Structural and Data Science-Driven Analysis to Assess Substrate Specificity of Diketopiperazine Reverse Prenyltransferase NotF: Cascade Biocatalytic Synthesis of (−)-Eurotiumin A

Samantha P. Kelly,†,§,∥ Vikram V. Shende,†,§,∥ Autumn R. Flynn,∥ Qingyun Dan,∥ Ying Ye,∥ Janet L. Smith,∥ Sachiko Tsukamoto,△ Matthew S. Sigman∥,* and David H. Sherman†,∥,*

†Life Sciences Institute, ∥Program in Chemical Biology, △Department of Biological Chemistry, †Department of Medicinal Chemistry, ‡Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA. §Department of Chemistry, University of Utah, Salt Lake City, UT, USA. ¶Graduate School of Pharmaceutical Sciences, Kumamoto University, Kumamoto, Japan. *These authors contributed equally: Samantha P. Kelly, Vikram V. Shende.

ABSTRACT: Prenyltransfer is an early-stage carbon–hydrogen bond (C–H) functionalization prevalent in the biosynthesis of a diverse array of biologically active bacterial, fungal, plant, and metazoan diketopiperazine (DKP) alkaloids. Towards the development of a unified strategy for biocatalytic construction of prenylated DKP indole alkaloids, we sought to identify and characterize a substrate-permissive C2 reverse prenyltransferase (PT). In the biosynthesis of cytotoxic notoamide metabolites, PT NotF is responsible for catalyzing the first tailoring event of C2 reverse prenyltransfer of brevianamide F (cyclo(L-Trp-L-Pro)). Obtaining a high-resolution crystal structure of NotF (in complex with native substrate and prenyl donor mimic dimethylallyl S-thiolodiphosphate (DMSPP)) revealed a large, solvent exposed substrate binding site, intimating NotF may possess significant substrate promiscuity. To assess the full potential of NotF’s broad substrate selectivity, we synthesized a panel of 30 tryptophan DKPs with a suite of sterically and electronically differentiated amino acids, which were selectively prenylated by NotF in often synthetically useful conversions (2 to >99%). Quantitative representation of this substrate library enabled the development of a descriptive statistical model that provided insight into the origins of NotF’s substrate promiscuity. Through this unique approach for understanding enzyme scope, we identified key substrate descriptors such as electrophilicity, size, and flexibility, that govern enzymatic turnover by NotF. Additionally, we demonstrated the ability to couple NotF-catalyzed prenyltransfer with oxidative cyclization using recently characterized flavin monooxygenase, BvnB, from the brevianamide biosynthetic pathway. This one-pot, in vitro biocatalytic cascade proceeds with exceptional substrate recognition, and enabled the first chemoenzymatic synthesis of the marine fungal natural product, (−)-eurotiumin A, in three steps and 60% overall yield.

INTRODUCTION

The incredible structural diversity and potent biological activity of C2 reverse prenylated diketopiperazine (DKP) natural products (NPs) has inspired the development of selective chemical methods for reverse prenyltransfer as an entry point for synthesis of prenylated NPs and their analogues. 1–3 State-of-the-art synthetic methods for C2 reverse prenylation of tryptophan substituents is typically performed at an early stage, requires multiple protecting groups, and fresh preparation of a prenylating reagent, such as prenyl-9-BBN. 4 The development of a general, biocatalytic method for prenyltransfer, i.e., a site-selective prenyltransferase (PT) with broad substrate scope, would provide an efficient method for installation of prenyl functionalities that circumvents the necessity of protecting groups and enables late-stage diversification with additional tailoring enzymes.

Widespread in primary and secondary metabolism of marine and terrestrial organisms, PTs are responsible for key transformations within biologically active NP biosynthesis, catalyzing the transfer of five carbon building blocks such as dimethylallyl (DMAPP), geranyl (GPP), farnesyl (FPP), and other pyrophosphates onto a variety of substrates. 9,10 Soluble aromatic PTs that share a conserved fold of five consecutive αβα repeats comprise the ABBA-type superfamily of PTs. 11 The dimethylallyl tryptophan synthase (DMATS)-type subfamily of ABBA PTs are often encoded in alkaloid biosynthetic gene clusters and catalyze both normal and reverse prenylation of indole and tryptophan derivatives, 11,12 with varying site-selectivities, 9,13–15 and often relaxed substrate specificity. 9,13,15–18 PTs often act early in the assembly of biologically active alkaloids, preceding a variety of oxidative tailoring events, including oxygenation (such as (+)-okaramine B 3), 19,20 spirocyclization ((+)-brevianamide A 1), 21–23 halogenation ((+)-malbrancheamides), 24 cycladdition ((+)-paraherquamides), 25 as well as many others (Figure 1A and S1). 26,27

We previously identified and characterized DMATS-type PT, NotF, which catalyzes C2 reverse prenylation of native substrate brevianamide F (cyclo(L-Trp-L-Pro) 5, Figure 1B) and found it to possess strict substrate selectivity in a limited scope analysis. 28 However, recent work using stereoisomeric cyclo(L/D-Trp-L/D-Pro) and cyclo(L/D-Trp-L/D-Ala) substrates suggested that NotF may possess broader flexibility in comparison to homologous DMATS-type PT, BrePT (83% sequence ID, Figure S3). 16–18 To gain molecular insight into how NotF may be able to accommodate a broad range of substrates, and toward our goal of developing a selective C2 reverse PT biocatalyst, we sought to further evaluate NotF’s substrate specificity through an interdisciplinary structural, biochemical, and computational approach.
RESULTS AND DISCUSSION

Structural characterization of NotF. We solved a ligand-free NotF structure at a 3.2 Å resolution, using molecular replacement with AnaPT as the search model. This NotF structure is consistent with the DMATS-type PTs within the ABBA PT superfamily, and possesses a classic PT fold with 10 inner antiparallel β strands surrounded by outer α helices (Figure 2A). A metal ion is absent in the active site due to the lack of a (N/D)DXXD Mg²⁺ coordination motif, also consistent with DMATS-type PTs. In the apo state of NotF, the central active site is open and solvent-accessible. A 3.0 Å NotF structure in complex with its natural substrate brevianamide F (5) and a non-hydrlysable DMAPP analog, dimethylallyl S-thiolodiphosphate (DMSPP), was determined by rescreening NotF using co-crystallization. In both the apo and holo structures, a flexible loop with 21 residues (G328–Y440) forms salt bridges with three positively charged residues (R122, K212, K282 and K369) forming salt bridges with the negatively charged phosphates (Figure 2C). Furthermore, conserved tyrosines (Y214, Y284, Y371 and Y440) also hydrogen bond with the DMSPP ligand, facilitating diphosphate dissociation, shielding the active site from solvent, and protecting the putative reactive carbocation intermediate. The prenyl acceptor brevianamide F (5) resides in a hydrophobic pocket, with the indole moiety tightly packed between a loop region (R97–S103) and the planar dimethylallyl group (Figure 2D, E). The conserved E108 forms a hydrogen bond with spatially constrained substrate indole N1, with the C2 acceptor facing the...
prenyl donor (Figure S4). Although DMSSP C1 is slightly closer to indole C3 (3.5 Å compared to 4.2 Å to indole C2), previous studies have shown that indole C2 is more favorable for Friedel-Crafts alkylation.\textsuperscript{34,35} Remarkably, no hydrogen bond or chargedcharged interaction involving the DKP core or fused proline side chain are present in the complex structure. The NotF active site pocket leaves substantial unoccupied space surrounding the fused DKP-proline bicycle, the binding site of which remains mostly solvent-exposed in the ligand-free structure. Based on these structures, we reasoned that NotF may be able to accommodate substrates with amino acid substitutions at the proline side chain without affecting the indole interaction mode. Thus, we set out to evaluate the scope of NotF’s substrate accommodation for potential biocatalytic production of NPs and their derivatives.

**Substrate scope of NotF.** Using a series of sterically and electronically differentiated tryptophanyl-DKP substrates (5, 8 – 36), we interrogated the substrate promiscuity of this PT. NotF was able to convert all 30 DKP substrates, often displaying complete consumption of substrate (Figure 3, Tables S2 – S31), to generate C2 reverse prenylated DKP products including NPs talathermophilin E (S1 from 8),\textsuperscript{36} preechinulin (37 from 9),\textsuperscript{37} fellutamine B (S2 from 35),\textsuperscript{38} and deoxybrevianamide E (6 from 5) (Figure S1).\textsuperscript{39} DKP substrates containing small, nonpolar side chains (5, 9 – 12, 15, 18, 23) were converted with high efficiency, reflected by ≥ 70% conversion in all cases. NotF also tolerates bulky (21 – 36) and polar (14, 15, 20, 21, 28) side chains with conversions ranging from 2 to 99%. The high substrate promiscuity of NotF offers an efficient biocatalytic method for selective chemoenzymatic synthesis of diverse C2 reverse prenylated DKPs. Furthermore, high conversion of alkyne-containing substrate 18 (76%) enables potential late-stage diversification via annulation reactions such as Larock indole synthesis\textsuperscript{40} or biorthogonal modification through one-pot diversification via click chemistry.\textsuperscript{41,42}

---

**Figure 3.** A) General reaction scheme and substrate scope of NotF using synthesized panel of tryptophanyl DKPs with % conversion in blue and initial rate (% conversion per hour) in purple; B) Reverse C2 prenylated products generated on preparatory scale and characterized by NMR.
Previous reports investigating PT substrate scope revealed that use of non-native substrates can induce a change of both chemico- and site-selectivity with respect to both the indole core and DMAPP.\textsuperscript{14,15,43} To confirm that NotF’s C2 reverse site-selectivity is irrespective of substrate structure, we performed preparative scale biocatalytic prenyltransfer on substrates varying in steric bulk including cyclo(L-Trp-L-Ala) (9), cyclo(L-Trp-L-Val) (13), cyclo(L-Trp-L-Pip) (23), and tetrahydrocarbazole-containing (36). Isolation and NMR structural characterization of all three reactions confirmed C2 reverse prenyltransfer with isolated yields (54–98%) consistent with those from analytical scale reactions (Figures 3, S141 – S162).

**Multivariate Linear Regression Modeling.** While NotF converted all 30 of these substrates, we were unable to identify clear trends between steric or electronic factors and the magnitude of catalytic turnover. To understand the relation of substrate structure and catalytic activity, we performed *in silico* rigid docking with a representative set of substrates in the holo NotF active site (29, 30, 35, and 36, Figures S37 – S40).\textsuperscript{44} Although productive binding poses were observed through docking of each substrate, binding scores did not correlate with the magnitude of catalytic turnover (Figure S40). In the absence of distinct structural factors (from both enzyme and substrates) which may govern substrate specificity of NotF, our access to a wide variety of structurally differentiated substrates and ability to efficiently collect kinetic data provided us a unique opportunity to gain insight into the molecular basis for substrate prenylation. We hypothesized that substrate structure and conformation may be related to kinetics of binding and dissociation and reasoned that a substrate-profiling approach, aided by computational tools, might clarify the basis for substrate turnover. In an alternative strategy to our typical use of molecular dynamics (MD) simulations to observe how substrate-protein interactions influence catalysis, we opted to use statistical modeling techniques, an approach traditionally used to parameterize small molecule catalysts,\textsuperscript{45,46} to gain insight into NotF’s substrate permissiveness. Three-dimensional unrestrained conformational ensembles of each substrate structure were collected, and quantitative chemical descriptors were acquired from density functional theory (DFT) calculations of the substrates at the B3LYP level of theory. These features included electronic (e.g., natural bond orbital (NBO) charges) and steric descriptors (e.g., Sterimol values).\textsuperscript{47} Regression of this library of descriptors to the log of empirically determined pseudo-first order initial rates (Figure 3), enabled the building of a statistical model (Figure 4A).

Our model is comprised of three terms, each with a negative coefficient, indicating that as magnitude of the descriptor values rise, the relative rate of reaction is reduced (Figure 4A). This model presents robust values for statistical evaluation (where the Training $R^2$ is high, and the Test $R^2$ and leave-one-out cross-validation $Q^2$ agree with both the Training and Test $R^2$ values, Figure 4A). Substrates containing unsaturated moieties, many of which are aromatic, possess high $\omega_{\text{min}}$ values (defined as the substrate electrophilicity index) and are prenylated at slower rates than those that are saturated (substrates with the 10 largest $\omega_{\text{min}}$ values are 18, 21, 24, 29 – 31, and 33 – 36). The active site of NotF consists of several aromatic residues (F216, Y266, F268, W424) that surround the binding pocket of native substrate 5 (Figure 4B). We reasoned substrates with an aromatic or unsaturated group in the variable side chain likely form noncovalent $\pi$-interactions with these aromatic residues, which may restrict product dissociation, resulting in decreased turnover (Figures S42 – S44). The Sterimol $L_{\text{max-min}}$ term (describing substrate flexibility, Figure 4C) also possesses an inverse relationship with the rate, indicating highly flexible substrates also convert more slowly. This is exemplified in substrate pairs 22 and 23, 29 and 30, and 35 and 36, which illustrate how restricting degrees of rotational freedom through tethering the side chain to the DKP (as in 23, 30, and 36) is reflected in increased rate of these substrates relative to their unconstrained pair. The third term, Sterimol $B_5$, indicates that a large maximum width of each substrate diminishes the rate of the reaction (Figure 4C). Analysis of the enzyme binding site bound to the native substrate 5 shows this is likely because of repulsive steric interactions within the active site in the case of bulkier substrates (such as 32, 35 and 36, Figure 3). These interactions may demand major active site repositioning for substrate acceptance, leading to slower substrate binding and ultimately decreased turnover.

While the model describes the differences observed in initial rates, we found that in many cases we cannot correlate this explanation to the total conversion achieved in these reactions. We therefore hypothesize that other factors, including product inhibition, may slow the rate once the substrates no longer exist in saturating concentration, effectively decreasing overall conversion.

---

**Figure 4.** Linear model and mechanistic interpretation of molecular feature relation to initial rate (% conversion/hour). A) The linear regression model shown indicates three terms which lead to decreased rate. The Training and Test $R^2$ values agree with each other, indicating a successfully descriptive model; B) The DKP substrate framework shown in a cartoon of the active site of NotF. Illustrated are key aromatic residues present within 5 Å of native substrate; C) The DKP substrate framework shown in a cartoon of the active site of NotF. In the crystal structure of native substrate 5 bound to NotF, active site residues tightly surround the binding pocket, indicating expansion in width or length coupled with increased flexibility would demand major residue repositioning.

---
By utilizing statistical modeling to correlate substrate structures to initial rates in this substrate-permissive reaction, we readily uncovered key substrate features that lead to differing rates of prenylation. Notably, this approach bypasses computationally intensive techniques that do not inherently reveal these features, such as molecular dynamics (MD) simulations. However, we sought to corroborate our model explicitly, and turned to induced-fit docking (IFD), which is a molecular docking protocol where both substrate and enzyme are flexible, to investigate our assertions further. In the case of substrates with an aromatic DKP moiety (such as 29) IFD shows π-interactions between the phenyl ring and Y266 (Figure S45). The binding pose of the non-asemic counterpart, 32, is similar to that of 29, however interaction with Y266 is lacking. This finding substantiates the ωmin term in our statistical model, indicating π-interactions between arene-containing DKPs and the binding pocket of NotF.

When 5 is bound to NotF, the residues in the pocket closely surround the substrate. But when 36, a substrate with high Lmax-min and B5max values, is docked via IFD we observe major repositioning of nearby residues Y266, I100, L193, P352 to accommodate the steric bulk of 36 (Figure S46). The RMSD of active-site residues (within 5 Å of ligand) when 36 is docked is 1.56 Å, relative to the docking pose of 5. When 12, a substrate with small Lmax-min and B5max values is docked, this repositioning does not occur, and the RMSD for active site residues is 1.06 Å (Table S33). These IFD results support the hypothesis that large Lmax-min and B5max values lead to major NotF residue repositioning upon substrate binding.

Overall, the results demonstrate a substrate-permissive method for selective C2 reverse prenylation of 30 tryptophenyl-DKPs representative of fungal indole alkaloid NP structures, suggesting NotF’s exceptional potential as a biocatalyst in the assembly of this class of NPs. Aside from the native NotF product deoxybrevianamide E (6) serving as a direct precursor to various fungal bicyclo[2.2.2]diazaoctane indole alkaloids, preechinulin (37) is a likely intermediate in the biosynthesis of several additional NPs, including variecolortide A (4), asprechinulin B (2), and (-)-uncarilin B (SS8), among many others (Figures 1 and S1). Similarly, C2 reverse prenylation of cyclo(L-Trp-L-Val) (13) is a likely precursor to cristiastatin F (SS5), cyclo(L-Trp-L-Trp) (35) to okaramine B (3, Figure 1), and cyclo(L-Trp-L-Pip) (23) has been utilized as a synthetic intermediate to the marfortines and related indole alkaloids (S10, Figure S1). The ability to selectively C2 reverse prenylate these substrates therefore represents a solution to a key transformation within these biosynthetic systems.

**Figure 5.** Biologically active prenylated pyrroloindoline DKP-containing NPs.

One-pot biocatalytic cascade to (-)-eurotium A. The fused pyrroloindoline scaffold is prevalent in a variety of prenylated DKG NPs with potent biological activity including anticancer ardeemin (43), antimicrobials roquefortine C (45) and drimetine A (46), and the structurally analogous diketomorpholine NP shornephine A (42, Figure 5). Given NotF’s broad substrate specificity and good to high turnover of DKG precursors to these scaffolds, we aimed to develop a biocatalytic cascade approach towards prenylated pyrroloindoline DKP.

We recently identified a stereoselective flavin monooxygenase (FMO), BvnB, as responsible for catalyzing indole 2,3 epoxidation and subsequent cyclization to 3-hydroxy-pyrroloindoline shunt product brevianamide E (47) (Figure 6, Figure S47). Therefore, we envisaged the use of BvnB in tandem with NotF might enable facile assembly of prenylated pyrroloindoline NPs in a one-pot reaction. After validating BvnB’s ability to catalyze in vitro epoxidation and cyclization of deoxybrevianamide E (6) to brevianamide E (47, Figure S48), we assessed potential compatibility between NotF and BvnB for the development of a biocatalytic cascade to generate prenylated hydroxy-pyrroloindoline NPs and analogues thereof. Gratifyingly, in a one-pot cascade with both biocatalysts, we observed high conversion of native substrate 5 to 47 (Figure S49). Notably, in control reactions lacking NotF, BvnB did not convert non-prenylated substrate 5, suggesting a requirement for molecular recognition of the C2 reverse prenyl substituent for productive catalysis.

Given the selective substrate recognition observed with native substrates 5 and 6, we envisioned the use of this NotF-BvnB cascade in a stereoselective, biocatalytic synthesis of fungal 3-hydroxy-pyrroloindoline NP, eurotiumin A (41), whose synthesis has not been previously reported. As in reactions with 5, we...
observed selective substrate recognition by each biocatalyst with no conversion of cyclo(L-Trp-L-Ala) (9) by BvnB in the absence of NotF (Figure 6, trace iii). Scale up of the NotF-BvnB cascade with 9 and structural characterization of final product confirmed the first biocatalytic, stereoselective synthesis of eurotumin A (41, 60% isolated yield over three steps, with a single, ultimate chromatographic purification), with no observed production of stereoisomers or shunt products, and enabled determination of optical rotation of 41 as [α]D35 −335.08 hereon referred to as (−)-eurotumin A. Thus, the "one-pot" use of NotF and BvnB provides an efficient route towards the hydroxypyrroloindoline scaffold present in NPs with a range of biological activities, which are amenable to further synthetic diversification.

CONCLUSION

Herein, we describe the biochemical, structural, and statistical characterization of the C2 reverse prenyltransferase (PT), NotF, which revealed the molecular basis for broad flexibility toward tryptophanyl-DKPs. We previously reported NotF’s high selectivity for brevianamide F (5) with no turnover of unnatural substrates cyclo(L-Trp-L-Trp) (35) and cyclo(L-Trp-L-Tyr) (33), and we reason that improved methods for NotF expression and purification resulted in more stable and active protein sample. We now demonstrate effective C2 reverse prenylation of 30 DKP substrates, and provide a straightforward, protecting group free, biocatalytic strategy to overcome a synthetically challenging C–H functionalization in indole alkaloid biosynthesis. NotF is capable of directly generating several prenylated NPs from simple DKP precursors (accessible in two synthetic steps), including talaterrhomelin E (S1), preechinulin (37), deoxybrevianamide E (6), and fellutanine B (S2), and precursors to additional NP classes including the bicyclo[2.2.2]diazaoctane indole alkaloids 1 and 7, polyprenylated cristatumin F (S5), oxidatively cyclized okaramine B (3), as well as the cycloechinulins and variecoloritide A (4, Figures 1 and S1).

Data science-driven approaches can provide powerful methods for enzyme characterization that supplement our current molecular tools for better understanding biocatalyst substrate scope. The statistical modeling of substrate features can be used to rationalize performance in biocatalytic reactions and enable engineering of improved biocatalysts with enhanced activity and broad applicability. In our study, the implementation of statistical modeling through multivariate linear regression (MLR) of steric and electronic molecular descriptors enabled our identification of substrate features that influence the rate of prenyltransfer by NotF. Our model found that substrates with high electrophilicity, flexibility, and steric bulk are turned over at slower rates, a result unlikely to be gleaned from more frequently used MD simulations.

Using computationally inexpensive induced-fit docking (IFD), we found that larger and more flexible substrates require rearrangement of active site residues that likely attenuate turnover to product. While previous studies have used statistical analysis to understand enzyme catalysis and guide engineering efforts, this work showcases the integration of substrate profiling using MLR with IFD as a computationally cost-effective and simple interrogation of substrate scope. To accompany random, site-saturation, structure-guided, and MD-guided mutagenesis strategies that our group has previously relied upon, we aim to harness the predictive power of machine learning and statistical modeling to further expand our protein engineering toolbox, and develop NotF variants and other NP biosynthetic enzymes with expanded substrate scope (Figure 7). We specifically envision engineering NotF to accommodate more bulk surrounding the indole substrate to access further functionalized biosynthetic intermediates and probe downstream tailoring events of numerous indole alkaloid biosynthetic pathways.

Finally, we developed a one-pot methodology using PT NotF and FMO BvnB that provides an efficient route towards the 3-hydroxypyrroloindoline NP scaffold from simple and readily accessible diketopiperazine building blocks. This biocatalytic cascade was applied towards the first synthesis of antioxidant NP (−)-eurotumin A as a single stereoisomer, in 60% yield over three steps with only one final chromatographic purification. With demonstrated crosstalk between notoamide and brevianamide biosynthetic enzymes, we envision further diversification of prenylated DKP-containing NPs. In summary, we have employed structural and statistical modeling approaches to demonstrate the molecular basis for NotF’s broad substrate scope, and its application in the biocatalytic assembly of the 3-hydroxypyrroloindoline scaffold.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ChemRxiv website.

Full experimental and computational details, NMR spectra, tables, and figures (PDF)

Substrate descriptors and induced fit docking materials (.zip)
ACKNOWLEDGMENTS

The authors thank the NSF CCI Center for Selective C–H Functionalization (CHE-1700982), NIH grant R35 GM118101, and the Hans W. Valtheich Professorship (to D.H.S.), NIH grant R01 DK042303, and the Margaret J. Hunter Professorship (to J.L.S.). M.S.S. and A.R.F. thank NIH grant R35 GM136271 for partial support. Computational resources were provided by the Center for High Performance Computing (CHPC) at the University of Utah. 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. GM/CA@APS (beamlines 23ID-B and 23ID-D) has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289.

REFERENCES


Bioactivity and Synthesis of Natural Products with...


Applications in synthesis:

Basis for broad selectivity identified via:

- screening 30 DKP substrates
- substrate bound high-resolution X-ray structure
- multivariate linear regression modeling

Train R² = 0.85
Test R² = 0.79
Training Q² = 0.85

one-pot diastereoselective biocatalytic cascade
first synthesis of euroultin A


Gago, F.; Gago, F.; A.; Prangé, T.; Pascard, C.; Moreau, S. Uncarili...-1021/jacs.0c06312.


