Discovery of new enzymes and metabolic pathways using structure and genome context

Suwen Zhao[1,8], **Ritesh Kumar**[2,8], **Ayano Sakai**[2,8], **Matthew W. Vetting**[3,8], **B. McKay Wood**[2,8], **Shoshana Brown**[4], **Jeffery B. Bonanno**[3], **Brandan S. Hillerich**[3], **Ronald D. Seidel**[3], **Patricia C. Babbitt**[4,*], **Steven C. Almo**[3,*], **Jonathan V. Sweedler**[2,5,*], **John A. Gerlt**[2,5,6,*], **John E. Cronan**[2,3,7,*], and **Matthew P. Jacobson**[1,*]

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, California, USA

[2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[3]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York, USA

[4]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California, USA

[5]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[6]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[7]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Abstract

Assigning valid functions to proteins identified in genome projects is challenging, with over-prediction and database annotation errors major concerns[1]. We, and others[2], are developing computation-guided strategies for functional discovery using "metabolite docking" to

[*]Correspondence and requests for materials should be addressed to: babbitt@cgl.ucsf.edu (P. C. B), steve.almo@einstein.yu.edu (S. C. A), jsweedle@illinois.edu (J. V. S), j-gerlt@illinois.edu (J. A. G), j-cronan@life.uiuc.edu (J. E. C), and/or Matt.Jacobson@ucsf.edu (M. P. J).
[8]These authors contributed equally to this work.

Competing financial interests

The authors declare a conflict of interest. M.P.J. is a consultant to Schrodinger LLC, which developed or licensed some of the software used in this study.

Data deposition

The atomic coordinates and structure factors for APO Hyp-B 2-epimerase (HpbD; PDBID:2PMQ) and tHyp-B-liganded HpbD (PDBID:4H2H) have been deposited in the Protein Data Bank, www.pdb.org.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

experimentally derived[3] or homology-based[4] three-dimensional structures. Bacterial metabolic pathways often are encoded by "genome neighborhoods" (gene clusters and/or operons), which can provide important clues for functional assignment. We recently demonstrated the synergy of docking and pathway context by "predicting" the intermediates in the glycolytic pathway in *E. coli*[5]. Metabolite docking to multiple binding proteins/enzymes in the same pathway increases the reliability of *in silico* predictions of substrate specificities because the pathway intermediates are structurally similar. We report that structure-guided approaches for predicting the substrate specificities of several enzymes encoded by a bacterial gene cluster allowed i) the correct prediction of the *in vitro* activity of a structurally characterized enzyme of unknown function (PDB 2PMQ), 2-epimerization of *trans*-4-hydroxy-L-proline betaine (tHyp-B) and *cis*-4-hydroxy-D-proline betaine (cHyp-B), and ii) the correct identification of the catabolic pathway in which Hyp-B 2-epimerase participates. The substrate-liganded pose predicted by virtual library screening (docking) was confirmed experimentally. The enzymatic activities in the predicted pathway were confirmed by *in vitro* assays and genetic analyses; the intermediates were identified by metabolomics; and repression of the genes encoding the pathway by high salt was established by transcriptomics, confirming the osmolyte role of tHyp-B. This study establishes the utility of structure-guide functional predictions to enable the discovery of new metabolic pathways.

We report application of this strategy to discover a novel reaction, 4*R*-hydroxyproline betaine 2-epimerase (Hyp-B 2-epimerase; Figure 1a), as well as the catabolic pathway by which tHyp-B is converted to α-ketoglutarate. The crystallographically determined unliganded structure of the uncharacterized "target" as well as homology models for binding proteins/enzymes encoded by its genome neighborhood were used to predict the Hyp-B 2-epimerase activity as well as those of downstream enzymes of the pathway.

The marine bacterium *Pelagibaca bermudensis* encodes an uncharacterized member of the enolase superfamily (GI: 114543141) in which two Lys residues of the TIM-barrel domain are positioned to function as acid/base catalysts[6–8]. The New York SGX Research Consortium determined its structure (PDB 2PMQ) because it shared <30% sequence identity with structurally characterized enolase superfamily members. The only ligand was the Mg^{2+} that stabilizes the enolate anion intermediate obtained by abstraction of the α-proton of a carboxylate substrate. The active site is sequestered from solvent by two closed loops so was suitable for virtual metabolite docking for substrate prediction (Supplementary Figure S1).

Figure 2 shows the genome neighborhoods of the gene encoding 2PMQ [*hpb*D; hydroxyproline betaine from its functional characterization (*vide infra*)] plus a putative *Paracoccus denitrificans* orthologue. The automated TrEMBL annotations (Supplementary Table S1) fail to assign the *in vitro* activity of HpbD or identify the metabolic pathway. *P. bermudensis* is not genetically tractable; hence, we studied *P. denitrificans* which encodes one HpbD orthologue and two sets of orthologues of most of the genes neighboring the *P. bermudensis hpbD* gene. Genome neighborhoods are "conserved" for other putative orthologues (~20 can be identified in the sequence databases; http://sfld.rbvi.ucsf.edu/).

The *in silico* ligand docking library (87,098 members) included the KEGG metabolite library[9] as well as other potential enolase superfamily substrates, e. g., dipeptides, N-acylated amino acids, acid sugars and the enolate anions obtained by abstraction of the α-

proton (high energy intermediates[10]) (**Methods**). The library was docked in the active site of HpbD using Glide SP, and energy scoring functions rank-ordered the members of the library according to binding affinity. The best-scoring molecules were enriched with amino acid derivatives, especially proline analogues and N-capped amino acid derivatives (Figure 3), allowing the prediction that HpbD is an amino acid racemase/epimerase, with the substrate likely having N-substitution.

The genome neighborhood includes an ABC transporter with a periplasmic binding protein (HpbJ) annotated as binding "glycine betaine/L-proline". The structure of a homologous binding protein with glycine betaine[11] (1R9L) was used as the homology model template (HpbJ and 1R9L share 48% sequence identity) (**Methods**). The binding site contains three Trp residues (Figure 1b) that form a pi-cation "cage" for a quaternary ammonium (betaine), which may also be electrostatically stabilized by Glu42, located 5.4 Å from the quaternary nitrogen. Thus, the homology model allowed the prediction that its ligand is a betaine. A library of 31 betaines was docked to the model; tHyp-B had the highest rank (Supplementary Table S2) so we predicted that HpbJ participates in transport of tHyp-B. In addition, the HpbD active site contains Trp320 and Asp292 that are similarly positioned relative to the predicted binding pose of betaines (Figure 1c). Therefore, the structural basis for the predicted specificity of HpbJ refined the prediction that the substrate for HpbD is a proline betaine, e.g., tHyp-B.

A homology model was constructed for the Rieske-type protein (HpbB1) using a homologue (3N0Q) as the template (60% sequence identity) (Figure 1d) (**Methods**). The active site resembled the binding sites in the betaine binding proteins (aromatic residues and Glu200); indeed, some Rieske-type proteins are betaine demethylases[12–14] (Supplementary Figure S2 and Table S3). Thus, we predicted that the substrate is a small betaine. [While this work was in progress, the X-ray structure of a Rieske-type Pro-B demethylase from *Sinorhizobium meliloti* (PDB 3VCP) was published[15]; its active site superimposed closely with our HpbB1 homology model.]

The results of library docking to the experimental *apo* structure of HpbD (2PMQ) and to homology models of HpbJ and HpbB1 allowed prediction that HpbD utilizes Hyp-B or Pro-B as substrate in a 1,1-proton transfer reaction. The betaines Gly-B and carnitine also were candidates (although these would be substrates for virtual reactions; the α-carbons are prochiral). Although >25 functions have been assigned to members of the enolase superfamily, including N-succinylamino acid racemases and dipeptide epimerases, no amino acid or amino acid betaine was known to be a substrate[8].

tHyp-B, L-Pro-B, Gly-B, and carnitine were incubated with HpbD in D_2O (**Methods**). The ^1H NMR spectra with tHyp-B, D-Pro-B, and Gly-B revealed exchange of the α-proton with solvent deuterium (the latter a virtual reaction); also, for tHyp-B, resonances associated with cHyp-B, the 2-epimer, were observed (Supplementary Figure S3). These results are expected for a 1,1-proton transfer reaction that equilibrates the configurations at carbon-2 of tHyp-B and Pro-B using two Lys acid/base catalysts.

The kinetic constants for tHyp-B and L-Pro-B were determined for both HpbDs (Figure 4b) (**Methods**). Although the k_{cat} values are large, the K_M values also are large so the k_{cat}/K_M values are modest. Betaines, including tHyp-B, are osmoprotectants accumulated by many bacteria, including pelagic (*P. bermudensis*) and plant-associated (*P. denitricans*), to survive osmotic stress[16–20]; their intracellular concentrations can approach molar levels[21,22]. We determined that the intracellular concentration of Hyp-B is 170 mM in *P. denitrificans* grown on glucose in the presence of 0.5 M NaCl and 20 mM tHyp-B (**Methods**). Therefore, Hyp-B 2-epimerase likely functions with a high intracellular concentration of tHyp-B, so the kinetic constants are both physiologically reasonable and expected[23]. That only four compounds were tested (tHyp-B, L-Pro-B, carnitine, and Gly-B) and two have physiologically relevant kinetic constants confirms that pathway docking enables efficient functional prediction.

The 1.70Å structure of HpbD was determined in the presence of tHyp-B (**Methods**; Supplementary Figure S4 and Table S4). The ligand electron density, with elevated B-factors, was interpreted as a mixture of tHyp-B (substrate) and cHyp-B (product) (Supplementary Figure S5). The betaine forms a pi-cation interaction with Trp320 and is proximal to Asp292, similar to the interactions in the Gly-B periplasmic binding protein (Figure 1c). The predicted pose with D-Pro-B superimposes closely with the experimental pose, explaining the correct computation-based prediction of substrate specificity.

We also hypothesized that the proteins/enzymes encoded in the HpbD genome neighborhoods constitute a catabolic pathway that degrades tHyp-B to α-ketoglutarate (Figure 4a), with HpbD catalyzing the first step in which tHyp-B is 2-epimerized to cHyp-B, i.e., the *in vivo* activity of HpbD is Hyp-B 2-epimerase. Subsequently, HpbB1/HpbC1, the Rieske-type protein, catalyzes demethylation of cHyp-B to N-methyl cHyp; HpbA, a flavin-dependent enzyme, converts N-methyl cHyp to cHyp; HpbE, a D-amino acid oxidase, catalyzes oxidation of cHyp to its iminoacid; HpbG, a member of the dihydrodipicolinate synthase superfamily[24] catalyzes dehydration of the 4-OH group and "hydrolysis" of the 5-amino group to α-ketoglutarate semialdehyde; and HpbF, an aldehyde dehydrogenase, catalyzes oxidation of α-ketoglutarate semialdehyde to α-ketoglutarate. This pathway would allow utilization of tHyp-B as a carbon and nitrogen source. The activities predicted for HpbE and HpbG were described recently in pathways for tHyp catabolism in *Pseudomonas aeruginosa*, *P. putida* and *S. meliloti*[25,26]; however, the sequences of HpbE and HpbG are sufficiently divergent (<35% sequence identity) that the cHyp oxidase and cHyp imino acid dehydratase/deaminase functions could not be assigned to the *P. bermudensis* and *P. denitrificans* enzymes without additional information (Supplementary Figure S6).

When tHyp-B is used as an osmoprotectant (sea water is ~0.6 M NaCl), its catabolism should be depressed to maintain high intracellular concentrations. However, in the absence of osmotic stress, bacteria should be able to catabolize tHyp-B as a carbon and nitrogen source. *P. denitrificans* utilizes both tHyp-B and cHyp-B as carbon and nitrogen sources in the presence of low salt, as expected if the genome encodes the proposed catabolic pathway (**Methods**). Moreover, tHyp-B alleviates growth inhibition in glucose medium by high salt (0.5 M NaCl), arguing that *P. denitrificans* uses tHyp-B as an osmoprotectant (Supplementary Figures S7–S9). Growth stimulation by tHyp-B in high salt glucose medium

could result from both osmoprotection and catabolism of tHyp-B. To address this possibility we utilized strain RPd4 which lacks the demethylases that convert the isomers of HypB to the isomers of N-methyl Hyp and, therefore, cannot utilize tHyp-B or cHyp-B as carbon source (Supplementary Table S8). Strain RPd4 grew almost as well on high salt glucose medium in the presence of tHyp-B or cHyp-B as the culture without salt supplementation, but growth on high salt glucose medium in the absence of tHyp-B or cHyp-B was strongly inhibited (Supplementary Figure S8). The results establish that tHyp-B and cHyp-B function as osmoprotectants.

We identified the metabolites obtained from tHyp-B under low salt conditions (**Methods**). In addition to Hyp-B (21 mM; Supplementary Table S10), N-methyl Hyp, and Hyp (the carbon-2 epimers cannot be distinguished), the predicted downstream Δ^1-pyrroline-4-hydroxy-2-carboxylate, α-ketoglutarate semialdehyde, and α-ketoglutarate were observed (Supplementary Figures S10 and S11). The metabolites were not detected with succinate as a carbon source. In high salt glucose medium containing tHyp-B, the intracellular concentration of Hyp-B was 170 mM (as expected for an osmolyte; Supplementary Table 10); however, its downstream metabolites were not detected. Thus, the flux through the pathway is regulated so tHyp-B is not catabolized when it is needed as osmoprotectant[19,20,27]. No Hyp-B was detected in cells grown on high salt glucose medium, establishing that *P. denitrificans* lacks an anabolic pathway for tHyp-B.

We used qRT-PCR to investigate expression of the genes encoding the catabolic pathway (**Methods and** Supplementary Table S6). *P. denitrificans* encodes one orthologue of Hyp-B 2-epimerase (HpbD) and the FAD-dependent N-methyl Hyp demethylase but two orthologues of the remaining proteins/enzymes involved in transport of tHyp-B and its catabolism (Figure 2). The genes encoding the pathway are upregulated by tHyp-B and cHyp-B, as expected if their encoded proteins are involved in the catabolic pathway.

The effects of high salt were determined using equimolar concentrations of glucose and either tHyp-B or cHyp-B. Salt (0.5 M NaCl) enhanced expression of the transporters (HpbN/HpbO/HpbH/HpbJ and HpbX/HpbY/HpbZ). In contrast, salt decreased expression of the genes encoding Hyp-B 2-epimerase (HpbD), both Hyp-B demethylases (HpbB1/HpbC1 and HpbB2/HpbC2) and the single N-methyl Hyp demethylase (HpbA) (Supplementary Table S6). Transport of tHyp-B/cHyp-B is required for uptake as both osmolytes and carbon/nitrogen sources; expression of their transporters is enhanced, whereas epimerization and demethylation are suppressed, thereby allowing tHyp-B/cHyp-B to be retained as osmolytes.

The genes encoding the *P. denitrificans* pathway were individually disrupted by insertion of antibiotic-resistance cassettes (**Methods and** Supplementary Table S7). The growth phenotypes are consistent with the predicted functions (Supplementary Discussion).

In summary, we used homology modeling and metabolite docking to several proteins encoded by a gene cluster to guide *in vitro* assignment of the novel Hyp-B 2-epimerase activity to 2PMQ, a structure determined by the Protein Structure Initiative. With knowledge of the catalytic capabilities of enzyme superfamilies, we also predicted the pathway that catabolizes cHyp-B to α-ketoglutarate. These predictions were verified by metabolomics

and genetics. Finally, we used transcriptomics to demonstrate that Hyp-B 2-epimerase is a "switch" that determines whether the tHyp-B is accumulated as an osmolyte or catabolized as carbon and nitrogen source.

Orthologues of HpbD can be identified in 20 microbial species (http://sfld.rbvi.ucsf.edu/), so both the *in vitro* activity and *in vivo* functional assignments identified in this study can be extended to these proteins/organisms. Moreover, we expect that the Hyp-B 2-epimerase activity assigned to HpbD will be used to leverage discovery of the *in vitro* activities and *in vivo* functions of uncharacterized homologues in the enolase superfamily.

We propose that pathway docking is an efficient strategy for predicting novel *in vitro* enzymatic activities and *in vivo* physiological functions. Additional refinements and applications of this strategy are in progress.

Methods

a. Homology modeling and docking

Sequence similarity network analysis—All sequences from the MLE subgroup in the Structure-Function Linkage Database (SFLD)[7] were used in the MLE subgroup network analysis. BLAST analyses were performed using these sequences as queries in an all-by-all fashion. The details have been described previously[4].

Sequences in the cHyp oxidase and Pyr4H2C deaminase networks were collected by BLAST, using red and blue dots in Supplementary Figure S3 as queries, and 10^{-100} as BLAST E-value cutoff. The Pythoscape v1.0 program[36] was used to make the two networks.

Homology modeling and docking—The models of HpbJ and HpbB1 were built by using our in-house software Protein Local Optimization (PLOP, marketed as Prime by Schrödinger LLC). The template PDBs used for HpbJ and HpbB1 are 1R9L and 3N0Q, respectively. The sequence alignment of each pair of target and template was made by the L-INS-i method in MAFFT v6.925b[37]. While constructing the models, we included both the metal ions and the cocrystallized ligands (if any) from the templates. For docking, 2PMQ, the 1.72 Å X-ray apo structure of HpbD, was used. The structures were processed by Protein Preparation Wizard in Schrödinger Suite 2009[38] prior to docking.

Two different libraries were used for docking in the active site of HpbD. The large metabolite library is the KEGG metabolite library plus potential substrates for members of the enolase superfamily not found in KEGG. The small library for focused docking to HpbD contained 31 betaines and betaine-like metabolites.

The KEGG metabolite library was generated by the following steps. First, we obtained 14,039 compounds from the KEGG COMPOUND database; then, we used LigPrep[39] in Schrödinger Suite 2009 to convert each compound from 2D to 3D, and enumerate up to 32 chiral forms. During this process, compounds with unspecified chemical groups (listed as "R"), polymers, and monatomic ions were automatically removed. Next, we removed compounds with molecular weights greater than 400 because we did not expect these would

fit into the active site of HpbD, as well as duplicates generated by LigPrep preparation. We obtained 82,952 unique KEGG ligands.

Potential substrates for the enolase superfamily proteins include all dipeptides (formed by 20 standard amino acids), several types of N-capped (N-succinyl, N-acyl, N-formimino, N-formyl and N-carbamoyl) amino acids, acid sugars (monoacid sugars, diacid sugars, uronate sugars, 6-deoxy acid sugars and phospho sugars) and their corresponding enolates (i.e., high energy intermediates); these also were processed by LigPrep. After combining the KEGG metabolite library with these additional potential substrates for members of the enolase superfamily, and removing duplicates, the library used for docking into the active site of HpbD contained 87,098 unique ligands.

The betaine library used for docking to the active site of HpbJ contains 31 betaines and betaine-like metabolites, including dimethylsulfoniopropionate (DMSP), ectoine, 5-hydroxyectoine and trigonelline; the compounds are listed in Supplementary Table S2. The members of this library also were processed by LigPrep.

Two docking methods were used. Glide SP docking followed by MM-GBSA was used with HpbD; the details have been described[40]. The Glide XP docking method[41] was used with HpbJ.

b. *In Vitro* Activity Measurements

Cloning, Expression, and Purification of the 2PMQ (HpbD)—The protein sample was provided by the NYSGXRC structural genomics center (PSI-2; U54GM074945).

Cloning, Expression, and Purification of the 2PMQ Orthologue (HpbD) from P. denitrificans—The protein sample was provided by the NYSGXRC structural genomics center.

Cloning, Expression, and Purification of the HypF from P. denitrificans—The *hypF* gene was amplified by PCR using primers P17 and P18 and genomic DNA of *P. denitrificans* as a template. The PCR product was digested with NdeI and BglII and ligated to pET15b expression vector, yielding plasmid pRK9. The cloned HypF was expressed in *E. coli* BL21 (DE3) cells for protein purification. 4 L of LB media was shaken at 20°C and induced with 0.5 mM IPTG when the culture reached 0.6 at OD_{600}. The cells were harvested after 24 hours by centrifugation. The cells were resuspended in 100 mL buffer containing 5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, and 0.1 mM dithiothreitol (DTT). The suspension was lysed by sonication and debris was cleared by centrifugation. The supernatant was applied to a Sepharose FF column charged with Ni^{2+} and eluted with a linear gradient (450 mL) of 60 mM to 1 M of imidazole buffered with 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9, and 0.1 mM DTT. The purest fractions were pooled and dialyzed against 20 mM Tris-HCl, pH 8.0, and 0.1 mM DTT.

Screening HpbD Activity by ^{1}H NMR—Epimerization/racemization of tHyp-B, L-Pro-B, Gly-B, and carnitine were screened by disappearance of the α-proton in a D_2O-containing buffer *via* ^{1}H NMR. The reaction mixture contained 10 mM compound, 50 mM Tris-DCl

(pD 8.), 10 mM MgCl$_2$, and 1 μM enzyme and was incubated at 30°C for 16 hours prior to acquistion of the 500 MHz ^{1}H NMR spectrum.

Polarimetric Assay for HpbD Activity—Hyp-B 2-epimerase and L-Pro-B racemase activities were measured at 25 °C by quantitating the change of optical rotation. The assay was performed in a 100 mm pathlength cell in a total volume of 0.8 mL and a Jasco P-1010 polarimeter with a Hg 405 nm filter. Buffer conditions for the assay were 50 mM Tris-HCl (pH 8.0) containing 10 mM MgCl$_2$.

Polarimetric Assay for HpbF Activity—Hyp epimerase activity was measured at 25 °C by quantitating the change of optical rotation. The assay was performed in a 100 mm pathlength cell in a total volume of 0.8 mL and a Jasco P-1010 polarimeter with a Hg 405 nm filter. Buffer conditions for the assay were 50 mM sodium phosphate buffer (pH 8.0) containing 1 mM DTT.

c. Structure Determination

Expression of HpbD—Plasmid 9437a2BNt21p1, obtained from NYSGRXC stock clones [42], consists of a codon optimized HpbD gene in pSGX2, a derivative of pET26b (NOVAGEN), with the N-terminal methionine of HpdD changed to the sequence MAHHHHHHSL. The vector was transformed into Rosetta2(DE3)pLysS Competent Cells (EMDMILLIPORE) and plated on LB-Agar plates. 5–10 colonies were added to 75 mL of LB with 0.5% glucose and grown overnight at 37° C. HpbD was expressed utilizing 4 L of autoinduction media at 25° C [43,44]. The starter culture and autoinduction media were distributed equally among ten 2L baffled flasks, and shaken at 300 rpm for approximately 24 hours to an OD >15. All growth media contained 100 μg/mL kanamycin (KAN) and 50 μg/mL chloramphenicol (CAM). Cells were pelleted and stored at −80 C.

Purification of HpbD—All purification was performed at 4° C. Cells were resuspended in 3X (*w/w*) buffer A (50 mM Hepes (pH 7.8), 150 mM NaCl, 20 mM imidazole, 10% (*w/v*) glycerol) supplemented with 0.1% (*v/v*) Tween-20 and disrupted by sonication. Cellular debris were removed by centrifugation, and the supernatant was applied to a 10 mL metal affinity column (Ni Sepharose High Performance:GE HEALTHCARE) pre-equilibrated with buffer A. The column was washed with five column volumes (CVs) of buffer A and subsequently eluted with 2 CV of the same buffer with 300 mM imidazole. Eluted protein was pooled and applied to a 120 mL Superdex 200 column (GE HEALTHCARE) equilibrated with buffer B (10 mM HEPES pH 7.5, 150 mM NaCl, 5% (*v/v*) glycerol). Fractions with greater than 95% purity by SDS-PAGE analysis were pooled, concentrated by centrifugal ultrafiltration, snap frozen in liquid N$_2$, and stored at −80 C.

Crystallization and structure solution of HpbD—Crystals were obtained by vapor diffusion at 18° C using the sitting-drop vapour diffusion method in 96-well IntelliPlates (ART ROBBINS). Equal volumes of protein [24.6 mg/mL in 10 mM HEPES (pH 7.5, 150 mM NaCl, 5% (*w/v*) glycerol, 5 mM EDTA, 2 mM NiCl$_2$] and crystallization buffer (70% (*v/v*) MPD, 0.1 M HEPES pH 7.5) were combined and equilibrated against 70 μL of crystallization buffer in the reservoir. Crystals grew as parallelograms measuring 0.05 × 0.15

mm over a 1–2 week time period. Crystals were soaked for 2 min in the reservoir solution supplemented with 200 mM trans-4-hydroxy-L-proline betaine (tHyp-B) and 50 mM $MgCl_2$. Crystals were flash cooled by immersion in liquid nitrogen, and subseqeuntly stored and shipped to the Advanced Photon Source beamline 31-ID (LILLY-CAT). Data were collected at 100 K and a wavelength of 0.97929 Å. Crystals were rotated through 180° in 1° increments and the data processed using MOSFLM[45] and scaled using SCALA[46] in space group $P2_1$. The unliganded structure (APO) was determined by selenomethionine SAD phasing by the NYSGXRC in 2007 from a C- terminally His-Tagged protein (2PMQ, Supplemental Table S4), with a dimer per asymmetric unit. A single subunit from the unliganded structure was used as a search model in molecular replacement for the structural determination of the liganded structure. PHASER[47] within the refinement package PHENIX[48] located eight subunits which subsequently could be assembled into the molecular octamer. Several rounds of manual rebuilding and ligand and water fitting within the molecular graphics program COOT[49] followed by refinement in PHENIX were performed to finalize the structure. Several iodine atoms (7–8/subunit), originating from the synthesis of the substrate, were modeled into difference density peaks with features suggestive of bound iodine. The geometry restraints for tHyp-B were produced using the PRODRG2 server[50]. The density was fit equally well by tHyp-B and the product cHyp-B. It is presumed that the protein is active in the crystalline form and that the density is most likely a mix of substrate and product; however, only tHyp-B was used in refinement. The final structure has 98.6% of its residues in favored regions of the Ramachandron plot, and 0.0% in dissallowed regions (4H2H, Supplemental Table S4). The liganded structure (4H2H) superimposes with the APO structure (subunit A vs. subunit A) with an RMSD of 0.25 Å over 366 aligned Cαs with no substantial changes to the structure upon ligand binding.

d. Microbiology

Bacterial Strains and Growth Conditions—*P. denitrificans* PD1222 wild-type and mutant strains were grown in minimal medium containing (g per liter) K_2HPO_4, 6.0; 4.0 g of KH_2PO_4 4;, sodium molybdate, 0,15; $MgS0_4{\cdot}7H_2O$, 0.2, $CaCl_2$,; 0.04; $MnSO_4{\cdot}2H_2O$, 0.001; and $FeSO_4{\cdot}7H_2O$ 1.1 g with or without 1.6 g of NH_4Cl as nitrogen source. *P. denitrificans* PD1222 was grown aerobically at 30°C in mineral medium supplemented with either glucose/succinate, t/c-Hyp-B or methanol at the same concentration (20 mM). *E. coli* strain TOP10 (Invitrogen) was used for plasmid maintenance, propagation and cloning purpose. *E. coli* strain S17-1 was used for conjugation[51]. Strains used are listed in Supplementary Table S7. To study the role of t/c-Hyp-B in osmoprotection, cultures were grown in minimal medium with glucose and +/− 500 mM NaCl. *E. coli* cultures were grown at 37°C in LB (Luria-Bertani media). Antibiotics were used at the following concentrations (in μg/ml: kanamycin sulfate, 50; chloramphenicol, 35 and sodium ampicillin, 100.

Construction of Disruption Mutants—Gene inactivation mutant strains were generated in *P. denitrificans* by conjugation[29] or electroporation[30].

Molecular Biology Protocols—Chromosomal DNA was isolated from 3–5 mL of *P. denitrificans* PD1222 cell cultures using a DNeasy Blood & Tissue Kit from Qiagen

(Hilden, Germany) or a Wizard Genomic DNA Purification Kit from Promega (Madsion, WI). Restriction enzymes, DNA polmerases, and T4 DNA ligases were purchased from New England Biolabs (Ipswich, 136 MA), Fermentas (St. Leon-Rot, Germany), Invitrogen (Carlsbad, CA), or Promega (Madison, WI). Plasmids were prepared from *E. coli* TOP10 cells using a Plasmid Mini Kit from Qiagen.

Gene disruption plasmids—Most plasmids for construction of gene disruptions were obtained by a standard protocol in which the appropriate chromosomal segments were amplified from *P. denitrificans* PD1222 genomic DNA using *Pfu* polymerase followed by insertion of the PCR products into the pGEM T Easy vector (Promega). The resulting plasmids were then digested with EcoRV (pRK1), NruI (pRK2) BmgBI (pRK4) and ligated to a 900 bp fragment blunted-ended chloramphenicol resistance (*cat*) cassette. These plasmids were then used as PCR templates with the same primers and the products were ligated to vector pSUP202, which had been digested with EcoRI and treated with the Klenow fragment of DNA polymerase I, plus the four dNTPs to give the plasmids used for gene disruption. Primer sets P1+P2, P3+P4 and P7+P8 gave rise to plasmids pRK1, pRK2 and pRK4, respectively. Plasmid pRK5 was obtained similarily using primers P9 and P10 except that the original PCR product was inserted into vector pCR2.1-TOPO (Invitrogen) and the *cat* cassette was inserted into the HincII site. Plasmid pRK3 was obtained similarly using primers P5 and P6 except that a 1400 bp kanamycin resistance cassette was inserted into the BmgBI site of the intermediate pGEM T Easy construct. In the case of plasmids pRK6 and pRK8, the products of PCR (primers P11+P12 and P15+P16, respectively) amplification from chromosomal DNA were inserted into pETBlue-1 (Novagen) and the kanamycin resisance cassette was inserted into the SfoI site of this plasmid. Plasmid pRK7 was obtained by the same manipulations using primers 13 and 14 except that the chloramphenicol cassette was inserted into the SfoI site.

Expression plasmids—These plasmids were constructed as above except that the PCR products obtained from chromosomal DNA contained the promoter and coding sequences and were directly ligated to vector pSUP404.2 which had been digested with EcoRI and treated with the Klenow fragment of DNA polymerase I plus the four dNTPs. Primers P19+P2, P20+P4 and P21+P10 gave rise to plasmids pRK10, pRK11 and pRK12.

High level protein expression plasmids pRK9 and pRK13 were obtained by insertion of the PCR products obtained from chromosomal DNA into pGEM T Easy. Plasmid pRK9 was obtained by ligation of the NdeI and BglII *hypF* fragment of the intermediate plasmid to pET15b digested with NdeI and BamHI whereas pRK13 resulted from ligation of the *hypO* NdeI-BplI fragment of the pGEM T Easy intermediate plasmid into pET15b cut with the same enzymes.

Cell preparation for gene expression analysis—*P. denitrificans* PD1222 wild-type or mutant cultures were grown in 10 mL of minimal media with 20 mM glucose as carbon source to an OD_{600} of 0.4. The cells were pelleted by centrifugation (4,000×*g*, 10 min, 4 °C). The cell pellet was washed twice and resuspended in 10 mL of minimal medium lacking a carbon source. The cultures were divided into two 5 mL aliquots. Glucose or succinate was added to one aliquot and Hyp-B (either *trans* or *cis*) or methanol was added to

the other as carbon source followed by aerobic growth at 30 °C for 1 h prior to cell harvest. For studying the effect of salt stress on gene expression, NaCl was added to the minimal medium to 500 mM; glucose or succinate were the carbon sources and Hyp-B was the osmoprotectant.

RNA Sample preparation—For preparation of RNA samples, 1.0 – 5.0 mL of an actively growing *P. denitrificans* PD1222 culture (OD_{600} of 0.5 – 0.6) were added to 2 volumes of RNA protect Bacteria Reagent (Qiagen). After vortex mixing for 10 s the solution was incubated for 5 min at room temperature. The cells were pelleted by centrifugation ($10,000 \times g$ for 5 min at 4°C), the supernatant was decanted, and the remaining liquid was removed. RNA isolation was performed on ice in an RNAse-free environment using an RNeasy Mini Kit (Qiagen), following the protocols for bacteria given by Qiagen. RNA concentrations were determined by absorption at 260 nm with one absorbance unit corresponding to 44 μg mL^{-1} RNA. Isolated RNA was analyzed by agarose gel electrophoresis, and spectrophotometrically in the Nanodrop (Thermo) using the ratios A_{260}/A_{280} and A_{260}/A_{230}, and the A_{350}/A_{220} absorption spectra to assess sample integrity and purity. RNA preparation purity and absence of DNA was validated by agarose gel electrophoresis and control PCR reactions. The RNA preparations were stored at −80°C until use.

qRT-PCR—Reverse transcription (RT) was performed on 1 μg of total RNA by using the RevertAid H Minus First Strand cDNA synthesis kit (Fermentas) according to the manufacturer's protocol. Of the cDNA, 1 μL was used in separate PCR reactions of 20 μL for each gene. Minus-RT controls were performed to test for genomic DNA contamination in each RNA sample. Primers were designed by Universal probe library system (Roche Applied Science). The primer length was 21 to 27 nucleotides, with a theoretical T_m of 58–60°C. The amplicon size ranged from 66 to 110 bp. Real-time PCR was carried out in 96-well plates using the Roche LightCycler 480 (LC480). The 20 μL PCR mix was prepared by adding to 1 μL cDNA template, 2 μl of forward and reverse primers (final concentration 150 nM of each primer) and 10 μL of SYBR 2X Concentration Green Master Mix (Roche). The PCR conditions were: one cycle at 95°C for 5 min, 40 cycles of amplification at 95°C for 15 s, 60°C for 1 min. Finally, a dissociation program was applied with one cycle at 95°C for 15 s, 60°C for 1 min and 95°C for 15 s. Efficiency of all the primers used in qRT PCR was calculated as 97% ± 2%. The gene expression data were expressed as Cp or cross point values. The 16S rRNA was used as a reference gene. The data was analysed by $2^{-\Delta\Delta CT}$ (Livak) method[52]. Data presented is the average of five biological replicates. Primer sequences are provided in Supplementry Table S9.

e. Metabolomics

Metabolomics analysis of the *P. denitrificans* tHPB degradation pathway—LC-FTMS metabolomics of whole cell extracts were carried out with samples of *Paracoccus denitrificans* fed with tHyp-B with or without osmotic stress. First, these experiments were performed as a time course in minimal medium with or without the addition of tHyp-B. Succinate minimal medium grown cells were diluted 1:500 into 250 mL minimal medium with tHyp-B as the sole carbon source and grown to an OD_{600} of ~ 0.7 (approximately 18

hours). The culture was harvested by centrifugation (4000 × g, 10 min, 4°C), washed, and resuspended in minimal medium without carbon source. The cell suspension was then depleted of catabolic metabolites by 15 minute incubation at 30°C before transferring them to ice. After verifying the concentrated cell density by multiple 50-fold dilutions into minimal medium (calculated $OD_{600} = 26.7 \pm 0.2$), 1 mL aliquots of OD_{600} 6 cell suspensions were prepared in eppendorf tubes on ice in minimal medium. 20 mM tHyp-B was added quickly to half of the samples before incubation in a 30°C water bath, and at time points of 0, 2, 5, and 15 minutes, samples were transferred on ice to a 4°C centrifuge for pelleting at 16,000 × g for 2 minutes. The supernatant was then removed, and the cell pellets were flash frozen in liquid nitrogen. The process of collecting time point samples from 30°C to liquid nitrogen took approximately 3 minutes. Samples were stored at −80°C before analysis.

To ascertain the effect of osmotic stress on the levels of the metabolites from tHyp-B catabolism, *P. denitrificans* was grown as described above except that a replicate culture with 0.5 M NaCl was also prepared. After these cultures reached OD_{600} ~ 0.5, each were concentrated by centrifugation, aliquoted into 1mL of cells with $OD_{600} = 6$, pelleted by centrifugation, separated from the supernatant, and frozen for analysis.

Metabolomics analysis of these samples followed the procedure from Erb *et al.*[53]. The cell pellets were extracted with 0.375 mL of 10 mM ammonium bicarbonate (pH 9.2) in 90% acetonitrile by pipetting followed by 15 minutes of vortexing at room temperature. The extraction was cleared of cell debris via two rounds of centrifugation at 16,000 × g before analysis. Samples were applied to a custom 11T LTQ-FT mass spectrometer (Thermo-Fisher Scientific) with an Agilent 1200 HPLC system equipped with a Sequant Zic-HILIC column (2.1 mm × 150 mm) previously equilibrated with the extraction buffer (solvent B). Solvent A was 10 mM ammonium bicarbonate, pH 9.2. Each extracted sample was injected at 100 μL for three separate chromatographic runs. The samples were eluted with a 200 μL/min flow rate using the following elution profile: 100% B for 17 minutes, a linear gradient from 100% B to 40% B over 3 minutes, and another linear gradient from 40% to 100% B over 15 minutes. Data were collected at a resolution of 50,000 with full scan set to m/z 100–1000, and duplicate samples were individually analyzed in either positive or negative mode.

Data analysis was performed with the Qualbrowser application of Xcalibur (Thermo-Fisher Scientific) (Supplementary Figures S10 and S11). For metabolites, tHyp-B, N-methyl cHyp, cHyp, and α-ketoglutarate, standards were run to verify retention time. Unfortunately, the *P. denitrificans* samples appeared to damage the Zic-HILIC column over time, as after dozens of runs, many of the peaks broadened and showed longer retention times. While HPLC analysis of betaine derivatives of biological origin have been reported[28,54–56], the metabolic analysis of bacteria utilizing betaines has not. It is assumed that the extremely large concentrations of tHyp-B accumulated in *P. denitrificans* cells were to blame due to overloading the column.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput Biol. 2009; 5:e1000605. [PubMed: 20011109]

2. Gerlt JA, et al. The Enzyme Function Initiative. Biochemistry. 2011; 50:9950–9962. [PubMed: 21999478]

3. Hermann JC, et al. Structure-based activity prediction for an enzyme of unknown function. Nature. 2007; 448:775–779. [PubMed: 17603473]

4. Song L, et al. Prediction and assignment of function for a divergent N-succinyl amino acid racemase. Nat Chem Biol. 2007; 3:486–491. [PubMed: 17603539]

5. Kalyanaraman C, Jacobson MP. Studying enzyme-substrate specificity *in silico*: a case study of the *Escherichia coli* glycolysis pathway. Biochemistry. 2010; 49:4003–4005. [PubMed: 20415432]

6. Babbitt PC, et al. The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the alpha-protons of carboxylic acids. Biochemistry. 1996; 35:16489–16501. [PubMed: 8987982]

7. Gerlt JA, Babbitt PC, Rayment I. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. Arch Biochem Biophys. 2005; 433:59–70. [PubMed: 15581566]

8. Gerlt JA, Babbitt PC, Jacobson MP, Almo SC. Divergent evolution in enolase superfamily: strategies for assigning functions. J Biol Chem. 2012; 287:29–34. [PubMed: 22069326]

9. Tanabe, M.; Kanehisa, M. Using the KEGG database resource. In: Baxevanis, Andreas D., et al., editors. Current protocols in bioinformatics. Vol. Chapter 1. 2012.

10. Hermann JC, et al. Predicting substrates by docking high-energy intermediates to enzyme structures. J Am Chem Soc. 2006; 128:15882–15891. [PubMed: 17147401]

11. Schiefner A, et al. Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli. J Biol Chem. 2004; 279:5588–5596. [PubMed: 14612446]

12. Goldmann A, et al. Symbiotic plasmid genes essential to the catabolism of proline betaine, or stachydrine, are also eequired for efficient nodulation by *Rhizobium-Meliloti*. FEMS Microbiol Lett. 1994; 115:305–311.

13. Burnet MW, et al. The stachydrine catabolism region in *Sinorhizobium meliloti* encodes a multi-enzyme complex similar to the xenobiotic degrading systems in other bacteria. Gene. 2000; 244:151–161. [PubMed: 10689197]

14. Wargo MJ, Szwergold BS, Hogan DA. Identification of two gene clusters and a transcriptional regulator required for Pseudomonas aeruginosa glycine betaine catabolism. J Bacteriol. 2008; 190:2690–2699. [PubMed: 17951379]

15. Daughtry KD, et al. Quaternary ammonium oxidative demethylation: X-ray crystallographic, resonance Raman, and UV-visible spectroscopic analysis of a Rieske-type demethylase. J Am Chem Soc. 2012; 134:2823–2834. [PubMed: 22224443]

16. Larsen PI, Sydnes LK, Landfald B, Strom AR. Osmoregulation in *Escherichia coli* by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose. Arch Microbiol. 1987; 147:1–7. [PubMed: 2883950]

17. Hanson AD, et al. Osmoprotective compounds in the Plumbaginaceae: a natural experiment in metabolic engineering of stress tolerance. Proc Natl Acad Sci U S A. 1994; 91:306–310. [PubMed: 8278383]

18. Amin US, Lash TD, Wilkinson BJ. Proline betaine is a highly effective osmoprotectant for *Staphylococcus aureus*. Arch Microbiol. 1995; 163:138–142. [PubMed: 7710327]

19. Bernard T, Pocard JA, Berrould B, Le Rudulier D. Variations in the response of salt-stressed Rhizobium strains to betaines. Arch Microbiol. 1986; 143:359–364.

20. Alloing G, Travers I, Sagot B, Le Rudulier D, Dupont L. Proline betaine accumulation and metabolism in alfalfa plants under sodium chloride stress. Exploring its compartmentalization in nodules. J Bacteriol. 2006; 188:6308–6317. [PubMed: 16923898]

21. Burg MB, Kwon ED, Kultz D. Regulation of gene expression by hypertonicity. Annu Rev Physiol. 1997; 59:437–455. [PubMed: 9074772]

22. Kempf B, Bremer E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch Microbiol. 1998; 170:319–330. [PubMed: 9818351]

23. Bar-Even A, et al. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry. 2011; 50:4402–4410. [PubMed: 21506553]

24. Babbitt PC, Gerlt JA. Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. J Biol Chem. 1997; 272:30591–30594. [PubMed: 9388188]

25. Watanabe S, et al. Identification and characterization of D-Hydroxyproline dehydrogenase and [1]-pyrroline-4-hydroxy-2-carboxylate deaminase involved in novel L-hydroxyproline metabolism of bacteria: metabolic convergent evolution. J Biol Chem. 2012

26. White CE, Gavina JM, Morton R, Britz-McKibbin P, Finan TM. Control of hydroxyproline catabolism in *Sinorhizobium meliloti*. Mol Microbiol. 2012; 85:1133–1147. [PubMed: 22804907]

27. Gloux K, Le Rudulier D. Transport and catabolism of proline betaine in salt-stressed Rhizobium meliloti. Arch Microbiol. 1989; 151:143–148.

28. Lenky CC, McEntyre CJ, Lever M. Measurement of marine osmolytes in mammalian serum by liquid chromatography-tandem mass spectrometry. Anal Biochem. 2012; 420:7–12. [PubMed: 21982861]

29. Van Spanning RJ, et al. A method for introduction of unmarked mutations in the genome of *Paracoccus denitrificans*: construction of strains with multiple mutations in the genes encoding periplasmic cytochromes c550, c551i, and c553i. J Bacteriol. 1991; 173:6962–6970. [PubMed: 1657872]

30. Matsson M, Ackrell BA, Cochran B, Hederstedt L. Carboxin resistance in Paracoccus denitrificans conferred by a mutation in the membrane-anchor domain of succinate:quinone reductase. Arch Microbiol. 1998; 170:27–37. [PubMed: 9639600]

31. Blunden G, et al. NMR-spectra of betaines from marine-algae. Magn Reson Chem. 1986; 24:965–971.

32. Blunden G, et al. Betaines and tertiary sulfonium compounds from 62 species of marine-algae. Biochem Syst Ecol. 1992; 20:373–388.

33. Rathinasabapathi B. Metabolic engineering for stress tolerance: installing osmoprotectant synthesis pathways. Ann Bot. 2000; 86:709–716.

34. Trinchant J, et al. Proline betaine accumulation and metabolism in alfalfa plants under sodium chloride stress. Exploring its compartmentalization in nodules. Plant Physiol. 2004; 135:1583–1594. [PubMed: 15235114]

35. Zhou S, Chen X, Zhang X, Li Y. Improved salt tolerance in tobacco plants by co-transformation of a betaine synthesis gene BADH and a vacuolar Na+/H+ antiporter gene SeNHX1. Biotechnol Lett. 2008; 30:369–376. [PubMed: 17968511]

36. Barber AE 2nd, Babbitt PC. Pythoscape: A framework for generation of large protein similarity networks. Bioinformatics. 2012

37. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 2005; 33:511–518. [PubMed: 15661851]

38. Schrödinger Suite 2009 Protein Preparation Wizard; Epik version 2.0; Impact version 5.5; Prime version 2.1. Schrödinger, LLC; New York: 2009.

39. Schrödinger Suite 2009 LigPrep, version 2.3. Schrödinger, LLC; New York: 2009.

40. Kalyanaraman C, Bernacki K, Jacobson MP. Virtual screening against highly charged active sites: Identifying substrates of alpha-beta barrel enzymes. Biochemistry. 2005; 44:2059–2071. [PubMed: 15697231]

41. Friesner RA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem. 2006; 49:6177–6196.10.1021/jm051256o [PubMed: 17034125]

42. Sauder MJ, et al. High throughput protein production and crystallization at NYSGXRC. Methods Mol Biol. 2008; 426:561–575. [PubMed: 18542890]

43. Fox BG, Blommel PG. Autoinduction of protein expression. Curr Protoc Protein Sci. 2009; (Unit 5):23. [PubMed: 19365792]

44. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005; 41:207–234. [PubMed: 15915565]

45. Leslie AG. The integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr. 2006; D62:48–57. [PubMed: 16369093]

46. Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006; 62:72–82. [PubMed: 16369096]

47. McCoy AJ, et al. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]

48. Zwart PH, et al. Automated structure solution with the PHENIX suite. Methods Mol Biol. 2008; 426:419–435. [PubMed: 18542881]

49. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; D60:2126–2132. [PubMed: 15572765]

50. Schuttelkopf AW, van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Cryst. 2004; D60:1355–1363.

51. Simon R, Priefer U, Puhler A. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Nat Biotechnol. 1983; 1:784–791.

52. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\triangle\triangle CT}$ method. Methods. 2001; 25:402–408. [PubMed: 11846609]

53. Erb TJ, et al. A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. Nat Chem Biol. 2012; 8:926–932. [PubMed: 23042035]

54. Chambers ST, Kunin CM. Isolation of glycine betaine and proline betaine from human urine. Assessment of their role as osmoprotective agents for bacteria and the kidney. J Clin Invest. 1987; 79:731–737. [PubMed: 3546377]

55. Oufir M, et al. Simultaneous measurement of proline and related compounds in oak leaves by high-performance ligand-exchange chromatography and electrospray ionization mass spectrometry for environmental stress studies. J Chromatogr A. 2009; 1216:1094–1099. [PubMed: 19135675]

56. Li C, Hill RW, Jones AD. Determination of betaine metabolites and dimethylsulfoniopropionate in coral tissues using liquid chromatography-time-of-flight mass spectrometry and stable isotope-labeled internal standards. J Chromatogr B Analyt Technol Biomed Life Sci. 2010; 878:1809–1816.
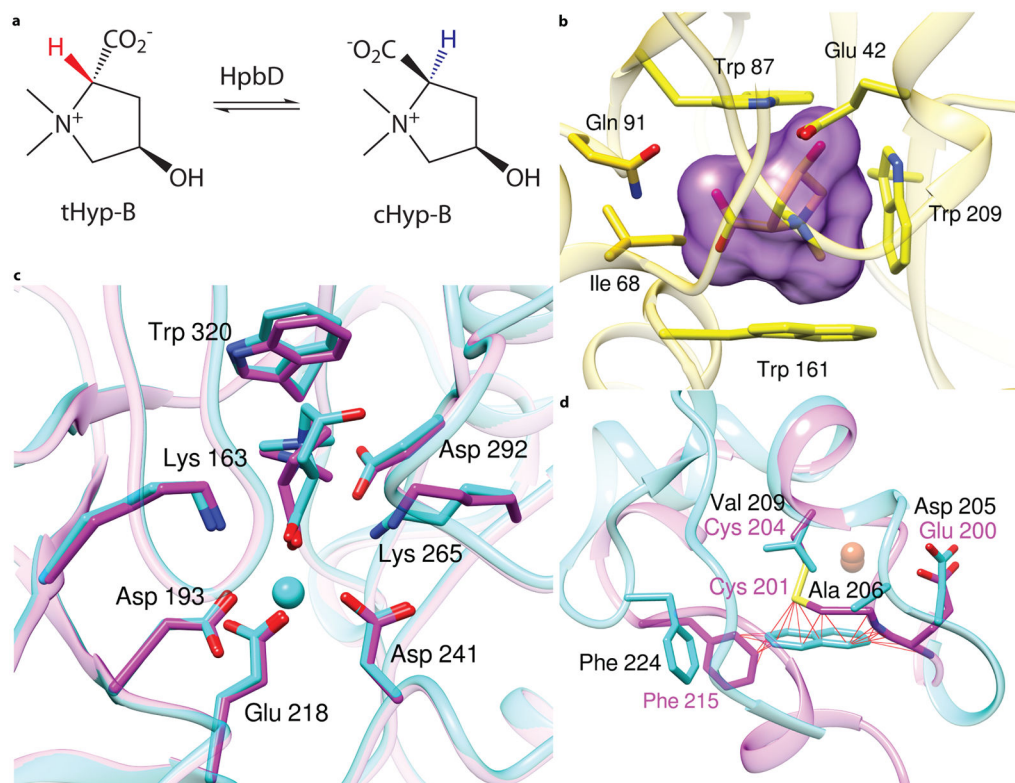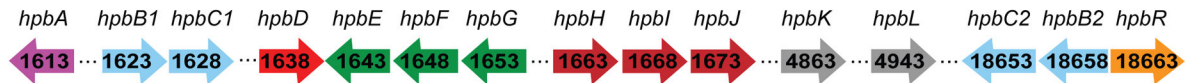
Figure 1. Homology modeling and docking results for HpbD, HpbJ and HpbB1

a, Reaction catalyzed by HpbD, the Hyp-B 2-epimerase. **b**, The binding site of the model of HpbJ, with the top ranked ligand tHyp-B docked. The ligand surface is shown in magenta. **c**, Comparison of HpbD top ranked docking pose of D-Pro-B (magenta) with the experimental pose of tHyp-B (cyan). The unliganded structure used in docking (PDB 2PMQ) and the later determined liganded structure (PDB 4H2H) are shown in magenta and cyan, respectively. **d**, Superposition of the model of HpbB1 (magenta) and the closest characterized Rieske-type protein (cyan, PDB 1O7G, a naphthalene dioxygenase), showing that the active site of the model is too small to accept naphthalene as a substrate. Steric clashes identified by using van der Waals overlap   0.6 Å are shown in red lines.

Pelagibaca bermudensis HTCC2601

hpbA hpbB1 hpbC1   hpbD   hpbE hpbF hpbG   hpbH   hpbI hpbJ   hpbK   hpbL   hpbC2 hpbB2 hpbR

1613 ··· 1623 1628 ···· 1638 1643 1648 1653 ····· 1663 1668 1673 ···· 4863 ··· 4943 ··· 18653 18658 18663

Paracoccus denitrificans PD1222

hpbK hpbL   hpbM hpbD hpbC1 hpbB1 hpbR   hpbX hpbY hpbZ   hpbA hpbC2 hpbB2

0015 0016 ··· 1186 1187 1188 1189 1190 ··· 1193 1194 1195 ··· 1197 2831 2832

78%   84%   64%   61%   59%   59%   83%   62%   58%

hpbN hpbO hpbH hpbJ   hpbE hpbF hpbG   hypF hypO   hypH   hypD

4746 4747 4748 4749 ··· 4751 4752 4753 ··· 4859 4860 ··· 4865 ··· 4867

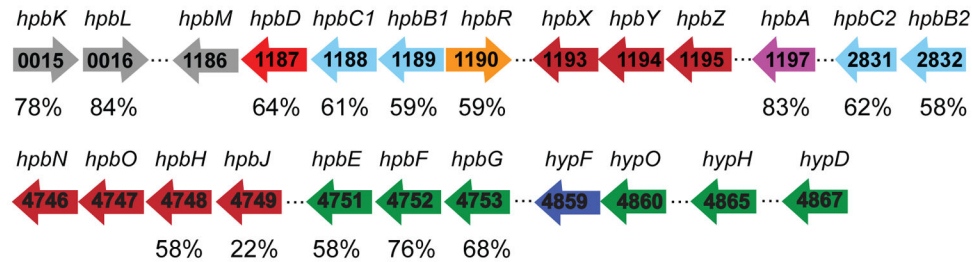   58%   22%   58%   76%   68%

Figure 2. Genome contexts of HpbD in *P. bermudensis* and the orthologous genes in *P. denitrificans*

The genes encoding orthologues are highlighted with the same color; the sequence identities relating orthologues in *P. bermudensis* and *P. denitrificans* are indicated. The ecological sources of tHyp-B would be seaweed (sargasso)[31,32] for the Sargasso Sea bacterium, *P. bermudensis* and plants[33–35] for the soil bacterium, *P denitrificans.*
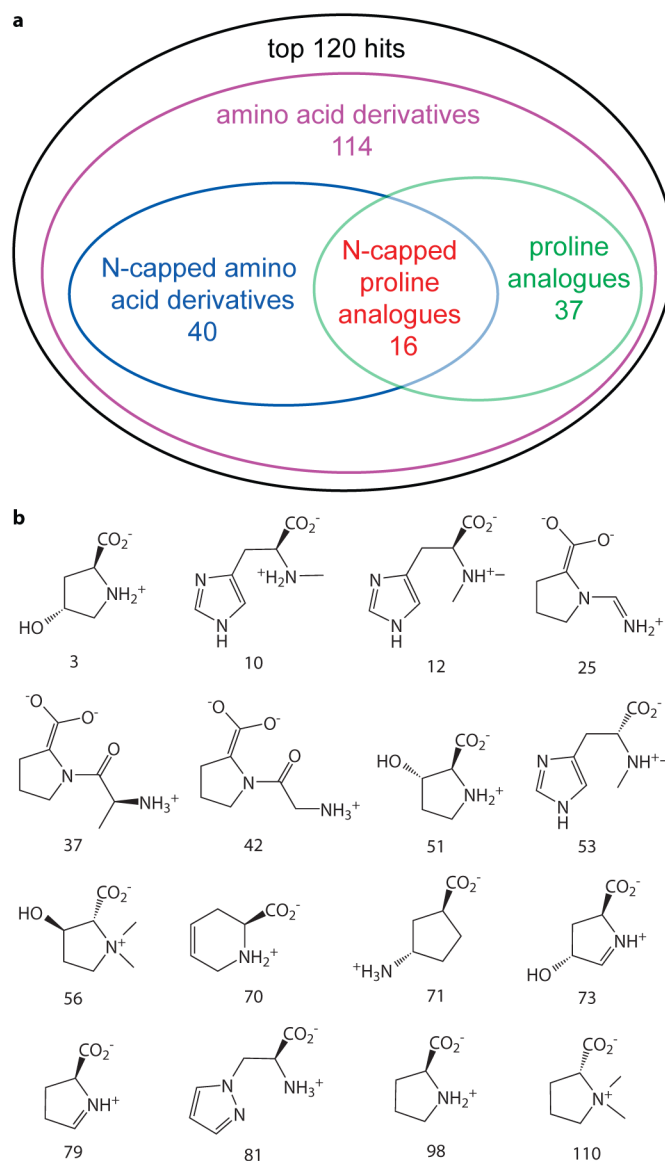
Figure 3. Chemotype analysis of HpbD docking results
a, Enriched chemotypes in the top 120 hits. Most of them are amino acid derivatives, in which N-capped amino acid derivatives and proline analogues are the two most common subtypes. **b**, Proline analogues in the rank ordered list of predicted ligands illustrating the frequent occurrence of N-modified proline analogues. Pro-B, a substrate for HpbD, ranks at number 110 in the list (top 0.12% of the docking library).
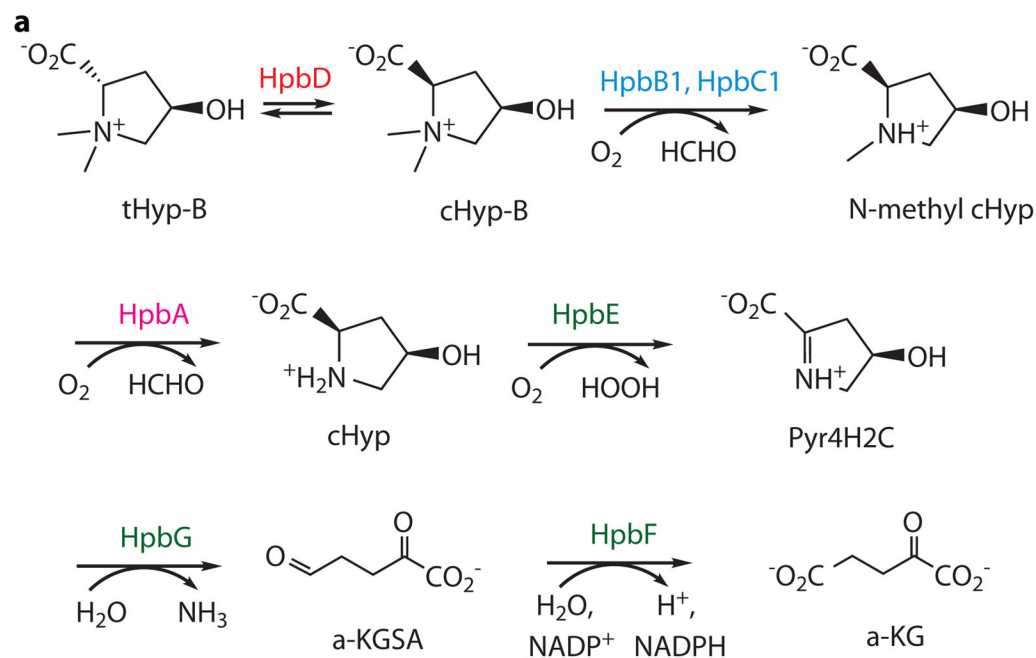
	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M $(s^{-1} M^{-1})$
P. bermudensis			
tHyp-B	330 ± 30	42 ± 10	7800
L-Pro-B	5.3 ± 0.3	8 ± 1	700
P. denitrificans			
tHyp-B	110 ± 9	64 ± 9	1700
L-Pro-B	68 ± 10	34 ± 10	2000

Figure 4. Catabolic pathway for tHyp-B and kinetic constants for HpbD

a, Catabolic pathway for tHyp-B. Based on the genome neighborhood contexts in *P. bermudensis* and *P. denitrificans*, tHyp-B is epimerized to cHyp-B that undergoes two N demethylation reactions to cHyp; cHyp is oxidized, dehydrated/deaminated, and finally oxidized to α-ketoglutarate (α-KG). Pyr4H2C and α-KGSA denote for ^{1}-pyrroline-4-hydroxy-2-carboxylate and α-ketoglutarate semialdehyde, respectively. The enzymes are colored as in Figure 2. b, Kinetic constants for HpbD from *P. bermudensis* and its orthologue from *P. denitrificans*.