Supplementary information for:

Insights into Thiotemplated Pyrrole Biosynthesis Gained from the Crystal Structure of Flavin-Dependent Oxidase in Complex with Carrier Protein

Hem R. Thapa[1], John M. Robbins[2,3], Bradley S. Moore[4,5], Vinayak Agarwal[1,6,*]

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta GA 30332

[2]School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332

[3]Krone Engineered Biosystems Building, Georgia Institute of Technology, Atlanta, GA 30332

[4]Center for Oceans and Human Health, Scripps Institution of Oceanography, University of California, San Diego, La Jolla CA 92093

[5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla CA 92093

[6]School of Biological Sciences, Georgia Institute of Technology, Atlanta GA 30332

*Correspondence: vagarwal@gatech.edu

Supplementary methods

Co-expression of *bmp1–4* for production of tetrabromopyrrole and construction of mutant strains

The construction of plasmid vectors for the coexpression of *bmp1–4* together with the *Marinomonas mediterranea* MMB-1 phosphopantetheinyl transferase has been described previously (1). Briefly, *bmp1* was inserted between BamHI and HindIII restriction sites of the RSFDuet-1 vector, followed by the phosphopantetheinyl transferase (GenBank accession number: WP_013660297) in the NdeI and XhoI sites of the same vector. The gene *bmp1* was inserted in the NcoI and HindIII sites for the vector CDFDuet-1 followed by *bmp3* in the NdeI and XhoI restriction sites of the same vector. The gene *bmp2* was introduced in the NdeI and XhoI restriction sites of the pETDuet-1 vector. The three vectors were co-transformed into *Escherichia coli* BL21Gold(DE3) competent cells and selected using three antibiotics (final concentrations: 50 μg/mL streptomycin, 50 μg/mL kanamycin and 100 μg/mL carbenicillin). Plasmids for mutant enzymes were generated by site directed mutagenesis using primers containing the desired mutations, verified by sequencing of plasmid DNA, and transformed similarly. All mutagenesis PCR were done using PrimerSTAR Max DNA Polymerase using following conditions: 98ºC initial denaturation for 2 min, 33 cycles of (denaturation at 98ºC for 30 sec, 55ºC annealing for 15 sec, 72ºC extension for 90 sec), and a final extension at 72ºC for 5 min.

A 1% v/v dilution of the overnight culture of BL21Gold(DE3) strains was used to inoculate LB medium supplemented with 1g/L KBr and antibiotics (50 μg/mL streptomycin, 50 μg/mL kanamycin and 200 μg/mL carbenicillin) and grown at 30ºC with vigorous shaking. After 2.5 h, cultures were induced with 5 μM isopropyl β-D-thiogalactopyranoside (IPTG). Five milliliters of cultures were harvested at indicated times, extracted twice with equal volumes of EtOAc and the organic layer was collected and removed *in vacuo*. The residue was dissolved in 500 μL of MeOH and 50 μL was injected on a Phenomenex C18 column (5 Å, 100 × 4.6 mm) operating at 0.7 mL/min flowrate on an Agilent Infinity 1260 liquid chromatography system coupled to a Bruker amaZon ion-trap mass spectrometer. HPLC solvents (A:water and B:MeCN) were supplemented with 0.1% (v/v) formic acid. The mass spectrometry data was collected from 5 min to 25 min in the negative ionization mode for the following liquid chromatography elution profile:

1. 0-5 min: 10% solvent B
2. 5-10 min: linear gradient to 70% solvent B
3. 10-20 min: linear gradient to 85% solvent B
4. 20-21 min: linear gradient to 100% solvent B
5. 21-23 min: 100% solvent B
6. 23-24 min: linear gradient to 10% solvent B
7. 24-26 min: 10% solvent B
8. 26-27 min: linear gradient to 100% solvent B
9. 27-30 min: 100% solvent B
10. 30-31 min: linear gradient to 10% solvent B
11. 31-33 min: 10% solvent B

Cloning, expression and purification of Sfp, Bmp3/Bmp1(CP) complex, and Bmp4

The construction of Sfp/pET24a and Bmp4/pET28b expression vectors have been described previously (2). The DNA sequence corresponding to N-terminal 77 residues encoding the CP domain of Bmp1 was cloned into MCS1 of CDFDuet-1 vector using BamHI and HindIII restriction sites. Bmp3 was then cloned into Bmp1(CP)/CDFDuet-1 vector using NdeI and XhoI restriction sites. The construct Bmp3/Bmp1(CP)/CDFDuet-1 encodes for Bmp3 without a tag and Bmp1(CP) with N-terminal His_6-tag. The *E.coli* BL21Gold(DE3) cells transformed with desired expression vector were used for protein expression. The cultures were grown in terrific broth medium supplemented with appropriate antibiotic at 37ºC until OD_{600} reached 0.6, and the protein expression was induced by adding 0.3 mM IPTG. For cultures expressing Bmp3/Bmp1(CP)/CDFDuet-1, cells were spiked with 10 mg of riboflavin after induction. The induced cultures were then grown for additional 18 h at 18ºC, cells harvested by centrifugation and stored at -80ºC for future use. Unless noted, all steps of protein purification were

conducted at 4ºC. Cell pellets were resuspended in binding buffer (20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 10% glycerol) and sonicated. The lysate was then centrifuged at 30,000×g for 40 min and supernatant loaded to a 5 mL His-Trap column using an AKTA Prime FPLC system. The column was washed extensively with washing buffer (20 mM Tris–HCl (pH 8.0), 1 M NaCl, 30 mM imidazole), and protein eluted with an increasing linear gradient of elution buffer (20 mM Tris–HCl (pH 8.0), 1 M NaCl, 250 mM imidazole). The eluant fractions containing desired protein were pooled followed by dialysis for 16 h against dialysis buffer (20 mM Tris–HCl (pH 8.9)). The dialyzed protein samples were loaded to a 5 mL Hi-Trap Q column, washed with two column volumes of buffer A (20 mM Tris–HCl (pH 8.9)) and protein eluted with an increasing linear gradient of buffer B (20 mM Tris–HCl (pH 8.9), 1 M KCl). Eluted protein fractions were concentrated using 3 kDa Amicon centrifugal filters, desalted into storage buffer (20 mM HEPES (pH 7.5), 50 mM KCl, 10% glycerol) using PD-10 gravity column, and stored in small aliquots at -80ºC.

Supplementary Tables

Table S1: Sequence similarity and RMSD for Bmp3 and Bmp1(CP) homologs in the PDB

Bmp3			**Bmp1(CP)**		
PDB	**Similarity**	**RMSD**	**PDB**	**Similarity**	**RMSD**
4IRN	43	0.92	2N5H	36	1.17
6AF6	41	1.11	5CZD	27	1.98
2VIG	32	1.21	2XZ1	13	1.99
1JQI	32	1.22	5KOF	23	2.07
4L1F	32	1.38	4IHG	23	2.18
2JIF	31	1.33	1F80	17	2.18
6FAH	30	1.40	5KP7	17	2.28
2CX9	32	1.37	2K0X	27	2.31
5OL2	33	1.37	2K93	23	2.32
2D29	32	1.39	3EBJ	21	2.34

Supplementary figures

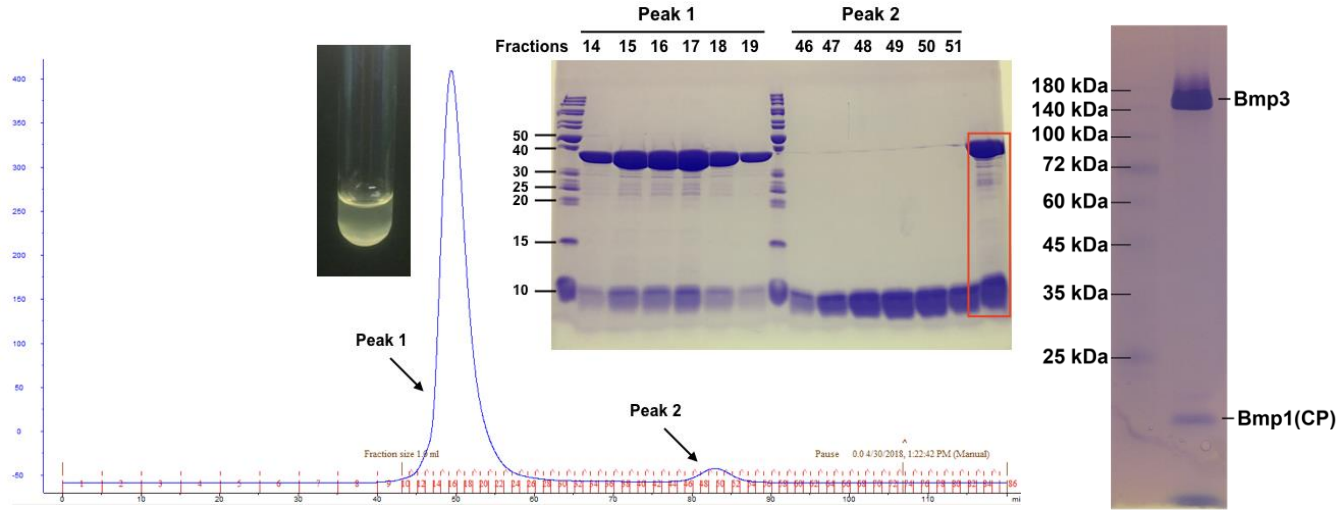

Figure S1. Size exclusion chromatography and PAGE for the Bmp3/Bmp1(CP) complex. (Left) Gel filtration through a S-75 gel filtration column leads to separation of the Bmp3 (Peak 1) from Bmp1(CP) (Peak 2), as visualized by denaturing SDS-PAGE. The red box denotes the complex that was loaded on to the gel filtration column, and used for crystallization. Eluted protein corresponding to Peak 1 precipitates, as shown for fraction 16. Bmp3 has a higher extinction coefficient leading to a higher peak than Bmp1(CP) on the elution chromatogram. Retention time for Peak 1 corresponds to ~150 kDa. (Right) Non-denaturing native PAGE for the Bmp3/Bmp1(CP) complex. Note that Bmp3 and Bmp1(CP) disassociate during native PAGE. Molecular weight of the Bmp3 monomer is 41.6 kDa, the observed bands indicate that the Bmp3/Bmp1(CP) complex disassociates into the Bmp3 tetramer and Bmp1(CP) monomers.

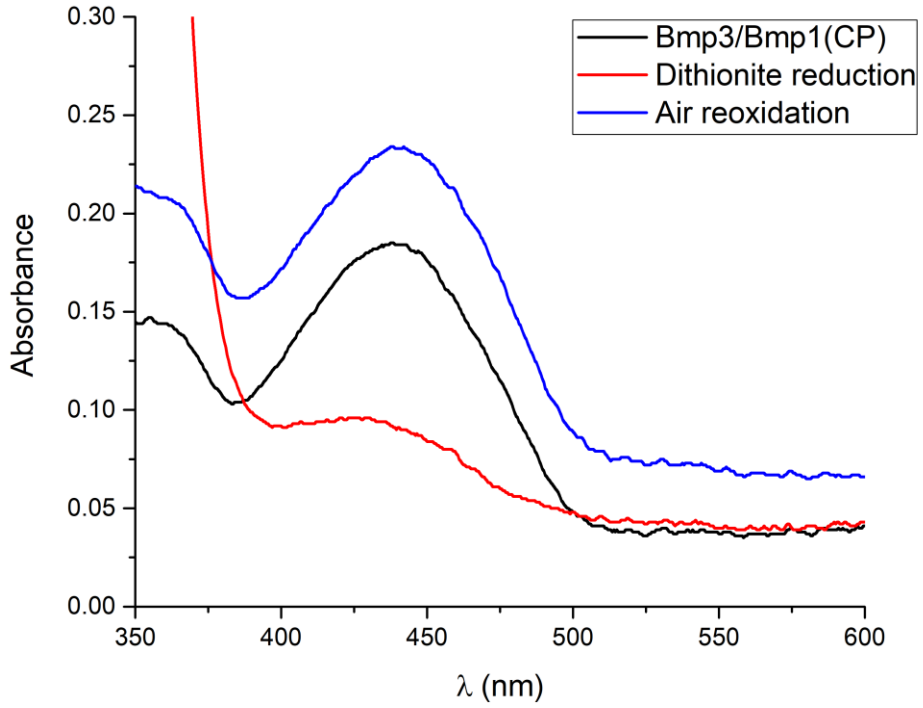

Figure S2. Bmp3 reduction by sodium dithionite and reoxidation by air. UV-Vis absorbance spectra of the Bmp3/Bmp1(CP) complex showing the change in oxidation state of FAD when treated with sodium dithionite followed by reoxidation by air.

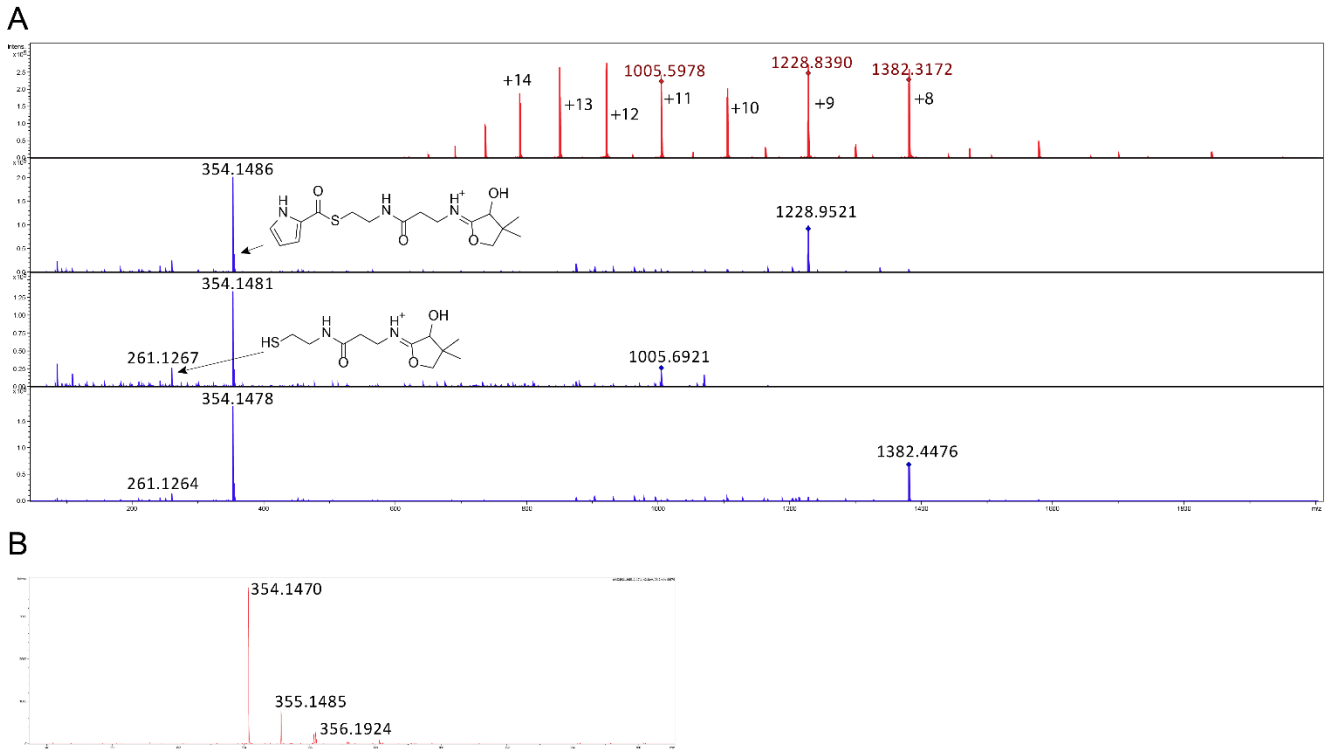

Figure S3. MS1 and MS2 spectra of the Bmp3/pyrrolyl-*S*-Bmp1(CP) complex. (**A**) Various different charged states for the pyrrolyl-*S*-Bmp1(CP) peptide were observed in the MS^1 spectra (top) which demonstrate the characteristic pyrrolyl-*S*-(cyclo)pantetheine and (cyclo)pantetheine MS^2 product ions (bottom). (**B**) Observed isotopic distribution of the pyrrolyl-*S*-(cyclo)pantetheine MS2 ion.

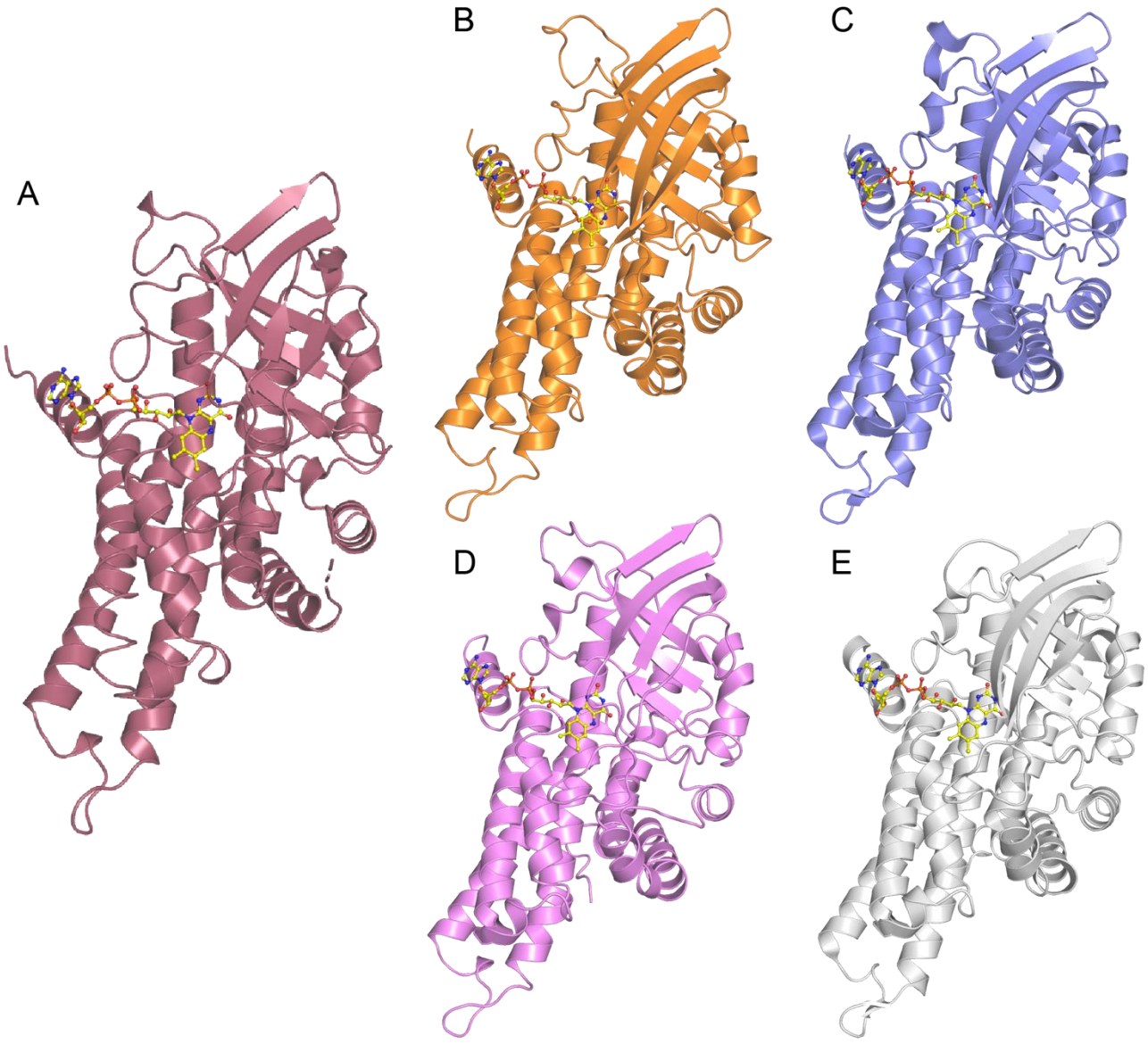

Figure S4. Structural similarity of Bmp3 to acyl-CoA dehydrogenases. (**A**) Tertiary structure of Bmp3 monomer. Structures of (**B**) medium chain (PDB:3MDE, RMSD 1.06 Å over 314 residues with side chains truncated at Cα) (3) and (**C**) short chain (PDB:1JQI, RMSD 0.98 Å) (4) fatty acid acyl-CoA dehydrogenases in complex with FAD. Structures of (**D**) isovaleryl-CoA (PDB:1IVH, RMSD 1.14 Å) (5) and (**E**) isobutyryl-CoA (PDB:1RX0, RMSD 1.11 Å) (6) dehydrogenases in complex with FAD. For simplicity, all structures are shown in complex with the cofactor FAD only (in stick-ball representation with carbon atoms colored yellow). Other examples of flavin-dependent oxidoreductases that are highly homologous to Bmp3 include KijD3 that is involved in the biosynthesis of the antibiotic kijanimicin, and FkbI that is involved in the biosynthesis of ascomycin (7, 8).

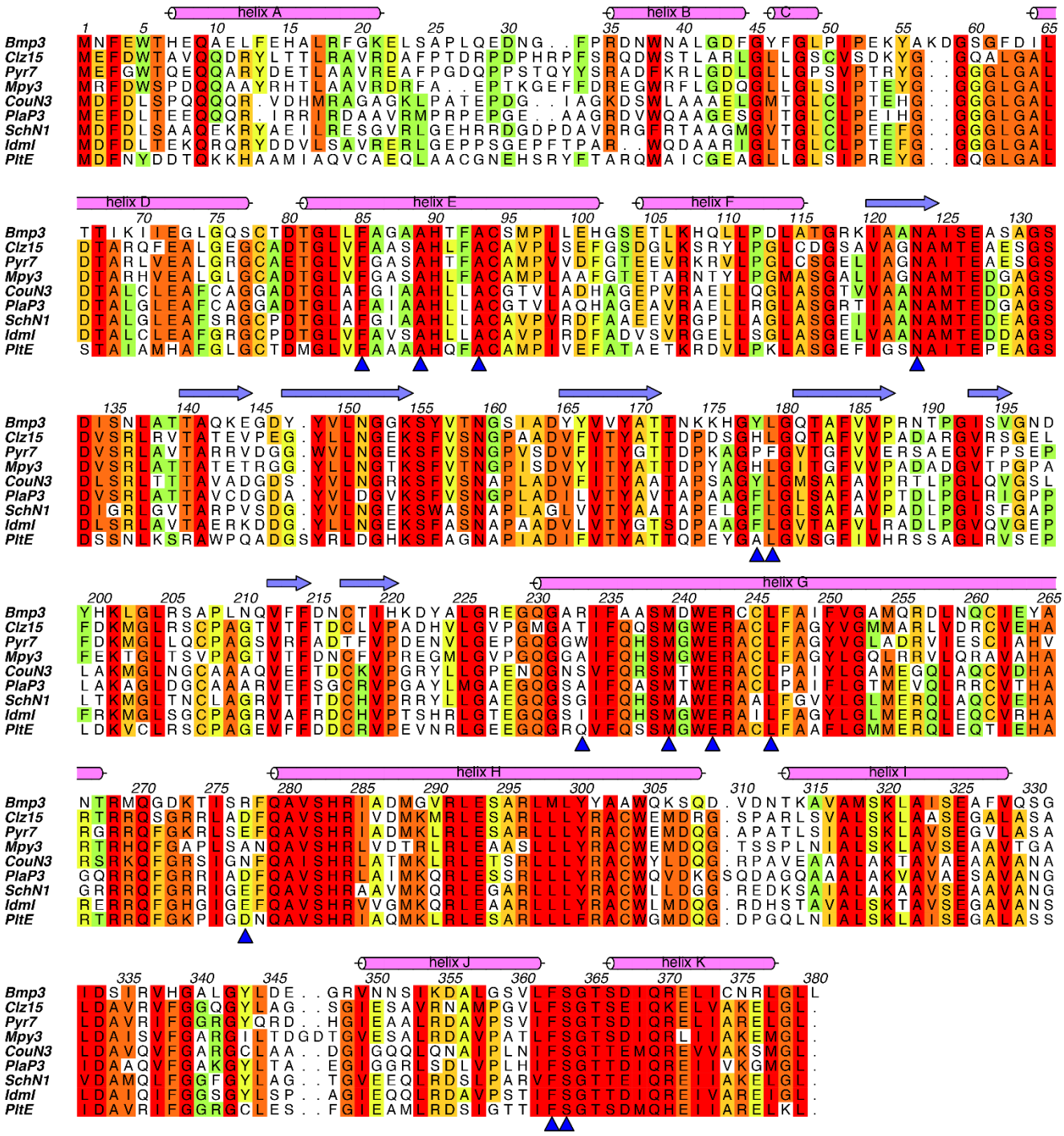

Figure S5. Sequence alignment of representative prolyl-*S*-dehydrogenases. Residue numbering and secondary structure elements are denoted for *Marinomonas mediterranea* MMB-1 Bmp3 (GenBank: WP_013663193). Residues discussed in text are highlighted by (▲). From top, the following sequences are included: *Streptomyces sp*. CNH-287 Clz15 (AHA12087.1), *Streptomyces vitaminophilus* Pyr7 *(*ABO15843.1), *Streptomyces sp*. CNQ-418

Mpy3 (AFP87520.1), *Streptomyces rishiriensis* CouN3 (AAG29788.2), *Streptomyces sp*. Tu6071 PlaP3 (ABB69752.1), *Streptomyces chartreusis* NRRL 3882 SchN1 (SOR83533.1), *Streptomyces antibioticus* IdmI (ACN69985.1), and *Pseudomonas protegens* PltE (ASE23711.1). Sequences were aligned using ClustalW and visualized using ALINE (9, 10).

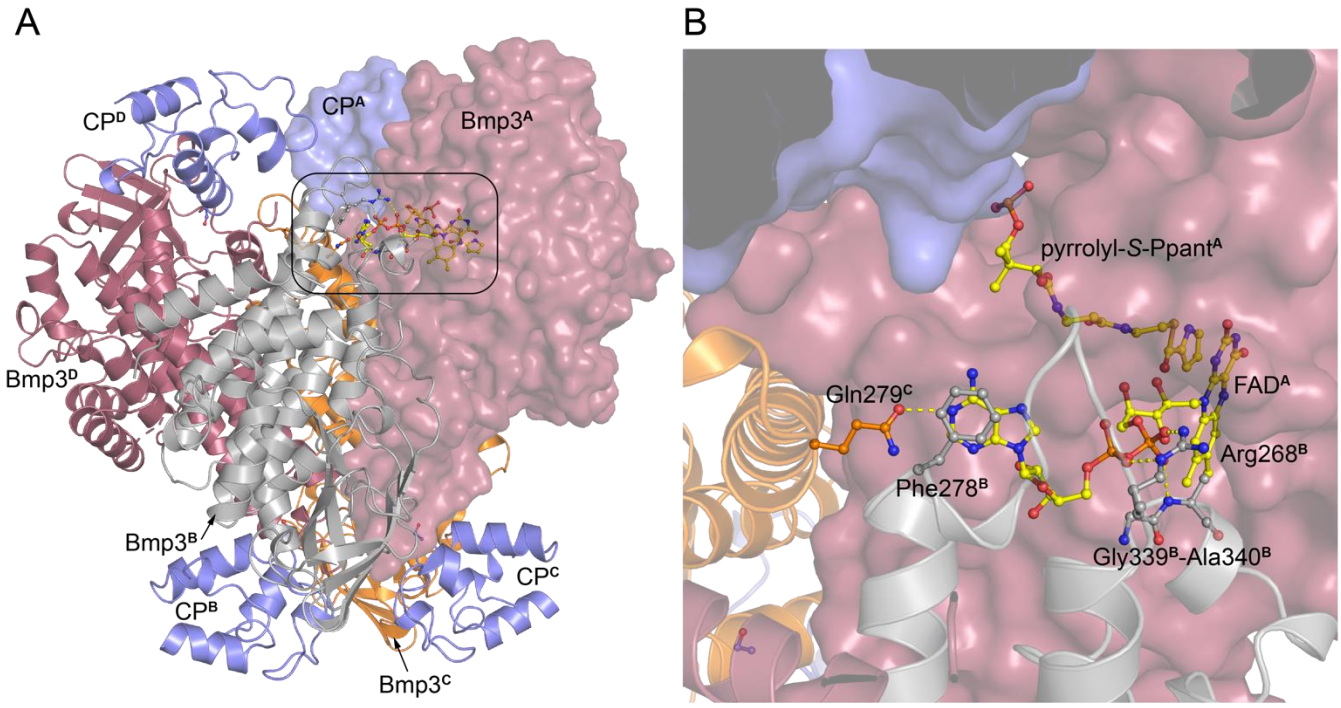

Figure S6. FAD binding at the interface of Bmp3 monomers in the biological tetramer. The four Bmp3/Bmp1(CP) pairs in the biologically relevant assembly are labeled A–D. (**A**) For clarity, FAD and pyrrolyl-*S*-Ppant for only $Bmp3^{A}$/$Bmp1(CP)^{A}$ unit are shown in stick-ball representation with carbon atoms colored yellow. The corresponding $Bmp3^{A}$/$Bmp1(CP)^{A}$ peptide chains are shown in surface representation. The other three $Bmp3^{B\text{-}D}$/$Bmp1(CP)^{B\text{-}D}$ subunits are shown in cartoon representation. The region for FAD binding, corresponding to the $Bmp3^{A}$, is boxed. (**B**) Zoomed-in view of the boxed region in panel A. The adenine ring for the FAD^{A} cofactor is stacked against the Phe278 side chain contributed by $Bmp3^{B}$. The adenine ring is also H-bonded to the Gln279 side chain amide contributed by $Bmp3^{C}$. The FAD^{A} pyrophosphate is bonded the $Arg268^{B}$ side chain guanidino side chain as well as to the $Gly339^{B}$-$Ala340^{B}$ main chain amide. Note that the FAD^{A} isoalloxazine ring is buried within $Bmp3^{A}$ and does not make contact with $Bmp3^{B}$.

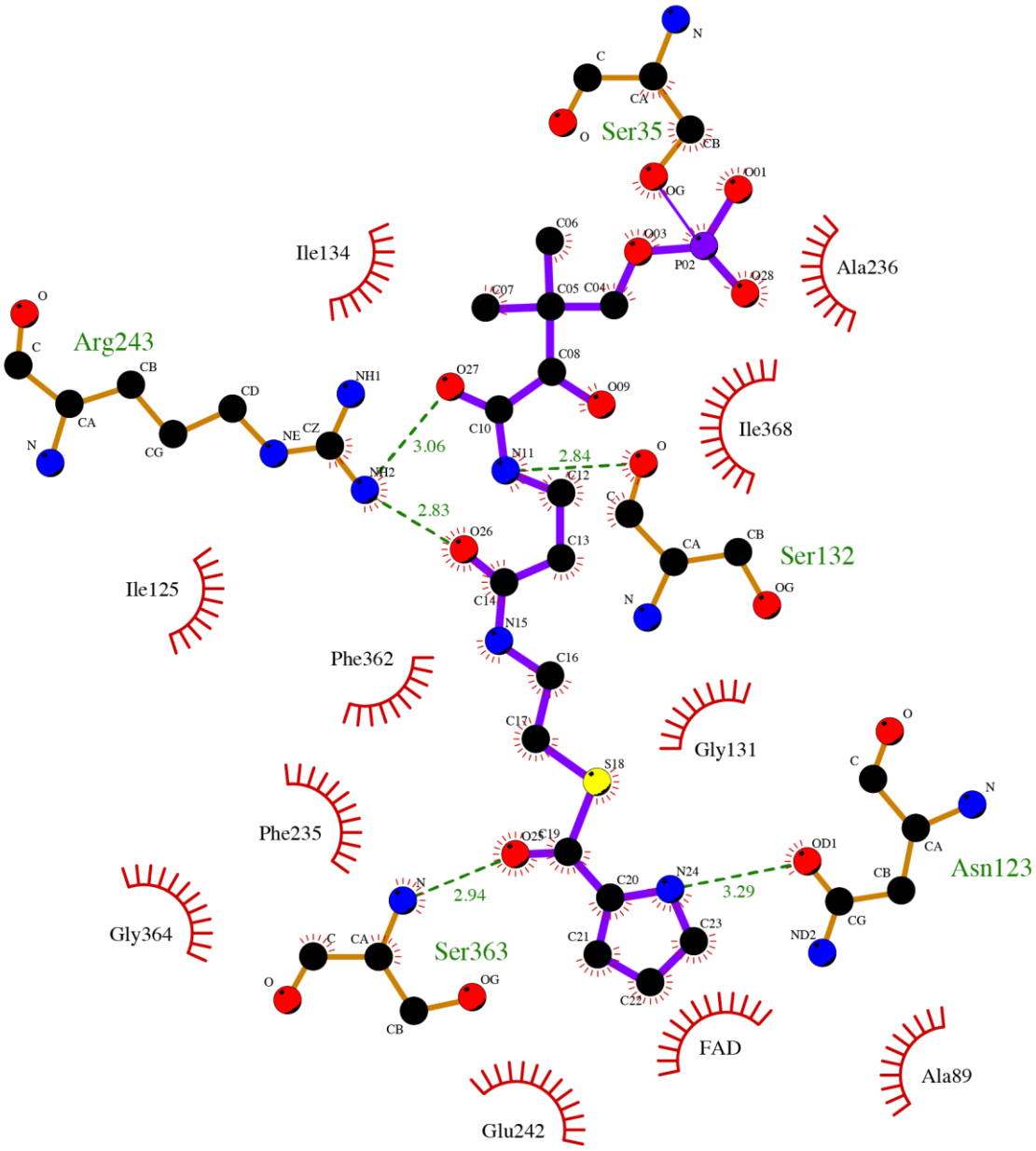

Figure S7. Interactions for pyrrolyl-*S*-phosphopantetheine. The two-dimensional rendering for the interactions of the acylated pyrrolyl-*S*-phosphopantetheine moiety was generated by LigPlot^{+} (11).

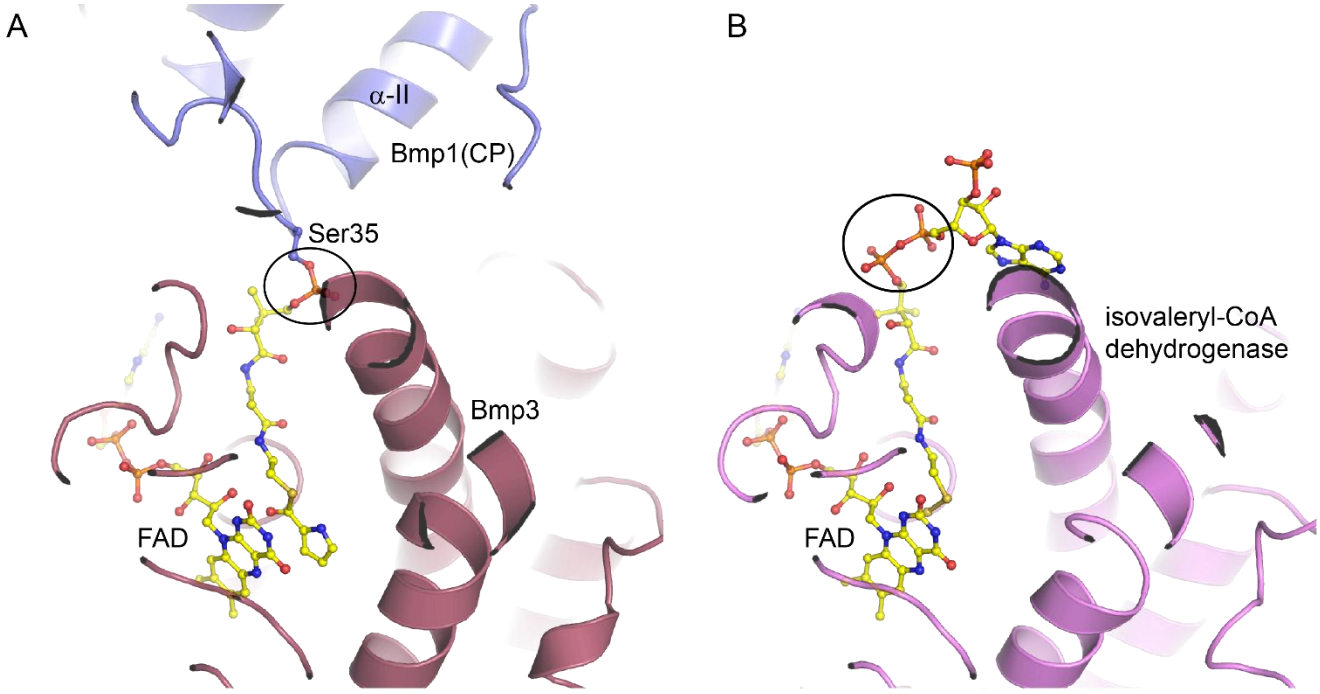

Figure S8. Comparison of the phosphopantetheine binding site for Bmp3 and isovaleryl-CoA dehydrogenase. Though the pantetheine chain bound in a similar fashion, note the difference in the placement of the phosphoryl moiety for (**A)** Bmp3 and the pyrophosphoryl moiety for (**B**) IVD (5). When superimposed, note that the adenine and the phosphorylated ribose for IVD would sterically clash with Bmp1(CP) helix-II.

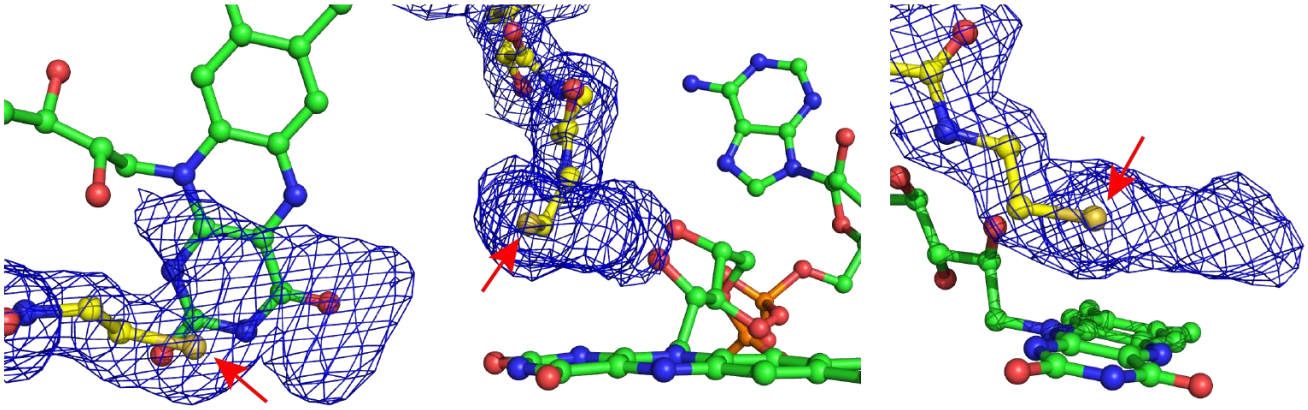

Figure S9. Electron density for the pyrrole ring. Three views of the difference Fourier electron density map contoured at 3.0σ in blue calculated with coefficients $|F_{obs}| - |F_{calc}|$ and phases from the final refined model with the coordinates of pyrrolyl-*S*-phosphopantetheine deleted prior to one round of refinement. For illustration purposes, atoms corresponding to the pyrrolyl moiety are manually omitted in the figure to illustrate the continuity of the omit density map from the phosphopantetheine chain to the pyrrole ring. Carbon atoms for the FAD cofactor are colored green. The phosphopantetheinyl thiol is marked by a red arrow.

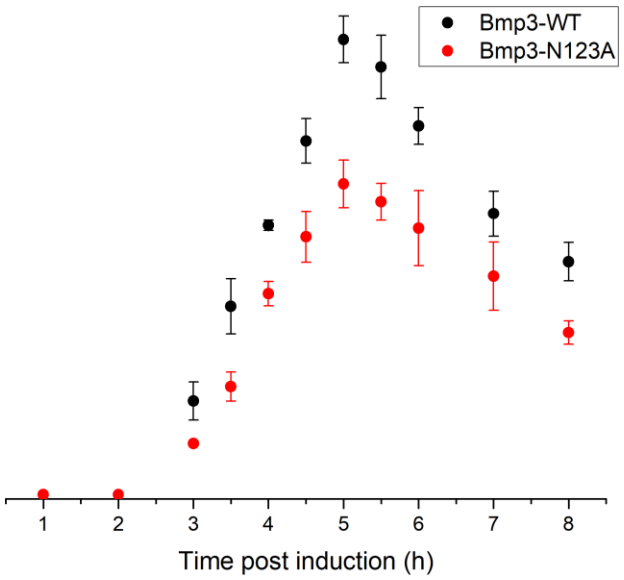

Figure S10. Time-dependent production of 2 by coexpression of *bmp1–4* in *E. coli*. Area under the extracted ion chromatogram for m/z = 381.75 ± 0.2 Da, the most abundant $[M-H]^{1-}$ isotopic mass corresponding to **2** is plotted. Mean and standard deviation from three independent measurements are reported. A continuous increase in production level of **2** is observed for 5 h post induction of protein expression.

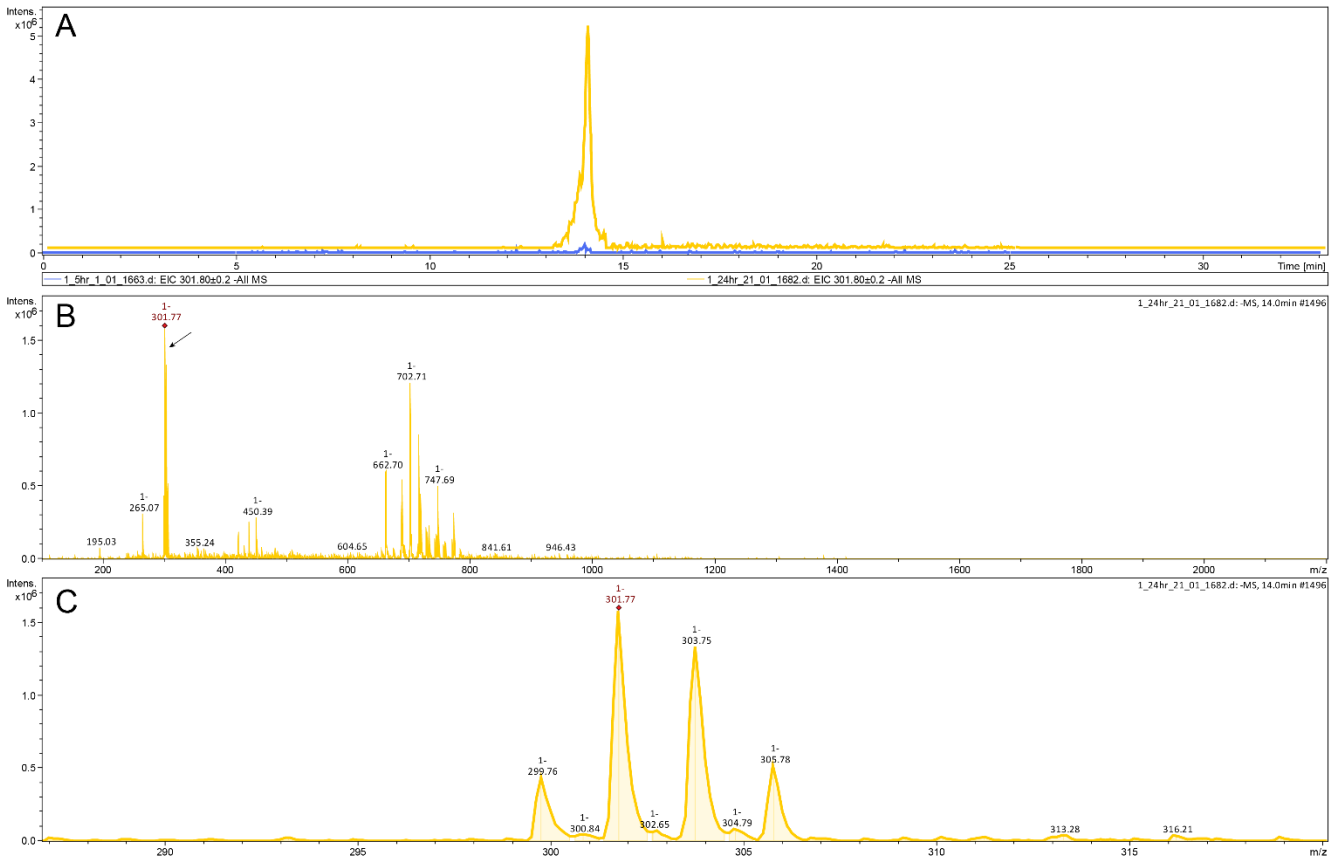

Figure S11. Time-dependent dehalogenative degradation of 2 in *E. coli*. (**A**) Extracted ion chromatograms (EICs) for m/z = 301.8, the most abundant $[M-H]^{1-}$ isotopic mass corresponding to tribromopyrrole (generated after loss of one bromine atom from **2**) are plotted for extracts generated 5 h (in blue) and 24 h (in yellow) post induction of protein expression for *bmp1–4* coexpression in *E. coli*. (**B**) MS1 spectra corresponding to the peak for the 24 h EIC shown in panel A with arrow pointing at the cluster of ions corresponding to tribromopyrrole. (**C**) Zoomed in view of the MS1 spectra showing the characteristic 3Br isotopic distribution for tribromopyrrole.

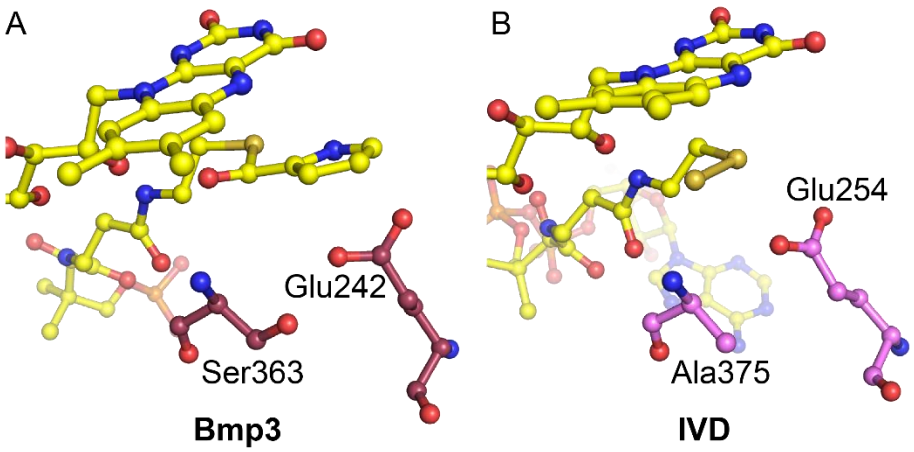

Figure S12. Comparison of actives sites of Bmp3 and isovaleryl-CoA dehydrogenase (IVD). The cofactor and bound ligands are shown in stick-ball representation with carbon atoms colored yellow. Note that Ser363 in Bmp3 (panel A) is replaced with Ala375 in the IVD active site (panel B).

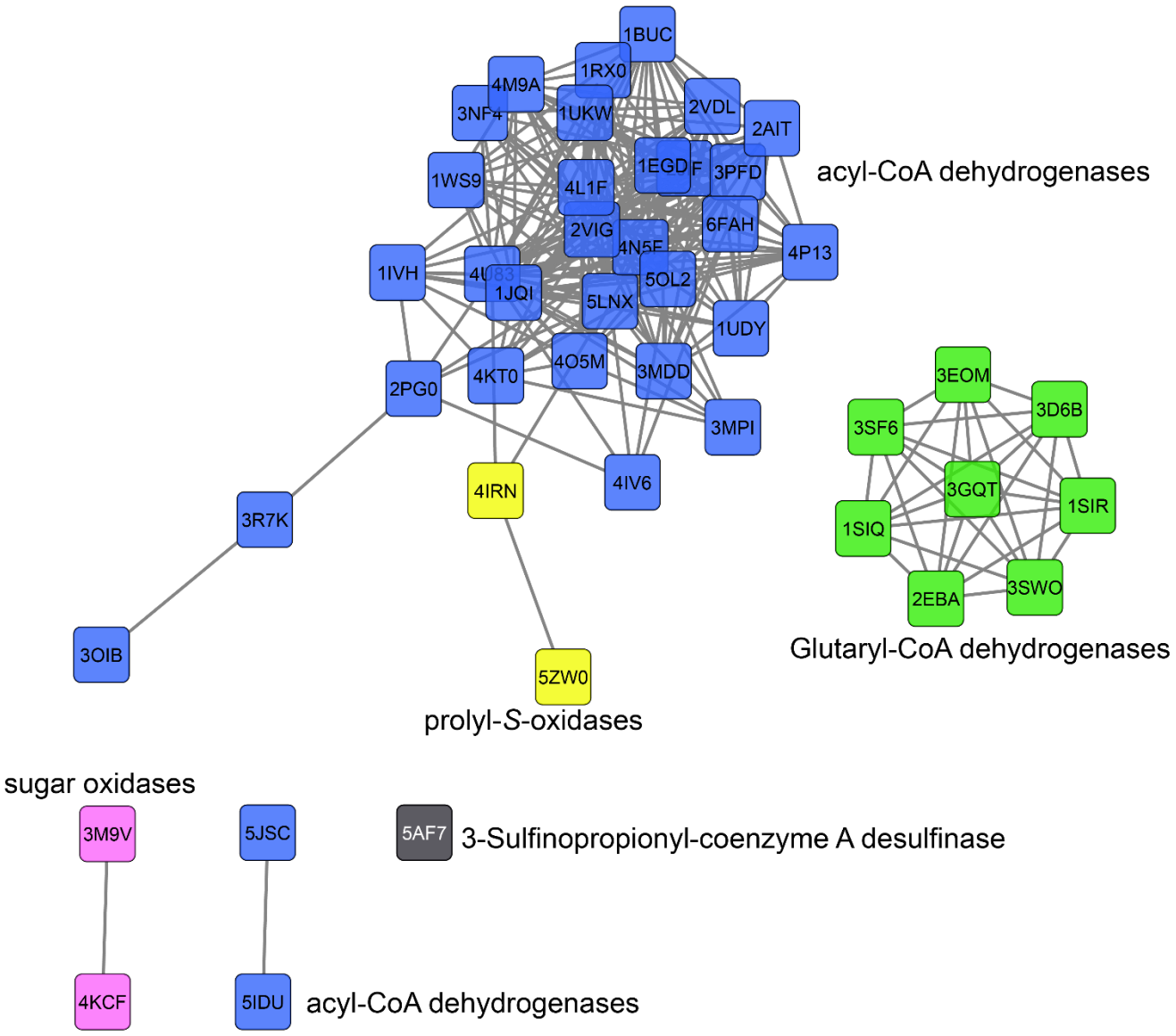

Figure S13. Network analysis of structural homologs of Bmp3. Nodes are labeled with the respective PDB ids. Top fifty-two primary sequences for Bmp3 homologs in the PDB database, as identified by BLAST, were provided as input to the EFI-EST analyses with E-value threshold of 10^{-5} (12). An alignment threshold of 70 was used to condense the network to 48 nodes that were illustrated as a network using Cytoscape. The prolyl-*S*-oxidases cluster closest to acyl-CoA dehydrogenases. Surprisingly, all glutaryl-CoA dehydrogenases, human and bacterial, cluster separately in a network distinct from acyl-CoA dehydrogenases.

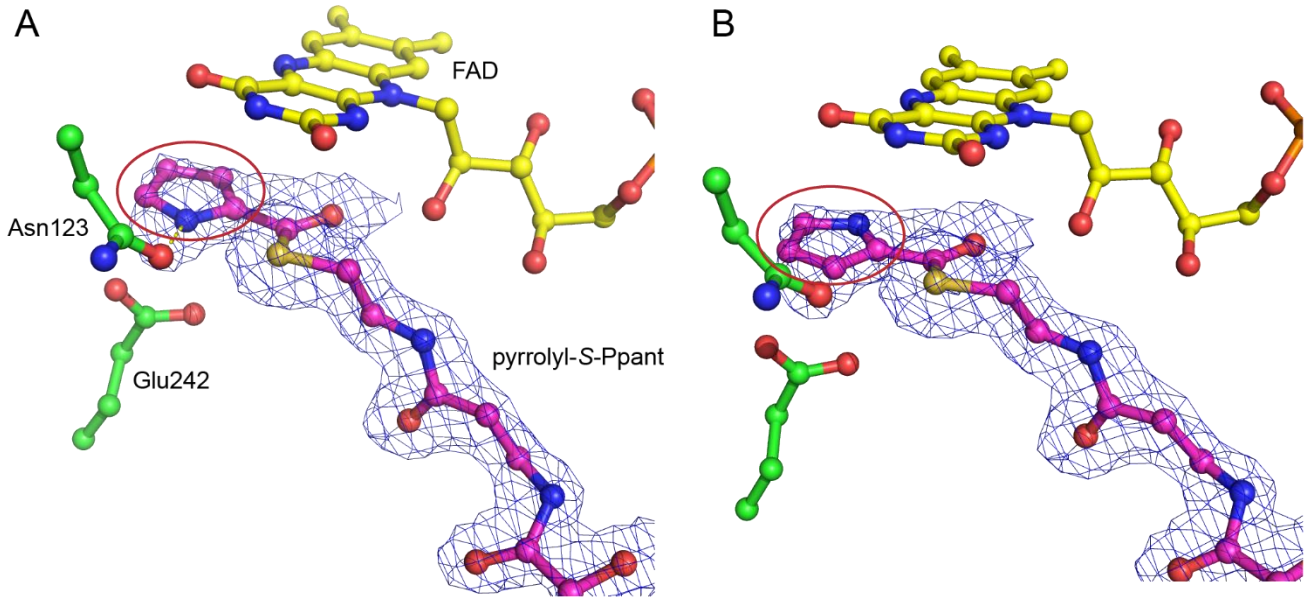

Figure S14. Alternate arrangements for the pyrrole ring. Based purely on the observed difference electron density (Fourier electron density map contoured at 3.0σ calculated with coefficients $|F_{obs}| - |F_{calc}|$ and phases from the final refined model with the coordinates of pyrrolyl-*S*-phosphopantetheine deleted prior to one round of refinement is shown in blue mesh), two alternate arrangements of the pyrrole ring are possible. The pyrrole ring is highlighted by the red oval. For the arrangement as shown in panel A, favored, the pyrrolyl nitrogen is within bonding distance with the Asn123 side chain amide whereas the pyrrolyl nitrogen in panel B, disfavored, does not have any stabilizing interactions. Note that both conformations for pyrrolyl-*S*-phosphopantetheine (abbreviated as pyrrolyl-*S*-Ppant in figure) have all atoms in full occupancy and at 1.9 Å resolution and satisfy the observed electron density equally well. The average B-factor for the pyrrolyl-phosphopantetheine is highly similar between the two conformations shown in panel A and B, 25.3 $Å^2$ and 27.4 $Å^2$, respectively.

Supplementary references

1. El Gamal A, *et al.* (2016). Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair. *Proc. Natl. Acad. Sci. U. S. A.* 113(14):3797-3802.
2. Agarwal V, *et al.* (2014). Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol* 10(8):640-647.
3. Kim JJ, Wang M, & Paschke R (1993). Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate. *Proc. Natl. Acad. Sci. U. S. A.* 90(16):7523-7527.
4. Battaile KP, *et al.* (2002). Crystal structure of rat short chain acyl-CoA dehydrogenase complexed with acetoacetyl-CoA: comparison with other acyl-CoA dehydrogenases. *J. Biol. Chem.* 277(14):12200-12207.
5. Tiffany KA, *et al.* (1997). Structure of human isovaleryl-CoA dehydrogenase at 2.6 A resolution: structural basis for substrate specificity. *Biochemistry* 36(28):8455-8464.
6. Battaile KP, Nguyen TV, Vockley J, & Kim JJ (2004). Structures of isobutyryl-CoA dehydrogenase and enzyme-product complex: comparison with isovaleryl- and short-chain acyl-CoA dehydrogenases. *J. Biol. Chem.* 279(16):16526-16534.
7. Watanabe K, Khosla C, Stroud RM, & Tsai SC (2003). Crystal structure of an Acyl-ACP dehydrogenase from the FK520 polyketide biosynthetic pathway: insights into extender unit biosynthesis. *J. Mol. Biol.* 334(3):435-444.
8. Thoden JB, Branch MC, Zimmer AL, Bruender NA, & Holden HM (2013). Active site architecture of a sugar N-oxygenase. *Biochemistry* 52(19):3191-3193.
9. Bond CS & Schuttelkopf AW (2009). ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D* 65:510-512.
10. Larkin MA, *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947-2948.
11. Laskowski RA & Swindells MB (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* 51(10):2778-2786.
12. Gerlt JA, *et al.* (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* 1854(8):1019-1037.