ACS Catalysis

pubs.acs.org/acscatalysis Research Article

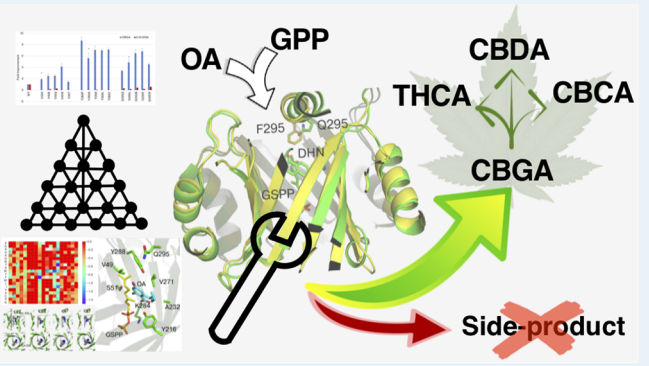
Structure-Guided Engineering of Prenyltransferase NphB for High-Yield and Regioselective Cannabinoid Production

Kevin Jie Han Lim,[⊥] Yossa Dwi Hartono,[⊥] Bo Xue,[⊥] Maybelle Kho Go,[⊥] Hao Fan,* and Wen Shan Yew*

Cite This: *ACS Catal.* 2022, 12, 4628–4639 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Synthetic biology efforts for cannabinoid research have seen a rapid expansion in recent years. This is in response to the increasing awareness and legalization of the secondary metabolites from *Cannabis sativa*, dubbed the green rush. In transgenic synthetic biology applications, NphB is a promiscuous prenyltransferase from *Streptomyces* sp. often used as a replacement in the prenylation step producing the cannabinoid cannabigerolic acid (CBGA), the key precursor to many other cannabinoids. However, its application as a CBGA synthase replacement is limited by its nonspecific regioselectivity in producing a side product along with CBGA. Herein, we demonstrated a detailed and extensive computational structure-guided approach in identifying target residues of mutation for engineering NphB for optimal CBGA production. Our comprehensive computational workflow has led to the discovery of several highly regiospecific variants that produce CBGA exclusively, with the best-performing V49W/Y288P variant having a 13.6-fold yield improvement, outperforming all previous work on NphB enzyme engineering. We subsequently investigated the effects of these mutations by X-ray crystallographic studies of the mutant variants and performed molecular dynamics simulations to uncover an interplay of a H-bonding network and an optimal ligand orientation that favors the CBGA production over the side product. Collectively, this study not only recapitulates the utility of computational tools in informing and accelerating experimental design but also contributes to a better understanding of molecular mechanisms that govern enzyme regioselectivity and readily aids in cannabinoid synthetic biology production for future research into maximizing their therapeutic potential.



KEYWORDS: *cannabinoid biosynthesis, computational protein engineering, synthetic biology, enzyme design, prenyltransferase*

INTRODUCTION

Cannabinoids are a class of natural products originally isolated from the plant *Cannabis sativa*. These meroterpenoids, characterized by a resorcinoyl polyketide commonly attached to a C10 geranyl moiety, have been used as a medicinal herb since the time of ancient civilizations.[1] However, clinical use and research into their therapeutic potential diminished following the criminalization of associated use in 1937 and their classification as a Schedule I drug under the Controlled Substances Act of 1970.[2] This is largely due to psychoactive components found in *C. sativa*, such as Δ^9-tetrahydrocannabinol (THC, the most studied phytocannabinoids), leading to a high potential for illicit use. Recently, there have been increasing awareness and legalization around the globe for its medicinal properties. In 2018, Epidiolex from GW Pharmaceuticals became the first cannabis-derived drug approved by the US Food and Drug Administration (FDA) for the treatment of seizures in patients with rare forms of epilepsy.[3]

At least 120 phytocannabinoids have been isolated to date, many of which are bioactive and may be therapeutically relevant. Unfortunately, many of these lesser-known cannabinoids are produced in low yield in *C. sativa* in comparison to major cannabinoids such as Δ^9-THC and cannabidiol (CBD). Microbial bioproduction is a viable alternative to traditional cannabis cultivation in producing these rare cannabinoids due to its significantly faster growth rate and relative ease in metabolic engineering. Additionally, the establishment of the cannabinoid biosynthetic pathway in microbial cell factories provides access to derivatized novel cannabinoid scaffolds for further studies into their unique bioactivities. Using an engineered yeast chassis, Luo *et al.* were able to produce acidic forms of THC and CBD in mg L^{-1} scale.[4] However, a significant bottleneck in the heterologous expression of the

Received: February 13, 2022
Revised: March 23, 2022

cannabinoid biosynthetic pathway, as identified by many groups, is producing cannabigerolic acid (CBGA), the central precursor to many other cannabinoids. The endogenous prenyltransferase from *C. sativa*, geranylpyrophosphate:olivetolate geranyltransferase (GOT), is a membrane-associated enzyme that is challenging to express in a microbial chassis.[4,5]

NphB is a soluble aromatic prenyltransferase from *Streptomyces* sp. first characterized by Kuzuyama *et al.* and has been shown to be able to prenylate a variety of aromatic substrates including olivetolic acid (OA) to form CBGA.[6] It serves as a suitable replacement to the CBGA synthase enzyme GOT in the synthetic pathway due to its highly soluble, well-characterized, and promiscuous nature.[7] However, previous studies have shown that it produces a side product, 2-*O*-geranyl-olivetolic acid (2-O-GOA), by prenylating OA with geranyl pyrophosphate (GPP) on a different site as shown in Figure 1.[5,8] This diverts flux away from the cannabinoid producing pathway through CBGA and hence reduces the yield of cannabinoids obtainable.

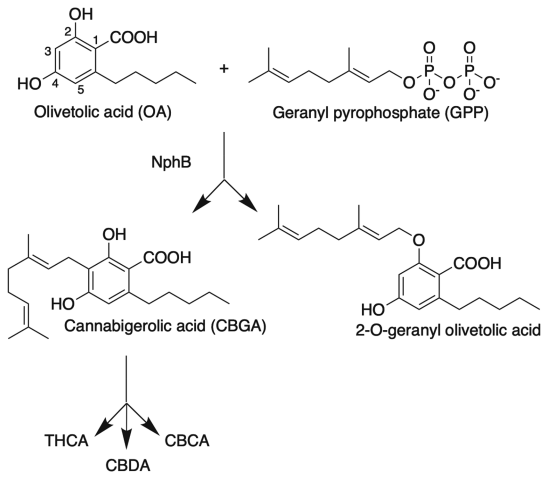


Figure 1. NphB produces two products at different prenylation sites when using OA and GPP as substrates.

Protein engineering can produce a high-yielding and regiospecific NphB enzyme variant for cannabinoid biosynthesis, as many other groups have previously demonstrated success in engineering enzyme regioselectivity. Frances Arnold's group used directed evolution to alter the enantiomeric selectivity of a D-selective hydantoinase to an L-selective enzyme and at the same time improve its activity fivefold.[9] However, protein engineering methods such as directed evolution require a robust selection method to pick out variants with the desirable trait. To narrow down the sequence space to survey, many groups have sought to introduce a predictive selection step using structure-guided computational mutagenesis prior to experimental testing.

Given the high level of commercial interest and popularity in the field of cannabinoid research, this has resulted in two separate accounts of protein engineering efforts by Valliere *et al.*[8] and Qian *et al.*[10] in obtaining variants of NphB with better regioselectivity and yield for the desired CBGA product. Both studies introduced a computational selection step to pick out favorable mutations in the NphB aromatic binding site and incorporated their favorable variants to produce CBGA. We would like to obtain a deeper understanding of the mechanism of the prenylation reaction, specifically on how target mutations favor prenylation on different sites of the substrate. Our group previously employed computational enzymatic evolution in altering the substrate specificity of the *Geobacillus kaustophilus* lactonase.[11] Herein, we applied similar engineering principles in altering NphB regioselectivity toward cannabinoid biosynthesis.

In this work, we performed a rigorous and extensive structure-guided rational mutagenesis to discover NphB variants that produce CBGA in high yield and regioselectivity. We first commenced with a careful computational step predicting single mutations to narrow down sequence space and subsequently performed focused experimental yield assays. After identifying favorable single variants, we expanded our search by testing saturation mutations at these residue positions and obtained even better-performing single variants. We then tested the double and triple mutant variants and arrived at the best-performing double variants that displayed a 13.6-fold CBGA yield improvement with highly specific regioselectivity and outperformed all previously established variants of NphB under the same experimental conditions. Subsequently, we characterized the best-performing variants in our study, determined their 3D structures through X-ray crystallography, and performed molecular dynamics (MD) simulations in an attempt to elucidate the effects of these mutations mechanistically. Observations from the X-ray structures and MD simulations suggested deleterious H-bonding interactions and specific ligand orientations that favor C-prenylation to form CBGA.

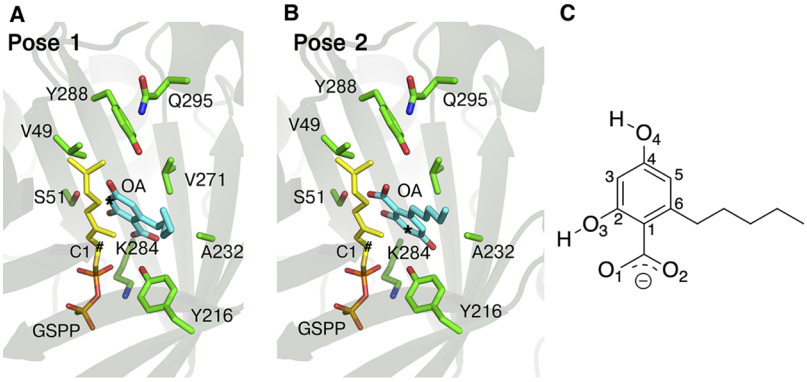


Figure 2. Docking (A) Pose 1 and (B) Pose 2 of OA in the NphB active site. The C3 atom on OA to be geranylated is marked with *, while C1 on GSPP is marked with #. (C) Atom numbering of OA.

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

RESULTS AND DISCUSSION

***In Silico* Screening and Identification of Favorable Mutations.** To identify variants of NphB with improved regioselectivity for CBGA product and yield, we started with docking and computational mutagenesis. The ligand-bound NphB X-ray crystal structure (PDB 1ZB6[6]) contains two ligands: a nonhydrolyzable analog of GPP named geranyl *S*-thiolodiphosphate (GSPP) and 1,6-dihydroxynaphthalene (1,6-DHN). The aromatic ligand 1,6-DHN was first removed from the enzyme structure and docked with OA. Thereafter, each of the 19 amino acid residues within 5 Å of the docked OA was mutated to other 19 possible residue types in the presence of the docked OA, resulting in 361 variants. Each variant was then redocked with OA, and the variant with the best docking score at each residue position was chosen for experimental testing.

The spacious nature of the aromatic substrate binding site of NphB, as observed by Kuzuyama *et al.*,[6] allows a diverse set of aromatic substrates to undergo prenylations. Docking of OA therefore resulted in diverse docking poses, many of which were excluded from consideration as they are not proximal enough to the geranyl moiety in GSPP for the catalytic reaction. We selected two viable docking poses, Pose 1 and Pose 2, that satisfied the catalytic constraint: the geranyl group in GSPP is proximal to the C3 carbon in OA (marked with * in Figure 2) to produce CBGA. Pose 1 was the only viable pose when docking to wild-type (WT) NphB and was the predominant pose (85% of cases) with the best docking score in most variants. In some variants, however, Pose 2 was observed (15% of cases) and gave a slightly better docking score than Pose 1.

In Pose 1 (Figure 2), the OA carboxyl group (O1, O2) points to the same direction as the GSPP pyrophosphate group and forms polar contacts with Y216 (3.0 and 3.9 Å; all distances mentioned henceforth are nearest heavy atom−heavy atom distances) and K284 (4.8 and 5.3 Å). O3 and O4 hydroxyl groups form polar contacts with S51 (3.0 Å) and Y288 (3.3 Å), respectively. O4 hydroxyl is also in the proximity of V49 (3.6 Å). The pentyl moiety points away from the polar side chains and toward hydrophobic residues V271 and A232. The C−C distance where the bond is to be formed between OA C3 and GSPP C1 is 3.9 Å. Pose 2 (Figure 2) is similar to Pose 1, with carboxyl group and O4 hydroxyl flipped, and the C−C prenyl bond distance is 4.5 Å. As Pose 1 is the dominant pose, we will only discuss Pose 1 in the following sections unless specifically mentioned. The best docking scores of the variants are shown in Figure 3.

Comparing our procedure with those in previous studies on the rational mutagenesis of NphB,[8,10] all computations started with the placement of OA in the aromatic binding site of the crystal structure of WT NphB (PDB 1ZB6[6]) in the presence of GSPP. Qian *et al.* used Autodock to dock orsellinic acid (OSA), an OA analog with a one-carbon instead of five-carbon alkyl chain, and identified crucial residues G286 and G273—with no further rotamer sampling—and found a beneficial G286S mutation. When OA was used as the substrate in their study, G286S produced CBGA with a yield about 2.5× that of WT. The regioselectivity of the variant G286S could not be determined as the 2-O-GOA side product yield was not reported; instead, the authors identified the main side product that was observed as 5-geranylated OLA.

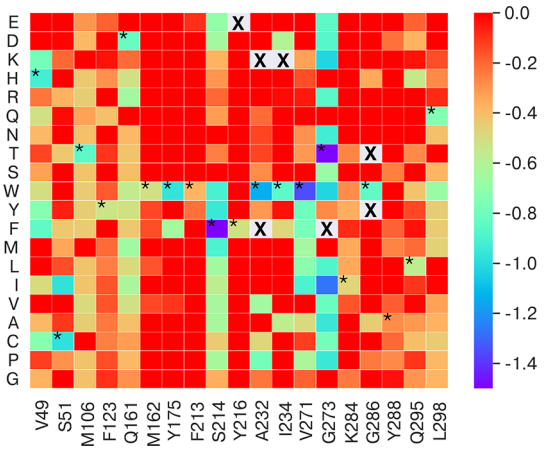


Figure 3. Heatmap of OA best docking scores from 361 NphB variants resulting from 19 mutation positions. Docking scores (kcal/mol) are shown relative to that of WT (0.00 is equal to, or worse than, the WT score). "X" indicates that docking did not yield any viable pose. "*" indicates the best dock score in each position.

Valliere *et al.* used the Rosetta design to simultaneously and stochastically sample side chain rotamers of 10 residues around manually placed OA, resulting in 22 variants in the first round, which include only two single variants (Y288A and Y288N). The best-performing double variants Y288A/G286S and Y288V/A232S were part of the more focused second library, which was rationally constructed after the observations from the testing of the first library. k_{cat} for the two variants was reported to be about 1000-fold that of WT. For the rest of the constructs, they exhibited less CBGA yield compared to these two and their 2-O-GOA yield was not reported. Additionally, the mechanistic advantage of these variants was not explored in their work as this was not their main focus.

In this work, we sampled side chain rotamers of variants at 19 residue positions, at one residue position and one variant at a time, around docked OA. OA was redocked at each variant, and for each position, the variant with the best GLIDE XP score was subjected to experimental testing. Subsequently, the best-performing single variants were then investigated for their synergistic activity of CBGA production.

Experimental Validation of Identified Mutations. The computationally predicted best variants from each of the 19 positions were then expressed and experimentally assayed for their CBGA/2-O-GOA yields. WT NphB produced CBGA and 2-O-GOA in roughly equal amounts.[5] From these 19 single variants, three stood out in the desired yield and regioselectivity, namely, V49H, Y288A, and Q295L (Figure 4), as they exclusively produced CBGA in 3.4-, 7.4-, and 7.1-fold yield improvement, respectively, compared to WT. Fractional yields of CBGA and 2-O-GOA and their fold changes with reference to those of the WT for all the variants in Figure 4 can be found in Table S1. All three variants were highly regioselective with nondetectable amounts of side product. Q161D, I234W, and L298Q exhibited only marginal improvement of CBGA yield over WT without improving regioselectivity as significant amounts of 2-O-GOA side product were also produced. We noted that Q161A was previously reported as a beneficial mutation in a patent;[12] however, the value of yield improvement was not reported. S51C, M106T, F123Y, F213W, and I234W produced more 2-O-GOA than the desired CBGA. The rest of the variants either produced CBGA

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

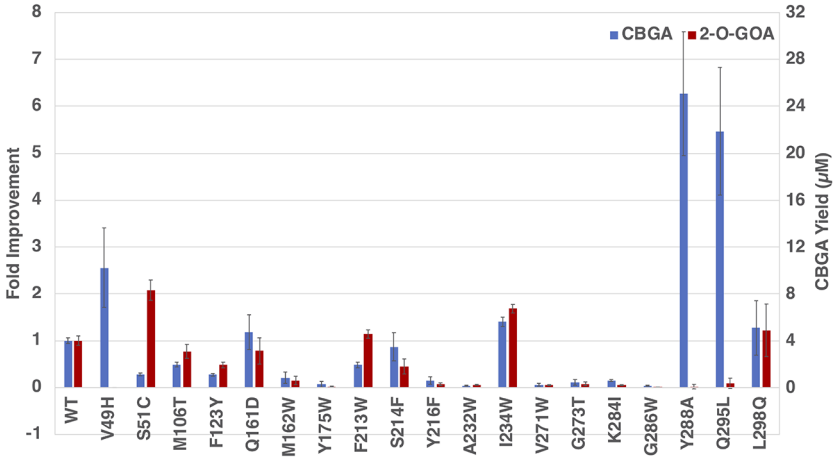


Figure 4. Fold improvement of CBGA and 2-O-GOA yields of NphB variants normalized to WT NphB. CBGA yields are shown on the right axis. Error bars are the standard deviation from triplicates.

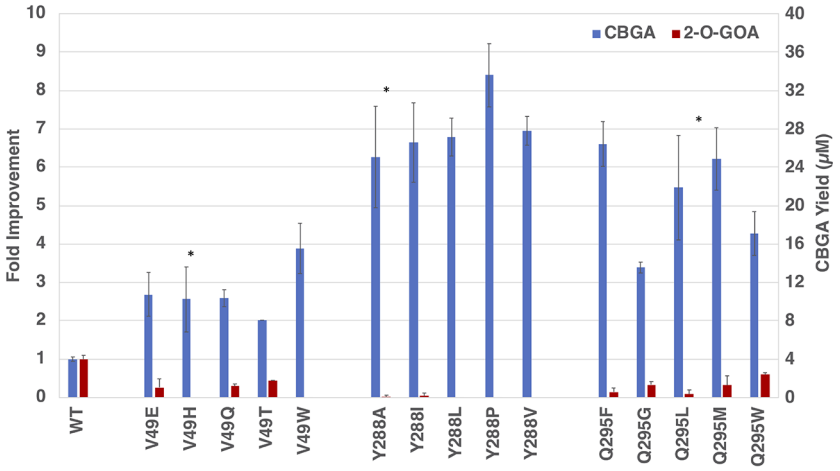


Figure 5. Fold improvement of CBGA and 2-O-GOA of the top five variants at positions V49, Y288, and Q295 normalized to WT NphB. CBGA yields are shown on the right axis. The error bar is the standard deviation from triplicates. The asterisk (*) denotes the computationally predicted best single variant.

at a lower yield than that of WT or were enzymatically inactive. Considering only single variants, besides Y288A as a favorable variant in agreement with Valliere *et al.*,[8] we have uncovered two new favorable single variants, V49H and Q295L, in this study.

Targeted Mutant Library Generation and Characterization. The three favorable variants, V49H, Y288A, and Q295L, evidently suggested that these three positions are important for the prenylation reaction and the corresponding mutations favor a CBGA-dominant product profile (refer to Figure 2 for active site residues). For proximity reference, the closest heavy atom−heavy atom distances of WT-docked OA to the side chains of V49, Y288, and Q295 are 3.6, 3.3, and 7.3 Å, respectively. To understand the mechanistic roles of the mutations at these positions and evaluate the performance of our computational workflow in capturing favorable mutations, we decided to exhaustively characterize all possible mutations at these three positions to see if better-performing variants could be uncovered (Figures S1−S3 and Table S2).

Indeed, variants with higher CBGA yields were discovered in all three sites. Figure 5 shows the top five CBGA-yielding mutations from each of these three positions from experimental assays. Each round of experimental testing was carried out separately using NphB-WT as the internal reference. The variants that produced the highest CBGA yield from these three positions are V49W, Y288P, and Q295F, producing 3.9-, 8.4-, and 6.6-fold yield improvement, respectively, compared to WT. It is worth noting that the computationally predicted variants (denoted with * in Figure 5) from these three positions (V49H, Y288A, and Q295L, which produced 2.6-, 6.3-, and 5.5-fold improvement, respectively) were within the top five CBGA-yielding variants out of all 19 mutations in each position, which demonstrated the accuracy of our computational predictions. Fractional yields and their fold differences of the two products reported in Figure 5 can be found in Table S3.

Combining Favorable Mutations. Having obtained high CBGA-yielding and regioselective single variants, we proceeded to combine these mutations to obtain double and triple variants to see if the beneficial effects of these mutations were synergistic. Parallel to the experimental effort, we also attempted a repeat of our computational step for a favorable double variant prediction. By performing a simultaneous rotamer search at two of the three positions identified (computational data not shown), we arrived at the computa-

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

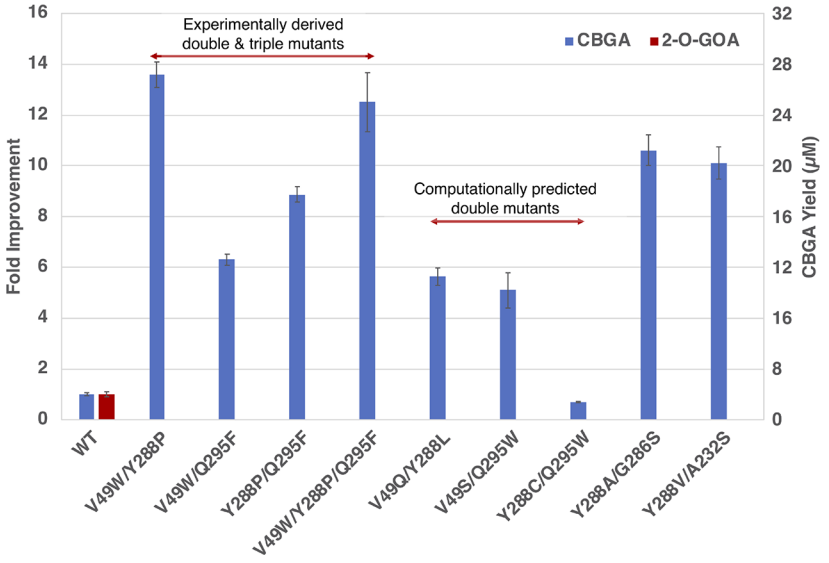


Figure 6. Fold improvement of CBGA and side-product yields of NphB double and triple variants. CBGA yields are shown on the right axis. The error bar is the standard deviation from triplicates.

tionally predicted combinations of V49Q/Y288L, V49S/Q295W, and Y288C/Q295W.

We tested all experimentally derived and computationally predicted double variants in a single run with triplicates along with WT NphB as the internal standard. Of the three double variants derived from combining the experimentally determined best single variants (V49W, Y288P, and Q295F), the V49W/Y288P double variant produced the highest CBGA yield (13.6-fold improvement over WT) as shown in Figure 6. The triple variant V49W/Y288P/Q295F (12.5-fold improvement over WT), generated by adding Q295F, did not improve CBGA yield further with respect to the V49W/Y288P double variant. The other two double variants V49W/Q295F and Y288P/Q295F (6.3- and 8.9-fold improvement over WT, respectively) only performed modestly in terms of the CBGA yield.

Compared to the above-mentioned experimentally derived double variants, the computationally predicted double variants did not produce a much higher CBGA yield (Figure 6 and Table S4). The computationally predicted double variants V49Q/Y288L, V49S/Q295W, and Y288C/Q295W produced 5.6-, 5.1-, and 0.7-fold change in CBGA, yield respectively, when compared to WT, all of which were lower than the CBGA yields from the experimentally derived double/triple variants. This highlights a limitation in our computational prediction workflow in predicting the performance of multiple mutations, as their active site conformations will have more deviations than single variants. Notably, all computationally predicted double variants were also highly regioselective in producing CBGA exclusively, which reinforce the notion that mutations in these three positions were key to the regioselectivity preference for CBGA.

Overall, the best-performing variant in this work, V49W/Y288P, notably has a higher CBGA yield (13.6-fold improvement over WT) when compared to the best variants from Valliere's Y288A/G286S (10.6-fold) and Y288V/A232S (10.1-fold)[8] when tested under the same experimental conditions (Figure 6). Finally, all double and triple mutant variants were found to be highly regioselective with a negligible side product.

Structural Insights from X-ray Structures of Enhanced NphB Variants. To understand how these mutations affect regioselectivity and product yield of NphB, we obtained structures of NphB variants by X-ray crystallography. Here, we have experimentally determined the structure of GPP-bound WT NphB and the non-ligand-bound structures of three single (V49W, Y288P, and Q295F) and one triple (V49W/Y288P/Q295F) variants (Table 1). For ease of reference, we will refer to the different structures with the following shorthand notations: WT: WT, VW: V49W, YP: Y288P, QF: Q295F, and YP3: V49W/Y288P/Q295F. Overall, our structures are quite similar to the published ligand-bound WT NphB structure (PDB 1ZB6), with RMSDs for all nonhydrogen atoms ranging from 0.5 to 1.0 Å. The identity and orientation of key residues at the active site are supported by unambiguous electron densities even for the medium-resolution structure for YP (Figure S10). The NphB structure comprises a cylindrical 10-strand β-barrel (β_1–β_{10}) with a solvent-filled core surrounded by solvent-exposed α-helices (α_1–α_{14}). A C-terminal α-helix forms a lid (α_{14}) covering the active site. The main variation observed among the different structures obtained is found in this C-terminal lid (residues 291–end), as well as two loops adjacent to it (Figures S4 and S11). The first loop is pre-β_1 (residues 42–47), while the second loop is pre-β_9 (residues 260–269). The remaining secondary structures and loops displayed little deviation. The variation also accounts for the differences between the actual structures and the docking models derived through virtual mutagenesis of the published wild-type structure. Since the disordered C-terminal lid in YP and QF could well become ordered upon ligand binding and adopt a similar conformation as predicted by the docking model, whereas the two loops have no direct involvement in ligand binding, the docking models remain as reliable surrogates for the NphB single variants with newly determined structures.

The V49 position is situated close to the ligands (3.6 Å to OA). Given the neighboring bulky Y288 (3.3 Å to OA), steric constraints are expected in accommodating the mutated tryptophan side chain. Our V49W structure revealed that the newly introduced W49 adopts a conformation in which its side

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628–4639

Table 1. X-ray Crystallographic Data Collection, Refinement, and Validation Statistics[a]

	WT + GPP	VW	YP	QF	YP3
PDB ID	7FHB	7FHC	7FHD	7FHE	7FHF
			Data collection		
wavelength (Å)	1.00000	0.95373	0.95373	0.95373	0.95373
resolution range (Å)	29.4−1.90 (1.97−1.90)	39.4−2.06 (2.13−2.06)	28.8−3.50 (3.63−3.50)	29.5−2.50 (2.59−2.50)	27.3−2.50 (2.59−2.50)
space group	*P* 21 21 21	*P* 21 21 21	*P* 1	*P* 21 21 21	*C* 1 2 1
unit cell (a, b, c [Å]; α, β, γ [°])	44.46, 78.36, 84.99; 90, 90, 90	44.34, 78.75, 85.60; 90, 90, 90	44.29, 75.16, 87.61; 89.94, 89.90, 90.00	43.83, 75.96, 79.51; 90, 90, 90	89.32, 44.00, 87.99; 90, 111.62, 90
total reflections	172,136 (17,089)	258,583 (24,574)	48,285 (5059)	125,091 (12,750)	78,273 (7811)
unique reflections	24,026 (1376)	19,182 (1879)	14,122 (1435)	9655 (957)	11,218 (1119)
multiplicity	7.2 (7.2)	13.5 (13.1)	3.4 (3.5)	13.0 (13.3)	7.0 (7.0)
completeness (%)	89.74 (58.18)	99.96 (99.95)	95.75 (98.02)	99.70 (99.79)	99.82 (99.91)
mean *I*/sigma(*I*)	18.18 (3.51)	9.19 (1.81)	2.30 (1.20)	15.44 (2.52)	7.95 (2.08)
Wilson B-factor (Å^2)	20.79	21.33	41.29	38.74	29.10
R-merge	0.086 (0.662)	0.252 (1.244)	0.597 (0.856)	0.156 (0.991)	0.207 (0.865)
R-meas	0.093 (0.714)	0.262 (1.295)	0.708 (1.008)	0.162 (1.031)	0.224 (0.935)
R-pim	0.035 (0.266)	0.071 (0.356)	0.377 (0.530)	0.044 (0.281)	0.084 (0.351)
CC1/2	0.996 (0.803)	0.995 (0.632)	0.392 (0.489)	0.998 (0.889)	0.989 (0.833)
CC*	0.999 (0.944)	0.999 (0.880)	0.750 (0.811)	1.000 (0.970)	0.997 (0.953)
			Refinement		
reflections used in refinement	21,627 (1376)	19,178 (1879)	13,648 (1433)	9634 (955)	11,209 (1118)
reflections used for *R*-free	1090 (64)	983 (97)	764 (76)	444 (52)	562 (70)
R-work	0.182 (0.231)	0.191 (0.223)	0.241 (0.239)	0.217 (0.300)	0.186 (0.259)
R-free	0.219 (0.326)	0.233 (0.289)	0.312 (0.283)	0.283 (0.364)	0.259 (0.354)
CC(work)	0.943 (0.862)	0.948 (0.776)	0.834 (0.817)	0.958 (0.843)	0.957 (0.908)
CC(free)	0.935 (0.793)	0.940 (0.612)	0.670 (0.856)	0.919 (0.826)	0.876 (0.860)
number of nonhydrogen atoms	2669	2588	8673	2271	2460
macromolecules	2362	2369	8673	2266	2347
ligands	19	0	0	0	0
solvent	288	219	0	5	113
protein residues	306	306	1121	293	303
RMS (bonds, Å)	0.008	0.007	0.005	0.008	0.007
RMS (angles, °)	1.05	0.93	0.91	1.11	0.87
			Validation		
Ramachandran favored (%)	98.68	98.03	96.99	96.89	94.98
Ramachandran allowed (%)	1.32	1.97	2.92	3.11	5.02
Ramachandran outliers (%)	0.00	0.00	0.09	0.00	0.00
rotamer outliers (%)	0.39	0.78	3.29	2.03	1.98
clashscore	2.97	2.99	12.00	7.13	4.30
average B-factor (Å^2)	20.76	22.02	29.07	42.23	32.60
macromolecules	19.61	21.26	29.07	42.23	32.54
ligands	31.38				
solvent	29.54	30.25		41.09	33.80

[a]Statistics for the highest-resolution shell are shown in parentheses.

chain forms almost parallel $\pi-\pi$ stacking with that of Y288 while partially occupying the space taken by the geranyl moiety of GSPP and 1,6-DHN in the WT NphB structure. Consequently, the GSPP docking pose in V49W is tilted toward the right side with respect to that in WT (Figure 7), without a significant shift in the Y288 side chain. A few reference views of WT NphB containing GSPP and docked OA (Pose 1) ligands in different rotational angles are depicted in Figure S5.

Presumably, C-prenylation of OA follows an S_N1-like dissociative mechanism featuring a weakly stable carbocation, similar to what has been proposed for that of 1,6-DHN.[13] In line with this presumption, V49W would provide a better shielded cation-π chamber for the carbocation, effectively lowering the energy barrier for its formation. Among the better-performing variants at position 49 are mutations to aromatic residues such as our computationally predicted V49H, which likely adopts a similar side chain conformation and exerts a similar effect as V49W. Correspondingly, the introduction of a negative charge by the V49E mutation or an electron-rich oxygen atom by V49Q to the vicinity of the active site may exert a similar stabilizing effect on the carbocation as the π chamber. In stark contrast to these substitutions, introducing a positively charged residue, Arg or Lys, at this or one of the other two exhaustively characterized positions (288 and 295) greatly reduces NphB enzymatic activity (Figure S1),

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

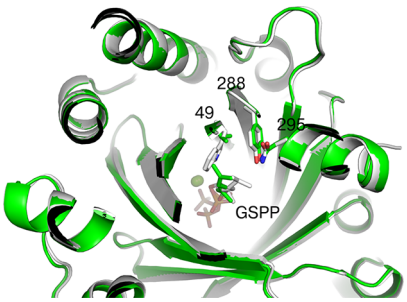


Figure 7. Superimposed structures of WT NphB (PDB 1ZB6, green) and variant V49W (white). GSPP ligand in WT is from the crystal structure (PDB 1ZB6); GSPP ligand in V49W is docked in.

further supporting the importance of carbocation stabilization to NphB-catalyzed prenylation.

Y288 is a residue on the 10th β-strand of NphB, with its side chain pointing toward the active site. Q295 resides in the C-terminal lid region, with its side chain ~4 Å away from that of Y288. Our experimentally determined single variant structures of Y288P and Q295F reveal the interplay of these two residues in defining the mobility of the lid. With either mutation, the lid becomes disordered and unresolvable in the crystal structure. For Y288P, this seems to be partly due to the loss of contacts with Q295, which likely is applicable to other Y288 substitutions with small side chain residues such as Y288A. For Q295F, the steric clash caused by the bulky side chain of Phe might be responsible for the release of the lid. Simultaneous mutations at these two positions, however, can fully restore the ordering of the lid, as seen in the crystal structure of the triple variant V49W/Y288P/Q295F. Comparing V49W/Y288P/Q295F with WT, we noticed that the lid, in the form of an α-helix in both structures, has a higher mobility according to its normalized B-factor (Figure S11) and shifts slightly in the triple variant. This shift is lateral, bringing F295 in closer contact with P288 and the GPP binding site, suggesting that the lid acts as a barrier for the otherwise solvent-exposed hydrophobic β-barrel (Figure 8). Further discussion on the structure–activity relationship of the mutations will be individually addressed in later sections with MD simulation results.

With the replacement of Gln by a more hydrophobic Phe side chain, the Q295F mutation would offer stronger hydrophobic interactions with the ligands and certain key hydrophobic/aromatic residues surrounding the ligands in the β-barrel active site, such as those at position 288, which may favor tighter ligand binding but at the same time disfavor product release. Since this mutation also has direct effects on the flexibility and movement of the lid, it seems plausible to anticipate dynamic factors at play for Q295F, which will affect the rate of ligand entry and exit. This hypothesis is supported by the kinetics for the double variant V49W/Y288P and the triple variants with the additional Q295F mutation, with a slight improvement in catalytic efficiency being observed for the triple variant. The k_{cat}'s of our double and triple variants were 280- and 337-fold higher than that of WT NphB (Table S5) while being about 1.5- and 1.8-fold higher than Valliere's Y288V/A232S variant, respectively, when tested under the same experimental conditions.

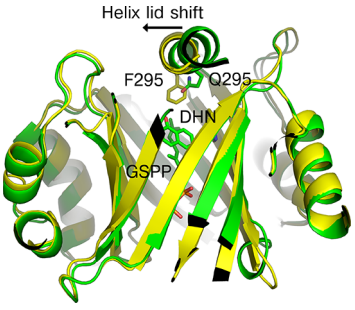


Figure 8. Superimposed structures of WT NphB (PDB 1ZB6, green) and variant V49W/Y288P/Q295F (yellow). GSPP and 1,6-DHN ligands are from WT NphB.

Structural and Dynamic Insights from MD Simulations of Enhanced Variants. In the absence of crystal structures with bound GSPP and OA, we sequentially redocked GSPP and OA to our resolved non-ligand-bound structures and ran MD simulations on them (VW, YP, QF and YP3; PDB 7FHC, 7FHD, 7FHE, and 7FHF, respectively) together with the docked WT structure (PDB 1ZB6) to gain further insights on OA binding and activity (Figure S6). The resulting OA docking poses resemble what we observed in our initial docking against WT structure. Pose 1 had the best docking scores in 85% cases and was ranked next best when Pose 2 was scored best in 15% cases. Hence, Pose 1 of OA was used in starting structures. GSPP was kept as is in the initial molecular dockings to preserve the X-ray crystal coordinates, but GPP was used for MD simulations to more accurately reflect the systems in experimental conditions.

From the analysis of all MD simulation results, two key observations on (1) the potential hydrogen bond (H-bond) networks of OA with residues in the active site and (2) the relative OA binding orientation for catalysis can be used to explain why the NphB variants we have uncovered in this study favor C-prenylation to form CBGA with respect to WT. Some additional properties were also evaluated from MD trajectories, such as the relative positions and orientations of GPP and OA represented in Figure S9. However, no significant differences in these parameters were observed between WT NphB and the regioselective variants. Here, we discuss the key H-bond networks of OA and the relative OA binding orientation in detail.

The H-Bond Networks of OA. In the initial docking Pose 1, a few residues in the NphB active site potentially make polar contacts, specifically hydrogen bonds (H-bonds), involving hydroxyl and carboxyl moieties of OA. We tracked all H-bond occupancies, i.e., percentage of all the MD snapshots within which the H-bond donor–acceptor pair was engaged, in the MD trajectories (Figure S7). Here, we discuss selected H-bond occupancies with implications shown in Figure 9. For OA O4, the WT Y288 hydroxyl group is the nearest H-bond partner

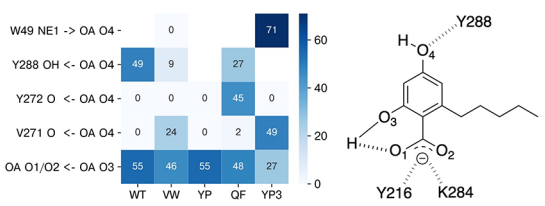


Figure 9. Selected H-bond occupancies (%) involving OA in MD systems. Arrows indicate donor -> acceptor H-bond directionality. Atom naming follows the PDB nomenclature (Y288 OH: hydroxyl O; O: backbone amide O; N: backbone amide N; W49 NE1: indole N).

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628–4639

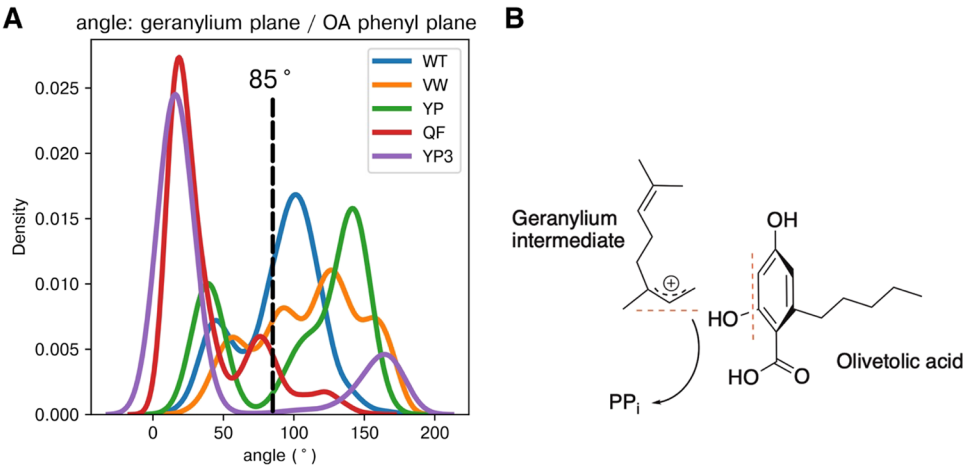


Figure 10. (A) Distributions of the angle between the geranylium carbocation plane and OA aromatic plane in all MD systems. The annotated dotted line is the initial value in WT. (B) Pictorial representation of the geranylium carbocation intermediate and olivetolic acid in WT. The red dotted line denotes the planes used to calculate angle.

and indeed is observed with 49% occupancy in WT. In VW, this H-bond is largely lost (only 9% occupancy) as OA is repelled by the bulky side chain in VW and tends to traverse deeper into the binding site during the simulation. In QF, the occupancy is also much reduced to 27% (Figure 9). Consequently, O4 hydroxyl forms a H-bond with the backbone carbonyl groups in V271 (24% in VW) or Y272 (45% in QF), both located in β_9 of the β-barrel. For YP and YP3, the Y288-O4 H-bond is eliminated due to the Y288P mutation. Taken together, the disruption of the Y288-OA H-bond interaction appears to be consistently observed in all variants that favor C-prenylation forming CBGA. Yang *et al.*, performing calculations on the WT NphB system with another aromatic substrate, 1,6-DHN, have proposed that Y288 helps stabilize certain substrate orientations via H-bonds and affects the product exit.[13] Similarly, Y288 seems to stabilize unfavorable orientation for CBGA production. Hence, mutating it away or disrupting its H-bond interactions with OA favors CBGA production. This is also supported by the yields of other Y288 variants. Y288F, where the only difference with WT is the absence of the hydroxyl moiety, saw an approximately 5-fold CBGA yield increase and no 2-O-GOA production (Figure S2). Most favorable mutations in producing CBGA in the Y288 position such as Pro, Ala, Ile, Leu, and Val (Figure 5) have side chains that abolished the ability of residue 288 to form a H-bond. Conversely, the only mutations at the Y288 position to form the 2-O-GOA side product (His and Gln shown in Figure S2) have polar NH groups with a similar length to form a H-bond. Besides residue 288, the indole moiety at residue 49 also forms a H-bond with O4 hydroxyl in YP3 (71%), possibly also contributing to CBGA selective production.

The variant W49 side chain, apart from presenting as a bulky hydrophobic side chain and sterically repelling OA deeper in the active site, possibly conferring more favorable kinetics, also provides a steric constraint to Y288 by $\pi-\pi$ stacking interaction, restricting its conformation. Following the earlier reasoning, this restricted range might have ruled out the Y288 conformations that favor 2-O-GOA as the product. This speculation is consistent with the observation that mutations of residue 49 to aromatic side chains His, Tyr, and Phe also selectively produce CBGA (Figure S1).

The Relative Orientation of Ligands for Catalysis. Since C-prenylation for CBGA production is presumed to occur via a carbocation intermediate, we also measured whether the to-be-formed geranylium carbocation plane is oriented favorably with respect to the OA aromatic plane, assuming that no drastic spatial rearrangement occurs to the geranylium intermediate after its detachment from the pyrophosphate leaving group. All variants indeed show a deviation of this angle compared to WT, where it is close to perpendicular (Figure 10A). Perpendicular orientation between the ligands is presumably not energetically favorable for C-prenylation since the respective sp^2 orbitals will not be favorably aligned (Figure 10B). Hence, the deviation from the perpendicular orientation that favors C-prenylation is consistent with their observed regioselectivity in all variant systems. Repulsion to OA by the bulkier Trp side chain in VW and YP3 could partly account for the deviation observed in these variants. However, it is not immediately clear why this deviation also occurs in YP and QF.

Further Computational Iteration: Machine Learning Model. Our first computational iteration yielded docking scores. While it was useful to guide the experimental efforts by prioritizing certain variants, it had arguably limited predictive power. Given that we had accumulated approximately 90 experimental CBGA yield values, we built a machine learning model as the next computational iteration.

From the 19 mutation sites, we had measured the CBGA yields of 93 single mutant variants, of which sites V49, Y288, and Q295 were saturated. An optimized, cross-validated regression model from a training set of 74 predicted correctly 17 out of 19 in the test set as having worse or better yield than WT (89.5% accuracy) (Figure S12A). None of the mutants in the 19 mutations sites were predicted to outperform the best single mutant variants in the previous three saturated sites V49W, Y288P, and Q295F (3.9-, 8.4-, and 6.6-fold CBGA yield compared to WT, respectively).

We then produced and measured CBGA yield for another 40 single mutant variants coming from sites S214, G273, and M162 (Table S6). Predicting for these 40 new variants, the model performed poorly when evaluated with the regression metric, coefficient of determination (R^2), but when evaluated with the classification metric, the model still predicted 32 out of 40 correctly (80% accuracy) (Figure S12B).

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

We then updated the model with the new data such that, out of 133, the updated model trained on 106 in the training set, predicting correctly 25 out 27 in the test set (93% accuracy) (Figure S12C). The updated model still did not predict any single mutant variant in the 19 sites to outperform the best single mutant variants in the previous three saturated sites. For both models, it is notable and intuitive that the site position and docking score are among the three most important features (Figure S12A,C).

While this machine learning model has limitations due to the small data set size, it provides more assurance that we are not likely to miss any other high-yielding single mutant variants in the mutation space. Retrospectively, we can also conclude that the first computational docking iteration, with its limitations, was useful enough to guide the subsequent iterations by identifying the mutation sites more likely to be crucial. The decision tree-based machine learning model also affirms the docking score to be one of the most important features to predict CBGA yield.

CONCLUSIONS

Enzyme engineering is a powerful tool to improve regioselectivity and yield. However, without a fast and accurate selection method to identify favorable mutations, it remains an arduous task to experimentally screen for the entire sequence space to look for the optimal mutations. This study has demonstrated the utility of a structure-guided modeling-docking approach in predicting mutations that will favor the desired reaction catalysis prior to experimental testing, thus significantly reducing the sequence space for experimental verification. This has allowed us to successfully arrive at a few highly regioselective and high CBGA-yielding NphB variants that will be useful in overcoming the bottleneck of the prenylation reaction in the heterologous reconstitution of the cannabinoid pathway.

To investigate the mechanistic roles of the mutations, we subsequently resolved X-ray crystal structures, performed kinetic assays, and performed MD simulations of WT and these variants that favor C-prenylation to form CBGA. This led to the uncovering of a complex interplay of different residues affecting the H-bonding network and orientation of the OA ligand that favors C-prenylation with respect to the geranylium carbocation intermediate. Apart from the implications in cannabinoid synthetic biology, understanding these mechanistic insights will facilitate future protein engineering efforts for prenyltransferase catalysis.

METHODS

Computational Mutagenesis. The computational mutagenesis includes the following steps. First, we docked olivetolic acid (OA) in the 1,6-DHN site of the ligand-bound structure of NphB with GSPP and 1,6-DHN (PDB 1ZB6[6]) with the Schrödinger 2018-4 GLIDE extra precision (XP).[14] Second, we took the best-scoring pose (denoted as Pose 1), amenable to the prenylation reaction mechanism, as the initial template for virtual mutagenesis with SCWRL4[15] and PLOP,[16] which perform side chain rotamer search. Nineteen residues having close contact with docked OA were mutated to every other possible amino acid side chain in the presence of OA, with the protein backbone fixed. Third, OA was then redocked to the generated mutant library and poses were evaluated. Most variants yielded Pose 1 as the best docking pose, but some yielded another pose (denoted as Pose 2), which is also amenable to the prenylation reaction but was not observed in the docking of OA to WT NphB.

For docking to newly resolved non-ligand-bound variant crystal structures, GSPP and OA from WT docking were placed in the aligned variant structures. GSPP and OA were then sequentially redocked: after GSPP redocking, the best-scoring GSPP pose was selected and the complex was used for OA docking.

Molecular Dynamics (MD) Simulation. The docking-generated NphB WT and variant structures in complex with GSPP and OA were used to start MD simulations after GSPP was changed to GPP. Missing loops and lid were added using Modeller.[17] Each system was prepared with CHARMM-GUI,[18,19] with solvation in a periodic 0.15 NaCl TIP3P water box with a minimum water height of 10 Å on each side. The protein was modeled with the CHARMM protein force field and ligands with the CGenFF force field.[20] All simulations were performed on a GPU cluster within GROMACS 2018.2.[21] The constructed system was first energy minimized for 5000 steps using the steepest descent method. After 125 ps NVT equilibration, 200 ns production simulations were performed at 298.15 K and 1 bar in the NPT ensemble for two replicates of each system. The particle mesh Ewald (PME) method[22] was used to treat all electrostatic interactions beyond a cutoff of 9 Å. The LINCS algorithm[23] was used for recording the length of bonds involving hydrogen during the simulation with an integration time step of 2 fs. Snapshots from each trajectory were saved every 100 ps during the production runs, and these trajectories were used for analysis with MDTraj.[24] Distribution plots are the kernel density estimation of snapshots from two MD replicates calculated by the Python seaborn package.[25] H-bond cutoffs are D-A distance <3.5 Å and D-H-A angle >150°. The angle between geranylium and OA phenyl planes is calculated by measuring the angle between the normals of the two planes. Geranylium plane is defined by the positions of GPP atoms C1, C2, and C3, and OA phenyl plane is defined by the positions of OA atoms C2, C4, and C6.

Machine Learning. The label vector was defined as CBGA fold-yield of a single mutant variant relative to WT. The mutation side chain was represented as a 17-dimensional amino-acid encoding vector.[26] The other features were site number (categorical), Glide XP docking score, other selected Glide outputs (e.g., distance of OA to GSPP), pose rescoring from other scoring functions, binding pocket volume, and surface area, totaling 69 features. Out of 93 data points, 74 were used in the training set and 19 in the test set. The 74 × 69-dimensional training matrix was used to train a machine learning model based on the XGBoost regressor algorithm as implemented in the Python package xgboost.[27] Cross-validations and hyperparameter optimization were done over five splits and model performance was evaluated on the coefficient of determination (R^2), achieving an out of fold score of 0.690. To evaluate with classification metrics, the CBGA yield of 1× WT was taken as the threshold to categorize the yield value as better or worse than WT. CBGA yield was predicted correctly to be better or worse than that of WT in 17 out of 19 variants in the test set (accuracy 89.5%).

The updated model was trained on a 106 × 69-dimensional training matrix. Cross-validations and hyperparameter optimization were done over five splits and the updated model performance was evaluated on the coefficient of determination

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

(R^2), achieving an out of fold score of 0.543. CBGA yield was predicted correctly to be better or worse than that of WT in 25 out of 27 variants in the test set (accuracy 93%). Members of train and test sets for the two models are listed in Table S7.

Cloning and Site-Directed Mutagenesis. To generate the mutants, site-directed mutagenesis was conducted using a polymerase chain reaction (PCR)-based method. Fragments were amplified using primers that carry the necessary mutation using the NphB wild-type in a pET-20b vector with the pelB tag removed as the template and PrimeSTAR Max DNA polymerase (Takara). The resulting fragments were then assembled using the NEBuilder HiFi DNA assembly master mix. The assembled product was then transformed into *E. coli* XL1-Blue cells (Stratagene) using heat-shock. Successfully transformed clones were picked and sequenced to check that the intended mutation has been effected.

Protein Expression and Purification. Protein expression was carried out using *E. coli* BL21 (DE3) cells (Novagen). Successful transformants were grown in 50 mL of LB broth with 100 μg/mL ampicillin to an OD600 nm absorbance of between 0.4 and 0.8 units before adding 0.2 mM isopropyl D-thiogalactopyranoside (IPTG) for induction at 20 °C for 16 h. Cell pellets were harvested by centrifugation and resuspended using 6 mL of the binding buffer (20 mM Tris−HCl (pH 7.9), 0.5 M NaCl, and 5 mM imidazole). Cell lysis was conducted using sonication, and the lysate was spun down to remove cellular debris. Purification of the His-tagged protein was conducted using CFS Protemist by loading the supernatant from the cell lysate. The eluent was dialyzed into a buffer containing 20 mM Tris−HCl (pH 8.0) and 100 mM NaCl and concentrated using the Amicon Ultra 0.5 mL centrifugal filters with a 3 kDa pore size.

Enzymatic Assays and Detection. Enzymatic assays to test the CBGA and side product yield of the variants were conducted using 100 mM Tris−HCl (pH 7.5), 10 mM $MgCl_2$, 2 mM olivetolic acid, 1 mM geranyl pyrophosphate, and 5.93 μM purified protein. Reactions were conducted in an end-point assay incubated at 30 °C for 18 h. The NphB wild-type was included in every round of testing as an internal reference. Samples were acidified after incubation using hydrochloric acid, and organic extraction was conducted three times using ethyl acetate in each round. The organic extract was then dried using a rotary vacuum concentrator and dissolved in methanol.

Detection was carried out using an Agilent 1290 Infinity HPLC coupled with an Agilent 6550 iFunnel Q-TOF mass spectrometer in the negative mode. The sample was separated using an Agilent InfinityLab Poroshell 120 C18 UHPLC column using 5 mM ammonium formate and acetonitrile as the mobile phases. The peak areas of both CBGA and side product of each construct were analyzed using an extracted ion count of 359.2228 m/z (corresponding to CBGA and side product mass as negative ions) and normalized using wild-type NphB as an internal reference. Fold-difference for similar variants across batches were averaged before comparison.

Kinetic Assays. Kinetic assays were performed using the following conditions: 50 mM Tris−HCl (pH 8.0), 5 mM $MgCl_2$, 2 mM GPP, and a range of olivetolic acid concentrations varying from 10 μM to 5 mM. The enzyme concentration used was varied for the different enzymatic constructs depending on catalytic efficiency and detection window. Five micromolars of the enzyme was used for all WT NphB reactions, 50 nM was used for the double variants, and 10 nM was used for the triple variants. Forty microliters of the reaction was incubated for 1, 3, 5, 10, 15, and 20 min at ambient temperature and subsequently quenched with 80 μL of acetonitrile and 0.1% formic acid. The quenched reaction mixture was then centrifuged at 21,000g for 5 min and analyzed by LC−MS using an Agilent LC-Triple-Quad 6495 system with an MRM profile looking for transitions of CBGA and CBGVA product ions. The Michaelis−Menten curve was fitted in GraphPad Prism using initial rates of product formation observed to obtain the kinetic parameters.

X-ray Crystallography. Crystals of NphB WT and selected variants were obtained by using the NT8 drop setter for protein crystallization (Formulatrix) to mix 100 nL of the protein stock solution (10−15 mg/mL) with 100 nL of the precipitant in 96-well 2-drop MRC sitting-drop vapor diffusion plates (Swissci) and incubating the plates at 20 °C. Before flash freezing, crystals were cryoprotected with the same precipitant supplemented with 30% glycerol. For the WT, crystals were soaked in the cryoprotectant supplemented with 1.0 mM GPP for 10 min before freezing. Diffraction data were collected at synchrotron beamlines in Australia (MX1[28] and MX2,[29] part of ANSTO) and Taiwan (13B1, NSRRC) and were processed using XDS[30] and HKL-2000.[31] To solve the structure, molecular replacement was carried out using the published NphB WT structure (PDB 1ZB6) as the search model in PHASER.[32] The solutions generated by PHASER were refined in REFMAC[33] and Phenix[34] and manually adjusted in Coot.[35] To minimize potential model bias, built models were cross-checked against their iterative-build OMIT maps.[36] MolProbity[37] was employed to monitor model quality. Pymol[38] was used for structure visualization and figure generation.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.2c00786.

Fold improvement of saturation variants at key positions (Figures S1−S3); reference views and composite omit maps of the experimentally determined structures (Figures S4−S5 and S10); additional results from the MD simulation experiments (Figures S6−S9); normalized B-factors (Figure S11); results from the machine learning models (Figure S12); fractions and fold differences of products produced for variants in the work (Table S1−S4 and S6); kinetic parameters of selected variants (Table S5); and variants in train and test sets of the machine learning models (Table S7) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Hao Fan − *Synthetic Biology for Clinical and Technological Innovation, National University of Singapore, Singapore 117456; Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599; Bioinformatics Institute, Singapore 138671;* orcid.org/0000-0003-0199-9752; Email: fanh@bii.a-star.edu.sg

Wen Shan Yew − *Synthetic Biology for Clinical and Technological Innovation, National University of Singapore, Singapore 117456; Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599; Department of*

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639

Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597; orcid.org/0000-0002-3021-0469; Email: wenshanyew@nus.edu.sg

Authors

Kevin Jie Han Lim − *Synthetic Biology for Clinical and Technological Innovation, National University of Singapore, Singapore 117456; Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599; Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597; Bioinformatics Institute, Singapore 138671*

Yossa Dwi Hartono − *Synthetic Biology for Clinical and Technological Innovation, National University of Singapore, Singapore 117456; Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599; Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597; Bioinformatics Institute, Singapore 138671*

Bo Xue − *Synthetic Biology for Clinical and Technological Innovation, National University of Singapore, Singapore 117456; Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599; Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597;* orcid.org/0000-0002-5172-2519

Maybelle Kho Go − *Synthetic Biology for Clinical and Technological Innovation, National University of Singapore, Singapore 117456; Synthetic Biology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599; Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597;* orcid.org/0000-0003-3894-2661

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.2c00786

Author Contributions

⊥K.J.H.L., Y.D.H., B.X., and M.K.G. contributed equally to the work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was undertaken in part using the MX1 and MX2 beamlines at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. We would also thank the technical services provided by the Synchrotron Radiation Protein Crystallography Facility of the National Core Facility Program for Biotechnology, Ministry of Science and Technology, and the National Synchrotron Radiation Research Center, a national user facility supported by the Ministry of Science and Technology of Taiwan, ROC. The computational work for this article was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg). This research was also supported by the Singapore National Research Foundation (to W.S.Y.) and the Biomedical Research Council of Agency for Science, Technology and Research (A*STAR), Singapore (to H.F.).

■ REFERENCES

(1) Pertwee, R. G. Cannabinoid pharmacology: the first 66 years. *Br. J. Pharmacol.* **2006**, *147*, S163−S171.

(2) Bridgeman, M. B.; Abazia, D. T. Medicinal Cannabis: History, Pharmacology, And Implications for the Acute Care Setting. *Pharm. Ther.* **2017**, *42*, 180−188.

(3) Felberbaum, M.; Walsh, S. *FDA approves first drug comprised of an active ingredient derived from marijuana to treat rare, severe forms of epilepsy*; FDA News Release, https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm611046.htm (Accessed 13 March 2019), 2018.

(4) Luo, X.; Reiter, M. A.; d'Espaux, L.; Wong, J.; Denby, C. M.; Lechner, A.; Zhang, Y.; Grzybowski, A. T.; Harth, S.; Lin, W.; Lee, H.; Yu, C.; Shin, J.; Deng, K.; Benites, V. T.; Wang, G.; Baidoo, E. E. K.; Chen, Y.; Dev, I.; Petzold, C. J.; Keasling, J. D. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **2019**, *567*, 123−126.

(5) Zirpel, B.; Degenhardt, F.; Martin, C.; Kayser, O.; Stehle, F. Engineering yeasts as platform organisms for cannabinoid biosynthesis. *J. Biotechnol.* **2017**, *259*, 204−212.

(6) Kuzuyama, T.; Noel, J. P.; Richard, S. B. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. *Nature* **2005**, *435*, 983−987.

(7) Lim, K. J. H.; Lim, Y. P.; Hartono, Y. D.; Go, M. K.; Fan, H.; Yew, W. S. Biosynthesis of Nature-Inspired Unnatural Cannabinoids. *Molecules* **2021**, *26*, 2914.

(8) Valliere, M. A.; Korman, T. P.; Woodall, N. B.; Khitrov, G. A.; Taylor, R. E.; Baker, D.; Bowie, J. U. A cell-free platform for the prenylation of natural products and application to cannabinoid production. *Nat. Commun.* **2019**, *10*, 565.

(9) May, O.; Nguyen, P. T.; Arnold, F. H. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nat. Biotechnol.* **2000**, *18*, 317−320.

(10) Qian, S.; Clomburg, J. M.; Gonzalez, R. Engineering Escherichia coli as a platform for the in vivo synthesis of prenylated aromatics. *Biotechnol. Bioeng.* **2019**, *116*, 1116−1127.

(11) Go, M. K.; Zhao, L. N.; Xue, B.; Supekar, S.; Robinson, R. C.; Fan, H.; Yew, W. S. Directed Computational Evolution of Quorum-Quenching Lactonases from the Amidohydrolase Superfamily. *Structure* **2020**, *28*, 635−642.e3. e3

(12) Mikheev, M.; Gao, D. Production of cannabinoids in yeast. WO2019014490A1, 17/01/2019, 2019.

(13) Yang, Y.; Miao, Y.; Wang, B.; Cui, G.; Merz, K. M., Jr. Catalytic mechanism of aromatic prenylation by NphB. *Biochemistry* **2012**, *51*, 2606−2618.

(14) Fellermeier, M.; Zenk, M. H. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Lett.* **1998**, *427*, 283−285.

(15) Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778−795.

(16) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. Force Field Validation Using Protein Side Chain Prediction. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.

(17) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf.* **2016**, *54*, 5−6.

(18) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.

(19) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., III; MacKerell, A. D., Jr.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405−413.

(20) Kim, S.; Lee, J.; Jo, S.; Brooks, C. L., III; Lee, H. S.; Im, W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* **2017**, *38*, 1879−1886.
(21) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(22) 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.
(23) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(24) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(25) Waskom, M. L. Seaborn: statistical data visualization. *J. Open Source Software* **2021**, *6*, 3021.
(26) Büchler, J.; Malca, S. H.; Patsch, D.; Voss, M.; Turner, N. J.; Bornscheuer, U. T.; Allemann, O.; Le Chapelain, C.; Lumbroso, A.; Loiseleur, O.; Buller, R. Algorithm-aided engineering of aliphatic halogenase WelO5* for the asymmetric late-stage functionalization of soraphens. *Nat. Commun.* **2022**, *13*, 371.
(27) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery: San Francisco, California, USA, 2016; pp 785−794.
(28) Cowieson, N. P.; Aragao, D.; Clift, M.; Ericsson, D. J.; Gee, C.; Harrop, S. J.; Mudie, N.; Panjikar, S.; Price, J. R.; Riboldi-Tunnicliffe, A.; Williamson, R.; Caradoc-Davies, T. MX1: a bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* **2015**, *22*, 187−190.
(29) Aragão, D.; Aishima, J.; Cherukuvada, H.; Clarken, R.; Clift, M.; Cowieson, N. P.; Ericsson, D. J.; Gee, C. L.; Macedo, S.; Mudie, N.; Panjikar, S.; Price, J. R.; Riboldi-Tunnicliffe, A.; Rostan, R.; Williamson, R.; Caradoc-Davies, T. T. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J. Synchrotron Radiat.* **2018**, *25*, 885−891.
(30) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125−132.
(31) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307−326.
(32) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658−674.
(33) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355−367.
(34) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213−221.
(35) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.
(36) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Adams, P. D.; Read, R. J.; Zwart, P. H.; Hung, L.-W. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 515−524.
(37) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12−21.
(38) Schrodinger LLC, *The PyMOL Molecular Graphics System*, Version 2.4.1.

https://doi.org/10.1021/acscatal.2c00786
ACS Catal. 2022, 12, 4628−4639