Resolving the hydride transfer pathway in oxidative conversion of proline to pyrrole

Atanu Acharya,[†] Dongqi Yi,[‡] Anna Pavlova,[†] Vinayak Agarwal,^{*,‡,¶} and James
C. Gumbart^{*,†,‡}

[†]*School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, United States*

[‡]*School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA
30332, United States*

[¶]*School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, United
States*

E-mail: vagarwal@gatech.edu; gumbart@physics.gatech.edu

Abstract

Thiotemplated pyrrole is a prevailing intermediate in the synthesis of numerous natural products where the pyrrole is tethered to a carrier protein (CP). Biosynthesis of the pyrrole requires oxidation of an L-proline side chain. Herein, we investigated the biocatalytic mechanism of proline-to-pyrrole synthesis using the recently reported (Thapa et al., *Biochemistry*, 2019, 58, 918) structure of a Type II non-ribosomal protein synthetase (NRPS) Bmp3-Bmp1 (Oxidase-CP) complex. The substrate (L-proline side chain) is attached to the Bmp1(CP) and the catalytic site is located inside the flavin-dependent oxidase (Bmp3). Interestingly, the FAD molecule is free (unbound) within the Bmp3 catalytic site. We show that the FAD isoalloxazine ring is stabilized in the catalytic site of Bmp3 by strong hydrogen bonding with Asn123, Ile125, Ser126, and Thr158. The stability of the dimeric Bmp3 system including one FAD molecule in our simulation suggests that the tetrameric Bmp3 assembly, as found in the crystal structure, is not functionally important. After the initial deprotonation followed by an enamine-imine tautomerization, oxidation of either the C2-C3 bond or the C2-N1 bond, through a hydride transfer (either from C3 or N1), is required for the pyrrole synthesis. Using molecular dynamics simulations, quantum mechanics/molecular mechanics simulations, and electronic structure calculations, we conclude that the hydride transfer is more likely to occur from C3 than N1. Additionally, we demonstrated the elasticity in the oxidase active site through enzymatic synthesis of proline derivatives.

Introduction

The pyrrole heterocycle is widespread in biological chemistry. In primary metabolism, pyrroles serve to bind metal ions in porphyrins. The metal binding capacity of pyrroles is exploited by bacterial secondary metabolites as well.¹ Pyrroles intercalate DNA and serve as a core synthon around which many pharmaceutical molecules are constructed.² Pyrroles in several secondary metabolite biosynthetic schemes are constructed via a type II non-ribosomal peptide synthetase (NRPS) adenylation-oxidation cascade which involves the $4e^-$ oxidation of the pyrrolidine side chain of L-proline (Fig. 1a).³ Prior to oxidation, the L-proline carboxylate is activated by adenylation, catalyzed by an adenytransferase, which then transfers L-proline to a carrier protein (CP) phosphopantetheinyl thiol.⁴⁻⁷ The phosphopantetheinyl thiol is derived from coenzyme A (CoA) and is appended post translationally to a serine side chain hydroxyl of the apo-CP to create the holo-CP.⁸ An oxidase then catalyzes the conversion of the pyrrolidine to pyrrole in a sequence of two $2e^-$ oxidations with electrons funneled to molecular oxygen via a flavin cofactor.^{9,10} The thiotemplated pyrrole thus developed can be delivered to type I polyketide synthases for chain elongation,¹¹ to type I NRPSs for ligation to other amino acids,¹² and to other modification enzymes such as halogenases¹³ or methyltransferases¹⁴ for further modification.

The biosynthesis of thiotemplated pyrroles using the abovementioned type II NRPS scheme relies on intramolecular protein-protein interactions, of the CP with the adenytransferase and of the CP with the oxidase, as well as catalysis in the respective adenytransferase and oxidase active sites. Biochemically validated atomistic detail for both sets of protein-protein interactions are now available.^{5,10} Catalysis in the adenytransferase active site follows the well accepted two-step process involving the ATP-mediated activation of the amino acid carboxylate followed by thioesterification of proline to the CP phosphopantetheinyl thiol. Conversion of this thiotemplated proline-derived pyrrolidine to pyrrole via a pyrroline intermediate⁹ in the oxidase active site remains less well resolved. The structure and overall mechanism of pyrroli(di)ne oxidases resembles the mitochondrial fatty

Figure 1: Chemical steps involved in pyrrole biosynthesis. (a) The adenylation-oxidation cascade illustrated as a chemdraw scheme formally showing the product of the adenylation. (b) An overall structural comparison of pyrrolyl-CP-Bmp3 (PDB: 6CXT) and acyl CoA-dehydrogenases (PDB: 3MDE). The flavin cofactor and the phosphopantetheinylated substrate/product are shown in ball-and-stick representation. (c) Key steps and plausible mechanisms in the Bmp3-catalyzed pyrrole biosynthesis. Catalytic state (CS) and product state are defined as well. Atom numbering schemes of pyrrolidine and isoalloxazine rings are shown in the CS-I state. The hydride-transfer step from C3 and from N1 are highlighted in orange and grey, respectively.

acid acyl-CoA dehydrogenases (Fig. 1b). Catalysis in fatty acid dehydrogenases proceeds by proton abstraction from the fatty acyl C_α by a catalytic glutamate residue followed by hydride transfer from the C_β to the flavin cofactor leading to the C_α-C_β oxidation.^{15,16} Pyrrolidine and pyrroline thioesters are unstable which has prevented a structural description of the substrate- or the intermediate-bound states of the pyrroli(di)ne oxidases.^{10,17,18} Furthermore, a lack of insight into substrate and intermediate binding in the oxidase active site has precluded deciphering the site of hydride transfer from the pyrroli(di)ne to the flavin

cofactor. Two mechanistic routes are conceivable, which converge to identical outcomes with interspersed imine-enamine tautomerizations (Fig. 1c).¹⁹ First is that the enzyme acts as an amine oxidase which would involve hydride transfer from ^{prolyl}N1 to ^{flavin}N5, as has been proposed previously.^{17,20} The second possibility is that the pyrroli(di)ne oxidase acts as a C $_{\alpha}$ -C $_{\beta}$ desaturase with the hydride being transferred from the ^{prolyl}C3 to ^{flavin}N5. At present, these two routes are indistinguishable.

Herein, we investigate the first two steps of pyrrole biosynthesis using molecular dynamics (MD), quantum mechanics/molecular mechanics (QM/MM) simulations, and electronic structure calculations. In particular, we assess the feasibility of the Glu242 residue to initiate the biosynthesis. More importantly, we provide mechanistic insights into the hydride transfer from the substrate to the FAD bound to the Bmp3 enzyme. We investigate the energy barriers for the unique hydride transfer process from the C3 atom of the substrate instead of from the N1 atom. Furthermore, by enzymatic synthesis of proline derivatives, we demonstrate elasticity in the oxidase active site, which supports the computational description of substrate binding.

Results and discussion

Model description and the substrate bound state

Prior genetic and enzymological description of a thiotemplated pyrrole biosynthetic pathway from marine bacteria had identified the L-proline-derived pyrrolidine/pyrrolyl carrier protein (CP) to be localized to the N-terminus of the Bmp1 polypeptide (henceforth referred to as Bmp1(CP)) and the pyrrolidine oxidase as the Bmp3 polypeptide.⁷ Subsequent crystallographic description of the Bmp1(CP)/Bmp3 complex identified a tetrameric Bmp3 core similar to that of acyl-CoA dehydrogenases.¹⁰ Each Bmp3 monomer binds one Bmp1(CP), thus organizing a 4:4 Bmp1(CP)/Bmp3 stoichiometric complex. Each Bmp1(CP) contacts two Bmp3 monomers in this tetrameric complex. One of these Bmp3 monomers is the one to

whose active site the phosphopantetheinylated molecular cargo is delivered to by Bmp1(CP). The second Bmp3 monomer only makes contact with the Bmp1(CP) without participating in catalysis. For computational analysis of catalysis occurring in the Bmp3 active site, our model system comprises a Bmp1(CP) peptide in complex with two Bmp3 peptides with the flavin cofactor preserved in the catalytically relevant Bmp3 monomer only. The model henceforth used for computation is illustrated in Figure 2a.

Structure of the pyrrolidine oxidase Bmp3 in complex with Bmp1(CP) was described in the product bound state.¹⁰ The substrate bound state, referred to as catalytic state-I (CS-I, Figure 1) was inaccessible. Other crystal structures of pyrrolidine oxidases were similarly not available with bound thiotemplated substrates.^{18,21} Hence, we first sought to recover the physiological substrate-bound CS-I state for Bmp3. To describe the CS-I state, we equilibrated the system where the thioesterified pyrrolidine heterocycle was modelled in the Bmp3 active site adjacent to the cofactor FAD isoalloxazine, followed by extended MD simulations. The initial position of the substrate in the CS-I state was set as found in the structure of the pyrrole in the product-bound state. We analyzed two replicas of 500-ns MD simulations of the CS-I model. For the occupancy analysis, we pooled data from both replicas.

In this CS-I model, we observed that the FAD isoalloxazine ring is stabilized by hydrogen bonding with Asn123, Ile125, Ser126, and Thr158 (Figures S2a-e). The percentage of all hydrogen bonding interactions between FAD, Bmp3, and water is provided in Table S1 and binding interactions of the FAD are illustrated in Figure S1. The pyrrolidine ring forms a hydrogen bond with the Asn123 side chain (Figure S2a), while the carbonyl group forms hydrogen bonds with the Ser363 backbone (Figure S2b,c). The substrate forms hydrogen bonds with Bmp3 predominantly through the Ppant moiety (Figure S2c-i), which attaches the pyrrolidine ring with the Bmp1(CP). In summary, the FAD isoalloxazine ring and the substrate form several hydrogen bonds with the Bmp3 protein to stabilize the catalytic site.

Similarity to acyl-CoA dehydrogenases and prior mutational analysis suggests that the

Figure 2: Distance and conformations of the substrate in the CS-I state. (a) Model Bmp1(CP)-Bmp3 system with one bound substrate (in blue) and FAD (in green). Active and inactive Bmp3 residues are shown in red and yellow, respectively. Substrate-bound Bmp1 is shown in a blue cartoon representation. The missing active site of chain C is shown as a green transparent sphere. (b) Definition of relevant distances. The center of the two oxygen atoms of Glu242 was used for the distance calculations. We defined $d1=d(\text{C2}-\text{Glu242})$ and $d2=d(\text{C2-H}-\text{Glu242})$, and $\Delta d = d1-d2$. (c) Representative structures of two conformations. C2-H atoms are shown as green spheres. (d) Distribution of the substrate conformations obtained from two independent MD simulations. Data from the two 500-ns simulations were pooled for this analysis.

Glu242 side chain carboxylate acts as the catalytic base for proton abstraction from the pyrrolidine side chain to initiate catalysis (Figure 1). We,¹⁰ and others¹⁸ had postulated that Glu242 abstracts the pyrrolidine C2-proton (pyrrolidine C2-H). Therefore, the distance between the Glu242 side chain carboxylate and pyrrolidine C2 and the pyrrolidine C2-H atoms is critical for catalysis. We calculated these distances along the MD trajectories (Figure S3).

We observed that the pyrrolidine C2 and pyrrolidine C2-H atoms stay close to the Glu242 side chain in MD simulations (Figure S3b). Since rotation is allowed around the pyrrolidine C2-

carbonyl carbon bond, the pyrrolidine ring can adopt two dominant conformations. To quantify the population distribution, we analyzed the differences between $\text{pyrrolidineC2-Glu242}$ and $\text{pyrrolidineC2-H-Glu242}$ distances as $\Delta d = d_1 - d_2$ (Figure 2b). One conformation brings the pyrrolidineC2-H proximal to the Glu242 carboxylate while the other conformation places the pyrrolidineC2-H distal to the Glu242 (Figure 2c). Since the two conformations can interconvert by rotation around a single bond, the rotational energy barrier should be small. This rotation does not change the $\text{pyrrolidineC2-Glu242}$ distance. If the C2-H atom is pointed towards Glu242, then $\Delta d > 0$. In contrast, $\Delta d < 0$ when the C2-H atom is pointed away from the Glu242 sidechain.

The combined data from two independent simulations show that the relative abundance of two rotamers are comparable (Figure 2d). The catalytically proficient conformation is represented by $\Delta d > 0$, since the deprotonated carboxylate of Glu242 side chain is geometrically close to the pyrrolidine C2-H atom. We observed that the catalytically proficient conformation was favored in the second simulation, while the catalytically deficient conformation was favored in the first simulation (Figure 2c,d).

We performed additional QM/MM simulations on the CS-I state to obtain more accurate dynamics of the catalytic site. Details of the QM/MM simulations are described in the Methods. From the QM/MM simulation, we observed that the most probable $\text{pyrrolidineC2-Glu242}$ distance is larger than the $\text{pyrrolidineC2-H-Glu242}$ distance, while both substrate atoms remain close to the Glu242 side chain (also see Figure S5a,b). Note that we do not observe the catalytically deficient $\Delta d < 0$ state from QM/MM simulation of the CS-I state, possibly because of the shorter time-scale (15 ps) of the QM/MM simulation. Within that time scale, QM/MM simulation only sampled conformations conducive to the proton transfer from the substrate to the Glu242 side chain (Figure S5a,b). While the substrate can bind in the active site in either conformation, catalysis can only proceed in one conformation and not in the other. Based on these results, we posit that the $\Delta d > 0$ state illustrated in Figure 2c is the catalytically proficient substrate bound state for pyrrolidine oxidase Bmp3. Mutations of

Glu242 in Bmp3 and the corresponding residue Glu244 in pyrrolidine oxidase AnaB render the enzymes inactive.^{10,17} Taken together, these results indicate the feasibility of the Glu242 protonation by proton transfer from the C2 atom of the substrate. More importantly, the active site is open enough for free rotation around the C-C2 bond allowing the pyrrolidine ring to fully rotate.

Hydride transfer in the CS-II intermediate

After thiotemplated substrate binding and abstraction of the pyrrolidineC2-H by the Glu242 side chain carboxylate, the catalytic state-II (CS-II, Figure 1) is defined by the protonated Glu242 side chain (Figure 1c). Pyrrolidine oxidation can proceed by tautomerization followed by hydride transfer from the pyrrolidineC3 to the flavinN5, or via hydride transfer from pyrrolidineN1 to flavinN5 (Figure 1c). Both of these routes have been proposed in literature, but evidence supporting either of the two routes has been lacking as the enolate intermediate was not accessible experimentally. To address this knowledge gap, we performed six independent 20-25 ps QM/MM simulations for the CS-II intermediate state using the same QM/MM scheme adopted for the CS-I simulations. In the CS-I state, all carbon and nitrogen atoms of the pyrrolidine ring are sp^3 hybridized. In contrast, the pyrrolidineC2 and pyrrolidineC3 atoms (for C2-C3 oxidation) of the substrate are in sp^2 hybridization states (Figure 1). In the case of plausible C2-N1 oxidation, C2 and N1 will become sp^2 hybridized. Both mechanisms have been proposed before in the flavin-catalyzed biosynthesis processes.^{9,17,20} We investigated which mechanism is more likely to occur using a total of 145 ps of QM/MM simulations along with transition state calculations.

We analyzed the pyrrolidineC3-flavinN5 and pyrrolidineN1-flavinN5 distances to identify the closer hydride donor between pyrrolidineN1 and flavinC3 (Figure 3a). We observed that the most probable pyrrolidineC3-flavinN5 distance was shorter (~ 3.0 Å) than the pyrrolidineN1-flavinN5 (~ 4.2 Å) (Figure 3a). Additionally, we calculated the distance of the flavinN5 atom from the pyrrolidineC3-H and pyrrolidineN1-H as illustrated in Figure 3b. The most probable pyrrolidineC3-

H-flavinN5 distances were shorter than the pyrrolidineN1-H-flavinN5 (1.0–2.0 Å and 4.0–4.5 Å, respectively) as shown in Figure 3b. Therefore, pyrrolidineC3 and pyrrolidineC3-H atoms are closer to the FAD cofactor as compared to pyrrolidineN1 and pyrrolidineN1-H, respectively. These results allow us to posit that the enolate intermediate tautomerizes such that the hydride transfer to the flavin cofactor occurs from pyrrolidineC3 and not from pyrrolidineN1. This then sets up the CS-III state in which the pyrrolidineC2–C3 bond should be unsaturated (Figure 1).

In support of the aforementioned assertion, in one of the QM/MM simulations, we captured the hydride transfer event from the CS-II pyrrolidineC3 to the flavinN5 (Figures 3c,d). Therefore, the pyrrolidineC3-H-flavinN5 distance distribution has two peaks for one of the simulations as illustrated in Figure 3b. We further queried the structural perturbation accompanying the hydride transfer in this simulation. We observed that the pyrrolidineC2–C3 bond length reduced from 1.50 Å to 1.37 Å, while the carbonyl C-pyrrolidineC2 bond length increased from 1.42 Å to 1.50 Å with the hydride transfer from the C3 atom (Figure 3e).

We optimized a model substrate in the CS-II and the CS-III states in a low-dielectric solvent ($\epsilon = 6.2528$) using the conductor-like polarizable continuum model (CPCM) implicit solvent model to obtain ideal bond lengths for pyrrolidineC2–C3 and pyrrolidineC–C2. The optimization was performed at the CAM-B3LYP/6-31+G(d) level of theory. The optimized pyrrolidineC2–C3 and carbonyl C-pyrrolidineC2 bond lengths are 1.51 Å and 1.35 Å in the CS-II state, respectively, and 1.34 Å and 1.35 Å in the CS-III state, respectively (Figure S7). The changes in these bond lengths match the probable bond lengths obtained from the QM/MM simulation of the states before and after the hydride transfer event. In contrast, pyrrolidineC2–N1 and carbonyl C–O distances remain similar before and after the hydride transfer event. Therefore, we observed only one peak in the most probable pyrrolidineC2–N1 distance (Figure S6c), indicating that pyrrolidineN1 did not participate in this hydride transfer event. Taken together, these results indicate that pyrrolidineC2–C3 oxidation during the hydride transfer is more likely to occur compared to pyrrolidineC2–N1 oxidation.

Figure 3: Distance distributions of the substrate in the CS-II state calculated from six independent QM/MM simulations. (a) Relative populations of distances between key pyrrolidine atoms (N1 and C3) and ^{flavin}N5 atom. (b) Relative populations of distances of the ^{flavin}N5 atom from ^{pyrrolidine}C3-H and ^{pyrrolidine}N1-H. Representative structures of the system (c) before and (d) after the hydride transfer. The relevant hydrogen atom is shown as an orange sphere. (e) Analysis of key bond distances calculated along the QM/MM simulation in which C2-C3 oxidation is explicitly observed.

Experimentally querying active site flexibility

Data presented above points towards the Bmp3 active site accommodating a large degree of flexibility in substrate binding with catalytically proficient and catalytically non-proficient

modes of substrate binding being energetically equivalent. To further explore substrate tolerance of the Bmp3 active site, we enzymatically generated the pyrrolidine C4-Me diastereomers and explored if they could be processed in the Bmp3 active site. Using previously described enzymatic synthesis schemes,^{22–24} L-leucine oxidative cyclization by the LdoA/NosE/NosF enzymes furnished the (2S,4S)4-methylproline (**1**, Figure 4) and the GriE/GriF/ProC enzymes furnished the (2S,4S)4-methylproline (**2**) diastereomer. The 4-methylproline isomers were purified from preparative scale enzyme reactions and confirmed by liquid chromatography/mass spectrometry after derivatization of the carboxylate (Figures S12a,b). Molecules **1** and **2** were used as substrates in a coupled adenylyltransfer-pyrrolidine oxidation reaction in which the adenylyltransferase Bmp4 transferred the proline derivatives to the phosphopantetheinyl thiol of holo-Bmp1(CP) *in vitro* followed by oxidation by wild type Bmp3.⁷

Figure 4: Oxidative conversion of L-leucine to 4-methylproline diastereomers **1** and **2** by the LdoA/NosE/NosF and GriE/GriF/ProC enzymes. After derivitization of the amino acids with Fmoc-Cl, the diastereomers could be chromatographically separated as illustrated by UV-absorbance chromatograms recorded at 254 nm.

As monitored by the formation of the characteristic cyclo-acylpantetheine fragmentation ejection ion,²⁵ full conversion of holo-Bmp1(CP) to 4-methylpyrrole-Bmp1(CP) was achieved in this reaction. While the activity of the adenylyltransferase Bmp4 would indeed be compromised as compared to activation of the native substrate L-proline, this result demonstrates elasticity in the Bmp3 active site, such that it can successfully incorporate diverse substrates.

Transition state and the free-energy barrier of hydride transfer

Although QM/MM simulations indicate the feasibility of the C2-C3 oxidation, they do not provide any information about the energy required for the hydride transfer event. To obtain additional insight into the energy required for hydride transfer, we performed transition state calculations using a cluster model. We discovered that the CS-II and CS-III intermediates are separated by 16.6 kcal/mol, with CS-III being lower in energy. Additionally, we observed that while the FAD isoalloxazine ring is planar in the CS-II intermediate, upon hydride transfer, the isoalloxazine ring bends out-of-plane around the short axis of the ring, leading to a reduction in conjugation in the CS-III intermediate (Figure 5). The bending of the isoalloxazine ring has been observed in prior studies²⁶ when FAD is reduced to FADH⁻. Furthermore, we observed that the HOMO and LUMO in the CS-II state are located on the FAD and the substrate (Figure 5), with a slightly higher contribution from the substrate. The overlap of the FAD and the substrate orbitals originates from the π - π^* overlap between them. However, the loss of planarity of the isoalloxazine ring in the CS-III intermediate leads to the abrogation of the π - π^* overlap. Hence, the HOMO and LUMO of the CS-III state are located only on FAD and the substrate, respectively. The localization of the HOMO on FAD in the CS-III intermediate confirms that the negative charge of the hydride is located on the FAD molecule in the CS-III intermediate.

The transition state (TS) is located 6.7 kcal/mol above the CS-II intermediate state. The TS is associated with a single imaginary frequency centered on the C3-H atoms, which captures the hydride transfer from C3 to N5 as indicated by the red arrow on the TS (Figure 5). Therefore, we conclude that the free energy barrier for the hydride transfer is 6.7 kcal/mol followed by structural rearrangements, which include slight bending of the isoalloxazine ring. In line with abovementioned findings, here, we additionally detect that the direct hydride transfer from N1 is geometrically constrained, and a few structural rearrangements must occur before the transfer. For example, the substrate must rotate around the C-C2 bond (see Figures 1 and S7 for definitions) to come closer to the ^{flavin}N5. A similar rotation

Figure 5: Transition state of hydride transfer. (Top left) The energy barrier for the plausible hydride transfer mechanism from the CS-II intermediate state. (Top right) Geometries of the CS-II and CS-III intermediates and the TS between them. (Bottom) Frontier molecular orbitals of the CS-II and CS-III states.

around the C–C2 bond was observed in the CS-I state (Figures 2c,d). In contrast to the CS-I state, C–C2 is a double bond in the CS-II state. Therefore, the rotational energy barrier is expected to be much higher in this case. We computed a potential energy surface (PES) along the O–C–C2–N1 dihedral angle (Figure S8) in a CPCM solvent model using a low dielectric-constant ($\epsilon = 6.2528$) to mimic the protein environment. The geometry of the molecule was optimized along the dihedral angle at 10° intervals starting from the optimized CS-II state geometry of the substrate. Consistent with other calculations using the cluster model, the relaxed PES scan was performed at the CAM-B3LYP/6-31+G(d) level of theory. We found that the aforementioned rotation has a 33.6 kcal/mol energy barrier, which is much higher than the free-energy barrier for FAD reduction by hydride transfer from the C3 atom. Although a complete rotation may not be required for C2–N1 oxidation, even a $\sim 70^\circ$ rotation is ~ 18 kcal/mol higher in energy than the initial state (Figure S8). To estimate the free energy of activation for this rotation, we also performed a TS calculation at the same

level of theory. The TS is confirmed by a single imaginary frequency, which represents the rotation around the dihedral angle. We observed that the free energy barrier for this rotation is 33.5 kcal/mol while the two end-states differ by only 0.5 kcal/mol. The free-energy barrier of the rotation is ~ 5 -fold higher than the hydride transfer from C3, which does not require additional structural rearrangements. Therefore, we conclude that the C2-C3 oxidation of the substrate is more likely to occur compared to C2-N1 oxidation during hydride transfer to the FAD.

Summary and Conclusions

We have simulated the initial state in the pyrrole biosynthesis in which the substrate protein is bound to an active and an inactive Bmp3-enzyme. We demonstrated that the tetrameric Bmp1-Bmp3 complex is not functionally important for the enzyme catalysis. One Bmp1-Bmp3 can perform the chemical reaction in the presence of a second inactive Bmp3. Furthermore, we demonstrated that the active site of Bmp3 can stabilize FAD and proline substrate. We showed that the Glu242 can initiate the biosynthesis process by capturing a proton from the substrate using MD and QM/MM simulations. Additionally, our results show a significant population of the substrate conformations that favors proton uptake by Glu242. Our simulations show that the C3 atom remains closer to the FAD hydride acceptor site than the N1 atom and one QM/MM simulation captured the hydride transfer event from C3 to FAD within 12 ps without any applied bias. We also determined that hydride transfer to FAD is coupled with C2-C3 oxidation of the substrate and must overcome a very small (6.7 kcal/mol) energy barrier. In contrast, the alternative mechanism, via C2-N1 oxidation, requires rotation around the C=C bond, which has a 33.5 kcal/mol activation free energy barrier. Our findings provide mechanistic insights for a reaction that is well represented in biological chemistry, but, has evaded experimental characterization due to instability of the substrate and intermediate species.

Methods

System preparation and MD simulations

We used the tetrameric complex structure (PDB ID: 6CY8) of the holo Bmp3-Bmp1 system at 2.7-Å resolution.¹⁰ To test the stability and the functional relevance of the Bmp3-Bmp1 complex, we prepared a system containing the substrate, a single bound Bmp1 (chain A), and a dimer of Bmp3 (chains B and C), where only chain B contains the catalytic FAD molecule (Figure 1A). Therefore, we call chains B and C active and inactive Bmp3, respectively. Bmp3 structures have eight missing residues (residues 24 to 31), which were modeled using SWISS model.²⁷ The Bmp1 structure is missing five N-terminal residues and 1 C-terminal residue, which were added using the VMD PSFgen plugin.²⁸ We solvated this complex in an explicit TIP3P water box²⁹ with at least 12 Å of solvent padding around the protein. The system was neutralized by adding 28 Na⁺ ions. We used the CHARMM36m force field³⁰ for protein residues. Initial force-field parameters for the FAD molecule and the substrate were taken from CGenFF,³¹ which assigns them by analogy to existing CHARMM parameters. Charges on the isoalloxazine ring of FAD and the pyrrolidine ring of the substrate (in the CS-I state) were further improved using the Force Field Toolkit (ffTK), as implemented in VMD.^{32,33} Parameters for the dihedral angles, where the pyrrolidine rings connect with the phosphopantetheine appendage, were also improved using ffTK (see Supporting Information for details). MD simulations were performed in the NPT ensemble using periodic boundary conditions at a constant temperature of 310 K and a constant pressure of 1 atm. Long-range electrostatics were computed using the Particle-mesh Ewald (PME) method.³⁴ We used a 12-Å cutoff with a switching function between 10-12 Å. All MD simulations were performed using a 2-fs time step.

We first minimized the modeled missing pieces followed by minimization of the entire system. We equilibrated the solvation box for 0.5 ns with the proteins restrained. After that, we equilibrated the system for another 0.5 ns with a 2 kcal/mol/Å² restraint on protein

backbone atoms, followed by a 50-ns equilibration without any restraints. The final structure from this equilibration was used to initiate all QM/MM simulations. Finally, we ran two independent production simulations for 500 ns, from which we collected trajectory data every 20 ps for subsequent analysis.

QM/MM simulations

The protein system was extracted from the 50-ns equilibration in the CS-I state. Since the active-site charge distribution changes significantly during the reaction, each system was re-solvated and neutralized before QM/MM simulations. The QM region included all residues directly involved in the reaction. The water molecule next to FAD is also included in the QM region to investigate its role in each step of the reaction. The resultant QM region included: FAD, substrate, Glu242, and one water molecule (Figure 6a). The rest of the system was treated as point charges.

Figure 6: Computational models used in this study. (A) QM/MM scheme adopted in the simulations of the CS-I state. QM-MM bonds at the boundary are shown in purple. (B) Cluster model adopted for the transition state calculations.

The link-atom scheme was used to satisfy the valency of the terminal QM atom, where a hydrogen atom was added as the link atom. The redistributed charge and dipole (RCD)

method was used to modify charges and dipoles in the QM-MM boundary region to avoid over-polarization. Additionally, we used the switch function to modify point charges and enforced net-zero electrostatic interactions between charges at the cutoff distance. The net-charge of the system was also set to zero. The QM part of the QM/MM calculation was performed at BP86/def2-SVP level of theory using ORCA.³⁵ The same level of theory was successfully used to model enzymatic hydride transfer from flavin mononucleotide (FMN) using QM/MM methodology previously.³⁶ Becke-Johnson damping (D3BJ) was used for dispersion corrections to the DFT.³⁷ QM/MM simulations were performed using the NAMD/ORCA interface³⁸ and a 0.5-fs time step. After minimization, we raised the temperature of the system to 310 K followed by equilibration for 100 fs with backbone restraints. Finally, we ran one production simulation for the CS-II state using the NPT ensemble at 310 K and 1 atm. The same procedure was followed for QM/MM simulations of the CS-I and CS-II states. The difference between the two states was included in the QM part of the QM/MM simulation.

QM calculations in a cluster model

The cluster model was constructed using the QM/MM equilibrated structure of the CS-II system. In addition to residues already included in the QM region, we added Ser363, Asn123, Thr158, and Phe85 residues, which are located close to the active site (Figure 6b). Phe85 and Thr158 can play important roles since the hydrogen bond of FAD with Thr158 and hydrophobicity near the active site are required to tune the catalysis.³⁹ In total, the cluster model included five residues, substrate, FAD, and one water molecule. All C $_{\alpha}$ -atom positions were frozen during subsequent QM optimizations to preserve the overall structure of the residues. Optimization of CS-II and CS-III states and the TS were performed at the CAM-B3LYP/6-31+G(d) with D3BJ dispersion corrections³⁷ using Gaussian16.⁴⁰ While the B3LYP functional has been used before to describe electronic states of FAD,²⁶ usage of additional long-range corrections is usually recommended for molecules with fused

rings.⁴¹ Therefore, we used the CAM-B3LYP functional, along with an implicit dispersion correction, to correctly describe non-bonded interactions within the cluster model. The average dielectric constant value inside a protein is usually in the 6-7 range.⁴² While some studies have estimated a slightly lower dielectric constant (3-4) inside a protein,^{43,44} it is unlikely that implicit solvents with a dielectric constant in 3-7 range will show discernible differences in molecular properties. Thus, we mimic the protein environment by using the CPCM implicit solvent model⁴⁵ with a low dielectric constant ($\epsilon = 6.2528$).

Conflicts of interest

There are no conflicts to declare.

Acknowledgement

The work was supported by the National Science Foundation (grant CHE-2004030 to VA and JCG). Computational resources were provided by the Extreme Science and Engineering Discovery Environment under grant number TG-MCB130173 and the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta. This work also used the Hive cluster, which is supported by the National Science Foundation under grant number 1828187 and is managed by PACE.

Supporting Information Available

Details of modeling steps, additional computational and experimental details, and hydride transfer movie are provided in the SI.

References

- (1) Hu, D. X.; Withall, D. M.; Challis, G. L.; Thomson, R. J. Structure, chemical synthesis, and biosynthesis of prodiginine natural products. *Chem. Rev.* **2016**, *116*, 7818–7853.
- (2) Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.
- (3) Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein–protein interactions. *Nat. Prod. Rep.* **2020**, *37*, 355–379.
- (4) Thomas, M. G.; Burkart, M. D.; Walsh, C. T. Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis. *Chem. Biol.* **2002**, *9*, 171–184.
- (5) Corpuz, J. C.; Podust, L. M.; Davis, T. D.; Jaremko, M. J.; Burkart, M. D. Dynamic visualization of type II peptidyl carrier protein recognition in pyoluteorin biosynthesis. *RSC Chem. Biol.* **2020**, *1*, 8–12.
- (6) Garneau-Tsodikova, S.; Dorrestein, P. C.; Kelleher, N. L.; Walsh, C. T. Protein assembly line components in prodigiosin biosynthesis: characterization of PigA, G, H, I, J. *J. Am. Chem. Soc.* **2006**, *128*, 12600–12601.
- (7) Agarwal, V.; El Gamal, A. A.; Yamanaka, K.; Poth, D.; Kersten, R. D.; Schorn, M.; Allen, E. E.; Moore, B. S. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol.* **2014**, *10*, 640–647.
- (8) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **2014**, *31*, 61–108.

- (9) Garneau, S.; Dorrestein, P. C.; Kelleher, N. L.; Walsh, C. T. Characterization of the formation of the pyrrole moiety during clorobiocin and coumermycin A1 biosynthesis. *Biochemistry* **2005**, *44*, 2770–2780.
- (10) Thapa, H. R.; Robbins, J. M.; Moore, B. S.; Agarwal, V. Insights into thiotemplated pyrrole biosynthesis gained from the crystal structure of flavin-dependent oxidase in complex with carrier protein. *Biochemistry* **2019**, *58*, 918–929.
- (11) Yi, D.; Acharya, A.; Gumbart, J. C.; Gutekunst, W. R.; Agarwal, V. Gatekeeping Ketosynthases Dictate Initiation of Assembly Line Biosynthesis of Pyrrolic Polyketides. *J. Am. Chem. Soc.* **2021**, *143*, 7617–7622.
- (12) Höfer, I.; Crüsemann, M.; Radzom, M.; Geers, B.; Flachshaar, D.; Cai, X.; Zeeck, A.; Piel, J. Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. *Chem. & Biol.* **2011**, *18*, 381–391.
- (13) Dorrestein, P. C.; Yeh, E.; Garneau-Tsodikova, S.; Kelleher, N. L.; Walsh, C. T. Dichlorination of a pyrrolyl-S-carrier protein by FADH₂-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc. Nat. Acad. Sci.* **2005**, *102*, 13843–13848.
- (14) Westrich, L.; Heide, L.; Li, S.-M. CloN6, a novel methyltransferase catalysing the methylation of the pyrrole-2-carboxyl moiety of clorobiocin. *ChemBioChem* **2003**, *4*, 768–773.
- (15) Kim, J.-J. P.; Miura, R. Acyl-CoA dehydrogenases and acyl-CoA oxidases: Structural basis for mechanistic similarities and differences. *Eur. J. Biochem.* **2004**, *271*, 483–493.
- (16) Ghisla, S.; Thorpe, C. Acyl-CoA dehydrogenases. *Eur. J. Biochem* **2004**, *271*, 494–508.
- (17) Mann, S.; Lombard, B.; Loew, D.; Méjean, A.; Ploux, O. Insights into the reaction mechanism of the prolyl-acyl carrier protein oxidase involved in anatoxin-a and homoanatoxin-a biosynthesis. *Biochemistry* **2011**, *50*, 7184–7197.

- (18) Lee, C.-C.; Ko, T.-P.; Chen, C.-T.; Chan, Y.-T.; Lo, S.-Y.; Chang, J.-Y.; Chen, Y.-W.; Chung, T.-F.; Hsieh, H.-J.; Hsiao, C.-D.; Wang, A. H.-J. Crystal Structure of PigA: A Prolyl Thioester-Oxidizing Enzyme in Prodigiosin Biosynthesis. *ChemBioChem* **2019**, *20*, 193–202.
- (19) Walsh, C. T.; Wencewicz, T. A. Flavoenzymes: versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* **2013**, *30*, 175–200.
- (20) Walsh, C. T.; Garneau-Tsodikova, S.; Howard-Jones, A. R. Biological formation of pyrroles: nature’s logic and enzymatic machinery. *Nat. Prod. Rep.* **2006**, *23*, 517–531.
- (21) Moncoq, K.; Regad, L.; Mann, S.; Méjean, A.; Ploux, O. Structure of the prolyl-acyl carrier protein oxidase involved in the biosynthesis of the cyanotoxin anatoxin-a. *Acta Cryst.* **2013**, *69*, 2340–2352.
- (22) Luesch, H.; Hoffmann, D.; Hevel, J. M.; Becker, J. E.; Golakoti, T.; Moore, R. E. Biosynthesis of 4-Methylproline in Cyanobacteria: cloning of nosE and nosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities. *J. Org. Chem.* **2003**, *68*, 83–91.
- (23) Lukat, P.; Katsuyama, Y.; Wenzel, S.; Binz, T.; König, C.; Blankenfeldt, W.; Brönstrup, M.; Müller, R. Biosynthesis of methyl-proline containing griselimycins, natural products with anti-tuberculosis activity. *Chem. Sci.* **2017**, *8*, 7521–7527.
- (24) Zwick III, C. R.; Renata, H. Remote C–H hydroxylation by an α -ketoglutarate-dependent dioxygenase enables efficient chemoenzymatic synthesis of manzacidin C and proline analogs. *J. Am. Chem. Soc.* **2018**, *140*, 1165–1169.
- (25) Dorrestein, P. C.; Bumpus, S. B.; Calderone, C. T.; Garneau-Tsodikova, S.; Aron, Z. D.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Kelleher, N. L. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl

- elimination reactions during tandem mass spectrometry. *Biochemistry* **2006**, *45*, 12756–12766.
- (26) Kabir, M. P.; Orozco-Gonzalez, Y.; Gozem, S. Electronic spectra of flavin in different redox and protonation states: a computational perspective on the effect of the electrostatic environment. *Phys. Chem. Chem. Phys.* **2019**, *21*, 16526–16537.
- (27) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L., et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- (28) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (30) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71.
- (31) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Jr, A. D. M. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comp. Chem.* **2010**, *31*, 671–690.
- (32) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *J. Comp. Chem.* **2013**, *34*, 2757–2770.

- (33) Pavlova, A.; Gumbart, J. C. Parametrization of macrolide antibiotics using the force field toolkit. *J. Comp. Chem.* **2015**, *36*, 2052–2063.
- (34) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (35) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.
- (36) Willistein, M.; Bechtel, D. F.; Müller, C. S.; Demmer, U.; Heimann, L.; Kayastha, K.; Schünemann, V.; Pierik, A. J.; Ullmann, G. M.; Ermler, U.; Boll, M. Low potential enzymatic hydride transfer via highly cooperative and inversely functionalized flavin cofactors. *Nat. Commun.* **2019**, *10*, 1–10.
- (37) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comp. Chem.* **2011**, *32*, 1456–1465.
- (38) Melo, M. C.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. NAMD goes quantum: an integrative suite for hybrid simulations. *Nat. Methods* **2018**, *15*, 351.
- (39) Mancini-Samuelson, G. J.; Kieweg, V.; Sabaj, K. M.; Ghisla, S.; Stankovich, M. T. Redox properties of human medium-chain acyl-CoA dehydrogenase, modulation by charged active-site amino acid residues. *Biochemistry* **1998**, *37*, 14605–14612.
- (40) Frisch, M. J. et al. Gaussian~16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (41) Acharya, A.; Chaudhuri, S.; Batista, V. S. Can TDDFT Describe Excited Electronic States of Naphthol Photoacids? A Closer Look with EOM-CCSD. *J. Chem. Theory Comput.* **2018**, *14*, 867–876.

- (42) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the dielectric “constant” of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.
- (43) Amin, M.; Küpper, J. Variations in Proteins Dielectric Constants. *ChemistryOpen* **2020**, *9*, 691.
- (44) Schutz, C. N.; Warshel, A. What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins: Struct., Funct., Bioinf.* **2001**, *44*, 400–417.
- (45) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comp. Chem.* **2003**, *24*, 669–681.

Graphical TOC Entry