Indigo Formation as a Predictor of Substrate Promiscuity in Cytochrome P450 BM3 Libraries

Donya Valikhani[1,2,4‡], Jonathan N. Besna[2,3,4‡], Olivier Rousseau[1,2,4‡], Claudèle Lemay-St-Denis[2,3,4], Guillaume Lamoureux[5] and Joelle N. Pelletier[1,2,3,4*]

[1] Department of Chemistry, Université de Montréal, 1375 Thérèse-Lavoie-Roux ave, Montréal, QC H2V 0B3, Canada

[2] PROTEO, the Québec Network for Protein Function, Engineering and Applications, Québec, G1V 0A6, Canada

[3] Department of Biochemistry and Molecular Medicine, Université de Montréal, 2900 Boulevard Édouard-Montpetit, Montréal, Québec, H3T 1J4, Canada

[4] CGCC, Center in Green Chemistry and Catalysis, Montreal, QC, Canada

[5] Department of Chemistry and Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, 08102, USA

‡These authors contributed equally.

*Corresponding author: Joelle N. Pelletier email : joelle.pelletier@umontreal.ca

KEYWORDS: High-throughput screening, indigo screening assay, P450 BM3, substrate promiscuity, protein engineering

ABSTRACT

Cytochrome P450 BM3 natively exhibits substrate promiscuity, mainly in hydroxylation and epoxidation reactions. Whereas directed evolution has broadened its scope of reactivity, screening for specific non-native enzymatic reactivity remains challenging, thereby limiting potential applications. In this study, we investigated the predictive potential of indigo production as an indicator of promiscuous hydroxylation in non-native aromatic compounds. Following site-saturation mutagenesis on 42 residues in the active-site region, we rapidly identified 97 previously unreported variants at 18 unique positions, significantly increasing the database of indigo-positive variants. From this expanded pool, we isolated 80 well-expressed indigo-positive variants and evaluated them alongside 46 well-expressed indigo-negative variants for their hydroxylation activity on the representative non-native aromatic compounds, anisole and naphthalene. We observed a strong correlation between indigo formation and aromatic compound hydroxylation, with 73-80% of indigo-positive variants hydroxylating these compounds, compared to 40-45% of indigo-negative variants. Notably, only indigo-positive variants showed greater than 3-fold higher activity than the wild-type P450 BM3. Furthermore, a positive correlation was obtained between promiscuous hydroxylation of anisole and naphthalene. The approach also allowed identification of new variants that hydroxylate a further aromatic compound, 4-phenyl-2-butanone, suggesting generality. Although we did not succeed in identifying the molecular basis underlying the indigo-positive phenotype, our findings demonstrate that indigo production is an effective screening tool, expediting the discovery of novel functional genes for promiscuous aromatic hydroxylation reactions. This method offers a cost-effective and efficient approach to identify and develop new biocatalysts, advancing the potential applications of P450 BM3 in scientific and industrial fields.

INTRODUCTION

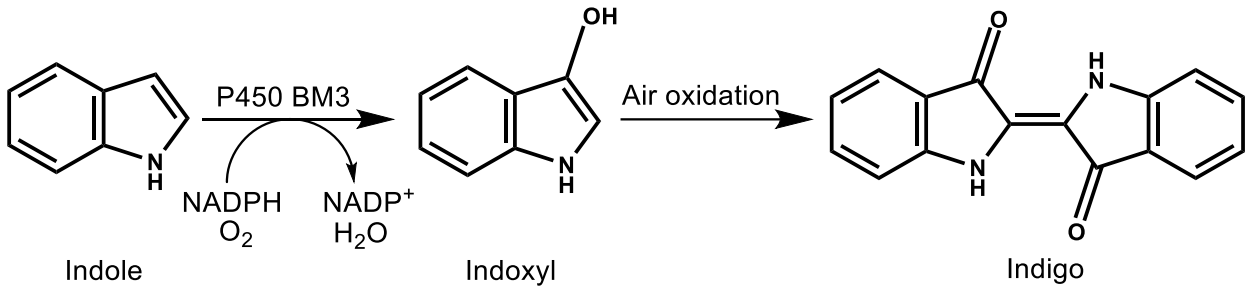
Hydroxylation reactions feature prominently in synthetic strategies, yet the selective hydroxylation of non-activated C-H bonds remains challenging. Traditional routes to C-H bond hydroxylation typically use oxidizing reagents and transition metals in organic solvents,[1-3] lacking in selectivity and running counter to the principles of green technology. Biocatalytic routes for C-H bond oxidation using atmospheric O_2 are increasingly being adopted. In addition to providing a greener alternative to traditional methods of oxidation, biocatalytic oxidations offer unprecedented selectivity.[4-6]

Cytochrome P450 monooxygenases (P450s or CYPs) are heme-containing enzymes that natively catalyze the hydroxylation of non-activated C-H bonds in a regio- and stereoselective manner. The native promiscuity of P450s has been broadened by engineering, giving access to a reactions including epoxidation, sulfoxidation, aryl−aryl coupling, ring cleavages, nitration, dealkylation and dehalogenation, on a wide variety of molecules.[7-10] In particular, P450 BM3 (CYP102A1) from *Bacillus megaterium* has been broadly investigated both for native and engineered reaction promiscuity.[11-15] One of its attractive features is that it is a self-sufficient P450: it contains both the monooxygenase and reductase domains required to complete the catalytic cycle. This contributes to its displaying one of the highest catalytic activities reported for P450 monooxygenases.[7, 10, 11, 14, 15] P450 BM3 natively hydroxylates long-chain fatty acids, and engineering has expanded its substrate scope for hydroxylation to include short-chain fatty acids, aromatic compounds, terpenes, cyclic and acyclic alkanes.[16-23] Those studies have demonstrated the potential for engineering P450 BM3 to further broaden the substrate scope for hydroxylation reactions. However, efficient screening of libraries of P450 BM3 variants for promiscuous hydroxylation is dependent on the availability of screening methods and, importantly, on the quality of the P450 BM3 variant libraries to be screened.

NADPH natively serves as the electron donor in the reductive oxidation reaction catalyzed by P450 BM3. NADPH depletion assays are widely applied to screen P450 variant libraries. However, this can lead to the detection of false positives because NADPH consumption can result from 'uncoupling', where the oxygen is non-productively diverted by reduction to the reactive oxygen species peroxide and superoxide anion.[11] Flow cytometry [24] and mass spectrometry[25-27]

can be applied to high-throughput screening for product formation, yet incur significant costs and preparation time. As a result, considerable effort has been invested in developing colorimetric and fluorescent screening assays based on product formation. Such colorimetric and fluorescent assays include screening for indigo,[23, 28-30] phenols,[31] epoxides,[32] ω-hydroxy fatty acids,[33] N-benzyl-3-hydroxypyrrolidines,[34] styrene derivatives[35] and hydroquinones.[36]

Direct screening of color or fluorescence formation in bacterial colonies on agar plates is the simplest approach to screening.[37] Accordingly, in this work, we propose a complementary approach to address the need for more effective variant screening, by creating a high-quality library of P450 BM3 variants having an increased propensity to hydroxylate aromatic compounds. The premise rests on the earlier observation by several research groups that a few P450 variants hydroxylate the non-native aromatic indole to indoxyl, followed by spontaneous oxidation to indigo (indigo-positive variants) (**Scheme 1)**; these variants promiscuously hydroxylate further non-native aromatic compounds. Specifically, engineering the *Pseudomonas putida* P450cam revealed a robust correlation between the variants' capacity to hydroxylate indole, diphenylmethane[30] and ethylbenzene derivatives.[29] In the case of P450 BM3, the indigo-positive F162I/M185T/L188P/M237I variant catalyzes the hydroxylation of coumarin, 7-ethoxycoumarin, chlorzoxazone and *p*-nitrophenol. Moreover, the indigo-positive F162I/E228K mutant displayed the most best performance in *O*-deethylation of 7-ethoxyresorufin *O*-deethylation and phenacetin.[38] We recently demonstrated that further indigo-positive P450 BM3 variants also hydroxylate 4-phenyl-2-butanone (4-PB) to the industrially relevant 4-(4-hydroxyphenyl)-2-butanone, or 'raspberry ketone', which is among the most widely used, high-value food additives.[23] In that study, point substitution at any of five active-site positions yielded three known and 13 previously-unreported indigo-active point-substituted variants, among which eight variants catalyzed oxidation of 4-PB. In contrast, none among the 37 indigo-negative variants tested hydroxylated 4-PB.[23] That work illustrated two foundational concepts: (1) there exist many solutions to engineer variants of P450 BM3 that produce indigo; (2) indigo production can serve as a predictor of promiscuous hydroxylation of a non-native aromatic substrate.



Scheme 1. Bioconversion of indole to indigo. The oxidation of indole to indoxyl by engineered variants of P450 BM3 is followed by the spontaneous dimerization of indoxyl to indigo.

Based on that further demonstration of the predictive capacity of indigo production in promiscuous hydroxylation of that non-native aromatic substrate, we established two goals: first, to expand the database of indigo-positive variants of P450 BM3, and then to determine if more highly diversified indigo-positive variants are generally predictive of promiscuous hydroxylation of non-native aromatic substrates.

Here, we targeted for site-saturation mutagenesis (SSM) all residues that lie above the plane of the heme, having at least one atom within 20 Å of the heme-iron. Screening for indigo-positive variants was facilitated by whole-cell colorimetric screening. By next-generation DNA sequencing (NGS), we immediately identified 141 unique, indigo-positive point-substituted variants distributed across 23 positions.

Having thus greatly expanded the diversity of indigo-positive variants, we individuated variants for assessment of hydroxylation of two representative aromatic compounds: anisole and naphthalene. By comparing 80 well-expressed indigo-positive variants with 46 well-expressed indigo-negative variants, we report a significant positive correlation between indigo formation and hydroxylation of anisole and of naphthalene. We envisage that this library of 80 well-expressed indigo-forming variants of P450 BM3 will serve to reduce early screening efforts and thereby accelerate discovery of further aromatic hydroxylation reactions of interest.

RESULTS AND DISCUSSION

Creation of point-substituted variants spanning the greater active-site region.

In our previous report, we had performed site-saturation mutagenesis (SSM) at five active-site residues of P450 BM3: V78, A82, F87, A264 and T268. Screening for indigo production in *E. coli* resulted in identification of 16 indigo-positive variants, 13 of which were previously unreported. Most resulted from substitution of residue A82 or V78. We further demonstrated that eight among the 16 indigo-positive variants were also active towards the hydroxylation of the non-native aromatic substrate, 4-PB, whereas none among the 37 indigo-negative variants tested, that were point-mutated at any of those five positions, hydroxylated 4-PB.[23]

Here, our objective was to determine whether indigo formation can serve more broadly to expand the substrate promiscuity of P450 BM3. This entails two hypotheses: that a greater diversity of indigo-positive P450 BM3 variants can be identified, and that the correlation between indigo formation and increased promiscuity for hydroxylation of non-native aromatic substrates applies to substrates other than indole. **Figure 1** illustrates the workflow for generating, identifying and assessing P450 BM3 variants for aromatic substrate promiscuity. To investigate whether a greater diversity of point-substituted variants of P450 BM3 result in indigo formation, SSM was performed. The 42 targeted positions were selected for having at least one atom within a 20 Å radius of the heme iron and lying near, or above, the heme plane, and exclude those where indigo formation has previously been reported (**Figure 2A**).

Each library was generated by megaprimer mutagenesis.[39] A defined primer mix was used to prevent codon redundancy and stop codons, theoretically encoding a uniform distribution of 20 amino acids at each of the 42 mutated codons (**Table S1**).

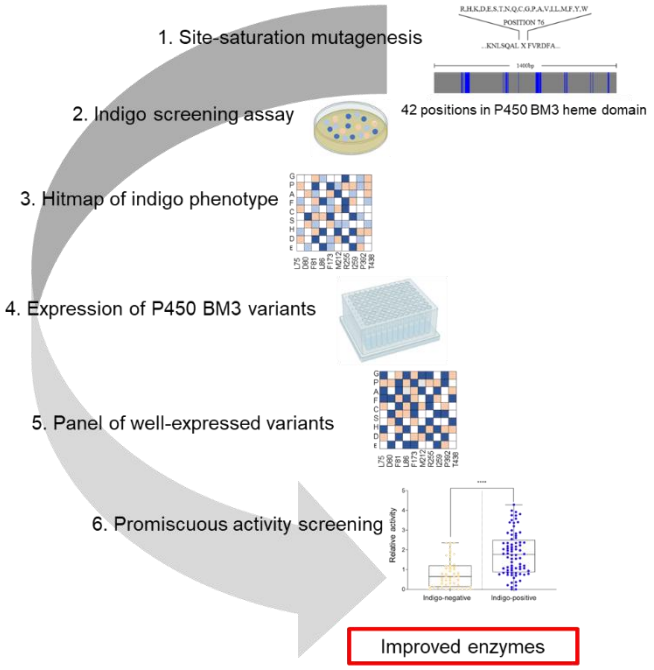


Figure 1. Workflow to diversify indigo-positive variants of P450 BM3 and verify prediction of promiscuous aromatic hydroxylation activity. 1) Site-saturation mutagenesis at each of 42 selected positions in the heme domain of P450 BM3. 2) Screening for indigo production on solid media containing indole. 3) Linking indigo-positive/negative phenotype to genotype. 4) Expression of individual variants under optimized conditions. 5) Hitmap of well-expressed indigo-positive variants. 6) Determining promiscuous activity of variants for hydroxylation of non-native aromatic substrates anisole and naphthalene.

The 42 selected residues are distributed over three elements of secondary structure known to contribute to the oxidation mechanism: the B'-, F/G- and I-helices, as well as sheets and loops in the vicinity of the heme that have the potential to modulate catalysis (**Figure 2B**). Specifically, the B'-helix plays a role in substrate recognition.[40] Notably, substitution of A74, V78, A82 in the B'-helix, alone or in combination, have been reported to improve indigo synthesis.[19, 23, 41-43] The F/G-helices undergo a large conformational change upon substrate binding, modulating the substrate access channel.[40, 44, 45] The I-helix is the longest of the P450 BM3 heme domain, spanning the entire active site cavity immediately above the heme. It contains residues that are critical to the catalytic mechanism and to the conformational changes observed in the F/G-helices.[40, 44, 45]

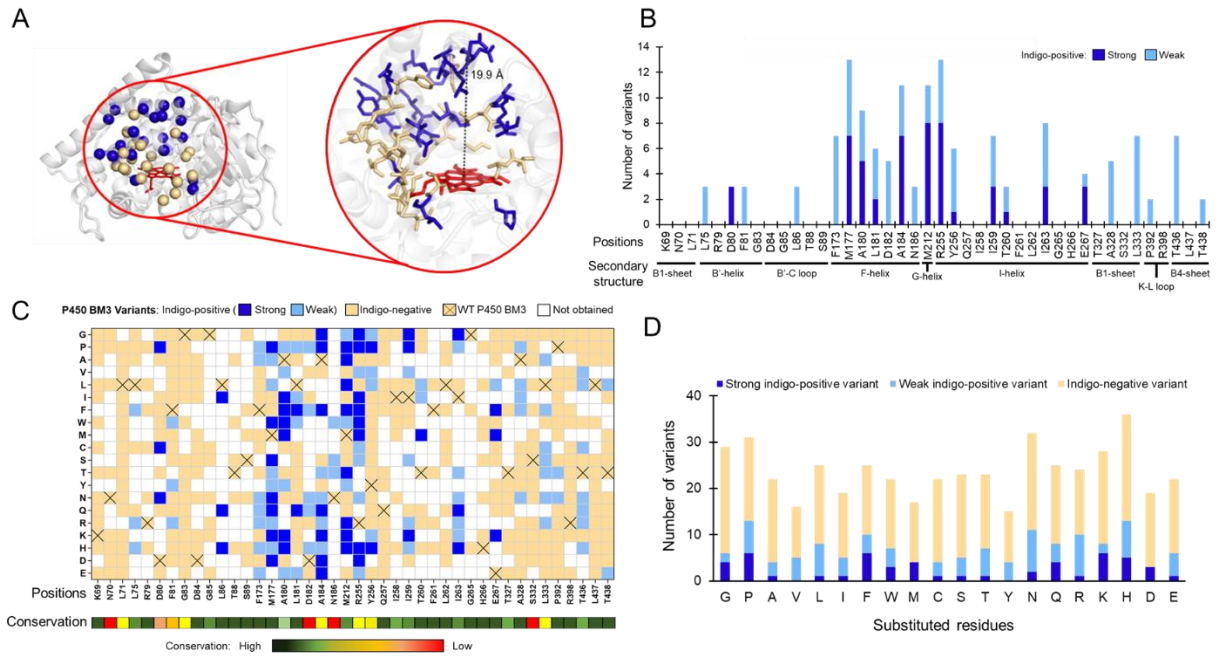


Figure 2. Creation and identification of previously unreported indigo-positive point-substituted variants. A) P450 BM3 monooxygenase domain (PDB ID: 1BU7) is represented in cartoon form and the heme in red sticks. The 42 substituted positions are shown as indigo spheres where at least one identified substitution resulted in indigo formation (n = 23) or beige spheres where only indigo-negative variants were identified (n = 19). The zoom illustrates the native sidechains at those positions, in the same colors. **B)** The number of indigo-positive variants identified at each diversified position. The secondary structure is indicated below. **C**) Phenotypic hitmap of point-substituted variants identified by NGS. Y-axis: Amino acid identity in one-letter code; X-axis: Sequence position. Indigo: strongly indigo-positive variant; pale blue: weakly indigo-positive variant; beige: indigo-negative variant; white: variant not identified by NGS. Crossed boxes: native sequence (WT P450 BM3). Sequence conservation as determined by ConSurf is represented below the plot according to the scale illustrated beneath. D) Distribution of substituted amino acids and their resulting phenotype according to the NGS results shown in panel B.

Identification of indigo-positive variants on solid medium.

The 42 site-saturated libraries were screened in *E. coli* on auto-inducing agar plates containing indole. Upon oxidation of indole to indoxyl and spontaneous oxidative dimerization to indigo, colonies expressing indole-active variants turn blue. Approximately 4500 colonies were picked into 3 distinct pools: 101 dark blue colonies, 441 pale blue and 4032 'not blue' (**Table S2**). The latter ranged from pinkish beige to white, a pink hue testifying to the incorporation of the iron-bound heme in P450 BM3; white colonies may express no P450 BM3 variant or express a variant

that is poorly folded and holds no heme.[23] Impressively, more than half of the libraries (24 of the 42 mutated positions) held blue colonies. Among these, dark blue colonies were found at 12 positions (D80, M177, A180, L181 A184, M212, R255, Y256, I259, T260, I263 and E267) and pale blue colonies at all but one of the 24 positions (**Table S2**), noting that two pale blue colonies were picked at position Q257, but no substitution yielding blue phenotype was identified by NGS.

The identity of the mutations in the three pools was deconvoluted by NGS. Prior to this work, only 20 point-substituted variants at any of six residues (V78, A82, F87, F162, L188, A264) had been reported to give rise to oxidation of indole (**Table S3**); those positions were excluded from this study.[23, 38, 46-50] Here, we identified a total of 141 previously unreported indigo-positive point-substituted variants: 51 unique variants from dark blue colonies and 90 from pale blue colonies, found at 23 positions (**Figure 2C**). We also identified 334 indigo-negative variants from beige colonies (**Figure 2C**). These results immediately demonstrate that readily observable oxidation of indole can be induced in P450 BM3 by point substitution of a variety of residues lying within 20 Å above the heme iron.

The amino acid coverage of the point-substituted libraries varied significantly according to the position (**Figure 2B**), averaging 60% coverage (**Figure S1**). This coverage largely sufficed to identify hot-spots: point substitution at residues 177 and 184 of the F-helix, residue 212 of the G-helix and residue 255 of the I-helix gave rise to more than half the total number of strong indigo-positive variants (30/51) (**Figure 2B**). Although no striking trend was observed, we note that substitution to proline, phenylalanine and lysine were the most frequently observed (each at 6 out of 42 positions) in strongly indigo-positive variants (**Figure 2D**). It is noteworthy that the incomplete coverage of substituted amino acids suggests that additional indigo-positive point-substituted variants could be identified at these hot-spot positions.

Structural distribution of indigo-positive variants.

Substitutions to the F/G-helices accounted for the largest proportion of indigo-positive variants detected (65 variants, 46% of indigo-positive variants) (**Figure 2B**). Among the five residues mutated in the B′ helix (L75, R79, D80, F81 and G83), only position D80 yielded dark blue variants. No strongly indigo-positive variants were identified upon mutation within the B1-sheet (K69, N70, L71, T327, A328, S332, L333) or within the K-L loop and B4-sheet P392, R398, T436,

L437, T438 despite good mutational coverage. Comparison of the propensity for each position to yield indigo-positive variants with the sequence conservation rate at each position, as determined by ConSurf did not provide any clear insights. Residues 177, 212, 259 and 263 are highly conserved whereas positions 180, 184 and 255 were moderately conserved, yet these seven positions yielded the highest number of indigo-positive variants and had high mutational coverage. [11] Residues mutated between 84-89 and between 259-267 are generally highly conserved and showed below average mutational coverage.

No significant difference in GC content was observed between regions with high and low mutation coverage, ruling out differences in GC content as contributing to mutagenic success. Other factors that can contribute to differences in mutational coverage include differences in primer annealing rate, codon usage and variation in tolerance to substitution at each residue (expression, folding, …). Overall, these results illustrate the breadth of sequence variation in P450 BM3 that can lead to a significant increase in indole hydroxylation.

Individuation of well-expressed P450 BM3 variants to investigate substrate promiscuity.

To investigate whether indigo production is correlated with oxidation of further aromatic compounds, variants were individuated from the initial libraries by plating on indole-containing solid medium. To obtain a balanced distribution of indigo-positive and indigo-negative variants according to the results of NGS, we included all mutated positions except for those in the C-terminal segment (from residue T237 onward), since no strong indigo producers were identified in that region. In total, 121 unique variants including 76 indigo-positive variants from blue colonies and 45 indigo-negative variants from beige (i.e., not blue) colonies were individuated by DNA sequencing. We further included nine indigo-positive and nine indigo-negative variants obtained in our earlier work[23], at positions V78, A82, F87, and A264 (**Figure 3A**). We thus considered 85 indigo-positive and 54 indigo-negative variants located at 31 positions for the next step **(Figure 3A)**.

Quantitation of P450 BM3 expression in the individuated variants.

When expressed in liquid medium in a high-throughput format, the concentration of well-folded, heme-containing P450 BM3 variants typically surpassed 0.2 μM and averaged 0.7 μM, as assessed

with the carbon monoxide (CO) binding assay[51] **(Figure 3B; Figure S2)**. Although variants in the F/G-helices were expressed at higher average concentrations (0.84 µM on average), there was no significant difference ($p > 0.05$, Student's t-test) between the average enzyme concentration of indigo-positive (0.75 µM) and indigo-negative (0.65 µM) variants. This demonstrates that indigo production is not a consequence of increased expression.

Variants expressed at concentrations lower than 0.2 µM were excluded from further study: B' helix variants V78D, R79F, A82G/P/R/V, B'-C loop variants L86H and F87S/V, as well as F-helix variants M177K and A180K, G-helix variant M212F and I-helix variant H266S. We thus retained 80 well-expressed indigo-positive variants and 46 well-expressed indigo-negative variants for further experimental characterization; they belong to three clusters of substituted positions (K69-F87, F173-M212 and R255-E267; **Figure S1**).

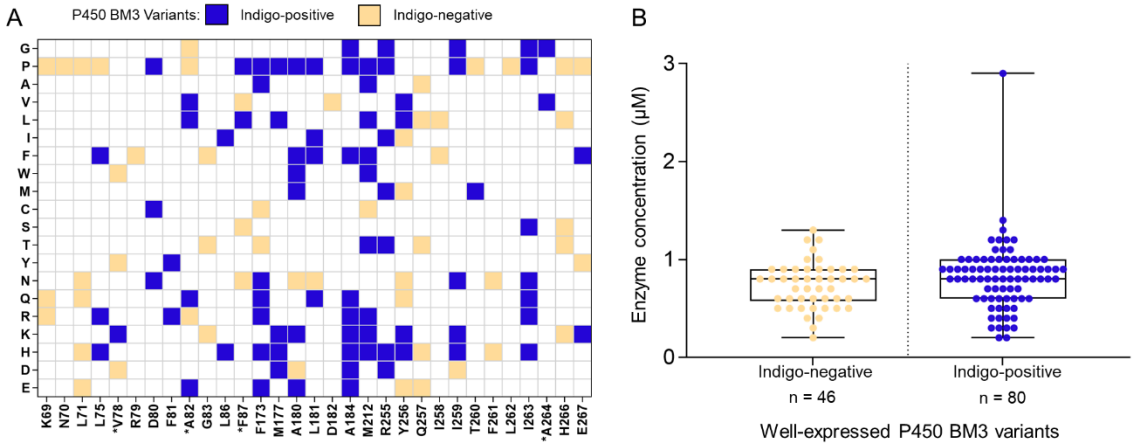


Figure 3. Well-expressed P450 BM3 point-substituted variants. A) Phenotypic hitmap of well-expressed indigo-positive (indigo) and indigo-negative (beige) P450 BM3 variants. Amino acid identity is on the vertical axis and positions are listed on the horizontal axis. Well-expressed indigo-positive variants are found at 22 positions and indigo-negative at 23 positions, with 14 positions having both. **B**) The concentration of the P450 BM3 variants was determined by CO binding assay (**Figure S2**). Variants above 0.2 µM were considered to be well-expressed and were further characterized. No significant difference was found between concentration of indigo-positive and indigo non-negative variants ($p > 0.05$). (*) indicates positions substituted in previous work.[23]

Evaluation of the indigo whole-cell screening assay.

Visual assays of whole cells producing colored compounds on solid medium offer high-throughput screening capacity, when automated colony-picking is applied. Nonetheless, they typically provide

only qualitative or semi-quantitative activity data. We assessed the sensitivity and accuracy of the rapid visual determination of blue phenotype in colonies by comparing to the spectrophotometric quantification of indigo in lysate of *E. coli* following overexpression of the P450 BM3 variants (**Figure S5**). Although colonies expressing WT P450 BM3 showed no trace of blue, trace indigo detected in the clarified lysate of WT P450 BM3 served as the baseline (2.7 μM, ~0.11% conversion).

The conditions applied for the whole-cell screening on solid media and the cellular lysate assay entailed important differences. First, levels of enzyme expression were not monitored in the whole cells whereas the CO binding assay was applied to normalize the enzyme concentration when using lysate. Second, substrate availability may differ, where cellular substrate uptake is required in whole cells whereas substrate is immediately available in lysate. Finally, substrate was exposed to the active enzyme for up to 40h in whole cells compared to only 3 min in lysate. We note that 3 min sufficed for the top variants to convert enough indole to indoxyl to observe indigo without precipitation, thus allowing determination of the full range of rates of substrate conversion.

As expected, the indigo-positive colonies consistently exhibited significantly higher indigo production in lysate compared to the indigo-negative colonies; 29 indigo-positive variants showed greater than 4-fold enrichment in indigo production relative to WT, reaching a maximum of 41-fold for variant I263G (**Figure S5, Table S4**). Discrepancies were only observed among the lowest indigo producers where several variants showed similar levels of activity to indigo-negative WT but exhibited the blue phenotype in whole cells. Conversely, some indigo-negative variants (5 % of expressed variants) exhibited slightly higher indigo production than WT in lysate yet were not visually detectable in colonies, indicating the limited sensitivity of the whole-cell indigo screening assay. Only one indigo-negative outlier (F173C) was observed with over 2-fold increase for indigo production (**Table S4**). No indigo-negative colony was among the top 40 indigo producers. Although of moderate sensitivity, the whole-cell indigo screening assay offers high-throughput for rapid differentiation of strong and weak indigo-producing variants.

Investigation of the molecular basis underlying indigo production in P450 BM3 variants

That gain-of-function substitutions have been identified at multiple positions included in different elements of secondary structure in the active site region suggests that diverse factors could be

involved in inducing indole oxidation; alternatively, a dominant mechanism may prevail. To investigate potential mechanisms that favor indole oxidation in point-substituted variants of P450 BM3, we turned to computational methods.

The WT P450 BM3 has been described in two major conformations: substrate-free (SF) and substrate-bound (SB).[45, 52] Upon substrate binding, a shift between the SF and SB states displaces an important water molecule.[53] This disrupts a hydrogen-bonding network between the heme, water molecules and catalytic residues A264 and T268, allowing oxygen binding and oxidation to proceed.[44] The shift from the SF to the SB state also induces a major conformational change.[40, 54] The I-helix, which holds a 167° kink in the SF state, becomes more linear (175° kink) in the SB state.[53] Since the kink pushes the F/G-helices and the loop or "lid domain" connecting them, substrate binding modulates the opening of the substrate access channel. [55] This acts as a gating system where bulk solvent fills the active site in the SF state but is excluded in the SB state, preventing unproductive use of electrons.[56, 57] Consequently, substitutions that alter the equilibrium between the SF and SB conformations have the potential to modulate indole uptake and oxidation.

With the goal of understanding the molecular basis underlying indigo production, we therefore investigated various conformations through molecular dynamics (MD) simulations of the active-site region of variants with indigo-positive and indigo-negative phenotypes. We previously performed biased MD simulations of WT P450 BM3, successfully evaluating the binding trajectory and energy of substrates entering the active site at the nanosecond scale.[55] Here, we applied the same robust Adaptive Biasing Force (ABF) approach[58, 59] to compare SF and SB states of indigo-positive and negative variants. We hypothesized that if the gain-of-function mechanism affects the binding path of indole in the dynamic P450 active site, we would observe differences between variants that hydroxylate indole and those that do not. To this end, we selected four representative indigo-positive/negative pairs of variants. The pairs bear a point substitution at neighbouring positions of different elements of secondary structure, to examine the impact of substitutions throughout the active site: A82F(+)/G83F(-) in the B'-helix, A184K(+)/L181N(-) in the F-helix, R255T(+)/Y256I(-) at the beginning of the I-helix, and I263G(+)/H266S(-) in the centre of I-helix (**Figure S6**).

Using the ABF approach, we mapped the trajectory of indole entry into the active site for each pair of variants via the previously described main substrate tunnels 2a and 2f [60, 61] over the course of a 200 ns MD simulation per tunnel. These conditions were selected because they allow full exploration of SF and SB conformations of the I-helix in the WT P450 BM3. In this approach, indole is placed at the protein surface near either tunnel mouth and is guided toward the active site. The ABF method calculates the mean force along the reaction coordinate that describes the substrate-binding path and, at each point of that reaction coordinate, applies an equal and opposite force. This allows the substrate to cross energy barriers in the free-energy landscape of the reaction coordinate. The process converges when substrate binding is governed by a flat free-energy landscape, where energy barriers and wells have been mapped and canceled.[58, 59]We thus calculated indole-binding paths into the active site, forcing exploration of varied kink angle of the I-helix (between 155° and 175°).

If indigo-positive substitutions favor indole entry into the active site, we expect to observe lower energy indole-binding paths than in the WT P450 BM3 or indigo-negative variants. However, upon comparing the energy maps describing the indole-binding paths of the four indigo-positive/indigo-negative pairs of variants for the two tunnels, we observed no significant trends that differentiate indole binding in the indigo-positive variants from that in their indigo-negative counterparts (**Figure S7**).

To investigate whether the substitutions that define the indigo-positive/indigo-negative pairs of variants trigger different conformational or dynamic changes in the absence of a substrate, we observed their dynamic behavior by performing unbiased ligand-free MD simulations, analyzing specific features that have been associated with catalysis (**Figure S8** - **Figure S10**). Specifically, we focused on dynamics of the F/G-helices and lid domain, and the I-helix kink. While indigo-positive variants generally display a less dynamic G-helix tail (**Figure S8**) and a larger proportion of conformations having a wider opening at the ligand access tunnel (**Figure S9**), no dynamic property was identified that clearly distinguished the indigo-positive from the WT P450 BM3 and indigo-negative variants in the absence of substrate.

That we failed to identify significant differences among the WT and four pairs of variants in indole-binding paths for two main substrate tunnels (2a, 2f), or in dynamics of the flexible F-helix, G-helix, I-helix region that promotes catalytic turnover over the 0.2 μs timescale, is informative. It

suggests that the mechanism(s) promoting indole hydroxylation may not be related to the substate's approach to the active site, but to any of a number of factors that we cannot readily simulate. These substitutions could potentially modify the electron transfer efficiency from NADPH to the ferric heme, or could facilitate access of the catalytical O_2 to the heme. In our simulations, the O_2 molecule is fixed on the heme iron, such that the approach of O_2 into the active site has not been monitored. Substrate and product inhibition have been reported in various reactions catalyzed by WT P450 BM3;[4–6] the substitutions could reduce potential inhibition by indole or indoxyl and favor indole binding or promote indoxyl release. Furthermore, the reductase domain of P450 BM3 is porous, making it difficult to establish substrate entry and product release paths throughout the extensive tunnel system. Small aromatic substrates may gain access to the active site by different tunnels, and their oxidized products exit through different one. We note that computationally resolving impact of substitutions on activity in this system will likely be facilitated by analyzing multiply-substituted variants, where effects should be greater. Our future efforts will address this.

Screening P450 BM3 variants for hydroxylation of further aromatic compounds

To verify whether indigo production resulting from the hydroxylation of the aromatic indole substrate can serve as a predictor of reactivity towards further non-native aromatic compounds, anisole (a single-ring aromatic compound) and naphthalene (the simplest polycyclic aromatic compound) were tested. [22, 62] The panel of 80 well-expressed indigo-positive and 46 well-expressed indigo negative point-substituted variants, defined above, were investigated. To increase screening throughput, we monitored formation of hydroxylated products by means of the colorimetric 4-aminoantipyrine (4-AAP) assay[31], using clarified *E. coli* lysate following P450 BM3 variant overexpression and quantification with the CO binding assay. We first verified that this assay, typically used for detection of phenols, also reacts with 1-naphthol and 2-naphtol to yield colored reaction products; similar to phenolic products, both conjugation products have an absorption peak near 485 nm **(Figure S14)**. We verified that neither anisole nor naphthalene give rise to significant color formation when reacted with 4-AAP, thus confirming that unreacted substrates will not cause significant background signal. We also verified that the background 4-AAP reaction with *E. coli* lysate lacking the P450 BM3 gene is insignificant. Standard curves for

each hydroxylated product were made to assess the sensitivity of the essay (**Figure S15**). The data collected from the colorimetric assay is provided in machine readable format (**Table S4**).

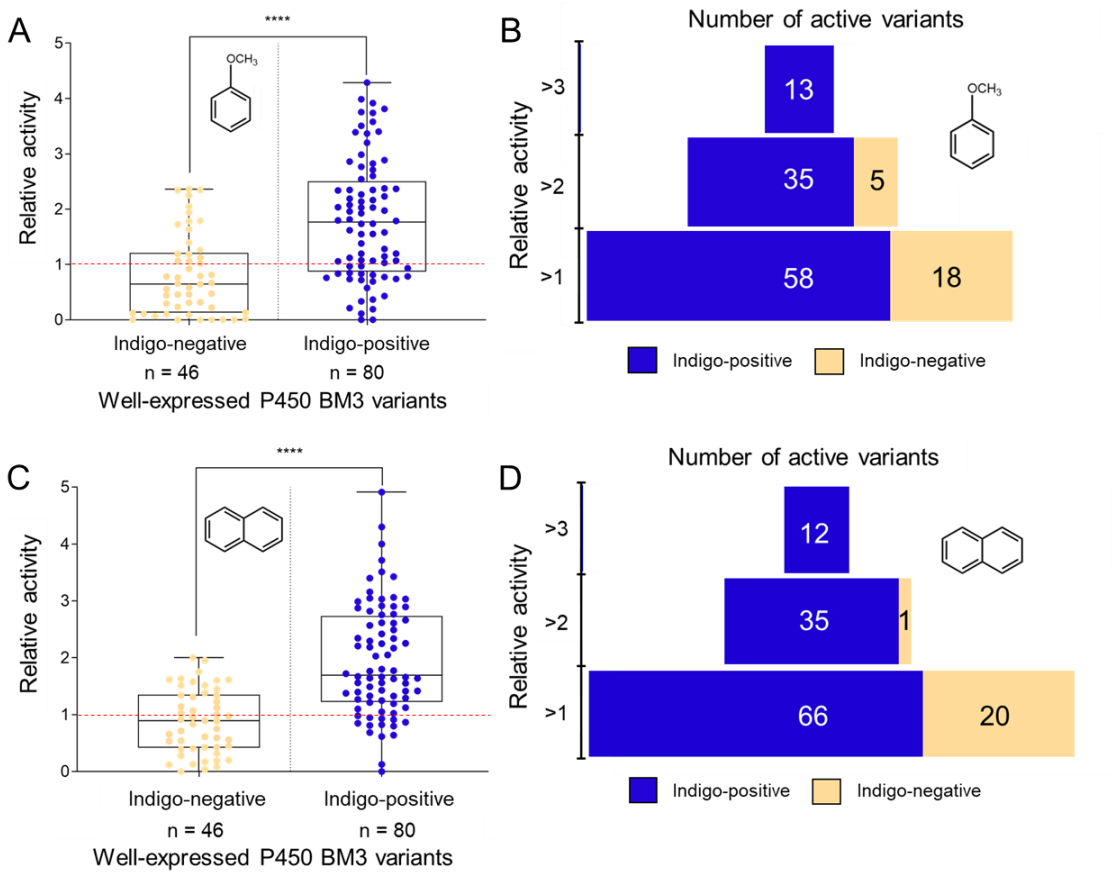


Figure 4. Hydroxylation activity for anisole and naphthalene of well-expressed P450 BM3 variants relative to WT. Promiscuous hydroxylation activity of P450 BM3 variants as determined by the 4-AAP assay towards (A) anisole and (C) naphthalene is given relative to WT activity (WT activity = 1). Distribution of indigo-positive and indigo-negative variants with higher activity than WT for the hydroxylation of (B) anisole and (D) naphthalene relative to WT activity.

We found higher anisole hydroxylation activity than WT in 73% of the indigo-positive population (58 of the 80 variants) compared to less than 40% of the indigo-negative variants (18 of the 46 variants) **(Figure 4A**). Variants displaying a greater than 2-fold increase in activity were predominantly indigo-positive (n = 35 *vs* n = 5 indigo-negative) and, remarkably, all giving a greater than 3-fold increase in anisole hydroxylation were indigo-positive variants (n=13/80) (**Figure 4B; Table S4**). The latter represent 16% of the indigo-positive population. A similar

enrichment pattern was observed for naphthalene hydroxylation, where more than 80% of the indigo-positive variants (n = 66/80) showed higher activity than WT compared to less than 45% for indigo-negative variants (n = 20/46) (**Figure 4C**). Variants displaying a greater than 2-fold increase in activity were overwhelmingly indigo-positive (n = 35/80, *vs* n = 1/46 indigo-negative) and, again, all giving a greater than 3-fold increase in naphthalene hydroxylation were indigo-positive variants (n=12/80) (**Figure 4D**), representing 15% of the indigo-positive population. Where indigo-negative variants showed aromatic hydroxylation activity, none were among the most active **(Figure 4)**.

The correlation between hydroxylation of indole and that of the two other aromatics is weakly positive (anisole: R = 0.425; naphthalene: R = 0.339) (**Figure S4A**, **B)**. Ordering by rank increases this positive correlation between indole and aromatic substrate hydroxylation (anisole: R = 0.561; naphthalene: R = 0.441) (**Figure S4C, D**). This enrichment of promiscuous hydroxylation activity in indigo-positive variants was statistically significant for both substrates throughout the three clusters of substitutions investigated (K69-F87, F173-M212 and R255-E267), except for the group of variants from positions R255-E267 for anisole hydroxylation **(Figure S3)**. Among the most active indigo-positive variants, A180F, A184F, R255D, R255M and I259K rank within the top 13 for both substrates **(Table S4)**, displaying > 3-fold increase. In contrast, substitutions L75H, M212K and Y256V procure a significant increase only towards anisole hydroxylation, whereas F81R, A184H and I263G mostly increase activity towards naphthalene hydroxylation. This illustrates the breadth of effects that position and identity of point-substitutions play in modifying substrate specificity.

Furthermore, a positive correlation was obtained (R = 0.5806, p < 0001) between promiscuous hydroxylation of anisole and naphthalene (**Figure 5**). Importantly, most of the top indigo producers (> 4-fold activity) are concentrated in the top right quadrant. They show increased fold-change activity towards indigo, anisole and naphthalene, suggesting some generality. Few indigo-negative variants occupy that quadrant; indeed, the poor/negative indigo producers cluster in the bottom-left quadrant, exhibiting low activity towards all three aromatic substrates. Overall, these results demonstrate that screening for indigo production enriches the pool of variants having increased promiscuity for hydroxylation of aromatic substrates.

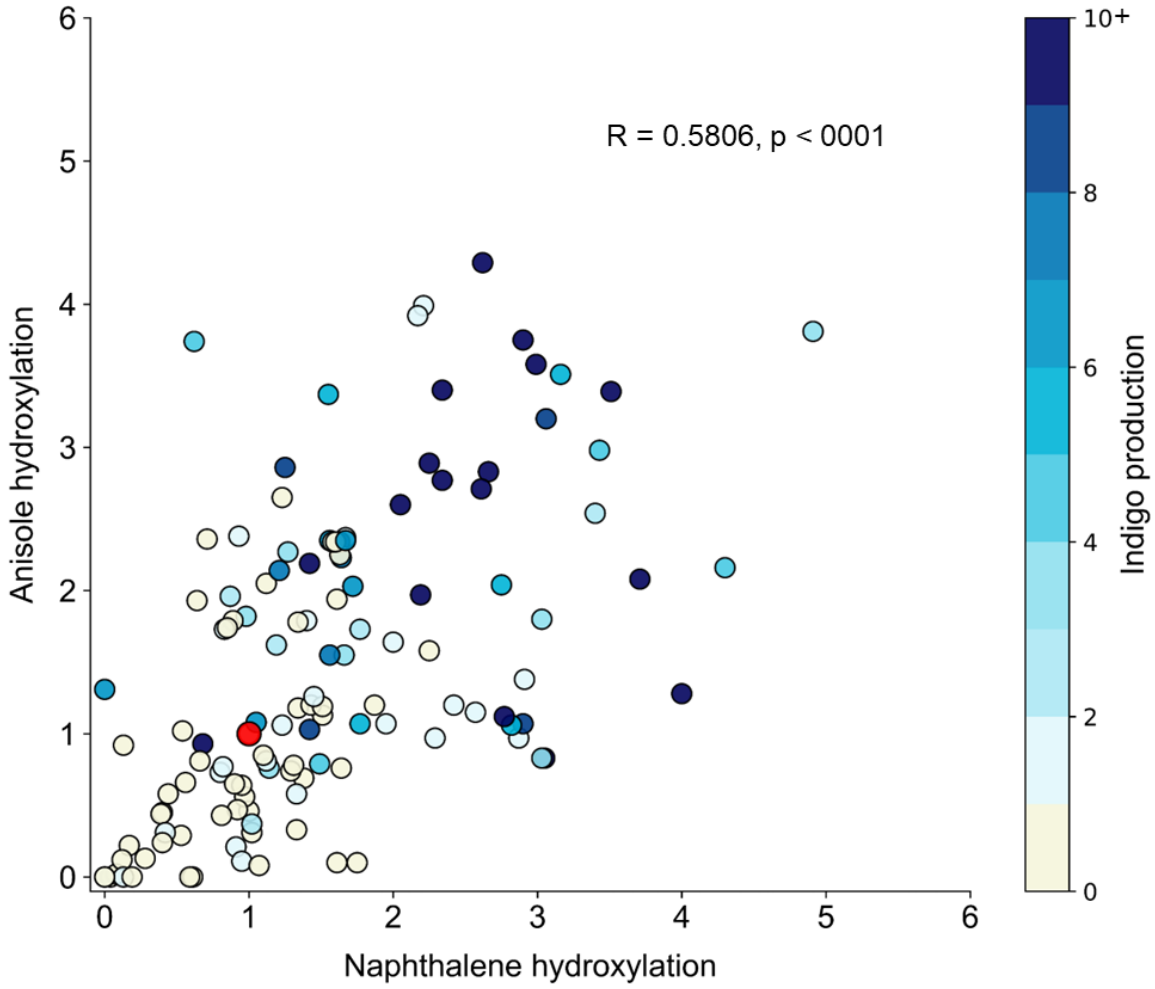


Figure 5. Correlation of promiscuous hydroxylation of anisole and naphthalene by P450 BM3 variants. Fold improvement of activity relative to the activity of WT P450 BM3 (= 1, shown in red) for expressed variants (n = 126) is plotted for hydroxylation of anisole *vs* naphthalene, and is shown in color-scale for production of indigo. The Pearson's r coefficient is given for anisole *vs* naphthalene; correlations with indigo production are plotted in **Figure S4**.

Impact of point substitutions on the activity of P450 BM3.

The top five variants for hydroxylation of anisole are at positions 75, 255 and 256, followed by a cluster of variants at positions 212, 180 and 184 (**Figure S11**). The WT P450 BM3 exhibits strong *ortho*-specific regioselectivity in the production of substituted phenols.[22, 62, 63] However, substitutions that alter ring regioselectivity have been reported.[23, 63, 64] In this context, the regioselectivity of P450 BM3 variants for the hydroxylation of anisole and naphthalene was investigated (**Figure S11; Table S5**). WT P450 BM3 showed predominant *ortho*-hydroxylation

of anisole to form guaiacol (86%), as previously reported.[22, 62] Most variants investigated here conserved the same product distribution pattern as WT. However, variants I263N, I263Q and I263G shifted the regioselectivity toward majority *para*-hydroxylation of anisole, affording up to 80% mequinol although with more than 2-fold total hydroxylation activity increase (**Figure S11A**).

Similarly, variant A82Q, the top producer of raspberry ketone (4-(*p*-hydroxyphenyl)-2-butanone) via *para*-hydroxylation of 4-PB identified in our previous work,[23] also yielded 80% *para*-hydroxylation of anisole, suggesting generality of *para*-selectivity. Indeed, to follow up on our earlier identification of raspberry ketone producing variants of P450 BM3,[23] the hydroxylation of 4-PB was also assessed here. Variant I263N is at least as active in hydroxylating 4-PB as is the top reported A82Q.[23] Further substitutions at position I263 (G, Q, H) were also among the top seven for this activity (**Figure S12**). Similar to anisole hydroxylation, variants at positions 75, 180 and 184 were also among the top 10 for hydroxylation of 4-PB, suggesting parallels in impact of those substitutions, yet variants at positions 212 and 255 were much less active for 4-PB, illustrating specificity.

Activity for hydroxylation of naphthalene greatly mirrored that of anisole, where variants at positions 75, 81, 180, 184, 255 and 263 made up the top 10, with variants at position 80, 87, 212, 260 following suit. The hydroxylation of naphthalene by WT P450 BM3 showed the expected high regioselectivity towards 1-naphthol (95%).[62] Only variant A184E showed a significant shift towards 2-naphthol formation (50%) among P450 BM3 variants (**Figure S11**). I263G, the top indigo producer, was the second-best producer of 1-naphthol. The substitution of isoleucine with glycine may provide the additional space in the binding pocket required to position indole and naphthalene relative to single-ring substrates. The promiscuous activity of variants at the previously underexploited positions 180, 184, 212, 255 and 263 may represent a promising evolutionary platform and merits further investigation. [65]

Finally, we investigated whether promiscuous hydroxylation of aromatic compounds had an impact on the native enzyme activity. WT P450 BM3 natively catalyzes the hydroxylation of fatty acids.[66] We selected 33 indigo-positive variants for this analysis; they are representative of high, mid-range and low hydroxylation activity for each of indole, anisole, and naphthalene. For these variants, we determined the total consumption of lauric acid (C_{12}) (**Figure S13**). The substrate

consumption ranged from 3% to 45% for active variants, compared to 61% lauric acid conversion for WT enzyme; we note that variant D80N was inactive with lauric acid. NADPH was the limiting substrate for the reaction (240 μM).

This demonstrates that the point substitutions made in the active-site region result in a broad range of effects on native reactivity, from negligeable impact to a severe reduction in hydroxylation of lauric acid. Variants F81R, M212W, M212K, A184E and R255M showed a hydroxylation activity above 50% of WT activity for lauric acid and ≥ 2-fold increase of promiscuous hydroxylation activity for both anisole and naphthalene, whereas others such as R255D, R255T, R255P and M212H showed low native activity (< 50% compared to WT) but higher promiscuity, with ≥ 2.2-fold increase of hydroxylation activity for both anisole and naphthalene. Substitutions (G, H, N, Q) at the highly conserved I263 position reduced the hydroxylation of lauric acid by 75% to 93% (**Table S7**). For example, variant I263G showed 7% lauric acid hydroxylation compared to WT while increasing the activity towards anisole hydroxylation by 2.4-fold, and the activity towards naphthalene by 15-fold (**Table S8**). These findings demonstrate that substitutions at this position result in a trade-off favoring substrate specificity towards small aromatics at the expense of the native long fatty acid. Furthermore, all substitutions analyzed at positions I263 inverted regioselectivity for hydroxylation of anisole but not naphthalene; for example, variant I263G favors *p*-hydroxylation of anisole (80% mequinol), contrary to WT P450 BM3 (86% guaiacol). These results highlight the potential for discovery of promiscuous reactivity afforded by investigating indigo-positive variants of P450 BM3.

CONCLUSION

Through an extensive screening process, we have significantly expanded the database of indigo-positive variants, identifying 97 previously unreported point-substituted variants at 18 distinct positions in the active-site area with 29 variants displaying over 4-fold increase in indigo formation compared to WT activity. We demonstrate that 70-80% of the 80 well-expressed indigo-positive variants demonstrated increased hydroxylation activity on two further non-native aromatic (anisole and naphthalene) compounds relative to the WT enzyme, clearly surpassing the performance of the 46 well-expressed indigo-negative variants where only 40% exhibited such activity increase. Notably, all top variants were indigo-positive, displaying at least 3-fold increases in promiscuous activity toward one or more non-native substrate, including indigo production (40 h) as measured

spectrophotometrically, as well as for anisole hydroxylation (top 18) and naphthalene hydroxylation (top 34), as demonstrated by the 4-AAP phenol conjugation assay.

The cost-effective and sensitive primary screening of indigo formation in *E. coli* colonies allowed reliable identification of all the strongest point-substituted P450 BM3 for indigo-positive variants. It also allowed for identification of several new variants with significantly increased activity towards 4-phenyl-2-butanone. While the breadth of reactivity with different substrate classes correlating with increased indigo formation is currently unknown, this examination of point-substituted variants of P450 BM3 sets the stage for accelerated discovery of combinatorial variants having new reactivities. The findings underscore the potential of visual screening as a powerful primary tool in expanding the functional repertoire of P450 BM3 variants.

ACKOWLEDGMENTS

This project was funded by the Natural Science and Engineering Research Council of Canada (NSERC) discovery grant RGPIN-N-2018-04686 and the Canada Research Chair in Engineering of Applied Proteins CRC-2020-00171 to JNP. We thank the workshop team of Université de Montréal MIL campus for engineering the 96-well plate carbon monoxide assay chamber as well as Ali Fendri and Adem Hadjabdelhafid-Parisien for fruitful discussion. JNB acknowledges the support of the NSERC-funded CREATE-APRENTICE program. O.R. held a PROTEO scholarships. CL-S-D was supported by scholarships from NSERC and Hydro-Québec.

MATERIAL AND METHODS

Materials: Mutagenic primers were purchased from Sigma-Aldrich. The Phusion Hot Start II polymerase, purification kit, DNA miniprep kit, DNA gel extraction kit, and electrocompetent *E. coli* DH5α were purchased from New England Biolabs. FastDigest restriction enzymes BamHI, EcoRI, FastAP and DpnI were purchased from Thermo Fisher Scientific. Ligations were performed with the TAKARA ligase from ClonTech. The pcWORI P450 BM3 WT vector was kindly provided by Prof. Frances Arnold (Caltech, USA) and used as the template for mutagenesis. Primers were designed using the SnapGene software and purchased from Sigma. Indole and mequinol were from Alfa Aesar. Other chemicals were purchased from Sigma-Aldrich or Thermo Fischer Scientific.

Enzyme library generation: Forty-two active-site residues having at least one atom within 20 Å above the heme-iron were individually substituted (**Table S2**). Site-saturation mutagenesis was performed using the megaprimer method with four sets of primers per position (**Table S1**) to encode each amino acid once and no stop codon: NDT, VMA, ATG and TGG.[67 68] PCR reactions were performed with a Bio-Rad T100 thermal cycler. Purified PCR products (Monarch® Plasmid Purification Kit, Miniprep) were digested using BamHI and EcoRI according to the manufacturer's instructions. DpnI was used to eliminate parental DNA. The gel-purified inserts were ligated into the BamHI, EcoRI and Fast-AP alkaline phosphatase-treated pcWORI P450 BM3 WT vector. Electrocompetent *E. coli* DH5α were transformed with each ligation mix and plated on LB agar with ampicillin (100 μg/mL). Following colony growth, colonies were recovered as a pool from each position-specific plate with 2 × 750 μL of LB with 100 μg/mL ampicillin. The resuspended cells were incubated for 1 h at 37 °C with agitation at 230 rpm and glycerol stocks were conserved at -80 °C.

Colony-based indigo screening: Overnight cultures of each mutagenic library were propagated in LB with μg/mL ampicillin and plated (100 μL of a 10000 × dilution) on ZYP5052 auto-inducing agar medium containing 0.5 mM indole in DMSO (0.2% final DMSO). Following incubation at 37 °C for 8 h then at 30°C for at least 40 h, plates were stored at 4 °C for at least 24 h to enhance the intensity of the indigo.

Sample preparation for next-generation DNA sequencing (NGS): Colonies were picked and pooled according to both indigo phenotype (dark blue, pale blue, or 'not blue' (beige or white) and the region of the mutated residue (in nucleotides: R1= 79-385, R2 = 444-720, R3 = 667-944, R4 = 837-1108, R5 = 1120-1390). Regions were approximately 250 bp in length to allow full sequencing by Illumina MiSeq NGS. Blue colonies were hand-picked (542 colonies) and other colonies were picked using a Pick-in Master PM-2s (Microtec) automated colony picker (4032 colonies). Overnight cultures in LB with 100 μg/mL ampicillin at 37 °C were prepared for each pool and miniprepped. PCR reactions were performed for each 250 bp region using specific primers containing Illumina CS1/CS2 sequences (**Table S6**). PCR products were kept in pools according to indigo phenotype and mutated region and submitted for barcoding and NGS using Illumina MiSeq 250bp technologies at the Génome Québec Innovation Center of McGill University.

NGS analysis: NGS was performed on five 250-bp regions of the gene (R1-R5) to provide reliable coverage of the 42 mutated positions. Illumina MiSeq sequencing delivered more than 12 million reads. An in-house Python script allowed distinguishing true mutations from sequencing errors and experimental biases, according to read quality and frequency of occurrence of that mutation relative to the background noise. Background was determined according to the frequency of incorrect bases at all positions that were not selected for mutation. We observed that the background noise was different for each pool (dark blue, pale blue, not blue) and region sequenced (R1-R5), illustrating the extent of experimental bias (**Figure S16**). We therefore defined the read frequency thresholds for each pool, for each number of mutations inserted and for each region (R1-R5) with a tolerance of 0.5% false positives. Mutations observed at a higher number of reads than the threshold determined at that position were considered valid in their respective pool.

Some variants were identified in more than one pool; for example, they could be identified in both the pale blue and dark blue pools. We performed Fisher's exact test to determine the pool in which each mutant was most highly represented. For each mutation, a p-value expressed the probability of finding the mutation by chance in that pool by comparing its frequency in the different pools. We filtered the mutations with p-values lower than 0.01 to retain only mutations giving rise to valid comparisons. Then we applied the odds ratio to assign each variant to a pool, thus defining its most likely phenotype.

Computational methods: To perform 2D ABF experiments, two reaction coordinates were used. The first was defined as the distance between the center of mass of indole and the heme iron, and ranged between 3 and 28 Å. The second was defined as the angle between the start, the center and the end of the I-helix. The angle varied between 155° and 175°, as these angles were explored during a 200 ns-unbiased MD simulation of WT P450 BM3. A dioxygen molecule was coordinated with the heme iron during the entire simulation. A total of 200 ns were recorded. Experiments were performed with substrate entering either the 2a or 2f tunnels, using the topology from.[55] The entry-point of tunnel 2a is defined by residues L14, N21 and A191, whereas the entry-point of tunnel 2f is defined by residues D23, N186, Q189 and T436. The width of the local free-energy bin was chosen to be 1 Å, and the adaptive biasing force was applied after each bin was visited at least 500 times. The potential of mean force (PMF) was graphed with R.

The unbiased MD simulations were performed for 200 ns for the same variant pairs. As for the ABF experiment, a dioxygen molecule was coordinated with the iron of the heme molecule. For the simulations, the αC RMSF, the I-helix kink angle and the distance between the side chains of P45 and A191 were calculated using a tcl script with VMD.[69 7] A191 defines the 2a tunnel entry and is located on the highly flexible region (on F/G loop) while P45 is located on a stable region on the B1 strand loop. Atoms P45/CG and A191/CB were chosen due to their proximity (**Figure S9**).

Individuation of P450 BM3 variants: To individuate P450 BM3 variants, the site-saturation mutagenesis libraries generated in *E. coli* for each of the 42 positions were transformed in electrocompetent *E. coli* DH5α according to each variant pool (R1-R5; dark blue, pale blue, not blue). Colonies from each library were picked, plated and numbered simultaneously on LB agar (for subsequent picking) and indigo screening plates (for phenotype affirmation). Following overnight colony formation at 37 °C for agar plates and, in parallel, following the steps outlined for indigo screening, colonies displaying identified color (blue or beige) on screening plates were picked from LB agar plates and inoculated into 1 mL LB media supplemented with ampicillin (100 µg/mL) in 96-deep well plates. Cells were propagated overnight at 37 °C with agitation at 230 rpm and glycerol stocks were conserved at -80 °C. Enzyme expression was carried out as previously reported and cell pellets were frozen at -20 °C for future use.[23] *E. coli* transformed with WT P450 BM3 or with the pUC18 vector with no insert were used as positive and negative controls for enzyme expression, respectively.

Enzyme expression in liquid media and quantification by CO assay: Enzyme expression from glycerol stock was carried out as previously reported and cell pellets were frozen at -20 °C for future use.[23] DH5α *E. coli* transformed with WT P450 BM3 or with the pUC18 vector with no insert were used as positive and negative controls for enzyme expression, respectively. Cells were thawed and lysed in 500 µL of lysis buffer, as described previously.[23] The concentration of well-folded, heme-containing enzyme was determined with the CO binding assay performed in a microplate format, as previously reported with the following changes: a custom aluminum frame equipped with a gasket, an adjustable inlet and outlet as well as a pressure gauge was used to incubate the 300 µL plate for 10 minutes with a CO pressure of approximately 10 PSI.[23]

Absorbance was quantified with a Beckman Coulter DTX 880 Multimode Detector microplate reader. Properly expressed enzyme cultures were sent for Sanger sequencing, as described below.

Sanger sequencing sample preparation: Overnight cultures in LB with 100 µg/mL ampicillin at 37 °C were prepared for only expressed variants (distinguished by CO assay) and 10 µL of grown cultures were sent for sequencing using defined primers for each region at the Génome Québec Innovation Center of McGill University, using the same primers as for NGS (Table S5).

Enzyme activity assays: For all assays, cell lysis was conducted as described previously, with 500 µL of lysis buffer.[23] All assays contained clarified cell lysate of a P450 BM3 variant at a concentration of ~ 0.3 µM, as determined by CO binding assay immediately prior to the activity assay. (A) Hydroxylation of lauric acid: Lauric acid (500 µM) and NADPH (240 µM) were added to the clarified cell lysates and made up to a final volume of 220 µL with potassium phosphate buffer (100 mM, pH 8.0). After 1 h of reaction time at room temperature, a 100 µL aliquot of the reaction mixture was quenched with 400 µL of methyl *t*-butyl ether containing 97 µM lauric acid-d5. Product formation was quantified by LC-MS. (B) Quantification of indigo formation: Indole (5 mM) in DMSO (2% final DMSO) was added to the clarified cell lysates and made up to 500 µL with NADPH (final concentration: 400 µM) in potassium phosphate buffer (100 mM, pH 8.0). Reactions were performed for 3 min at room temperature, with intense agitation using a Lab-Line 4625 titer plate shaker, then quenched with 2 volumes of $CHCl_3$ with thorough vortexing. The spontaneous oxidation of indoxyl to indigo was allowed to take place overnight at room temperature, shielded from light. Following centrifugation at 3000 *g* for 20 min at 4 °C, the organic layer was extracted, and product formation was quantified by measuring absorbance at 620 nm using a Cary 100 Bio UV-Vis spectrophotometer (Agilent Technologies). (C) 4-AAP assay for hydroxylation of anisole and naphthalene: Enzyme assays were performed in 300 µL 96-well microtiter plates. Each well included clarified cell lysate and substrate (anisole, 22 mM; naphthalene, 2.2 mM) made up to a final volume of 220 µL with potassium phosphate buffer (100 mM, pH 8.0). Substrates were dissolved in DMSO to a final concentration of 2-5% (v/v) in the reaction mixture. Reaction mixtures were incubated for 5 min on a microplate shaker at room temperature. Reactions were initiated by addition of NADPH (400 µM). After of 3 min reaction time, reactions were stopped as specified, according to method of analysis. Reactions were analyzed either with the colorimetric 4-aminoantipyrine (4-AAP) phenol reaction, by GC-MS or

by LC-MS, as specified. (D) Hydroxylation of 4-phenylbutanone (4-PB): Reactions were performed and analyzed as in (C) but using 5 mM 4-phenylbutanone and a reaction time of 1 h.

Colorimetric 4-aminoantipyrine (4-AAP) phenol reaction. For colorimetric detection of phenol formation, enzyme activity assays were stopped with addition of 25 µL quenching buffer (4 M urea in 0.1 M NaOH) followed by addition of 4-aminoantipyrine (20 µL, 5mg/mL stock) and $K_2S_2O_8$ (20 µL, 5 mg/mL stock) for color development as described elsewhere.[31] After 30 min incubation, reaction plates were centrifuged (3000 rpm, 22 °C). The supernatant was subsequently transferred to another plate, and its absorbance was measured at 492 nm using a Beckman Coulter DTX 880 Multimode Detector microplate reader. Control reactions using *E. coli* (pUC18) lysate expressing no P450 BM3 were included in each plate.

LC-MS analysis: A 100 µL aliquot of the reaction mixture was quenched with 400 µL of ethyl acetate. LC-MS was performed with a Waters Acquity UPLC system coupled to a Waters Xevo-TQ detector. Samples (1µL) were injected on a Waters BEH C18 column, 50 × 2.1 mm, 1.8 µm with a flow rate of 0.4 mL/min using a gradient of 10 mM ammonium formate (A) and acetonitrile (B), from 40% B to 100% B in 4.5 min with holding time of 1 min. Single Ion Monitoring (SIM) in negative mode was used. For lauric acid, m/z = 199.15 and for lauric acid-d5, m/z = 204.18.

GC-MS analysis: A 100 µL aliquot of the reaction mixture was quenched with 200 µL of ethyl acetate containing 5.7 M t-butyl phenol. GC-MS analysis was performed with a 7890-5975C GC-MS from Agilent Technologies. A ZB-WAXplus 30 m × 250 µm × 0.25 µm column was used. Samples (1 µL) were injected with an inlet temperature of 280 °C. A temperature gradient was initiated from 150 °C (hold for 1 min) to 240 °C (hold for 8 min) at a rate of 18 °C/min. Product masses were detected under electron ionization with SIM of standards guaiacol (124.1 m/z), mequinol (109 m/z),1-naphthol (144.1 m/z), 2-naphthol (115.1 m/z).

REFERENCES

1. Iqbal, Z.; Joshi, A.; Ranjan De, S., Recent Advancements on Transition-Metal-Catalyzed, Chelation-Induced ortho-Hydroxylation of Arenes. *Adv. Synth. Catal.* **2020,** *362* (23), 5301-5351.
2. Kitamura, T.; Fujiwara, Y., Synthetic Reactions via C–H Bond Activation: Oxidation of C–H Bonds. In *Comprehensive Organometallic Chemistry III*, Mingos, D. M. P.; Crabtree, R. H., Eds. Elsevier: Oxford, 2007; pp 213-250.
3. Kudrik, E. V.; Sorokin, A. B., Oxidation of aliphatic and aromatic CH bonds by t-BuOOH catalyzed by μ-nitrido diiron phthalocyanine. *J. Mol. Catal. A: Chem.* **2017,** *426*, 499-505.
4. Dong, J., et al., Biocatalytic Oxidation Reactions: A Chemist's Perspective. *Angew Chem Int Ed Engl* **2018,** *57* (30), 9238-9261.
5. Chakrabarty, S.; Wang, Y.; Perkins, J. C.; Narayan, A. R. H., Scalable biocatalytic C–H oxyfunctionalization reactions. *Chem. Soc. Rev.* **2020,** *49* (22), 8137-8155.
6. Charlton, S. N.; Hayes, M. A., Oxygenating Biocatalysts for Hydroxyl Functionalisation in Drug Discovery and Development. *ChemMedChem* **2022,** *17* (12), e202200115.
7. Urlacher, V. B.; Girhard, M., Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology. *Trends Biotechnol.* **2019,** *37* (8), 882-897.
8. Wei, Y.; Ang, E. L.; Zhao, H., Recent developments in the application of P450 based biocatalysts. *Curr. Opin. Chem. Biol.* **2018,** *43*, 1-7.
9. Zhang, X.; Li, S., Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **2017,** *34* (9), 1061-1089.
10. Li, Z., et al., Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *J. Biol. Chem.* **2020,** *295* (3), 833-849.
11. Whitehouse, C. J. C.; Bell, S. G.; Wong, L.-L., P450BM3 (CYP102A1): connecting the dots. *Chem. Soc. Rev.* **2012,** *41* (3), 1218-1260.
12. Thistlethwaite, S.; Jeffreys, L. N.; Girvan, H. M.; McLean, K. J.; Munro, A. W., A Promiscuous Bacterial P450: The Unparalleled Diversity of BM3 in Pharmaceutical Metabolism. *Int. J. Mol. Sci.* **2021,** *22* (21), 11380.
13. Zuo, R., et al., Engineered P450 biocatalysts show improved activity and regio-promiscuity in aromatic nitration. *Sci. Rep.* **2017,** *7* (1), 842.
14. Munro, A. W., et al., P450 BM3: the very model of a modern flavocytochrome. *Trends in Biochemical Sciences* **2002,** *27* (5), 250-257.
15. Narhi, L. O.; Fulco, A. J., Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in Bacillus megaterium. *J. Biol. Chem.* **1987,** *262* (14), 6683-6690.
16. Neufeld, K.; Marienhagen, J.; Schwaneberg, U.; Pietruszka, J., Benzylic hydroxylation of aromatic compounds by P450 BM3. *Green Chem.* **2013,** *15* (9), 2408-2421.
17. Santos, G. d. A.; Dhoke, G. V.; Davari, M. D.; Ruff, A. J.; Schwaneberg, U., Directed Evolution of P450 BM3 towards Functionalization of Aromatic O-Heterocycles. *Int. J. Mol. Sci.* **2019,** *20* (13), 3353.
18. Jung, S. T.; Lauchli, R.; Arnold, F. H., Cytochrome P450: taming a wild type enzyme. *Curr. Opin. Biotechnol.* **2011,** *22* (6), 809-817.

19. Li, H. M.; Mei, L. H.; Urlacher, V. B.; Schmid, R. D., Cytochrome P450 BM-3 evolved by random and saturation mutagenesis as an effective indole-hydroxylating catalyst. *Appl Biochem Biotechnol* **2008,** *144* (1), 27-36.
20. Janocha, S.; Schmitz, D.; Bernhardt, R., Terpene Hydroxylation with Microbial Cytochrome P450 Monooxygenases. In *Biotechnology of Isoprenoids*, Schrader, J.; Bohlmann, J., Eds. Springer International Publishing: Cham, 2015; pp 215-250.
21. Farinas, Edgardo T.; Schwaneberg, U.; Glieder, A.; Arnold, Frances H., Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation. *Adv. Synth. Catal.* **2001,** *343* (6-7), 601-606.
22. Dennig, A.; Lülsdorf, N.; Liu, H.; Schwaneberg, U., Regioselective o-Hydroxylation of Monosubstituted Benzenes by P450 BM3. *Angew. Chem. Int. Ed.* **2013,** *52* (32), 8459-8462.
23. Rousseau, O., et al., Indigo Formation and Rapid NADPH Consumption Provide Robust Prediction of Raspberry Ketone Synthesis by Engineered Cytochrome P450 BM3. *ChemCatChem* **2020,** *12* (3), 837-845.
24. Ruff, A. J.; Dennig, A.; Wirtz, G.; Blanusa, M.; Schwaneberg, U., Flow Cytometer-Based High-Throughput Screening System for Accelerated Directed Evolution of P450 Monooxygenases. *ACS Catal.* **2012,** *2* (12), 2724-2728.
25. Furuya, T., et al., Characterization of Orphan Monooxygenases by Rapid Substrate Screening Using FT-ICR Mass Spectrometry. *Chem. Biol.* **2008,** *15* (6), 563-572.
26. Lim, K. B.; Ozbal, C. C.; Kassel, D. B., High-Throughput Mass Spectrometric Cytochrome P450 Inhibition Screening. In *Cytochrome P450 Protocols*, Phillips, I. R.; Shephard, E. A.; Ortiz de Montellano, P. R., Eds. Humana Press: Totowa, NJ, 2013; pp 25-50.
27. de Rond, T., et al., A High-Throughput Mass Spectrometric Enzyme Activity Assay Enabling the Discovery of Cytochrome P450 Biocatalysts. *Angew. Chem. Int. Ed.* **2019,** *58* (30), 10114-10119.
28. Park, S.-H., et al., Engineering Bacterial Cytochrome P450 (P450) BM3 into a Prototype with Human P450 Enzyme Activity Using Indigo Formation. *Drug Metab. Dispos.* **2010,** *38* (5), 732-739.
29. Kelly, P. P., et al., Active site diversification of P450cam with indole generates catalysts for benzylic oxidation reactions. *Beilstein J. Org. Chem.* **2015,** *11*, 1713-1720.
30. Çelik, A.; Speight, R. E.; Turner, N. J., Identification of broad specificity P450CAM variants by primary screening against indole as substrate. *Chem Commun (Camb)* **2005,** (29), 3652-3654.
31. Wong, T. S.; Wu, N.; Roccatano, D.; Zacharias, M.; Schwaneberg, U., Sensitive Assay for Laboratory Evolution of Hydroxylases toward Aromatic and Heterocyclic Compounds. *J. Biomol. Screen.* **2005,** *10* (3), 246-252.
32. Alcalde, M.; Farinas, E. T.; Arnold, F. H., Colorimetric High-Throughput Assay for Alkene Epoxidation Catalyzed by Cytochrome P450 BM-3 Variant 139-3. *J. Biomol. Screen.* **2004,** *9* (2), 141-146.
33. Meng, S., et al., High Throughput Screening Method for Engineering P450 Towards Terminal Hydroxylation of Fatty Acids. *J. Biobased Mater. Bioenergy* **2019,** *13* (1), 79-85.
34. Tang, W. L.; Li, Z.; Zhao, H., Inverting the enantioselectivity of P450pyr monooxygenase by directed evolution. *ChemComm* **2010,** *46* (30), 5461-5463.
35. Markel, U., et al., A Photoclick-Based High-Throughput Screening for the Directed Evolution of Decarboxylase OleT. *Chemistry* **2021,** *27* (3), 954-958.

36. Weingartner, A. M., et al., A hydroquinone-specific screening system for directed P450 evolution. *Appl Microbiol Biotechnol* **2018,** *102* (22), 9657-9667.
37. Turner, N. J., Agar Plate-based Assays. In *Enzyme Assays*, 2005; pp 137-161.
38. Park, S. H., et al., Engineering bacterial cytochrome P450 (P450) BM3 into a prototype with human P450 enzyme activity using indigo formation. *Drug Metab. Dispos.* **2010,** *38* (5), 732-9.
39. Tyagi, R.; Lai, R.; Duggleby, R. G., A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step. *BMC Biotechnol.* **2004,** *4* (1), 2.
40. Li, H.; Poulos, T. L., The structure of the cytochrome p450BM-3 haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat. Struct. Biol.* **1997,** *4* (2), 140-146.
41. Huang, W.-C., et al., Filling a Hole in Cytochrome P450 BM3 Improves Substrate Binding and Catalytic Efficiency. *J. Mol. Biol.* **2007,** *373* (3), 633-651.
42. Li, Q. S.; Schwaneberg, U.; Fischer, P.; Schmid, R. D., Directed evolution of the fatty-acid hydroxylase P450 BM-3 into an indole-hydroxylating catalyst. *Chemistry* **2000,** *6* (9), 1531-6.
43. Ma, L., et al., Enzymatic synthesis of indigo derivatives by tuning P450 BM3 peroxygenases. *Synth Syst Biotechnol.* **2023,** *8* (3), 452-461.
44. Ravichandran, K. G.; Boddupalli, S. S.; Hasermann, C. A.; Peterson, J. A.; Deisenhofer, J., Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* **1993,** *261* (5122), 731-6.
45. Li, H.; Poulos, T. L., Modeling protein-substrate interactions in the heme domain of cytochrome P450(BM-3). *Acta Crystallogr. D Biol. Crystallogr.* **1995,** *51* (Pt 1), 21-32.
46. Ma, L., et al., Enzymatic synthesis of indigo derivatives by tuning P450 BM3 peroxygenases. *Synth Syst Biotechnol* **2023,** *8* (3), 452-461.
47. Ma, L., et al., Development of MEMS directed evolution strategy for multiplied throughput and convergent evolution of cytochrome P450 enzymes. *Sci China Life Sci* **2022,** *65* (3), 550-560.
48. Kong, F., et al., Evolving a P450BM3 Peroxygenase for the Production of Indigoid Dyes from Indoles. *ChemCatChem* **2022,** *14* (24), e202201151.
49. Li, Q.-S.; Schwaneberg, U.; Fischer, P.; Schmid, R. D., Directed Evolution of the Fatty-Acid Hydroxylase P450 BM-3 into an Indole-Hydroxylating Catalyst. *Chem. Eur. J.* **2000,** *6* (9), 1531-1536.
50. Li, Q. S.; Ogawa, J.; Schmid, R. D.; Shimizu, S., Indole Hydroxylation by Bacterial Cytochrome P450 BM-3 and Modulation of Activity by Cumene Hydroperoxide. *Bioscience, Biotechnology, and Biochemistry* **2005,** *69* (2), 293-300.
51. Omura, T.; Sato, R., The Carbon Monoxide-binding Pigment of Liver Microsomes: I. Evidence for its hemoprotein nature. *J. Biol. Chem.* **1964,** *239* (7), 2370-2378.
52. Capoferri, L., et al., Insights into regioselective metabolism of mefenamic acid by cytochrome P450 BM3 mutants through crystallography, docking, molecular dynamics, and free energy calculations. *Proteins: Struct., Funct., Bioinf.* **2016,** *84* (3), 383-396.
53. Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A., Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **2001,** *40* (45), 13456-65.
54. Urban, P.; Lautier, T.; Pompon, D.; Truan, G., Ligand Access Channels in Cytochrome P450 Enzymes: A Review. *Int. J. Mol. Sci.* **2018,** *19* (6).

55. Ebert, M. C. C. J. C.; Guzman Espinola, J.; Lamoureux, G.; Pelletier, J. N., Substrate-Specific Screening for Mutational Hotspots Using Biased Molecular Dynamics Simulations. *ACS Catal.* **2017,** *7* (10), 6786-6797.
56. Daff, S. N., et al., Redox control of the catalytic cycle of flavocytochrome P-450 BM3. *Biochemistry* **1997,** *36* (45), 13816-23.
57. Sevrioukova, I. F.; Li, H.; Zhang, H.; Peterson, J. A.; Poulos, T. L., Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (5), 1863-8.
58. Darve, E.; Rodríguez-Gómez, D.; Pohorille, A., Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **2008,** *128* (14).
59. Comer, J., et al., The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B* **2015,** *119* (3), 1129-1151.
60. Cojocaru, V.; Winn, P. J.; Wade, R. C., The ins and outs of cytochrome P450s. *Biochim Biophys Acta* **2007,** *1770* (3), 390-401.
61. Ebert, M. C. C. J. C.; Dürr, S. L.; A. Houle, A.; Lamoureux, G.; Pelletier, J. N., Evolution of P450 Monooxygenases toward Formation of Transient Channels and Exclusion of Nonproductive Gases. *ACS Catal.* **2016,** *6* (11), 7426-7437.
62. Li, Q. S.; Ogawa, J.; Schmid, R. D.; Shimizu, S., Engineering cytochrome P450 BM-3 for oxidation of polycyclic aromatic hydrocarbons. *Appl Environ Microbiol* **2001,** *67* (12), 5735-9.
63. Whitehouse, C. J. C., et al., Evolved CYP102A1 (P450BM3) variants oxidise a range of non-natural substrates and offer new selectivity options. *ChemComm* **2008,** (8), 966-968.
64. Dietrich, M.; Do, T. A.; Schmid, R. D.; Pleiss, J.; Urlacher, V. B., Altering the regioselectivity of the subterminal fatty acid hydroxylase P450 BM-3 towards γ- and δ-positions. *J. Biotechnol.* **2009,** *139* (1), 115-117.
65. Fansher, D. J.; Besna, J. N.; Fendri, A.; Pelletier, J. N., Choose Your Own Adventure: A Comprehensive Database of Reactions Catalyzed by Cytochrome P450 BM3 Variants. *ACS Catal.* **2024,** *14* (8), 5560-5592.
66. Di Nardo, G.; Gilardi, G., Optimization of the Bacterial Cytochrome P450 BM3 System for the Production of Human Drug Metabolites. *Int. J. Mol. Sci.* **2012,** *13* (12), 15901-15924.
67. Tang, L., et al., Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques* **2012,** *52* (3), 149-58.
68. Sanchis, J., et al., Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates. *Appl Microbiol Biotechnol* **2008,** *81* (2), 387-97.
69. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-38.