from autoSHARP and TLS refinement with one TLS group per protein molecule. Data and refinement statistics are summarized in Supporting Information Table S1.

Validation and deposition

The quality of the crystal structures was analyzed using the JCSG Quality Control server (http://smb.slac.stanford.edu/jcsg/QC). This server verifies: the stereochemical quality of the model using AutoDepInputTool,[80] MolProbity,[81] and Phenix[82] agreement

between the atomic model and the data using RESOLVE[75]; the protein sequence using CLUSTALW[83]; and the ADP distribution using Phenix, and differences in R_{cryst}/R_{free}, expected R_{free}/R_{cryst} and various other items including atom occupancies, consistency of NCS pairs, ligand interactions, special positions related by symmetry etc. using in-house scripts to analyze refinement log file and PDB header. Protein quaternary structure analysis was performed using the PISA server.[14] Figures 1–3, 5, and 6 were prepared with either PyMOL[84] or UCSF Chimera.[85] Atomic coordinates and experimental structure factors for EUBREC_0560, EUBREC_2651 and PG_1327 were deposited in the Protein Data Bank (http://www.wwpdb.org) with accession codes 3ele, 3f0h, and 3g0t, respectively.

Protein expression and purification for biochemical characterization

E. coli BL21 (DE3) transformed with expression vector (pSpeedET) was grown at 37°C in LB broth supplemented with kanamycin (50 μg/mL) shaking at 225 rpm. Protein expression was induced with IPTG (0.5 m*M*) when cultures reached an optical density at 600 nm of 0.5 to 0.7. Expression continued for six hours and the cells were then harvested by centrifugation at 4°C (5,500 rpm) and stored at −80°C. Cells were resuspended in lysis buffer (50 m*M* NaH_2PO_4 pH 8.0, 300 m*M* NaCl, 10 m*M* imidazole) with lysozyme (1 mg/mL). Cells were lysed by sonication at a temperature not exceeding 10°C. The crude lysates were clarified by centrifugation at 4°C (18,000 rpm), and the cleared lysate was loaded onto a Ni^{2+} affinity column pre-equilibrated with lysis buffer. The column was washed with 100 mL lysis buffer. A gradient to 100% elution buffer (50 m*M* NaH_2PO_4 pH 8.0, 300 m*M* NaCl, 250 m*M* imidazole) was performed at a flow rate of 1 mL/min for 450 min. The fractions containing the target proteins were identified by SDS-PAGE. Fractions containing the desired proteins were pooled into an ultrafiltration cell (Amicon) and concentrated with stirring under N_2. The concentrated sample was dialyzed at 4°C against 6 L purification buffer (50 m*M* HEPES pH 7.5, 100 m*M* KCl, 1 m*M* DTT, 1 m*M* EDTA).

For EUBREC_0560, a secondary purification was performed. Following elution from the Ni^{2+} affinity chromatography, EUBREC_0560 was buffer exchanged by dialysis against 6 L anion exchange loading buffer (10 m*M* TEA-HCl pH 7.8, 1 m*M* DTT, 0.5 m*M* EDTA). The buffer exchanged sample was then loaded onto a FastQ anion exchange column pre-equilibrated with anion exchange buffer. A gradient to 100% anion exchange elution buffer (10 m*M* TEA-HCl pH 7.8, 1 m*M* DTT, 0.5 m*M* EDTA, 300 m*M* KCl) was performed at a flow rate of 1.0 mL/min for 500 min. The fractions containing the target protein were identified by SDS-PAGE. The dialyzed samples were further concentrated by centrifugal ultrafiltration (Millipore). Final protein concentrations were determined either by absorbance at 280 nm, or the DC (Bio-Rad) protein assay using IgG as a standard.

Qualitative half-reaction activity

Various potential substrates (Table III), each at 1 m*M*, were incubated one at a time for 10 min with 0.2 mg/mL enzyme in 50 m*M* pH 7.1 HEPES, 100 m*M* KCl to determine qualitative half-reaction activity. The reaction was followed spectrally between 260 and 500 nm to monitor the appearance of the pyridoxamine phosphate (PMP) intermediate (330 nm). For EUBREC_2651, spectra were taken at various time intervals (Fig. 4) using 3 mg/mL EUBREC_2651 in 100 m*M* pH 7.1 HEPES, 100 m*M* NaCl, and 1 m*M* PS.

Pre-steady-state kinetics assays

The kinetics of the transamination half-reaction of the PLP-bound form of PG_1327 and EUBREC_2651 were monitored by measuring spectral changes between 340 and 420 nm using a diode array detector. The reaction mixture contained 50 m*M* pH 7.5 HEPES, 300 m*M* NaCl, 1 mg/mL enzyme and the amino acid substrates varied between 0.5 to 30 m*M*. Measurements were taken on an Applied Photophysics SX.18MV-R stopped flow spectrophotometer. The data were globally fit with Specfit to a two-exponential model.

Steady-state kinetics assays

All reactions were kept at a constant temperature of 25°C. For PG_1327, the rate of the steady-state transamination between L-Phe and α-ketoglutarate was monitored by coupling the production of phenylpyruvate (PP) to its reduction by D−2-hydroxyisocaproate dehydrogenase (DHDH). The rate of PP reduction by DHDH was measured by monitoring the concomitant oxidation of NADH at 340 nm. The 1 mL reaction mixture contained 50 m*M* pH 8.1 HEPES with 300 m*M* NaCl, 1–30 m*M* α-ketoglutarate, 0.5–10 m*M* L-phenylalanine, 200 μ*M* NADH, 40 μ*M* PLP, 0.5 U/mL DHDH, and 5.2 n*M* PG_1327.

For EUBREC_0560, the rate of the steady-state transamination between L-Trp and α-ketoglutarate was measured in a similar fashion as described above. The 1 mL reaction mixture contained 100 m*M* pH 7.2 HEPES with 200 m*M* KCl, 0.025–1 m*M* α-ketoglutarate, 0.05–5 m*M* tryptophan, 200 μ*M* NADH, 40 μ*M* PLP, 1 m*M* DTT, 1 m*M* EDTA, 0.5 U/mL DHDH, and 22.4 n*M* EUBREC_0560.

For EUBREC_2651, the steady-state rate of β-elimination of phosphate from PS to produce ammonia and pyruvate was measured by coupling the production of pyruvate to lactate dehydrogenase (LDH).

The rate of pyruvate reduction was measured by monitoring the concomitant oxidation of NADH at 340 nm. The 1 mL reaction contained 50 m*M* pH 7.2 HEPES, 100 m*M* NaCl, 0.25–10 m*M* PS, 20 μ*M* PLP, 2.5 U/mL LDH, and 2 μ*M* EUBREC_2651. The rate of the steady-state transamination reaction between glutamate and oxaloacetate was measured by coupling the production of α-ketoglutarate to glutamate dehydrogenase (GDH). The concomitant oxidation of NADH in the GDH reaction was followed at 340 nm. The 1 mL reaction contained 50 m*M* pH 7.0 HEPES with 100 m*M* KCl, 0.05–2 m*M* L-glutamate 0.5–20 m*M* oxaloacetate, 200 μ*M* NADH, 50 μ*M* PLP, 1 m*M* EDTA, 1 m*M* DTT, 2.5 U/mL GDH, 5 m*M* ammonium acetate, and 121 n*M* EUBREC_2651. The reaction between α-ketoglutarate and aspartate was followed by fixing L-aspartate and α-ketoglutarate at constant concentrations of 25 m*M*, and 100 μ*M*, respectively. The 1 mL reactions contained 50 m*M* pH 7.2 HEPES with 100 m*M* NaCl, 1–100 μ*M* α-ketoglutarate or 1–50 m*M* L-aspartate, 200 μ*M* NADH, 5 U/mL malate dehydrogenase (MDH), and 244 nM EUBREC_2651.

HPLC detection of EUBREC_2651 reaction products

A 200 μL reaction mixture containing 20 m*M* *O*-phospho-L-serine, 10 m*M* α-ketoglutarate, 100 m*M* NaCl, 50 m*M* pH 7.3 HEPES, and either 10 μ*M* EUBREC_2651 or no enzyme (negative control) was allowed to react for 2.5 h at which point the protein was denatured with 2 μL glacial acetic acid and the precipitated protein was removed by centrifugation. The reaction mixture was derivatized with *o*-phthaldehyde (OPA) and analyzed by HPLC as previously described.[86]

Molecular dynamics

The program YASARA[87] (http://www.yasara.org) was used to perform slow-cooling molecular dynamics simulations using the AMBER03 force field. External aldimines of potential substrates *O*-phospho-L-serine, phenylalanine, and tryptophan were constructed, energy minimized, and manually docked into the active site of EUBREC_2651, PG_1327, and EUBREC_0560, respectively. The crystallographically observed bound cofactor was used as a guide for placement of the external aldimine intermediate into the active site prior to simulation. Waters, ions, alternative side-chain conformations, and other nonstandard residues were removed. Selenomethionine residues in the crystal structure were represented as methionines for the simulation. To prepare the complex for simulation, the "md_run macro" supplied with the program was used. This macro performs a series of tasks including energy minimizing the complex, adding water and counter ions to the simulation cell, assigning force field parameters, predicting pK_a values and assigning protonation states, assigning initial atom velocities according to a Boltzmann distribution, and initiating molecular dynamics. The structures were equilibrated at 300 K for 25 ps, and then slowly cooled to 20 K over 40 ps.

Acknowledgments

The authors thank the members of the JCSG high throughput structural biology pipeline for their contribution to this work. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). Use of the Advanced Photon Source (APS), an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. Genomic DNA from Eubacterium rectale was a gift of Professor Jeffrey I. Gordon, Washington University in St. Louis, School of Medicine. Genomic DNA from Porphyromonas gingivalis (ATCC Number BAA-308D-5) was obtained from the American Type Culture Collection (ATCC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

References

1. Norvell JC, Berg JM (2007) Update on the protein structure initiative. Structure 15:1519–1522.
2. Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wooley J, Wüthrich K, Wilson IA (2010) The JCSG high-throughput structural biology pipeline. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:1137–1142.
3. Krishna SS, Weekes D, Bakolitsa C, Elsliger MA, Wilson IA, Godzik A, Wooley J (2010) TOPSAN: use of a collaborative environment for annotating, analyzing and disseminating data on JCSG and PSI structures. Acta Crystallogr Sect F Struct Biol Cryst Commun 66: 1143–1147.
4. Grishin NV, Phillips MA, Goldsmith EJ (1995) Modeling of the spatial structure of eukaryotic ornithine decarboxylases. Protein Sci 4:1291–1304.
5. Jansonius JN (1998) Structure, evolution and action of vitamin B6-dependent enzymes. Curr Opin Struct Biol 8:759–769.
6. Schneider G, Kack H, Lindqvist Y (2000) The manifold of vitamin B6 dependent enzymes. Structure 8:R1–6.
7. Eliot AC, Kirsch JF (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. Annu Rev Biochem 73:383–415.
8. Denessiouk KA, Denesyuk AI, Lehtonen JV, Korpela T, Johnson MS (1999) Common structural elements in the architecture of the cofactor-binding domains in

unrelated families of pyridoxal phosphate-dependent enzymes. Proteins 35:250–261.
9. Percudani R, Peracchi A (2009) The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. BMC Bioinformatics 10:273.
10. UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res 38: D142–148.
11. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 60:2256–2268.
12. Ye Y, Godzik A (2005) Multiple flexible structure alignment using partial order graphs. Bioinformatics 21: 2362–2369.
13. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217.
14. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.
15. Singh R, Spyrakis F, Cozzini P, Paiardini A, Pascarella S, Mozzarelli A (2013) Chemogenomics of pyridoxal 5'-phosphate dependent enzymes. J Enzyme Inhib Med Chem 28:183–194.
16. Kleywegt GJ (1999) Recognition of spatial motifs in protein structures. J Mol Biol 285:1887–1897.
17. Donini S, Percudani R, Credali A, Montanini B, Sartori A, Peracchi A (2006) A threonine synthase homolog from a mammalian genome. Biochem Biophys Res Commun 350:922–928.
18. Donini S, Ferrari M, Fedeli C, Faini M, Lamberto I, Marletta AS, Mellini L, Panini M, Percudani R, Pollegioni L, Caldinelli L, Petrucco S, Peracchi A (2009) Recombinant production of eight human cytosolic aminotransferases and assessment of their potential involvement in glyoxylate metabolism. Biochem J 422:265–272.
19. Di Salvo ML, Scarsdale JN, Kazanina G, Contestabile R, Schirch V, Wright HT (2013) Structure-based mechanism for early PLP-mediated steps of rabbit cytosolic serine hydroxymethyltransferase reaction. Biomed Res Int 2013:458–571.
20. Tao Y, Ferrer JL, Ljung K, Pojer F, Hong F, Long JA, Li L, Moreno JE, Bowman ME, Ivans LJ, Cheng Y, Lim J, Zhao Y, Ballare CL, Sandberg G, Noel JP, Chory J (2008) Rapid synthesis of auxin via a new tryptophan-dependent pathway is required for shade avoidance in plants. Cell 133:164–176.
21. Han Q, Fang J, Li J (2001) Kynurenine aminotransferase and glutamine transaminase K of *Escherichia coli*: identity with aspartate aminotransferase. Biochem J 360:617–623.
22. Hosono A, Mizuguchi H, Hayashi H, Goto M, Miyahara I, Hirotsu K, Kagamiyama H (2003) Glutamine:phenylpyruvate aminotransferase from an extremely thermophilic bacterium, *Thermus thermophilus* HB8. J Biochem (Tokyo) 134:843–851.
23. Okamoto A, Nakai Y, Hayashi H, Hirotsu K, Kagamiyama H (1998) Crystal structures of *Paracoccus denitrificans* aromatic amino acid aminotransferase: a substrate recognition site constructed by rearrangement of hydrogen bond network. J Mol Biol 280:443–461.
24. Donini S. Molecular cloning and biochemical characterization of pyridoxal 5'-phosphate dependent enzymes of unknown function. (2004–2007) PhD Thesis, Department of Biochemistry and Molecular Biology. University of Parma, Parma, Italy.
25. Matsui I, Matsui E, Sakai Y, Kikuchi H, Kawarabayasi Y, Ura H, Kawaguchi S, Kuramitsu S, Harata K (2000) The molecular structure of hyperthermostable aromatic aminotransferase with novel substrate specificity from *Pyrococcus horikoshii*. J Biol Chem 275:4871–4879.
26. Itoh H, Dempsey WB (1970) Purification of 3-phosphoserine-α-ketoglutarate transaminase from *Escherichia coli* B. Life Sci II 9:1289–1294.
27. Basurko MJ, Marche M, Darriet M, Cassaigne A (1999) Phosphoserine aminotransferase, the second step-catalyzing enzyme for serine biosynthesis. IUBMB Life 48:525–529.
28. Ali V, Nozaki T (2006) Biochemical and functional characterization of phosphoserine aminotransferase from *Entamoeba histolytica*, which possesses both phosphorylated and non-phosphorylated serine metabolic pathways. Mol Biochem Parasitol 145:71–83.
29. Hester G, Stark W, Moser M, Kallen J, Markovic-Housley Z, Jansonius JN (1999) Crystal structure of phosphoserine aminotransferase from *Escherichia coli* at 2.3 Å resolution: comparison of the unligated enzyme and a complex with α-methyl-l-glutamate. J Mol Biol 286:829–850.
30. Passera E, Campanini B, Rossi F, Casazza V, Rizzi M, Pellicciari R, Mozzarelli A (2011) Human kynurenine aminotransferase II–reactivity with substrates and inhibitors. FEBS J 278:1882–1900.
31. Katsura Y, Shirouzu M, Yamaguchi H, Ishitani R, Nureki O, Kuramitsu S, Hayashi H, Yokoyama S (2004) Crystal structure of a putative aspartate aminotransferase belonging to subgroup IV. Proteins 55:487–492.
32. Nowicki C, Hunter GR, Montemartini-Kalisz M, Blankenfeldt W, Hecht H, Kalisz HM (2001) Recombinant tyrosine aminotransferase from *Trypanosoma cruzi*: structural characterization and site directed mutagenesis of a broad substrate specificity enzyme. Biochim Biophys Acta 1546:268–281.
33. Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, Raushel FM (2007) Structure-based activity prediction for an enzyme of unknown function. Nature 448:775–779.
34. Lukk T, Sakai A, Kalyanaraman C, Brown SD, Imker HJ, Song L, Fedorov AA, Fedorov EV, Toro R, Hillerich B, Seidel R, Patskovsky Y, Vetting MW, Nair SK, Babbitt PC, Almo SC, Gerlt JA, Jacobson MP (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. Proc Natl Acad Sci USA 109:4122–4127.
35. Dam P, Olman V, Harris K, Su Z, Xu Y (2007) Operon prediction using both genome-specific and general genomic information. Nucleic Acids Res 35:288–298.
36. Mao F, Dam P, Chou J, Olman V, Xu Y (2009) DOOR: a database for prokaryotic operons. Nucleic Acids Res 37: D459–463.
37. Mao X, Ma Q, Zhou C, Chen X, Zhang H, Yang J, Mao F, Lai W, Xu Y (2014) DOOR 2.0: presenting operons and their functions through dynamic and integrated views. Nucleic Acids Res 42:D654–659.
38. Park YS, Sweitzer TD, Dixon JE, Kent C (1993) Expression, purification, and characterization of CTP: glycerol-3-phosphate cytidylyltransferase from *Bacillus subtilis*. J Biol Chem 268:16648–16654.
39. Schertzer JW, Brown ED (2003) Purified, recombinant TagF protein from *Bacillus subtilis* 168 catalyzes the polymerization of glycerol phosphate onto a membrane acceptor in vitro. J Biol Chem 278:18002–18007.
40. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA,

Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M (2009) The NIH Human Microbiome Project. Genome Res 19:2317–2323.

41. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. Nature 449:804–810.
42. Lozupone CA, Hamady M, Cantarel BL, Coutinho PM, Henrissat B, Gordon JI, Knight R (2008) The convergence of carbohydrate active gene repertoires in human gut microbes. Proc Natl Acad Sci USA 105:15076–15081.
43. Turnbaugh PJ, Gordon JI (2009) The core gut microbiome, energy balance and obesity. J Physiol 587:4153–4158.
44. Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI (2011) Human nutrition, the gut microbiome and the immune system. Nature 474:327–336.
45. Murgas Torrazza R, Neu J (2011) The developing intestinal microbiome and its relationship to health and disease in the neonate. J Perinatol 31:S29–34.
46. Tilg H, Kaser A (2011) Gut microbiome, obesity, and metabolic dysfunction. J Clin Invest 121:2126–2132.
47. Wallace TC, Guarner F, Madsen K, Cabana MD, Gibson G, Hentges E, Sanders ME (2011) Human gut microbiota and its relationship to health and disease. Nutr Rev 69:392–403.
48. Turnbaugh PJ, Henrissat B, Gordon JI (2010) Viewing the human microbiome through three-dimensional glasses: integrating structural and functional studies to better define the properties of myriad carbohydrate-active enzymes. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:1261–1264.
49. Koropatkin NM, Martens EC, Gordon JI, Smith TJ (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16:1105–1115.
50. Koropatkin N, Martens EC, Gordon JI, Smith TJ (2009) Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont. Biochemistry 48:1532–1542.
51. Bakolitsa C, Xu Q, Rife CL, Abdubek P, Astakhova T, Axelrod HL, Carlton D, Chen C, Chiu HJ, Clayton T, Das D, Deller MC, Duan L, Ellrott K, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Lam WW, Marciano D, McMullan D, Miller MD, Morse AT, Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Tien HJ, Trame CB, van den Bedem H, Weekes D, Hodgson KO, Wooley J, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) Structure of BT_3984, a member of the SusD/RagB family of nutrient-binding molecules. Acta Crystallogr Sect F Struct Biol Cryst Commun 66: 1274–1280.
52. Das D, Finn RD, Carlton D, Miller MD, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Chen C, Chiu HJ, Chiu M, Clayton T, Deller MC, Duan L, Ellrott K, Ernst D, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Marciano D, McMullan D, Morse AT, Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Rife CL, Sefcovic N, Tien HJ, Trame CB, van den Bedem H, Weekes D, Wooten T, Xu Q, Hodgson KO, Wooley J, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) The structure of BVU2987 from *Bacteroides vulgatus* reveals a superfamily of bacterial periplasmic proteins with possible inhibitory function. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:1265–1273.
53. Das D, Kozbial P, Han GW, Carlton D, Jaroszewski L, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Chen C, Chiu HJ, Chiu M, Clayton T, Deller MC, Duan L, Ellrott K, Elsliger MA, Ernst D, Farr CL, Feuerhelm J, Grzechnik A, Grant JC, Jin KK, Johnson HA, Klock HE, Knuth MW, Krishna SS, Kumar A, Marciano D, McMullan D, Miller MD, Morse AT, Nigoghossian E, Nopakun A, Okach L, Oommachen S, Paulsen J, Puckett C, Reyes R, Rife CL, Sefcovic N, Tien HJ, Trame CB, van den Bedem H, Weekes D, Wooten T, Xu Q, Hodgson KO, Wooley J, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) The structure of KPN03535 (gi|152972051), a novel putative lipoprotein from *Klebsiella pneumoniae*, reveals an OB-fold. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:1254–1260.
54. Tan K, Tesar C, Wilton R, Keigher L, Babnigg G, Joachimiak A (2010) Novel α-glucosidase from human gut microbiome: substrate specificities and their switch. FASEB J 24:3939–3949.
55. Xu Q, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Cai X, Carlton D, Chen C, Chiu HJ, Chiu M, Clayton T, Das D, Deller MC, Duan L, Ellrott K, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Marciano D, McMullan D, Miller MD, Morse AT, Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Sefcovic N, Tien HJ, Trame CB, van den Bedem H, Weekes D, Wooten T, Yeh A, Zhou J, Hodgson KO, Wooley J, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) A conserved fold for fimbrial components revealed by the crystal structure of a putative fimbrial assembly protein (BT1062) from *Bacteroides thetaiotaomicron* at 2.2 Å resolution. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:1281–1286.
56. Yeh AP, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Cai X, Carlton D, Chen C, Chiu HJ, Chiu M, Clayton T, Das D, Deller MC, Duan L, Ellrott K, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Lam WW, Marciano D, McMullan D, Miller MD, Morse AT, Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Tien HJ, Trame CB, van den Bedem H, Weekes D, Wooten T, Xu Q, Hodgson KO, Wooley J, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) Structure of *Bacteroides thetaiotaomicron* BT2081 at 2.05 Å resolution: the first structural representative of a new protein family that may play a role in carbohydrate metabolism. Acta Crystallogr Sect F Struct Biol Cryst Commun 66:1287–1296.
57. Mozzarelli A, Bettati S (2006) Exploring the pyridoxal 5'-phosphate-dependent enzymes. Chem Rec 6:275–287.
58. Amadasi A, Bertoldi M, Contestabile R, Bettati S, Cellini B, di Salvo ML, Borri-Voltattorni C, Bossa F, Mozzarelli A (2007) Pyridoxal 5'-phosphate enzymes as targets for therapeutic agents. Curr Med Chem 14: 1291–1324.
59. Muller IB, Wu F, Bergmann B, Knockel J, Walter RD, Gehring H, Wrenger C (2009) Poisoning pyridoxal 5-phosphate-dependent enzymes: a new strategy to target the malaria parasite *Plasmodium falciparum*. PLoS One 4:e4406.

60. Conti P, Tamborini L, Pinto A, Blondel A, Minoprio P, Mozzarelli A, De Micheli C (2011) Drug discovery targeting amino acid racemases. Chem Rev 111:6919–6946.
61. Wu F, Christen P, Gehring H (2011) A novel approach to inhibit intracellular vitamin B6-dependent enzymes: proof of principle with human and plasmodium ornithine decarboxylase and human histidine decarboxylase. FASEB J 25:2109–2122.
62. Prakash S, Rodes L, Coussa-Charley M, Tomaro-Duchesneau C (2011) Gut microbiota: next frontier in understanding human health and development of biotherapeutics. Biologics 5:71–86.
63. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 99:11664–11669.
64. Klock HE, Koesema EJ, Knuth MW, Lesley SA (2008) Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins 71:982–994.
65. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol 229:105–124.
66. Santarsiero BD, Yegian DT, Lee CC, Spraggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens RC (2002) An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr 35: 278–281.
67. Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP (2002) An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot. J Appl Crystallogr 2002:720–726.
68. Miller MD, Deacon AM (2007) An X-ray microsource based system for crystal screening and beamline development during synchrotron shutdown periods. Nucl Instrum Methods Phys Res A 582:233–235.
69. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat 9: 401–406.
70. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography 26.
71. Collaborative Computing Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Cryst D 50:760–763.
72. Evans P (2006) Scaling and assessment of data quality. Acta Cryst D 62:72–82.
73. Sheldrick GM (2008) A short history of SHELX. Acta Cryst A 64:112–122.
74. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. Methods Mol Biol 364:215–230.
75. Terwilliger TC (2000) Maximum-likelihood density modification. Acta Cryst D 56:965–972.
76. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallogr 26:795–800.
77. Abrahams JP, Leslie AG (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Cryst D 52:30–42.
78. Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat Protoc 3:1171–1179.
79. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 374:300–321.
80. Yang H, Guranovic V, Dutta S, Feng Z, Berman HM, Westbrook JD (2004) Automated and accurate deposition of structures solved by X-ray diffraction to the Protein Data Bank. Acta Cryst D 60:1833–1839.
81. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, III, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–383.
82. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D 66:213–221.
83. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.
84. DeLano WL (2008) The PyMOL Molecular Graphics System. Palo Alto, CA: DeLano Scientific LLC.
85. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.
86. Fogle EJ, Toney MD (2011) Analysis of catalytic determinants of diaminopimelate and ornithine decarboxylases using alternate substrates. Biochim Biophys Acta 1814:1113–1119.
87. Krieger E, Joo K, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins 77 Suppl 9:114–122.