LETTER

doi:10.1038/nature12576

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

Suwen Zhao[1]*, Ritesh Kumar[2]*, Ayano Sakai[2]*, Matthew W. Vetting[3]*, B. McKay Wood[2]*, 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[2,6,7] & Matthew P. Jacobson[1]

**Assigning valid functions to proteins identified in genome projects is challenging: overprediction and database annotation errors are the principal concerns[1]. We and others[2] are developing computation-guided strategies for functional discovery with 'metabolite docking' to experimentally derived[3] or homology-based[4] three-dimensional structures. Bacterial metabolic pathways often are encoded by 'genome neighbourhoods' (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 *Escherichia coli*[5]. Metabolite docking to multiple binding proteins and enzymes in the same pathway increases the reliability of *in silico* predictions of substrate specificities because the pathway intermediates are structurally similar. Here we report that structure-guided approaches for predicting the substrate specificities of several enzymes encoded by a bacterial gene cluster allowed 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 also 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 concentrations was established by transcriptomics, confirming the osmolyte role of tHyp-B. This study establishes the utility of structure-guided functional predictions to enable the discovery of new metabolic pathways.**

We have applied a structure-guided strategy using metabolite docking to multiple proteins and enzymes in a metabolic pathway to discover a previously undocumented reaction, 4*R*-hydroxyproline betaine 2-epimerase (Hyp-B 2-epimerase; Fig. 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 a binding protein and a second enzyme encoded by its genome neighbourhood 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 (National Center for Biotechnology Information GI number 114543141) in which two lysine 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 less than 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 and was therefore suitable for virtual metabolite docking for substrate prediction (Supplementary Fig. 1).

Figure 2 shows the genome neighbourhoods of the gene encoding 2PMQ (*hpb*D; Hyp-B 2-epimerase from its functional characterization (see below)) plus a putative *Paracoccus denitrificans* orthologue. The automated TrEMBL annotations (Supplementary Table 1) fail to assign the *in vitro* activity of HpbD or to identify the metabolic pathway. *P. bermudensis* is not genetically tractable; we therefore studied *P. denitrificans*, which encodes one HpbD orthologue and two sets of orthologues of most of the genes neighbouring the *P. bermudensis hpbD* gene. Genome neighbourhoods are 'conserved' for other putative orthologues (about 20 can be identified in the sequence databases at 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 such as 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 by 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 (Fig. 3), permitting the prediction that HpbD is an amino acid racemase/epimerase, with the substrate probably having *N*-substitution.

The genome neighbourhood 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] (PDB 1R9L) was used as the homology model template (HpbJ and 1R9L share 48% sequence identity) (Methods). The binding site contains three tryptophan residues (Fig. 1b) that form a π-cation 'cage' for a quaternary ammonium (betaine), which may also be electrostatically stabilized by Glu 42, located 5.4 Å from the quaternary nitrogen. Thus, the homology model permitted 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 2), so we predicted that HpbJ participates in the transport of tHyp-B. In addition, the HpbD active site contains Trp 320 and Asp 292, which are similarly positioned relative to the predicted binding pose of betaines (Fig. 1c). The structural basis for the predicted specificity of HpbJ therefore refined the prediction that the substrate for HpbD is a proline betaine, for example tHyp-B.

A homology model was constructed for the Rieske-type protein (HpbB1) using a homologue (PDB 3N0Q) as the template (60% sequence identity; Fig. 1d) (Methods). The active site resembled the binding sites in the betaine-binding proteins (aromatic residues and Glu 200); indeed, some Rieske-type proteins are betaine demethylases[12–14] (Supplementary Fig. 2 and Supplementary Table 3). We therefore predicted that the

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, USA. [2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, USA. [4]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94143, USA. [5]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [6]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [7]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.
*These authors contributed equally to this work.

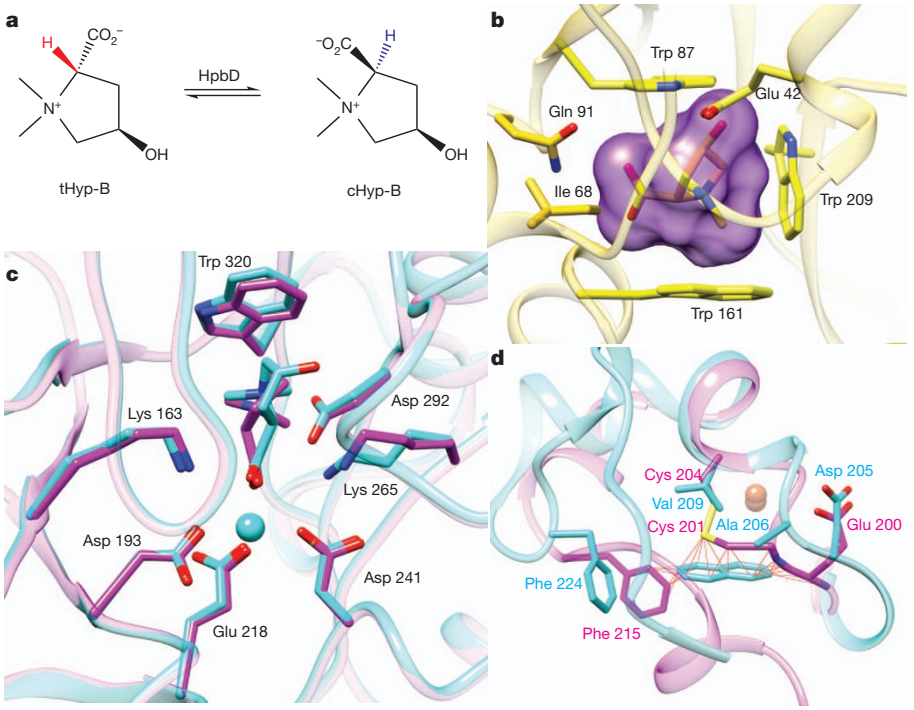



**Figure 1 | Homology modelling and docking results for HpbD, HpbJ and HpbB1.** **a**, The reaction catalysed by HpbD, the Hyp-B 2-epimerase. **b**, The binding site of the model of HpbJ, with the top-ranked ligand tHyp-B docked. The ligand surface is shown in magenta. **c**, Comparison of HpbD top-ranked 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 subsequently 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 a van der Waals overlap of 0.6 Å or more are shown in red lines.

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 enabled us to predict that HpbD uses Hyp-B or Pro-B as a 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 more than 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 being a virtual reaction); in addition, for tHyp-B, resonances associated with cHyp-B, the 2-epimer, were observed (Supplementary Fig. 3). 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 lysine acid–base catalysts.

The kinetic constants for tHyp-B and L-Pro-B were determined for both HpbD orthologues (Fig. 4b) (Methods). Although the $k_{cat}$ values are large, the $K_m$ values are also 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. denitrificans*) species, 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). Hyp-B 2-epimerase therefore probably 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 1.70-Å structure of HpbD was determined in the presence of tHyp-B (Methods, Supplementary Fig. 4 and Supplementary Table 4).

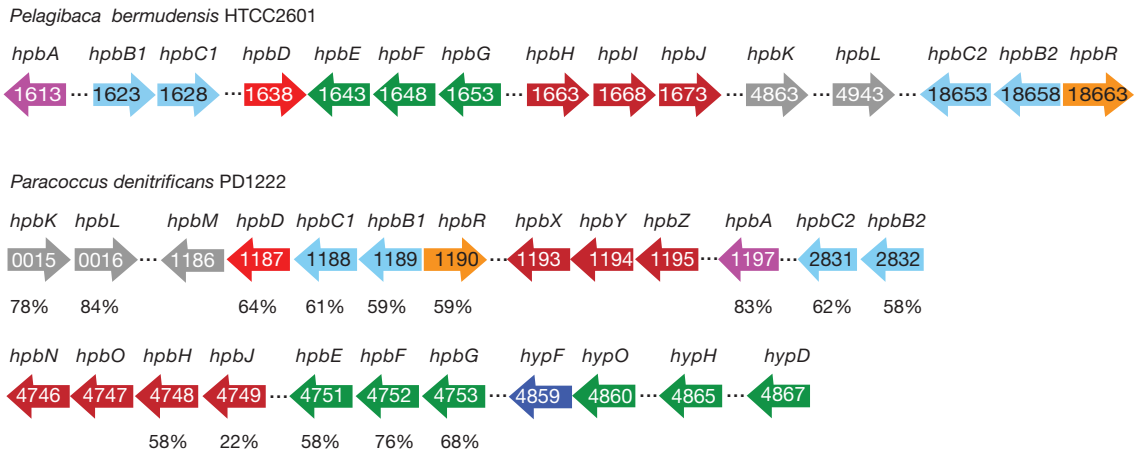



**Figure 2 | Genome contexts of HpbD in *P. bermudensis* and the orthologous genes in *P. denitrificans*.** The genes encoding orthologues are highlighted with the same colour; the sequence identities relating orthologues in *P. bermudensis* and *P. denitrificans* are indicated. The ecological sources of tHyp-B would be seaweed (sargasso) for the Sargasso Sea bacterium *P. bermudensis*, and plants for the soil bacterium *P. denitrificans*.

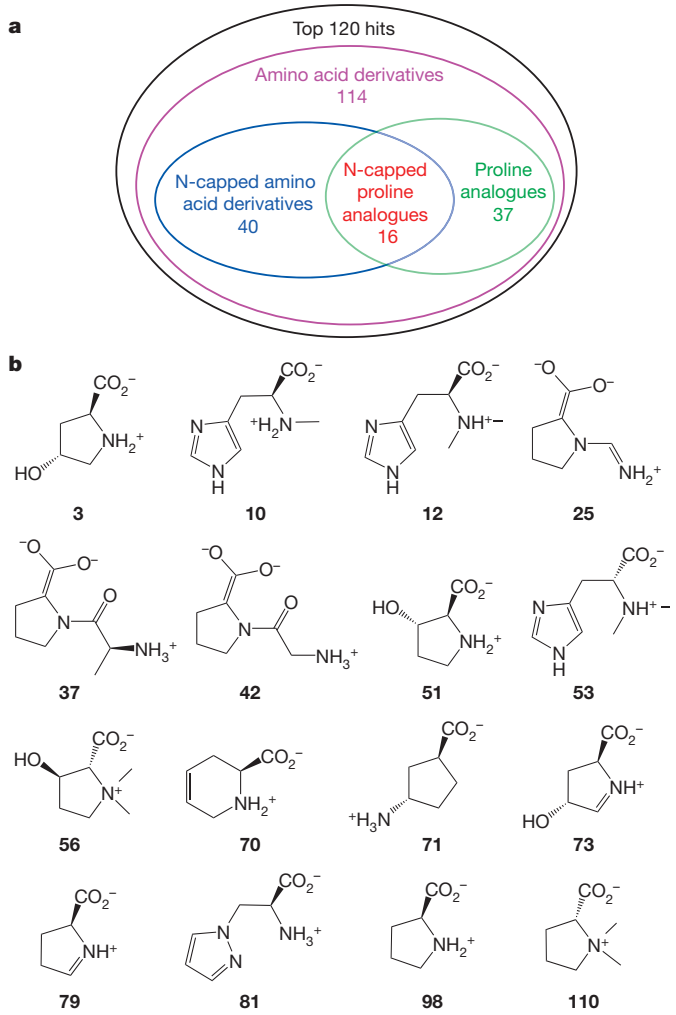
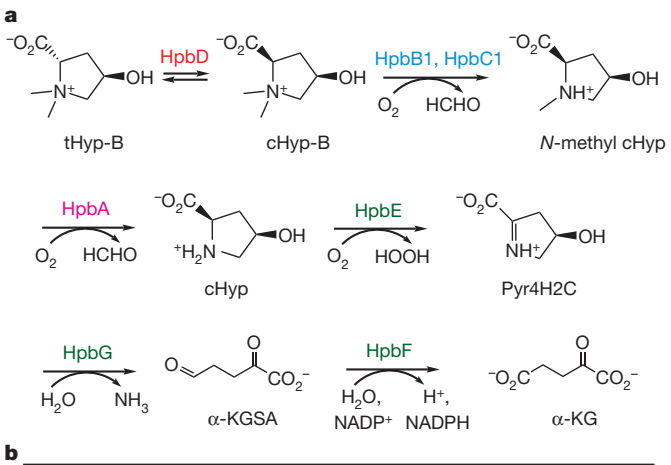



**Figure 3 | Chemotype analysis of HpbD docking results.** **a**, Enriched chemotypes in the top 120 hits. Most of them are amino acid derivatives, in which N-capped amino acid derivatives and proline analogues are the two most common subtypes. **b**, Proline analogues in the rank-ordered list of predicted 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).




b

| Species | | $k_{cat}$ ($s^{-1}$) | $K_m$ (mM) | $k_{cat}/K_m$ ($s^{-1}$ $M^{-1}$) |
|---|---|---|---|---|
| *P. bermudensis* | tHyp-B | 330 ± 30 | 42 ± 10 | 7,800 |
| | L-Pro-B | 5.3 ± 0.3 | 8 ± 1 | 700 |
| *P. denitrificans* | tHyp-B | 110 ± 9 | 64 ± 9 | 1,700 |
| | L-Pro-B | 68 ± 10 | 34 ± 10 | 2,000 |

**Figure 4 | Catabolic pathway for tHyp-B and kinetic constants for HpbD.** **a**, Catabolic pathway for tHyp-B. On the basis of the genome neighbourhood contexts in *P. bermudensis* and *P. denitrificans*, tHyp-B is epimerized to cHyp-B which undergoes two *N*-demethylation reactions to cHyp; cHyp is oxidized, dehydrated and deaminated, and finally oxidized to α-ketoglutarate (α-KG). Pyr4H2C, $\Delta^1$-pyrroline-4-hydroxy-2-carboxylate; α-KGSA, α-ketoglutarate semialdehyde. The enzymes are coloured as in Fig. 2. **b**, Kinetic constants for HpbD from *P. bermudensis* and its orthologue from *P. denitrificans*.

The ligand electron density, with elevated *B*-factors, was interpreted as a mixture of tHyp-B (substrate) and cHyp-B (product) (Supplementary Fig. 5). The betaine forms a π-cation interaction with Trp 320 and is proximal to Asp 292, similar to the interactions in the Gly-B periplasmic binding protein (Fig. 1c). The predicted pose with D-Pro-B superimposes closely on the experimental pose, explaining the correct computation-based prediction of substrate specificity.

We also speculated that the binding proteins and enzymes encoded in the HpbD genome neighbourhoods constitute a catabolic pathway that degrades tHyp-B to α-ketoglutarate (Fig. 4a), with HpbD catalysing the first step in which tHyp-B is 2-epimerized to cHyp-B; that is, the *in vivo* activity of HpbD is Hyp-B 2-epimerase. Subsequently, HpbB1/HpbC1, the Rieske-type protein, catalyses the 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, catalyses the oxidation of cHyp to its imino acid; HpbG, a member of the dihydrodipicolinate synthase superfamily[24], catalyses the dehydration of the 4-OH group and 'hydrolysis' of the 5-amino group to α-ketoglutarate semialdehyde; and HpbF, an aldehyde dehydrogenase, catalyses the oxidation of α-ketoglutarate semialdehyde to α-ketoglutarate. This pathway would permit the 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*, *Pseudomonas putida* and *S. meliloti*[25,26]; however, the sequences of HpbE and HpbG are so divergent (less than 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 Fig. 6).

When tHyp-B is used as an osmoprotectant (sea water is about 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 at low salt concentrations, as expected if the genome encodes the proposed catabolic pathway (Methods). Moreover, tHyp-B alleviates growth inhibition at high salt concentration (0.5 M NaCl) in glucose medium, arguing that *P. denitrificans* uses tHyp-B as an osmoprotectant (Supplementary Figs 7–9). 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 used strain RPd4, which lacks the demethylases that convert the isomers of Hyp-B to the isomers of *N*-methyl Hyp and therefore cannot utilize tHyp-B or cHyp-B as a carbon source (Supplementary Table 8). Strain RPd4 grew almost as well on high-salt glucose medium in the presence of tHyp-B or cHyp-B as did 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 Fig. 8). The results establish that tHyp-B and cHyp-B function as osmoprotectants.

We identified the metabolites obtained from tHyp-B at low salt concentrations (Methods). In addition to Hyp-B (21 mM; Supplementary Table 10), *N*-methyl Hyp and Hyp (the carbon-2 epimers cannot be distinguished), the predicted downstream $\Delta^1$-pyrroline-4-hydroxy-2-carboxylate, α-ketoglutarate semialdehyde and α-ketoglutarate were observed (Supplementary Figs 10 and 11). 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 that 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 quantitative PCR with reverse transcription (qRT–PCR) to investigate expression of the genes encoding the catabolic pathway (Methods and Supplementary Table 6). *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 and enzymes involved in the transport of tHyp-B and its catabolism (Fig. 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 concentration were determined using equimolar concentrations of glucose and either tHyp-B or cHyp-B. Salt (0.5 M NaCl) enhanced the expression of the transporters (HpbN/HpbO/HpbH/HpbJ and HpbX/HpbY/HpbZ). In contrast, salt decreased the 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 6). Transport of tHyp-B/cHyp-B is required for uptake as osmolytes as well as carbon and 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 the insertion of antibiotic-resistance cassettes (Methods and Supplementary Table 7). The growth phenotypes are consistent with the predicted functions (Supplementary Discussion).

Here we have used homology modelling and metabolite docking to several proteins encoded by a gene cluster to guide the *in vitro* assignment of the previously undocumented 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 the *in vivo* functional assignments identified in this study can be extended to these proteins and organisms. Moreover, we expect that the Hyp-B 2-epimerase activity assigned to HpbD will be used to facilitate the discovery of the *in vitro* activities and *in vivo* functions of uncharacterized homologues in the enolase superfamily.

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

## METHODS SUMMARY

The metabolite library was docked into the unliganded structure of HpbD (2PMQ) with program Glide SP followed by rescoring with MM-GBSA. Homology models for HpbJ and HpbB1 were made by PLOP v. 25.0; the betaine library was docked into the model of HpbJ with Glide XP. Kinetic constants for HpbD were measured by quantifying the change in optical rotation. HpbD was expressed with a carboxy-terminal hexahistidine tag; crystals were grown by sitting-drop vapour diffusion and flash-cooled in liquid nitrogen. Data were collected at the Advanced Photon Source beamline 31-ID (Lilly-CAT); the structure was determined by molecular replacement using the unliganded structure (2PMQ).

Metabolomics analyses by liquid chromatography Fourier transform mass spectrometry were performed with tHyp-B either as a sole source of carbon or as both an osmoregulant and the sole source of carbon[28]. Gene expression profiles were studied with a Roche LightCycler 480. *P. denitrificans* gene disruption mutants were generated by conjugation[29] or electroporation[30].

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 22 January; accepted 15 August 2013.**
**Published online 22 September 2013.**

1. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLOS Comput. Biol.* **5,** e1000605 (2009).
2. Gerlt, J. A. *et al.* The Enzyme Function Initiative. *Biochemistry* **50,** 9950–9962 (2011).
3. Hermann, J. C. *et al.* Structure-based activity prediction for an enzyme of unknown function. *Nature* **448,** 775–779 (2007).
4. Song, L. *et al.* Prediction and assignment of function for a divergent *N*-succinyl amino acid racemase. *Nature Chem. Biol.* **3,** 486–491 (2007).
5. Kalyanaraman, C. & Jacobson, M. P. Studying enzyme-substrate specificity *in silico*: a case study of the *Escherichia coli* glycolysis pathway. *Biochemistry* **49,** 4003–4005 (2010).
6. Babbitt, P. C. *et al.* The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry* **35,** 16489–16501 (1996).
7. Gerlt, J. A., Babbitt, P. C. & Rayment, I. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys.* **433,** 59–70 (2005).
8. Gerlt, J. A., Babbitt, P. C., Jacobson, M. P. & Almo, S. C. Divergent evolution in enolase superfamily: strategies for assigning functions. *J. Biol. Chem.* **287,** 29–34 (2012).
9. Tanabe, M. & Kanehisa, M. Using the KEGG database resource. *Curr. Protocols Bioinformatics* **38,** 1.12.1–1.12.43 (2012).
10. Hermann, J. C. *et al.* Predicting substrates by docking high-energy intermediates to enzyme structures. *J. Am. Chem. Soc.* **128,** 15882–15891 (2006).
11. Schiefner, A. *et al.* Cation-π 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.* **279,** 5588–5596 (2004).
12. Goldmann, A. *et al.* Symbiotic plasmid genes essential to the catabolism of proline betaine, or stachydrine, are also required for efficient nodulation by *Rhizobium meliloti*. *FEMS Microbiol. Lett.* **115,** 305–311 (1994).
13. Burnet, M. W. *et al.* The stachydrine catabolism region in *Sinorhizobium meliloti* encodes a multi-enzyme complex similar to the xenobiotic degrading systems in other bacteria. *Gene* **244,** 151–161 (2000).
14. Wargo, M. J., Szwergold, B. S. & Hogan, D. A. Identification of two gene clusters and a transcriptional regulator required for *Pseudomonas aeruginosa* glycine betaine catabolism. *J. Bacteriol.* **190,** 2690–2699 (2008).
15. Daughtry, K. D. *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.* **134,** 2823–2834 (2012).
16. Larsen, P. I., Sydnes, L. K., Landfald, B. & Strom, A. R. Osmoregulation in *Escherichia coli* by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose. *Arch. Microbiol.* **147,** 1–7 (1987).
17. Hanson, A. D. *et al.* Osmoprotective compounds in the Plumbaginaceae: a natural experiment in metabolic engineering of stress tolerance. *Proc. Natl Acad. Sci. USA* **91,** 306–310 (1994).
18. Amin, U. S., Lash, T. D. & Wilkinson, B. J. Proline betaine is a highly effective osmoprotectant for *Staphylococcus aureus*. *Arch. Microbiol.* **163,** 138–142 (1995).
19. Bernard, T., Pocard, J.-A., Berrould, B. & Le Rudulier, D. Variations in the response of salt-stressed *Rhizobium* strains to betaines. *Arch. Microbiol.* **143,** 359–364 (1986).
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.* **188,** 6308–6317 (2006).
21. Burg, M. B., Kwon, E. D. & Kultz, D. Regulation of gene expression by hypertonicity. *Annu. Rev. Physiol.* **59,** 437–455 (1997).
22. Kempf, B. & Bremer, E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch. Microbiol.* **170,** 319–330 (1998).
23. Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50,** 4402–4410 (2011).
24. Babbitt, P. C. & Gerlt, J. A. Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272,** 30591–30594 (1997).
25. Watanabe, S. *et al.* Identification and characterization of D-hydroxyproline dehydrogenase and $\Delta^1$-pyrroline-4-hydroxy-2-carboxylate deaminase involved in novel L-hydroxyproline metabolism of bacteria: metabolic convergent evolution. *J. Biol. Chem.* **287,** 32674–32688 (2012).
26. White, C. E., Gavina, J. M., Morton, R., Britz-McKibbin, P. & Finan, T. M. Control of hydroxyproline catabolism in *Sinorhizobium meliloti*. *Mol. Microbiol.* **85,** 1133–1147 (2012).
27. Gloux, K. & Le Rudulier, D. Transport and catabolism of proline betaine in salt-stressed *Rhizobium meliloti*. *Arch. Microbiol.* **151,** 143–148 (1989).
28. Lenky, C. C., McEntyre, C. J. & Lever, M. Measurement of marine osmolytes in mammalian serum by liquid chromatography–tandem mass spectrometry. *Anal. Biochem.* **420,** 7–12 (2012).

29. Van Spanning, R. J. *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 $c_{550}$, $c_{551i}$, and $c_{553i}$. *J. Bacteriol.* **173,** 6962–6970 (1991).
30. Matsson, M., Ackrell, B. A., Cochran, B. & Hederstedt, L. Carboxin resistance in *Paracoccus denitrificans* conferred by a mutation in the membrane-anchor domain of succinate:quinone reductase. *Arch. Microbiol.* **170,** 27–37 (1998).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This research was supported by cooperative agreements from the US National Institutes of Health (U54GM093342, U54GM074945 and U54GM094662). Molecular graphics and analyses were performed with the University of California, San Francisco (UCSF) Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at UCSF (supported by National Institutes of Health P41-GM103311). Use of the Advanced Photon Source, an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the US DOE under contract no. DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access Team (LRL-CAT) beamline at Sector 31 of the Advanced Photon Source was provided by Eli Lilly Company, which operates the facility.

**Author Contributions** S.Z., R.K., A.S., M.W.V., B.M.W., S.B., J.B.B., B.S.H., R.D.S., P.C.B., S.C.A., J.V.S., J.A.G., J.E.C. and M.P.J. designed the research. S.Z., R.K., A.S., M.W.V., B.M.W., J.B.B., B.S.H. and R.D.S. performed the research. S.Z., R.K., A.S., M.W.V., B.M.W., S.B., J.B.B., B.S.H., R.D.S., P.C.B., S.C.A., J.V.S., J.A.G., J.E.C. and M.P.J. analysed data. S.Z., R.K., A.S., M.W.V., B.M.W., S.B., J.B.B., B.S.H., R.D.S., P.C.B., S.C.A., J.V.S., J.A.G., J.E.C. and M.P.J. wrote the paper.

**Author Information** The atomic coordinates and structure factors for APO Hyp-B 2-epimerase (HpbD) and tHyp-B-liganded HpbD are deposited in the Protein Data Bank under accession numbers 2PMQ and 4H2H, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.C.B. (babbitt@cgl.ucsf.edu), S.C.A. (steve.almo@einstein.yu.edu), J.V.S. (jsweedle@illinois.edu), J.A.G. (j-gerlt@illinois.edu), J.E.C. (j-cronan@life.uiuc.edu) or M.P.J. (matt.jacobson@ucsf.edu).

## METHODS

**Homology modelling 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 with 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 Fig. 6 as queries, and $10^{-100}$ as BLAST E-value cutoff. The Pythoscape v. 1.0 program[31] was used to make the two networks.

Homology modelling and docking. The models of HpbJ and HpbB1 were built with our in-house software Protein Local Optimization (PLOP, marketed as Prime by Schrödinger LLC). The template PDBs used for HpbJ and HpbB1 were 1R9L and 3N0Q, respectively. The sequence alignment of each pair of target and template was made by the L-INS-i method in MAFFT v. 6.925b (ref. 32). While constructing the models we included both the metal ions and the co-crystallized 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 (ref. 33) before 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[34] in Schrödinger Suite 2009 to convert each compound from two dimensions to three dimensions and to 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 masses greater than 400 Da because we did not expect these to 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 (that is, 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 dimethylsulphoniopropionate (DMSP), ectoine, 5-hydroxyectoine and trigonelline; the compounds are listed in Supplementary Table 2. 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 previously[35]. The Glide XP docking method[36] was used with HpbJ.

***In vitro* activity measurements.** Cloning, expression, and purification of the 2PMQ (HpbD). The protein sample was provided by the NYSGXRC structural genomics centre (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 centre.

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 *Nde*I and *Bgl*II and ligated to pET15b expression vector, yielding plasmid pRK9. The cloned HypF was expressed in *E. coli* BL21 (DE3) cells for protein purification. Luria–Bertani medium (4 l) was shaken at 20 °C and induced with 0.5 mM isopropyl β-D-thiogalactoside when the culture reached a $D_{600}$ of 0.6. The cells were harvested after 24 h by centrifugation. The cells were resuspended in 100 ml of 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 imidazole buffered with 0.5 M NaCl, 20 mM Tris-HCl pH 7.9, 0.1 mM DTT. The purest fractions were pooled and dialysed against 20 mM Tris-HCl pH 8.0, 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 by means of $^1$H NMR. The reaction mixture contained 10 mM compound, 50 mM Tris-DCl pD 8.0, 10 mM $MgCl_2$ and 1 μM enzyme and was incubated at 30 °C for 16 h before acquisition 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 quantifying the change in optical rotation. The assay was performed in a total volume of 0.8 ml in a cell with a path length of 100 mm, using 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 quantifying the change in optical rotation. The assay was performed in a total volume of 0.8 ml in a cell with a path length of 100 mm, using 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.

**Structure determination.** Expression of HpbD. Plasmid 9437a2BNt21p1, obtained from NYSGRXC stock clones[37], consists of a codon optimized HpbD gene in pSGX2, a derivative of pET26b (Novagen), with the amino-terminal methionine of HpdD changed to the sequence MAHHHHHHSL. The vector was transformed into Rosetta2 (DE3)pLysS competent cells (EMD Millipore) and plated on Luria–Bertani agar plates. Five to ten colonies were added to 75 ml of Luria–Bertani medium with 0.5% glucose and grown overnight at 37 °C. HpbD was expressed using 4 l of autoinduction medium at 25 °C (refs 38, 39). The starter culture and autoinduction medium were distributed equally among ten 2 l baffled flasks, and shaken at 300 r.p.m. for about 24 h to $D_{600} > 15$. All growth media contained 100 μg ml$^{-1}$ kanamycin and 50 μg ml$^{-1}$ chloramphenicol. Cells were pelleted and stored at −80 °C.

Purification of HpbD. All purification was performed at 4 °C. Cells were resuspended in 3× (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 was removed by centrifugation, and the supernatant was applied to a 10-ml metal-affinity column ($Ni^{2+}$ Sepharose High Performance; GE Healthcare) pre-equilibrated with buffer A. The column was washed with five column volumes of buffer A and subsequently eluted with two column volumes of the same buffer containing 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 more than 95% purity by SDS–PAGE analysis were pooled, concentrated by centrifugal ultrafiltration, snap-frozen in liquid nitrogen and stored at −80 °C.

Crystallization and structure solution of HpbD. Crystals were obtained by vapour 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$^{-1}$ 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) 2-methyl-2,4-pentanediol, 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 mm × 0.15 mm over a 1–2-week 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 subsequently 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 were processed with MOSFLM[40] and scaled with SCALA[41] in space group $P2_1$. The unliganded structure (APO) was determined by selenomethionine single-wavelength anomalous diffraction phasing by the NYSGXRC in 2007 from a carboxy-terminally hexahistidine-tagged protein (2PMQ; Supplementary Table 4), with one 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[42] within the refinement package PHENIX[43] located eight subunits, which could subsequently be assembled into the molecular octamer. Several rounds of manual rebuilding and ligand and water fitting within the molecular graphics program COOT[44] followed by refinement in PHENIX were performed to finalize the structure. Several iodine atoms (seven or eight per subunit), originating from the synthesis of the substrate, were modelled into difference density peaks with features suggestive of bound iodine. The geometry restraints for tHyp-B were produced with the PRODRG2 server[45]. The density was fitted 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 probably a mix of substrate and product; however, only tHyp-B was used in refinement. The final structure has 98.6% of its residues in favoured regions of the Ramachandran plot, and 0.0% in disallowed regions (4H2H; Supplementary Table 4). The liganded structure (4H2H) superimposes with the APO structure (subunit A on subunit A) with a root mean squared deviation of 0.25 Å over 366 aligned Cα atoms with no substantial changes to the structure on ligand binding.

**Microbiology.** Bacterial strains and growth conditions. *P. denitrificans* PD1222 wild-type and mutant strains were grown in minimal medium containing (in grams per litre) $K_2HPO_4$ 6.0, $KH_2PO_4$ 4.0, sodium molybdate 0.15, $MgSO_4.7H_2O$ 0.2,

$CaCl_2$ 0.04, $MnSO_4.2H_2O$ 0.001, $FeSO_4.7H_2O$ 1.1, 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/cHyp-B or methanol at the same concentration (20 mM). *E. coli* strain TOP10 (Invitrogen) was used for plasmid maintenance, propagation and cloning purposes. *E. coli* strain S17-1 was used for conjugation[46]. Strains used are listed in Supplementary Table 7. To study the role of t/cHyp-B in osmoprotection, cultures were grown in minimal medium with glucose, in the presence or absence of 500 mM NaCl. *E. coli* cultures were grown at 37 °C in Luria–Bertani medium. Antibiotics were used at the following concentrations (in μg ml$^{-1}$): kanamycin sulphate 50, chloramphenicol 35, 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 with a DNeasy Blood and Tissue Kit (Qiagen) or a Wizard Genomic DNA Purification Kit (Promega). Restriction enzymes, DNA polmerases and T4 DNA ligases were purchased from New England Biolabs, Fermentas, Invitrogen or Promega. Plasmids were prepared from *E. coli* TOP10 cells with a Plasmid Mini Kit (Qiagen).

Gene disruption plasmids. Most plasmids for the construction of gene disruptions were obtained by a standard protocol in which the appropriate chromosomal segements 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 *Eco*RV (pRK1), *Nru*I (pRK2), *Bmg*BI (pRK4) and ligated to a 900-base-pair (bp) fragment blunt-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 *Eco*RI 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 similarly by 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 *Hin*cII site. Plasmid pRK3 was obtained similarly by using primers P5 and P6 except that a 1,400-bp kanamycin resistance cassette was inserted into the *Bmg*BI site of the intermediate pGEM T Easy construct. For 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 *Sfo*I site of this plasmid. Plasmid pRK7 was obtained by the same manipulations with primers 13 and 14 except that the chloramphenicol cassette was inserted into the *Sfo*I 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 that had been digested with *Eco*RI 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, respectively.

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 *Nde*I and *Bgl*II *hypF* fragment of the intermediate plasmid to pET15b digested with *Nde*I and *Bam*HI, whereas pRK13 resulted from ligation of the *hypO Nde*I–*Bpl*I 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 medium with 20 mM glucose as carbon source to a $D_{600}$ of 0.4. The cells were pelleted by centrifugation (4,000*g* for 10 min at 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 before cell harvesting. For studying the effect of salt stress on gene expression, NaCl was added to the minimal medium to 500 mM; glucose or succinate was the carbon source, 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 ($D_{600}$ 0.5–0.6) was added to two volumes of RNAprotect Bacteria Reagent (Qiagen). After vortex-mixing for 10 s the solution was incubated for 5 min at 22 °C. The cells were pelleted by centrifugation (10,000*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 analysed by agarose gel electrophoresis, and spectrophotometrically in the Nanodrop (Thermo) using the ratios $A_{260}/A_{280}$ and $A_{260}/A_{230}$; the $A_{350}/A_{220}$ absorption spectra were used 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 was performed on 1 μg of total RNA by using the RevertAid H Minus First Strand complementary DNA synthesis kit (Fermentas) in accordance with 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 the Universal probe library system (Roche Applied Science). The primer length was 21–27 nucleotides, with a theoretical $T_m$ of 58–60 °C. The amplicon size ranged from 66 to 110 bp. Real-time PCR was performed in 96-well plates with a Roche LightCycler 480. The 200-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 2× 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. The efficiency of all the primers used in qRT–PCR was calculated as 97 ± 2%. The gene expression data were expressed as crossing point (CP) values. The 16S rRNA was used as a reference gene. The data was analysed by the $2^{-\Delta\Delta C_T}$ (Livak) method[47]. Data presented are the average of five biological replicates. Primer sequences are provided in Supplementary Table 9.

**Metabolomics.** Metabolomics analysis of the *P. denitrificans* tHPB degradation pathway. Liquid chromatography–Fourier transform mass spectrometry metabolomics of whole-cell extracts were performed with samples of *P. 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 of minimal medium with tHyp-B as the sole carbon source and grown to a $D_{600}$ of about 0.7 (about 18 h). The culture was harvested by centrifugation (4,000*g* for 10 min at 4 °C), washed, and resuspended in minimal medium without carbon source. The cell suspension was then depleted of catabolic metabolites by incubation for 15 min at 30 °C before being transferred to ice. After verification of the concentrated cell density by multiple 50-fold dilutions into minimal medium (calculated $D_{600}$ 26.7 ± 0.2), 1-ml aliquots of cell suspensions with a $D_{600}$ of 6 were prepared in Eppendorf tubes on ice in minimal medium. tHyp-B (20 mM) was added quickly to half of the samples before incubation in a 30 °C water bath; at time points of 0, 2, 5 and 15 min, samples were transferred on ice to a centrifuge at 4 °C for pelleting at 16,000*g* for 2 min. 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 about 3 min. 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 a $D_{600}$ of about 0.5, each was concentrated by centrifugation, aliquoted into 1 ml of cells with a $D_{600}$ of 6, pelleted by centrifugation, separated from the supernatant, and frozen for analysis.

Metabolomics analysis of these samples followed the procedure from ref. 48. 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 min of vortex-mixing at 22 °C. The extraction was cleared of cell debris by two rounds of centrifugation at 16,000*g* before analysis. Samples were applied to a custom 11-T liquid-trap quadrupole Fourier transform mass spectrometer (Thermo-Fisher Scientific) with an Agilent 1200 high-performance liquid chromatography (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 in 100 μl for three separate chromatographic runs. The samples were eluted at a flow rate of 200 μl min$^{-1}$ with the following elution profile: 100% B for 17 min, a linear gradient from 100% B to 40% B over 3 min, and another linear gradient from 40% to 100% B over 15 min. Data were collected at a resolution of 50,000 with full scan set to $m/z$ 100–1,000, and duplicate samples were analysed individually in either positive or negative mode.

Data analysis was performed with the Qualbrowser application of Xcalibur (Thermo-Fisher Scientific) (Supplementary Figs 10 and 11). For metabolites, tHyp-B, *N*-methyl cHyp, cHyp and α-ketoglutarate, standards were run to verify retention time. Unfortunately, the *P. denitrificans* samples seemed to damage the Zic-HILIC column over time, because after dozens of runs many of the peaks broadened and showed longer retention times. Although HPLC analyses of betaine derivatives of biological origin have been reported[28,49–51], the metabolic analysis of bacteria with the use of betaines has not. It is assumed that the extremely large concentrations of tHyp-B accumulated in *P. denitrificans* cells were to blame as a result of overloading the column.

31. Barber, A. E. & Babbitt, P. C. Pythoscape: a framework for generation of large protein similarity networks. *Bioinformatics* **28,** 2845–2846 (2012).
32. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
33. Suite, S. 2009 *Protein Preparation Wizard; Epik version 2.0; Impact version 5.5; Prime version 2.1* (Schrödinger LLC, 2009).
34. Suite, S. 2009 *LigPrep, version 2.3* (Schrödinger LLC, 2009).
35. Kalyanaraman, C., Bernacki, K. & Jacobson, M. P. Virtual screening against highly charged active sites: identifying substrates of α–β barrel enzymes. *Biochemistry* **44,** 2059–2071 (2005).
36. Friesner, R. A. *et al.* Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **49,** 6177–6196 10.1021/jm051256o (2006).
37. Sauder, M. J. *et al.* High throughput protein production and crystallization at NYSGXRC. *Methods Mol. Biol.* **426,** 561–575 (2008).
38. Fox, B. G. & Blommel, P. G. Autoinduction of protein expression. *Curr. Protocols Protein Sci.* 5.23.1–5.23.18 (2009).
39. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
40. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 48–57 (2006).
41. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 72–82 (2006).
42. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
43. Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426,** 419–435 (2008).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **D60,** 2126–2132 (2004).
45. Schuttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 1355–1363 (2004).
46. Simon, R., Priefer, U. & Puhler, A. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Nature Biotechnol.* **1,** 784–791 (1983).
47. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25,** 402–408 (2001).
48. Erb, T. J. *et al.* A RubisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nature Chem. Biol.* **8,** 926–932 (2012).
49. Chambers, S. T. & Kunin, C. M. 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.* **79,** 731–737 (1987).
50. 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* **1216,** 1094–1099 (2009).
51. Li, C., Hill, R. W. & Jones, A. D. 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.* **878,** 1809–1816 (2010).