Insights into substrate recognition by the unusual nitrating enzyme RufO

Benjamin D. Dratch, Kirklin L. McWhorter, Tamra C. Blue*, Stacey K. Jones*, Samantha M. Horwitz, and Katherine M. Davis*

Department of Chemistry, Emory University, 1515 Dickey Dr., Atlanta, GA 30322, USA

*Correspondence: katherine.davis@emory.edu

αT.B. and S.J. contributed equally to this paper
Abstract

Nitration reactions are crucial for many industrial syntheses; however, current protocols lack site-specificity and employ hazardous chemicals. The non-canonical cytochrome P450 enzymes RufO and TtxtE catalyze the only known direct aromatic nitration reactions in nature, making them attractive model systems for the development of analogous biocatalytic and/or biomimetic reactions that proceed under mild conditions. While the associated mechanism has been well characterized in TtxtE, much less is known about RufO. Herein, we present the first structure of RufO alongside a series of computational and biochemical studies investigating its unusual reactivity. We demonstrate that free L-tyrosine is not readily accepted as a substrate, despite previous reports to the contrary. Instead, we propose that RufO natively modifies L-tyrosine tethered to the peptidyl carrier protein of a non-ribosomal peptide synthetase encoded by the same biosynthetic gene cluster and present both docking and molecular dynamics simulations consistent with this hypothesis. Our results expand the scope of direct enzymatic nitration reactions and provide the first evidence for such a modification of a peptide synthetase-bound substrate that may aid in the downstream development of biocatalytic approaches to synthesize rufomycin analogs and related drug candidates.

Main Text

Nitroaromatics are important molecular building blocks for the production of industrial commodities, such as polymers, dyes, pesticides, explosives, and pharmaceuticals. Current synthetic routes to generate them, however, lack precise control over diastereoselectivity and typically require highly acidic conditions. The discovery of two non-canonical cytochrome P450s (CYPs) from Streptomyces, TtxtE and RufO, capable of installing nitro groups (NO₂) during the biosynthesis of non-ribosomal peptide natural products has renewed interest in developing more sustainable biocatalytic methods to produce such industrially relevant nitro compounds. In fact, these CYPs are the only enzymes discovered to date capable of efficiently catalyzing direct and regioselective nitration reactions, yet relatively little is known about the mechanism by which they achieve their novel reactivity.

More specifically, TtxtE is responsible for production of 4-nitro-L-Trp as a precursor to the dipeptide phytotoxin thaxtomin A, while RufO supplies 3-nitro-L-Tyr for incorporation into the heptapeptide tuberculostatic agent rufomycin. The latter is a promising lead compound in the treatment of various multidrug-resistant mycobacterial infections and even some cancers. Previous studies suggest that these transformations diverge from the standard CYP catalytic cycle upon reaction of the ferric-superoxo with nitric oxide (NO), likely afforded by an NO synthase directly preceding the CYPs within their respective biosynthetic gene clusters (BGCs), to yield a peroxyxinitrite adduct. Subsequent cleavage of the O-O bond would result in a ferryl heme and NO₂ radical; however, the mechanism of nitration that follows remains unclear (Fig. S1). Crystal structures of TtxtE have been solved in complex with substrate, L-Trp, but the observed orientation of the indole does not favor nitration at the C4 position, making the basis for site selective modification equally difficult to ascertain.

In an effort to gain insight into the basis for the remarkable regioselectivity of nitrating CYPs, we sought to structurally characterize RufO for comparison. RufO from S. atratus was, therefore, expressed and purified according to methods adapted from Tomita et al., before crystallization via vapor diffusion approaches (see SI for details). Herein, we report the first crystallographic model of the enzyme, which we determined to a resolution of 1.87 Å (PDB: 8SPC, Table S1). Generally speaking, RufO recapitulates the highly conserved CYP fold and is comprised of 12 α-helices (A-L) and five β-sheets (β1-4 and β6) (Fig. 1A). Comparison with TtxtE (PDB: 4TPO) reveals minimal differences despite relatively low sequence homology (29%, Fig. S2A). More specifically, alignment of α-carbons in the peptide backbone yields a root-mean-square deviation (rmsd) of 3.26 Å over 365 atoms. The largest divergence can be found in the...
BC-loop, which is commonly implicated in substrate binding by CYPs. \(^\text{15}\) Previous structures of TxtE depict ordering of this region (residues 59-90) upon complexation with L-Trp, resulting in the formation of two short \(\alpha\)-helices (B’ and B’’). The BC-loop of RufO (residues 69-86), by contrast, displays minimal secondary structure with helices B’ and B’’ almost undetectable. Not only does this comparatively abbreviated loop increase solvent access to the active site, but our data further indicate that it likely remains ordered in the absence of substrate (Fig. 1B). Instead, we observe unusual destabilization of the F-helix (residues 153-166), located at the distal face of the Cys-coordinated heme. Like the BC-loop, the F-helix and nearby structural motifs often play a role in substrate recognition within the CYP superfamily. \(^\text{15}\) The ensuing FG-loop is highly disordered, as is seen in all structures of wild-type TxtE solved to date, and molecular dynamics (MD) studies of the L-Trp nitrating enzyme suggest that it may play a crucial role in orientating substrate during catalysis. \(^\text{16}\)

Beyond differences in large-scale structural features, the active site environment of RufO is substantially more hydrophobic than its only known nitrating counterpart (Fig. 1C,D). Many of the residues that form charged interactions directly with the amino and carboxylate moieties of L-Trp to promote substrate binding in TxtE (e.g. Arg59, Asn293, and Thr296) are replaced by nonpolar amino acids (e.g. Phe69, Val283, and Leu286) in RufO. \(^\text{5, 13, 17}\) An exception to this observation is the hydrophilic residue Gln84 from RufO, which occupies the same position as Met88 in TxtE, where it is thought to work alongside Phe395 to bind the indole moiety of L-Trp. \(^\text{5}\) In addition to changes in polarity, residues lining the substrate binding pocket of RufO are generally less bulky than those in TxtE (Fig. 1C,D), thereby increasing the volume of the active site by approximately 25% (Fig. S2B). The physiological relevance of this altered binding pocket is unclear given the proposed free amino acid substrates are of a comparable size. Unfortunately, we were unable to obtain structures of RufO complexed with L-Tyr or 3-nitro-L-Tyr that might help to rationalize these differences, despite numerous attempts.

We therefore turned to UV-visible absorption spectroscopy to further investigate features of the RufO reactant complex. Reduction of the ferric enzyme resulted in a blue-shift of the Soret peak from \(~421\) nm to \(~411\) nm. We additionally identify two Q-bands at 543 and 570 nm in the spectrum of the ferric enzyme that coalesce to a single peak located at 547 nm upon reduction (Fig. 2A). Given the potential for short-lived intermediates, we employed stopped-flow methods to study complexation with the co-substrates \(O_2\) and NO. Rapid mixing of ferrous RufO with \(O_2\) resulted in the formation of a new species within the dead-time of the instrument that is characterized by a Soret peak at \(~425\) nm and a Q-band at 561 nm (Fig. 2B, S3). Such features are consistent with previously reported spectra of the short-lived ferric-superoxo intermediate in TxtE and other CYPs. \(^\text{13}\) Formation of a putative NO-bound complex, by contrast, occurred within \(~400\) ms and yielded peaks at 440 nm and 561 nm (Fig. 2B, S4). This red-shift in the Soret upon interaction of RufO with NO is similar to, but slightly more significant (~5 nm) than, that observed upon NO binding to TxtE. \(^\text{13}\) Nonetheless, the spectral features and kinetics of co-substrate binding are remarkably consistent between the two enzymes, supporting the hypothesis that peroxynitrite formation in RufO likely occurs via the same mechanistic pathway as in TxtE.

We were consequently quite surprised when our attempts to reproduce a blue-shift of the Soret peak, reported previously upon incubation of ferric RufO with L-Tyr, were unsuccessful. \(^\text{8}\) Such a shift is typically indicative of substrate binding and displacement of the axial water ligand from the heme cofactor in CYPs, however, the spectrum of RufO remained unchanged following the addition of up to 20 equivalents of L-Tyr dissolved in water (Fig. 3A). We were only able to reproduce a blue-shift to \(~370\) nm when Tyr was dissolved in \(1\) M HCl. Unfortunately, the same 50 nm blue-shift is also observed with HCl alone, suggesting the change is due to inclusion of HCl rather than binding of the free amino acid reported previously. \(^\text{8}\) To further assess the ability of RufO to accept free L-Tyr as a substrate, we performed activity assays under a
series of well-established conditions (see Supporting Information for more details). All reactions were quenched in HCl for subsequent analysis via direct injection or HPLC-coupled mass spectrometry (LC-MS) (Fig. 3B, S5-11). Extracted ion chromatograms were monitored for a mass-to-charge ratio (m/z) of 227.0695, corresponding to 3-nitro-L-Tyr, however, no detectable peaks were observed that correlate to putative product formation in agreement with our UV-vis data. Altogether, the aforementioned results prompted reevaluation of RufO’s predicted function with particular emphasis on substrate identity. Both RufO and TxtE are produced by BGCs that also encode non-ribosomal peptide synthetases (NRPSs). These large multi-modal enzymes recognize, activate, and sequentially link small molecule substrates, commonly proteinogenic or modified amino acids, to generate complex natural products. The tet BGC, for example, encodes two single-module NRPSs, TxtA and TxtB, which bind L-Phe and 4-nitro-L-Trp, respectively, to produce the dipeptide core of thaxtomin A. The ruf BGC, by contrast, encodes a seven-module NRPS, RufT, the third module of which was previously hypothesized to incorporate 3-nitro-L-Tyr within the polypeptide scaffold of rufomycin following modification of the free amino acid by RufO. Our results run counter to this interpretation, as we see no evidence for significant RufO reactivity with free L-Tyr. Instead, we propose that RufO modifies the NRPS-bound amino acid during peptide assembly.

To interrogate this claim, we first performed a sequence analysis of conserved NRPS binding motifs. Substrate incorporation studies have identified an eight-residue core sequence that dictates specificity of NRPS adenylation (A) domains. Using the prototypical A-domain from GrsA as a reference, we identified the composition of this motif from the modules of RufT and TxtB, respectively. Comparison of the eight-residue core from the third RufT A-domain reveals high similarity (≥75% sequence identity) to 10 of the 22 A-domains identified by the NRPS substrate predictor database as L-Tyr binding (Table S2). The TxtB specificity motif is much more distinct, with most known L-Trp-binding A-domains ≤50% identical (Table S3). This lack of sequence conservation in TxtB likely arises due to tailoring of the substrate binding pocket to preferentially accommodate 4-nitro-L-Trp over L-Trp. By contrast, the similarity of the relevant RufT A-domain core sequence to known L-Tyr binding motifs is more consistent with incorporation of the unmodified amino acid. We hypothesize that a certain degree of promiscuity remains inherent to substrate recognition by the relevant RufT A-domain, as previous studies report rufomycin production by ΔrufO mutants is restored upon the inclusion of 3-nitro-L-Tyr in the growth media.

It is not unusual for substrates selected by an A-domain to be modified by tailoring enzymes, including many CYPs, following covalent attachment to the phosphopantetheinyl arm of the peptidyl carrier protein (PCP) domain. Because the substrate is esterified under these conditions, we first sought to repeat activity assays with the L-Tyr methyl ester (TME). To interrogate the possibility that nitration occurs after peptide bond formation, we also attempted assays with an N-acetylated L-Tyr ethyl ester (TEEM). In both cases, mass spectra yielded no peaks corresponding to a nitrated product, nor did the position of the Soret peak change upon incubation with the L-Tyr analogs that could indicate binding to the active site (Fig. 3C-E, S12). It seems likely that these small molecules were insufficient to mimic interactions with the phosphopantetheinyl arm or other features of the PCP-domain that may be required to stably orientate the substrate in the active site.

Due to experimental challenges associated with generating PCP-bound L-Tyr, we pursued molecular docking studies to predict its binding mode to RufO. As no crystal structures exist for any portion of RufT, a model of the third PCP-domain (residues 3029-3104), predicted to incorporate L-Tyr, was first generated using AlphaFold. Missing residues of RufO were also built in silico to obtain a complete structure of the CYP (see Supporting Information for details). The resultant docking model of the complex depicts the PCP-
domain directly over the solvent-exposed active site of RufO (Fig. 4A). Furthermore, the hydroxyl group of Ser36, which ultimately hosts the phosphopantetheine moiety, faces directly towards the heme cofactor. Such a conformation would likely position a PCP-tethered L-Tyr within the active site pocket. We subsequently appended the phosphopantetheinyl arm to the PCP-domain in silico and again attempted molecular docking to simulate this binding mode. Unfortunately, the modeled conformation of the FG-loop sterically occluded placement within the active site. As this loop is a dynamic structure, we propose it serves a gatekeeping role, controlling access to the active site, however, such behavior is difficult to capture computationally.

Thus, to predict interactions with the phosphopantetheine moiety, we docked the holo-PCP-domain to the incomplete crystallographic model of RufO. With much of the FG-loop missing due to disorder, the phosphopantetheinyl arm was able to extend into the active site toward the heme cofactor. The physiological relevance and stability of this complex was subsequently evaluated after rebuilding the FG-loop and covalently appending L-Tyr in silico via a series of MD simulations. Note that the putative PCP-bound substrate was truncated at the β-carbon of Ser36 to reduce computational cost. Intriguingly, the FG- and BC-loops were displaced from the active site opening over the course of these simulations, suggesting that, while they appear to interact with the apo-PCP-domain (Fig. S13), they may not directly associate with the phosphopantetheine moiety.

The L-Tyr-tethered phosphopantetheinyl arm, by contrast, remains stably positioned in the active site, as assessed by the time evolution of average RMSD values (Fig. 4B). Furthermore, its configuration is remarkably consistent across all three trajectories, with the largest changes occurring near the entrance to the active site, likely due to the greater degree of freedom associated with the removal of the PCP-domain (Fig. 4C). A hydrogen-bonding interaction with the backbone of Ala277 is consistently observed, thereby restricting heterogeneity closer to the metalloc cofactor (Fig. 4D). The hydrophobic nature of the active site also appears to play a crucial role in facilitating placement, making numerous contacts with the putative substrate (Fig. S14). In particular, Leu158 and Leu373 form a hydrophobic pocket to accommodate the phenyl group of the L-Tyr moiety (Fig. 4D). The exact orientation of L-Tyr, however, varies between the different simulations. The hydroxyl group either interacts with the heme-Fe or a nearby glutamine sidechain (Fig. 4C, D). Excitingly, hydrogen bond formation with Gln74 places the site of nitration, C3, within 6.6 Å of the metal center in an orientation seemingly ideal for modification.

In conclusion, we report the first crystal structure of RufO. Careful analysis of the presented model reveals subtle but key differences between the enzyme and its L-Trp-nitrating counterpart, including a more open, hydrophobic active site, as well as changes to lid loops that are commonly implicated in CYP substrate binding. While spectroscopic characterization of the enzyme is consistent with the hypothesis that RufO performs a nitration reaction similar to TxtE, via the formation of a peroxynitrite intermediate, we were unable to obtain any evidence of reactivity with free L-Tyr reported previously. Instead, we propose that RufO modifies L-Tyr bound to the PCP-domain associated with the third RufT module. Docking models support this revised order of peptide assembly and provide unique insights into the basis for substrate recognition by RufO. Although biochemical validation is still required, our results hint at unexpected versatility within nitrating CYPs and provide crucial insights into the biosynthesis of important lead compounds relevant for the treatment of drug-resistant mycobacterial infections.

**Associated Content**

Detailed description of materials and methods; MS data for enzymatic assays; crystallographic and refinement statistics; A-domain sequence analysis; stopped-flow data; docking simulations; and other supplemental figures. These data are available free of charge at:
Acknowledgements

This work was financially supported by the National Science Foundation Graduate Research Fellowship Program (no. 1937971 to S.K.J. and S.M.H.), the National Institutes of Health (grant R35 GM147557 to K.M.D.) and Emory University start-up funds. GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. Mass spectrometry analyses were conducted by Dr. Siming Wang and Wen Lu at the Georgia State University Mass Spectrometry Facilities, which are partially supported by an NIH grant for the purchase of a Waters Xevo G2-XS Mass Spectrometer (1S10OD026764-01). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.
Figure 1. Structural comparison of RufO (PDB: 8SPC) and L-Trp-bound TxtE (PDB: 4TPO). (A) Overall fold of RufO with annotated α-helices and β-sheets. Note that β6 is too small to depict. (B) Loop regions facilitating access to the active site. (C) RufO and (D) TxtE active site residues surrounding the catalytic heme cofactor. RufO α-helices/β-sheets/loops are shown in blue/gold/grey and TxtE in red/green/tan. The heme cofactor is colored white and L-Trp in light green.
Figure 2. UV-visible absorption spectra of (A) substrate-free RufO in the ferric (black) and ferrous (grey) states. Q-bands are magnified in the insert. (B) Spectral components representing the ferric-superoxo (red) and NO-bound (blue) states of RufO determined via singular value decomposition and global fitting of stopped-flow data.
Figure 3. Absorption spectra and activity assays demonstrating the lack of significant L-Tyr binding to Rufo. (A) Soret peaks and Q-hands are unaltered by L-Tyr binding (black), and only shift following the addition of 1 M HCl (purple). (B) Direct injection mass spectra of Rufo reactions with L-Tyr (red peak). The red diamond indicates where 3-nitro-L-Tyr is expected (m/z [m+H]⁺ = 227.032). (C) Absorption spectra of Rufo bound to L-Tyr analogs selected to mimic the amino acid bound to the phosphopantetheinyl arm of the PCP-domain. Spectral traces are offset for clarity. Corresponding mass spectra from direct injection assessing reactivity with (D) TME or (E) TEEM. Red peaks correspond to these small molecules, while red diamonds indicate where nitrated product is expected (TME-NO₂: m/z [m+H]⁺ = 242.082, TEEM-NO₂: m/z [m+H]⁺ = 298.109).
Figure 4. Computational prediction of the Rufo substrate complex. (A) Surface model of the apo-PCP-domain bound to Rufo. We highlight the serine that is ultimately phosphopantetheinylated in red. (B) Time-evolution RMSD plots of three independent 50 ns MD simulations of Rufo in complex with the docked phosphopantetheinyl arm covalently linked to L-Tyr. (C) Snapshots of the three trajectories bound in the Rufo active site pocket. Variation is primarily observed where the arm would attach to the PCP-domain. (D) Representative configuration of the PCP-bound L-Tyr primed for the nitration reaction. The distance between the Fe ion and C3 is highlighted, as are key hydrogen bonds and hydrophobic contacts. See Fig. S14 for a more complete interaction diagram. Secondary structure elements are colored as in Fig. 1.
References


For Table of Contents Only