Directed Evolution of the NdmB Enzyme for the Synthesis of Paraxanthine

Mengjin Zhang, Wanbin Xing, Lingli Chen, Shutong Tan, Narna, Yaqi Han, Xuanyu Meng, Wenling Zhu, Yijun Liu, Siyu Wang, Lu Zhang, Xiyu Zhao, Chang Liu, Wei Meng, Guangchao Sui, Pengchao Wang

Abstract:

Paraxanthine (PX) is a naturally occurring dietary ingredient with a wide range of applications in promoting physical performance and treating psychiatric disorders. However, its production still relies on low-yield and poorly specific chemical synthesis. The current study modified the NdmB enzyme through directed evolution to enhance its catalytic efficiency in PX synthesis. An efficient screening system based on a whole-cell biosensor was designed and machine learning models were employed to discover that alterations at glutamine 289 (Q289) significantly enhanced the catalytic efficiency of NdmB in promoting the N3-demethylation of caffeine. Site-directed saturation mutagenesis experiments showed that the $NdmB^{Q289T}$ mutant increased PX production from 0 g/L to 1.22 g/L. Further optimization of the promoter region boosted PX yield to 5.42 g/L, a 29-fold increase over the highest reported yield. This research shows the potential of combining directed evolution, biosensors, and computational tools to optimize biocatalysis and guide future PX production.

Keywords: **paraxanthine, NdmB, directed evolution, whole-cell biosensor, machine learning, metabolic engineering**

1. Introduction

Paraxanthine (PX) is a naturally occurring dietary component and the primary metabolite of caffeine in the human body[1]. Compared to caffeine, PX has several advantages, including lower mutagenicity, genotoxicity, and cytotoxicity, while exhibiting stronger effects on physical stimulation. Due to its excitatory properties and endurance-enhancing effects, PX holds potential applications in performance enhancement and may improve mental state by modulating central nervous system activity, showing promise in the treatment of mental disorders[2,3,4]. However, the high price, currently at 1,522 $/g, and extremely low natural presence of PX have significantly hindered its further research and applications. As PX is rarely found naturally, its extraction methods from natural resources are rare. Chemical synthesis of PX primarily involves methylation reactions of selective sites, but the uncontrollable nature of the process makes it difficult to stop at the desired stage, which poses challenges for efficient PX production[5].

Therefore, biosynthetic methods with higher yields and environmental friendliness have gradually been applied to PX production. It has been reported that a strain of *Pseudomonas putida* CBB5 (*P. putida* CBB5) isolated from soil can utilize caffeine as its sole carbon and nitrogen source[6]. This strain contains a gene cluster that degrades caffeine to xanthine through stepwise demethylation. The enzymes NdmA and NdmB catalyze demethylation at the N1 and N3 positions, respectively, while the NdmCDE

reductase complex catalyzes demethylation at the N7 position, ultimately degrading caffeine into xanthine[7,8,9]. Most current methods for the biosynthesis of methylxanthines are based on this gene cluster, and efficient production of 1-methylxanthine (1-MX), 3-methylxanthine (3-MX) and 7-methylxanthine (7-MX) has already been achieved[10,11,12]. As a similar demethylase, NdmB shares high homology with NdmA. Previous studies have constructed NdmA double mutants and replaced loop structures near the active site with the corresponding structures of NdmB to enable N3-demethylation of caffeine, leading to PX synthesis. However, the biosynthetic yield was still low, with a maximum yield of only 183.8 mg/L, accompanied by a large amount of by-product (7-MX). The primary reasons for this are the low recognition ability of NdmB for caffeine and poor substrate specificity, leading to inefficiencies in the N3-demethylation reaction[13,14,15].

Directed evolution, which harnesses the power of genetic diversity and selection, has been widely used to engineer proteins with desired functions[16]. The semi-rational design, which introduces mutations at key sites and conducts efficient screening in a broad sequence space, offers advantages over many purely rational or non-rational design approaches[17]. Moreover, whole-cell biosensors, which couple enzyme activity with bacterial growth, can directly link mutant function to biological fitness, providing an effective screening strategy[18].

In this study, we constructed a screening system based on a whole-cell biosensor that links the catalytic activity of NdmB mutants to bacterial growth. By integrating machine learning models, we systematically analyzed the relationship between

mutation sites and product formation, and performed site-directed saturation mutagenesis on the key residue Q289 to assess its impact on PX production. Additionally, we employed promoter optimization to further enhance PX yield, and ultimately we developed innovative strategies and tools for the efficient biosynthesis of PX.

2. Results

2.1 Development of a Whole-Cell Biosensor and a Mutant Screening Strategy

2.1.1 Biosensor Development and Evaluation

Transcription factors and nucleic acid aptamers are often utilized in the development of high-throughput screening tools due to their high specificity and tunability. In this study, we initially attempted to construct a screening platform based on these two elements, but the results were not as expected (Figure S1 and S2)[19,20,21]. We then shifted to developing a whole-cell biosensor by knocking out the *guaB* gene in *E. coli*, thereby blocking the DNA and RNA synthesis pathways and rendering the bacteria unable to grow in the absence of exogenous xanthine. Additionally, the *ndmDCEA* gene cluster was introduced, where *ndmA* catalyzes N1-demethylation and *ndmC* catalyzes N7-demethylation, which can promote the conversion of PX into xanthine. Bacterial survival depends on efficient PX production by NdmB-catalyzed N3-demethylation of caffeine (**Figure 1A and B**). To validate the functionality of the pYB1s-ndmDCEA plasmid-driven biosensor system, the plasmid was introduced into

an *E. coli* strain with the *guaB* gene knocked out. The transformed strain was then inoculated into the medium containing xanthine, PX, theobromine (Tb), caffeine, theophylline (Tp), or a control medium. We observed a significant increase in OD_{600} only in the medium containing xanthine and PX, while OD_{600} remained low under the other conditions. This confirmed that both xanthine and PX were essential for bacterial growth (Figure 1C). Additionally, we observed bacterial growth at different concentrations of PX, further demonstrating that the constructed sensor system could well respond to the presence of different substrate concentrations (Figure 1D).

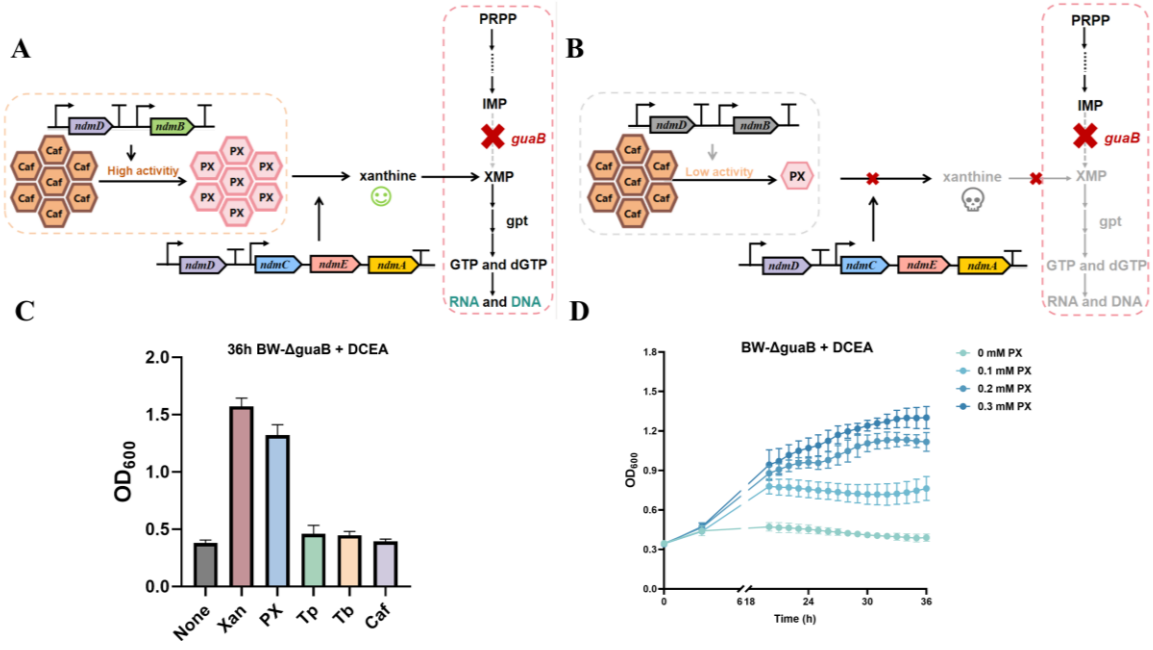

Figure 1. Construction and validation of the whole-cell biosensor for PX.

(A) High-activity strains catalyze the demethylation of caffeine to produce PX through the action of NdmB. Under the regulation of the pYB1s-ndmDCEA plasmid, PX is further converted into xanthine, restoring the growth of the *guaB* knockout

strain. **(B)** Low-activity strains fail to efficiently produce PX, resulting in no bacterial growth. **(C)** OD_{600} values of strain growth under different substrate conditions. The most significant growth was observed in the medium containing PX. **(D)** The growth of the strain positively correlated with PX concentration, demonstrating that the sensor could successfully detect and utilize PX.

2.1.2 Screening of ndmB Mutants Using the Whole-Cell Biosensor

After confirming the effectiveness of the whole-cell biosensor, we used it to screen *ndmB* mutants with enhanced capabilities to degrade caffeine. Previous studies have analyzed the substrate binding interactions within the SRPBCC domain of the NdmA enzyme for both caffeine and theobromine, identifying key residues (Q289 and L293) as critical for substrate specificity[19]. Homology modeling of NdmA and NdmB also revealed non-conserved residues within the active site pocket (W256, C267, and M271), which could likely contribute to substrate binding orientation and specificity[13]. Based on this prediction, we performed combinatorial mutations at two sets of residues and employed the biosensor plasmid to screen for clones based on their viability.

In the screening process, *E. coli* cells carrying mutant *ndmB* genes were inoculated into the medium, and after a series of steps, we screened for clones with higher OD_{600} values (**Figure 2**). The reaction products were then analyzed by high-performance liquid chromatography (HPLC), and the sequence of the mutants were determined by DNA sequencing.

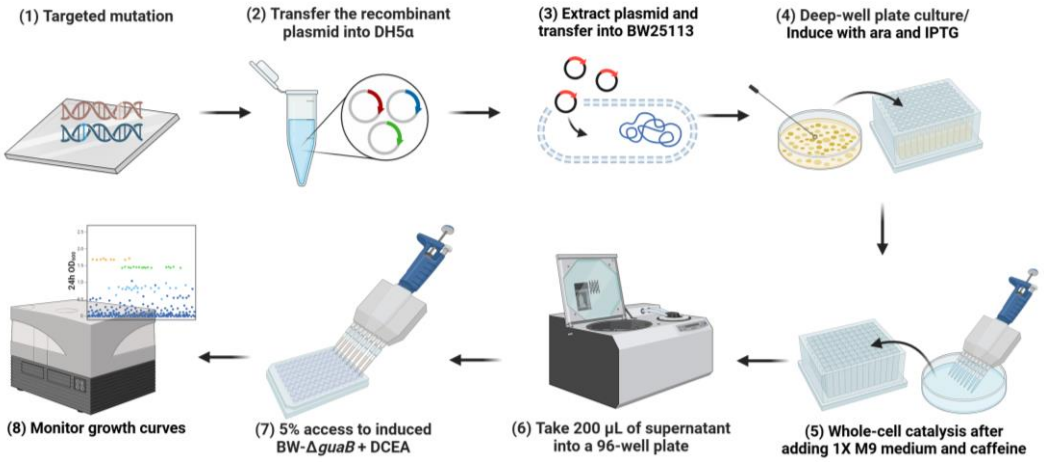

Figure 2. Schematic diagram of the directed evolution and mutant screening process for NdmB.

The process includes: (1) Site-directed mutagenesis of key residues in NdmB using degenerate primers; (2) Transformation of recombinant plasmids into *E. coli* DH5α competent cells for amplification; (3) Plasmid extraction and transformation into the *E. coli* BW25113 strain; (4) Induction of mutant NdmB expression in deep-well plates containing arabinose and IPTG; (5) Addition of 1 × M9 medium and caffeine for whole-cell catalysis; (6) Collection of the supernatant by centrifugation, transfer to a 96-well plate; (7) Addition of the whole-cell biosensor; (8) Monitoring of bacterial growth curves to evaluate the catalytic efficiency of each mutant in producing PX.

2.1.3 Screening Results Analysis

After conducting HPLC to detect the reaction products and DNA sequencing to determine the mutant sequences, we analyzed the mutation types and their associated products (Figure S3)**.** Surprisingly, the strains with high OD_{600} values produced not only PX but also other methylxanthine derivatives, such as 1-MX and 7-MX. These

compounds were also converted into xanthine in the sensing module, allowing the strains to survive. This indicates that NdmB mutants gained demethylation activity not only at the N3 position but also at the N1 and even N7 positions (Figure 3). The diversity of products generated by these NdmB mutants significantly increased the complexity of our screening, and further investigation was necessary to identify mutants that could selectively promote PX production.

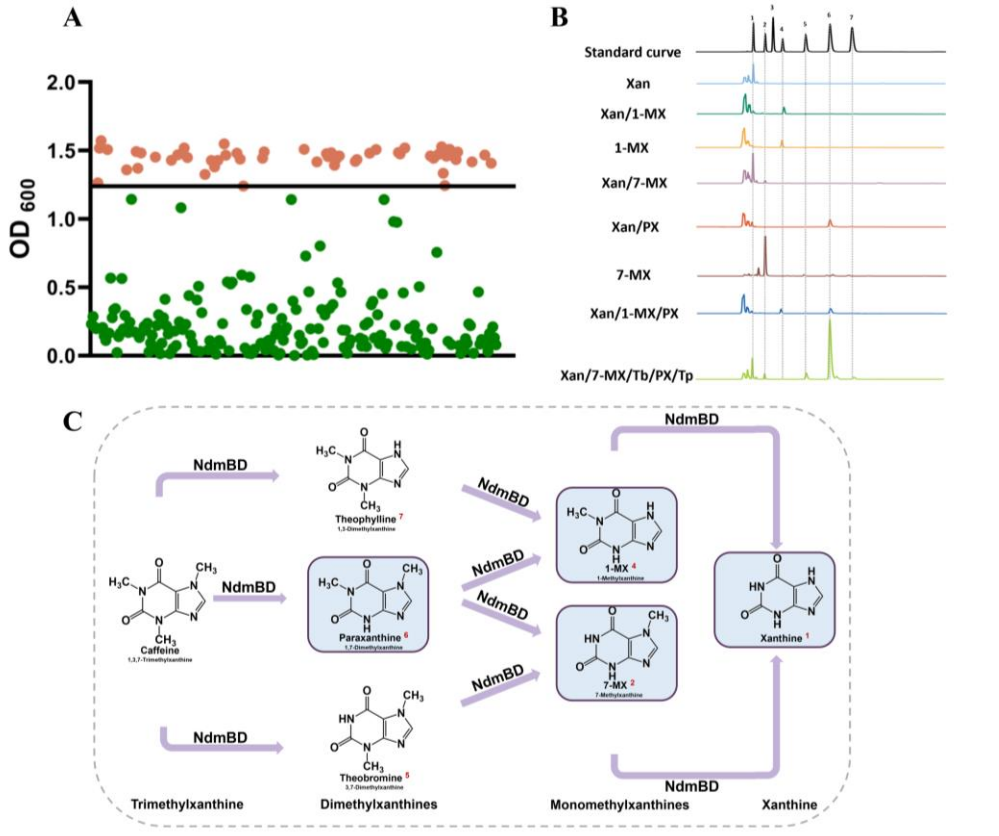

Figure 3. Screening results and by-product analysis of NdmB mutants.

(A) Screening results of mutants using the PX whole-cell biosensor. The figure highlights positive clones with higher OD_{600} values (in orange), indicating that these strains likely have a stronger capability for PX production. **(B** and **C)** HPLC analysis of the standard curves for products from different strains. The analysis confirmed that the mutants produced not only PX but also other demethylated by-products, such as 1-

MX and 7-MX.

2.2 Machine Learning-Based Protein Optimization and Molecular Mechanism Study

2.2.1 Machine Learning

The integration of machine learning with evolutionary engineering can significantly enhance the efficiency and effectiveness of protein optimization[22,23,24]. To identify mutation sites favoring PX production, we employed a Ridge regression model for machine learning analysis. We first compiled a dataset of experimental results from protein mutants, encoding the mutation types using one-hot encoding to generate feature matrices. Product yields served as the target variable. The dataset was randomly divided into training and test sets, which were used for model fitting and evaluating the model's generalizability, respectively. Mean squared error (MSE) was used as the performance metric to assess the prediction accuracy and error distribution (**Figure 4A**). Analysis of the regression coefficients revealed that mutations at Q289 showed a strong positive effect on the synthesis of both PX and 7-MX, while their impact on other by-products was relatively minor. The experimental results provide critical insights for subsequent site-directed mutagenesis experiments at Q289 (**Figure 4B**).

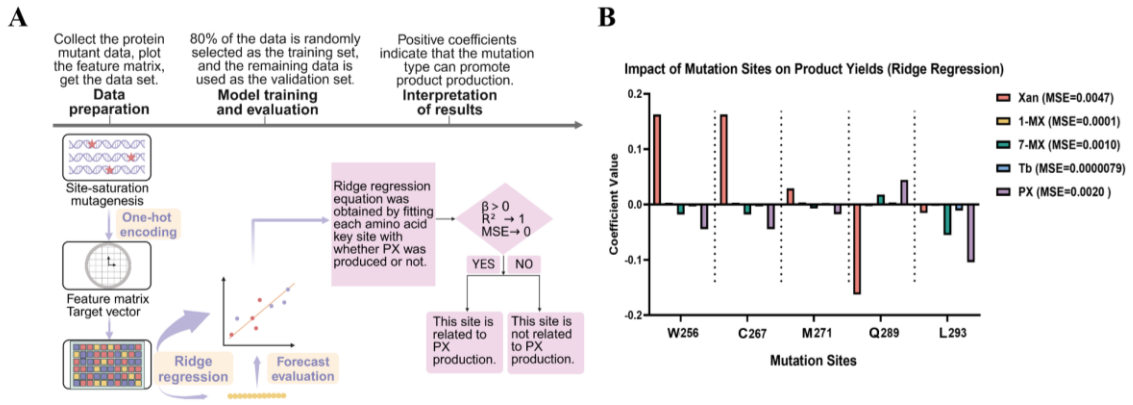

Figure 4. Product analysis of NdmB mutants based on the Ridge regression model.

(A) Schematic diagram of the machine learning workflow. The regression coefficients' signs were used to determine the impact of each mutation site on product formation. **(B)** The Ridge regression model analyzed the influence of five mutation sites (W256, C267, M271, Q289, L293) on the generation of different products. Positive regression coefficients indicated that the mutations had a favorable effect on methylxanthine production, with Q289 mutations significantly enhancing the production of PX and 7-MX, while other sites tended to promote the formation of different demethylated products.

2.2.2 Site-Directed Saturation Mutagenesis and Functional Optimization of the Q289 Residue

Based on the predictions from the machine learning model, we designed degenerate primers to perform site-directed saturation mutagenesis at Q289. The mutants were screened using the whole-cell biosensor, and strains with higher OD_{600}

values were selected for further analysis through HPLC and DNA sequencing. The experimental data revealed two key findings: first, when Q289 was mutated to threonine ($NdmB^{Q289T}$), PX production increased significantly, reaching 1.22 g/L (6.76 mM), which is 7.3 times higher than the previously reported maximum yield. Second, as the side chain (R-group) decreased in size (threonine → serine → alanine), a distinct shift in the types and ratios of products was observed: smaller R-groups resulted in a gradual decrease in PX production, while the yield of 7-MX increased accordingly. This pattern indicates that the size of the R-group at the position 289 significantly influences the enzyme's substrate selectivity: as the R-group becomes smaller, NdmB not only catalyzes N3-demethylation but also exhibits N1-demethylation activity (**Figure 5**).

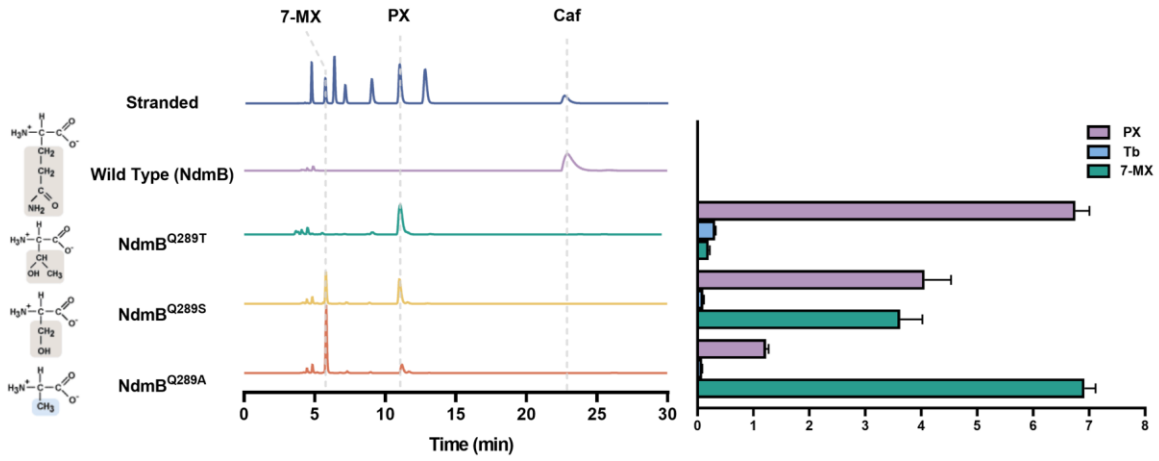

Figure 5. HPLC analysis of catalytic products from NdmB mutants.

The $NdmB^{Q289T}$ mutant exhibited a significant increase in PX production, while the $NdmB^{Q289A}$ mutant showed a preference for producing 7-MX. Larger R-groups (such as threonine) favor PX production, while smaller R-groups (such as alanine) tend to generate 7-MX.

2.2.3 Molecular Docking and Molecular Dynamics Simulations

Site-directed mutagenesis may have a profound impact on the selection of enzymes with desired activities[25]. To precisely simulate the behavior of the target protein under whole-cell catalysis conditions, we optimized the NdmB structure obtained from the PDB using molecular dynamics (MD) simulations. MD simulations allow us to track the conformational dynamics of proteins at the atomic level[26,27,28]. After generating the models for wild-type NdmB and its mutants ($NdmB^{Q289T}$, $NdmB^{Q289S}$, and $NdmB^{Q289A}$), we performed molecular docking of caffeine with NdmB[29]. The interactions between NdmB and methylxanthine structures, including steric hindrance and hydrogen bonding, jointly regulated substrate positioning within the active site.

The docking results indicated that the steric hindrance of Q289 affected PX production. (**Figure 6A and B**). Furthermore, when PX was used as a substrate, its carbonyl group interacted with the Thr residue, preventing PX from entering the active pocket further. This simulation likely explains the $NdmB^{Q289T}$ mutant's specificity for N3-demethylation (**Figure 6C**). In the $NdmB^{Q289S}$ mutant, once PX entered the active site, a part of it was further converted to 7-MX, while the rest remained unreacted due to steric hindrance from the Ser residue's flexible side chain (**Figure 6D and E**). In the $NdmB^{Q289A}$ mutant, PX did not interact with A289, allowing the N1-methyl group of PX to approach closer to the active center, thereby increasing 7-MX production (**Figure 6F**). Overall, the site-directed saturation mutagenesis of Q289 significantly produced

multiple mutants with different efficiency of demethylation reactions, which offered valuable insights for future enzyme engineering.

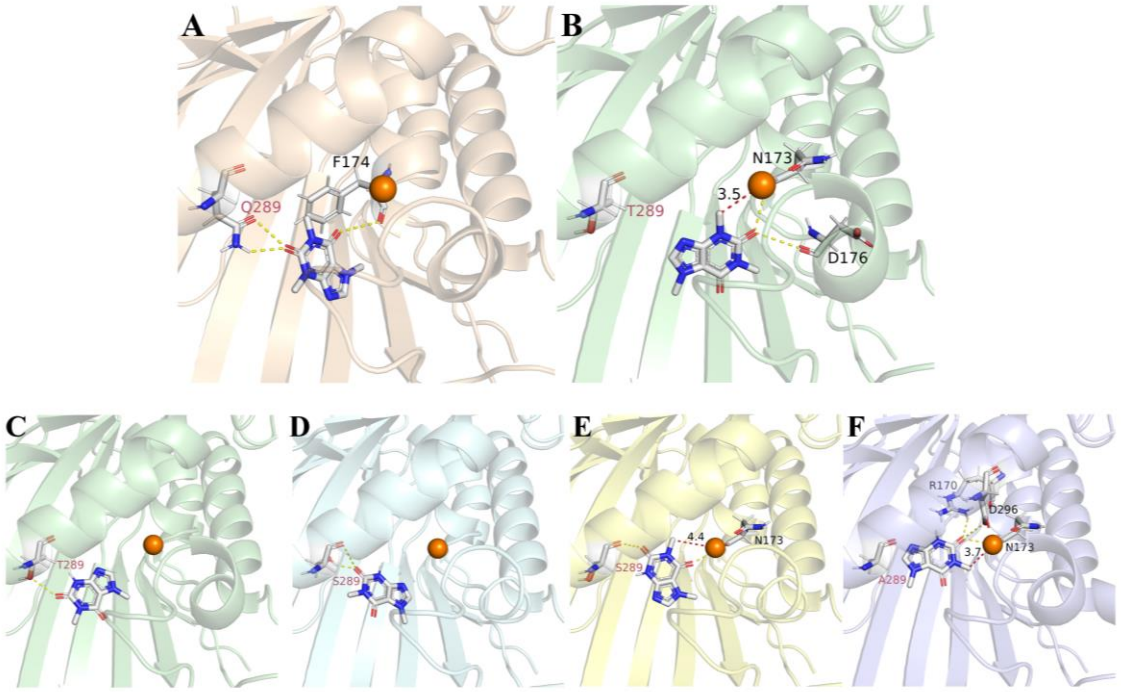

Figure 6. Molecular docking analysis of NdmB mutants with substrates.

(A) Docking analysis of wild-type NdmB with caffeine, showing that the key residue Q289 blocks caffeine from entering the active pocket. **(B)** Docking analysis of $NdmB^{Q289T}$ with caffeine indicates that reduced steric hindrance allows the N3 methyl group of caffeine to approach the active center at a distance of 3.5 Å. **(C)** Docking analysis of $NdmB^{Q289T}$ with PX reveals that the absence of the N3 methyl group exposes a carbonyl group that interacts with the Thr residue, which prevents further demethylation. **(D** and **E)** Docking analysis of $NdmB^{Q289S}$ with PX demonstrates steric effects from the Ser side chain, partially converting PX to 7-MX. **(F)** Docking analysis of $NdmB^{Q289A}$ with PX indicates that lack of A289 interaction with PX allows the N1 methyl group to approach the active center, leading to increased

production of 7-MX.

2.3 Optimization of Metabolic Pathways to Increase PX Yield

Promoter optimization is a key strategy to improve metabolic pathways, particularly the -10 region, which plays a crucial role in RNA polymerase binding and transcription initiation[30,31]. In this study, we performed random mutagenesis in the six bases of the -10 region of the pJ23107 promoter, followed by screening using a whole-cell biosensor. Next, we selected the strains with the highest activity for larger-scale fermentation and analyzed their methylxanthine production. The results demonstrated that under conditions of 200 OD and 40 mM substrate concentration, the PX yield of the E8 strain reached 5.42 g/L, which represents a 4.4-fold increase compared to the $NdmB^{Q289T}$ mutant without metabolic engineering, and a 29-fold improvement over the highest reported PX yield to date. These findings highlight the success of optimizing metabolic pathways (**Figure 7**).

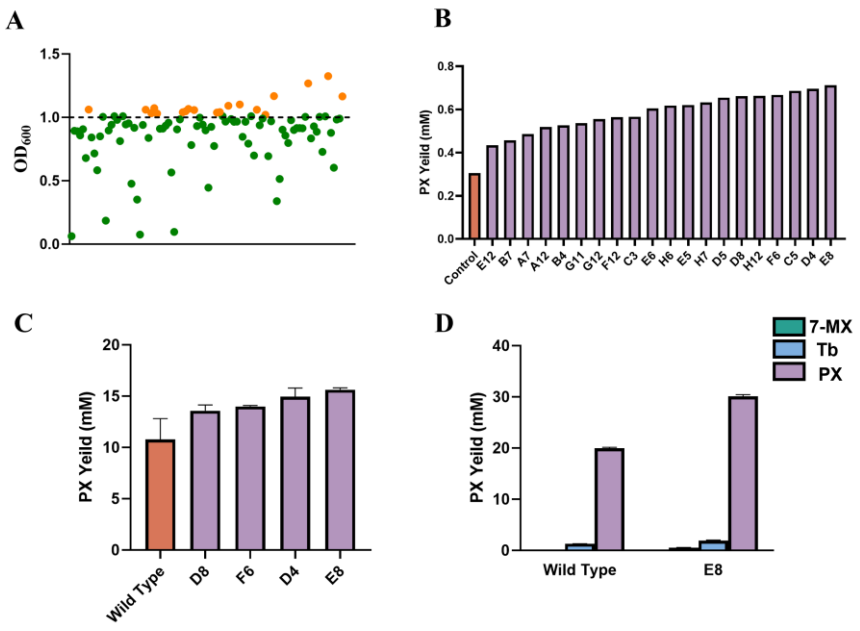

Figure 7. Screening of promoter -10 region mutants and PX yield analysis.

(A) Screening results of the promoter -10 region mutants. Strains with OD_{600} values higher than 1.0 are shown in orange shade. Since $NdmB^{Q289T}$ was used as the template to create additional mutants, the overall OD_{600} values in the screening were overall higher than it. **(B)** Twenty strains with OD_{600} values greater than 1.0 were selected, and their PX production was measured using HPLC. **(C)** Further large-scale cultivation of the top four mutants (D8, F6, D4, E8) was performed. Under conditions of 100 OD and a substrate concentration of 20 mM, the PX yield of strain E8 reached 15.62 ± 0.13 mM. **(D)** When OD_{600} was increased to 200 and the substrate concentration raised to 40 mM, the PX yield of strain E8 further increased to 30.11 ± 0.23 mM.

3. Discussion

This study successfully enhanced PX production through directed evolution of the NdmB enzyme. First, we employed semi-rational design to generate beneficial mutants. Given the complexity of mutant products, we utilized a Ridge regression model for machine learning, which identified Q289 as a key residue influencing substrate promiscuity in NdmB. Subsequent targeted mutations at Q289 significantly improved the enzyme's ability to selectively catalyze N3-demethylation of caffeine, demonstrating the potential of computational tools in guiding enzyme design[32]. Notably, optimization of the promoter region increased the PX yield of the $NdmB^{Q289T}$ mutant to 5.42 g/L, representing a 29-fold improvement over the highest previously reported

yield.

While previous research has primarily focused on the modifications to the NdmA enzyme, the novelty of this study lies in structural modifications to NdmB[33]. By combining directed evolution with computational methods, we were able to significantly enhance catalytic efficiency, demonstrating a forward-thinking approach in the field of protein engineering. Additionally, this study applied a whole-cell biosensor, coupled with an *E. coli* strain lacking the *guaB* gene, to directly link PX production to the bacterial growth, providing a specific and efficient screening method.

Despite achieving a significant increase in PX yield through directed evolution, the broad specificity of NdmB's products remains a challenge. The generation of multiple methylxanthine derivatives complicates the identification of the most effective mutants. Future research should focus on further refining the active site structure through molecular dynamics simulations and crystallographic studies, which would elucidate the molecular mechanisms by which different R-groups affect enzyme activity and substrate specificity. Additionally, machine learning could be adapted to explore non-conserved residues in the active site or nearby regions, potentially uncovering more efficient mutation combinations[34,35].

In conclusion, this study demonstrates the substantial potential of integrating directed evolution, whole-cell biosensors, and computational modeling in optimizing industrial biocatalytic processes. We anticipate further integration of metabolic engineering and fermentation technologies to enhance the commercial viability of PX

biosynthesis.

4. Materials and Methods

4.1 Chemicals, Strains, and Materials

4.1.1 Compounds and Standards

7-MX (purity 97%) was purchased from Aladdin Biotech Co., Ltd. (Shanghai, China). 1-MX (purity 97%), 3-MX (purity 99%), PX (purity 97%), and theobromine (purity 99%) were obtained from Macklin Biochemical Co., Ltd. (Shanghai, China), while theophylline (purity 99%) was sourced from Coolaber Co., Ltd. (Beijing, China). All other reagents used in the experiments were standard molecular biology reagents. Caffeine was derived from guarana extract containing approximately 22% caffeine, as pure caffeine was not directly obtained. The standard curve for caffeine was established using the indirect method with 7-MX and theobromine (Figure S4).

4.1.2 Materials

During the screening process using the whole-cell biosensor, strains were cultured in LB medium. The LB liquid medium which was prepared by dissolving 10 g tryptone, 5 g yeast extract, and 10 g NaCl in 1 L deionized water. For LB agar plates, 15 g/L agar was included inadded to the formulation. In constructing and screening the mutant library, *E. coli* strains were cultured in M9 medium composed of 17 g $Na_2HPO_4 \cdot 7H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 4 g/L glucose. Cultures were grown at 37°C with shaking at 200 rpm, while cultivation in 96-well plates occurred at 30°C with shaking at 850 rpm. Streptomycin, chloramphenicol, and

kanamycin were used as antibiotics at a working concentration of 100 mg/L. The ZYM5052 medium used for protein expression contained 10 g/L tryptone, 5 g/L yeast extract, 50 mM phosphate, 50 mM ammonium chloride, 5 mM sodium sulfate, 0.5% glycerol, 0.05% glucose, 0.2 mM magnesium sulfate, and various essential metal ions. The transformation solution was prepared using 50 mM Tris-HCl buffer (pH 9.0).

4.1.3 Strains

The bacterial strains and plasmids used in this study are listed in Supplementary Table 1. Competent cells of *E. coli* DH5α, BW25113, and BW-Δ*guaB* used in the experiments were prepared in the laboratory, with some competent cells purchased from Weidi Biotech Co., Ltd. (Shanghai, China). All chemicals were of analytical grade and purchased from reputable commercial suppliers. The primary strain used in this study was the gene-edited *E. coli* BW-Δ*guaB*. Primers for plasmid construction and site-directed mutagenesis were synthesized by Ruibiotech (Harbin, China); details are available in Supplementary Table 2.

Gene synthesized in this study involved *ndmA* (JQ061127), *ndmB* (JQ061128), *ndmC* (JQ061129), *ndmD* (JQ061130), and *ndmE* (KC778191) genes. All genes were codon-optimized and synthesized by Integrated DNA Technologies (IDT). Plasmid construction was carried out using the Gibson Assembly method (Gibson Assembly, C116) and cloning was completed with the Gibson Assembly Cloning Kit (New England Biolabs, NEB)[36]. Arabinose induced the expression of the pBAD promoter, while the lac operon regulated the plac promoter. To construct a cofactor regeneration system, the *frmA* and *frmB* genes were amplified from *E. coli* K12, and the *formate*

dehydrogenase (FDH) gene was optimized and synthesized by Qingke Biotechnology Co., Ltd.

To knock out the *guaB* gene in the *E. coli* genome, RED recombination was employed, utilizing pKD46 plasmid for gene deletion[37]. Initially, the targeted fragments were prepared: guaB-up and guaB-down fragments were amplified by PCR using BW25113 genome as the template with specific primers. The kanamycin resistance gene was amplified using pccdK2-up-kan-down as the template; then, the plasmid pccdK2-O was digested with KpnI to obtain the pccdK vector fragment. These three fragments were assembled using the Gibson Assembly method to obtain the pccdK-up-kan-down targeting fragment. Next, gene deletion experiments were conducted: the pKD46 plasmid was introduced into *E. coli* BW25113 to prepare electrocompetent cells. The competent cells (100 ul) were incubated on ice with more than 200 ng of the targeting fragment for 10-30 min before being electroporated. After electroporation, 1000 μL of pre-warmed LB medium (37°C) was immediately added, gently mixed, and transferred to a 1.5 mL centrifuge tube, which was incubated at 30°C with shaking at 150 rpm for 45-60 min. The cells were then spread onto dual-antibiotic selective plates containing kanamycin (Kana) and streptomycin (Str) for screening. Finally, colony PCR was performed to verify gene deletion. Since the targeting fragment contained the kanamycin resistance gene, only successfully integrated strains survive, while others failed to form colonies on the selective plates.

4.2 Construction of the Mutant Library

4.2.1 Site-Directed Mutagenesis

Site-directed saturation mutagenesis was employed to construct the mutant library. Target sequences were selected, focusing on key residues in the NdmB protein, such as W256 and Q289, for mutagenesis. Degenerate primers (NNN, where N represents any nucleotide) were designed to introduce all possible amino acid substitutions at the chosen sites. The PCR reaction conditions were as follows: an initial denaturation at 98°C for 5 min, followed by 25 cycles of 98°C denaturation for 30 sec, x°C annealing for 30 sec, and 72°C extension for y sec. This was followed by a final extension at 72°C for 5 min, and the reaction was held at 25°C.

4.2.2 Vector Construction

The mutated *ndmB* gene was ligated into the vector using the Golden Gate assembly method to construct the mutant library[38]. The recombinant plasmids were transformed into *E. coli* DH5α to generate the mutant strains. Following established methods[39], the transformed cultures were plated on LB agar containing 50 mg/mL of the appropriate antibiotics and incubated at 37°C for 12 h. Colony PCR was performed to screen the clones, and plasmids from positive clones were extracted, amplified, and sequenced to verify mutation accuracy.

4.3 Mutant Library Screening

4.3.1 Induction and Whole-Cell Catalysis

Constructed mutant strains were inoculated into ZYM-5052 medium, consisting of 10 g/L tryptone, 5 g/L yeast extract, 50 mM phosphate, 50 mM ammonium chloride,

5 mM sodium sulfate, 0.5% glycerol, 0.05% glucose, 0.2 mM magnesium sulfate, and various essential metal ions. Cultures were grown at 37°C with shaking at 200 rpm. When the OD_{600} reached 0.5 to 0.8, IPTG and arabinose were added to induce *ndmB* expression, and cultures were incubated further at 25°C with shaking at 200 rpm for 18 h. After cultivation, cells were collected by centrifugation at 4000 rpm for 10 min, and the supernatant was discarded. The cells were resuspended in a filtered 1 mM caffeine solution, and the whole-cell catalysis reaction was carried out at 20°C and 200 rpm for 18 h.

Simultaneously, the pYB1s-ndmDCEA plasmid was transformed into the *E. coli* BW-Δ*guaB* strain, which grows slowly due to the knockout of the *guaB* gene. To aid growth, double the amount of yeast extract was added to the ZY medium. After the culture achieved sufficient turbidity, induction was initiated under the same conditions, with incubation at 25°C and 200 rpm for 18 h. Following induction, cells were harvested by centrifugation, and the supernatant was removed. The cells were then washed with deionized water to eliminate any residual xanthine.

4.3.2 Screening of Mutants Using the Whole-Cell Biosensor for Paraxanthine

Washed cells were resuspended and centrifuged at 4000 rpm for 10 min, and 200 μL of the supernatant was collected. To this, 10 μL of an additive solution (comprising glucose, arabinose, IPTG, trace elements, magnesium sulfate, calcium chloride, and streptomycin) was added, along with 10 μL of the pYB1s-ndmDCEA solution. The mixture was transferred into a 96-well plate, and the initial OD_{600} was measured. The OD_{600} was subsequently monitored every 12 h for a total of three readings over 36 h to

assess bacterial growth. Strains exhibiting robust growth were selected for further liquid culture testing and analysis.

4.4 High-Performance Liquid Chromatography (HPLC) Analysis

4.4.1 Sample Preparation

The whole-cell catalysis reaction mixture was diluted with sterile water to a concentration at which the product was fully dissolved. A 1 mL aliquot of the diluted mixture was centrifuged at 10,000 g for 10 min. The supernatant was collected into a new 1.5 mL microcentrifuge tube and filtered through a 0.22 μm membrane filter. The filtered solution was then transferred to an HPLC vial for analysis.

4.4.2 HPLC Detection

The analysis was conducted using an Agilent 1100 series HPLC system equipped with an Agilent SB-C8 column (4.6 × 300 mm², 5 μm particle size). The injection volume was set to 10 μL, with a column temperature of 35°C and a flow rate of 0.7 mL/min. Methylxanthine compounds were detected at a wavelength of 280 nm. The mobile phase consisted of a mixture of methanol, water, and acetic acid at a ratio of 15:85:0.5 (v/v/v).

4.5 Molecular Docking and Molecular Dynamics (MD) Simulations

To predict interactions between the NdmB enzyme and its substrates, molecular docking an MD simulations were employed. The standard structure of NdmB was downloaded from the Protein Data Bank. Common protein structure prediction methods, such as homology modeling and AlphaFold, did not accurately predict NdmB's

structure, prompting the use of MD simulations for more precise modeling under catalytic conditions.

For MD simulations, the NdmB structure was processed using GROMACS. As the GROMACS library lacked force field parameters for the Fe^{3+} ion, relevant parameters were imported from the literature to simulate the behavior of Fe^{3+} in an aqueous environment. The AMBER99SB-ILDN force field was selected for its improved side-chain torsion potentials, enhancing simulation accuracy[40]. Although the [2Fe-2S] cluster lacked available force field data, it was omitted from the simulations since the study focused on the interactions between the SPRBCC domain and the substrates.

To replicate the catalytic conditions, the environment was set to pH 9.0, a temperature of 293.15 K, and a pressure of 1 bar, mimicking experimental reaction conditions. According to the *E. coli* intracellular pH gradient model, a pH of 9.0 outside the cell corresponds to an intracellular pH of 7.781[41]. The protein structure was protonated using PDB2PQR, and Na^+ and Cl^- ions were added to neutralize the system. Energy minimization was performed using a combination of steepest descent and conjugate gradient methods, reducing the maximum force to less than 50. After energy minimization, NVT (constant temperature and volume) and NPT (constant temperature and pressure) simulations were conducted. The system was considered equilibrated when the potential energy and temperature stabilized near target values (Figure S5). The same method was used to analyze NdmB mutants, ensuring consistency across all simulation parameters.

Once accurate enzyme structures were obtained, AutoDockTools (ADT) was used for molecular docking of caffeine with NdmB. The receptor was prepared by removing water molecules and adding hydrogen atoms, and the structure was saved as a PDBQT file. The caffeine structure was downloaded from PubChem, and hydrogen atoms were added before saving the structure in PDBQT format. Molecular docking analysis was performed using AutoDock Vina 1.1.2, to evaluate the binding modes of substrates like caffeine and PX within the active site of the NdmB enzyme[42]. Multiple docking runs were performed, and the lowest energy conformations were selected for detailed analysis, shedding light on the effects of different mutations on substrate binding. Subsequently, we conducted a detailed comparison of the structural dynamics and interactions of wild-type NdmB and the $NdmB^{Q289T}$ mutant in the presence of caffeine. The analysis focused on key aspects such as the distance variations relative to the non-heme iron, radial distribution function (RDF), principal component analysis (PCA), and covariance mapping. Particular attention was given to hydrogen bond formation, aiming to further elucidate the interaction mechanisms between the mutant and caffeine during the binding process (Figure S6 and S7).

4.6 Random Mutagenesis of the -10 Region of the Promoter

Degenerate primers were designed to perform random mutagenesis on the -10 region of the *ndmB* gene promoter, generating a diverse mutant library. The PCR products were digested with restriction enzymes and purified. Using the Golden Gate assembly method, the mutated promoter fragments were ligated with a linearized vector, and the assembly products were transformed into selective media containing Str for

screening to obtain multiple promoter mutant strains. These selected mutants were then induced for expression and subjected to whole-cell catalysis experiments. The screening system carrying the pYB1s-ndmDCEA plasmid was employed to evaluate differences in catalytic activity among the various mutants.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

This research was supported by funding from Northeast Forestry University. We sincerely thank the strong support from the College of Life Sciences and the Aulin College, Northeast Forestry University. We are grateful for the experimental platform provided by the university, the guidance of our teachers, and the assistance of our mentors.

References

1. Benowitz, N. L., Jacob, P., Mayan, H., & Denaro, C, Sympathomimetic effects of paraxanthine and caffeine in humans*. *Clinical Pharmacology & Therapeutics*, *58*(6), 684–691, (1995).

   DOI: https://doi.org/10.1016/0009-9236(95)90025-x

2. Orrú, M., Guitart, X., Karcz-Kubicha, M., Solinas, M., Justinova, Z., Barodia, S. K., Zanoveli, J., Cortes, A., Lluis, C., Casado, V., Moeller, F. G., & Ferré, S,

Psychostimulant pharmacological profile of paraxanthine, the main metabolite of caffeine in humans. *Neuropharmacology*, *67*, 476–484,(2013).

DOI: https://doi.org/10.1016/j.neuropharm.2012.11.029

3. Yoo, C., Xing, D., Gonzalez, D. E., Jenkins, V., Nottingham, K., Dickerson, B., Leonard, M., Ko, J., Lewis, M. H., Faries, M., Kephart, W., Purpura, M., Jäger, R., Wells, S. D., Liao, K., Sowinski, R., Rasmussen, C. J., & Kreider, R. B, Paraxanthine provides greater improvement in cognitive function than caffeine after performing a 10-km run. *Journal of the International Society of Sports Nutrition*, *21*(1),(2024).

   DOI: https://doi.org/10.1080/15502783.2024.2352779

4. Jäger, R., Purpura, M., Wells, S. D., Liao, K., & Godavarthi, A, Paraxanthine supplementation increases muscle mass, strength, and endurance in mice. *Nutrients*, *14*(4), 893, (2022).

   DOI: https://doi.org/10.3390/nu14040893

5. Algharrawi, K. H. R., & Subramanian, M, Production of 7-methylxanthine from Theobromine by Metabolically Engineered E. coli. *Iraqi Journal of Chemical and Petroleum Engineering*, *21*(3), 19–27,(2020).

   DOI: https://doi.org/10.31699/ijcpe.2020.3.3

6. Summers, R.M., Metabolism, enzymology, and genetic characterization of caffeine degradation by pseudomonas putida CBB5. PhD thesis, University of Iowa, (2011).

DOI: https://doi.org/10.17077/etd.qto8ggr4

7. Algharrawi, K. H. R., Summers, R. M., Gopishetty, S., & Subramanian, M, Direct conversion of theophylline to 3-methylxanthine by metabolically engineered E. coli. *Microbial Cell Factories,* 14(1), (2015).

   DOI: https://doi.org/10.1186/s12934-015-0395-1

8. Blecher, R., & Lingens, F, The metabolism of caffeine by APseudomonas putidaStrain. *Hoppe-Seyler´S Zeitschrift Für Physiologische Chemie*, *358*(2), 807–818, (1977).

   DOI: https://doi.org/10.1515/bchm2.1977.358.2.807

9. Summers, R. M., Seffernick, J. L., Quandt, E. M., Yu, C. L., Barrick, J. E., & Subramanian, M. V, Caffeine Junkie: an Unprecedented Glutathione S-Transferase-Dependent Oxygenase Required for Caffeine Degradation by Pseudomonas putida CBB5. *Journal of Bacteriology*, *195*(17), 3933–3939, (2013).

   DOI: https://doi.org/10.1128/jb.00585-13

10. Summers, R., Gopishetty, S., Mohanty, S., & Subramanian, M, New genetic insights to consider coffee waste as feedstock for fuel, feed, and chemicals. *Open Chemistry*, *12*(12), 1271–1279, (2014).

    DOI: https://doi.org/10.2478/s11532-014-0550-2

11. McKeague, M., Wang, Y., Cravens, A., Win, M. N., & Smolke, C. D, Engineering a microbial platform for de novo biosynthesis of diverse methylxanthines.

Metabolic Engineering, *38*, 191–203, (2016).

DOI: https://doi.org/10.1016/j.ymben.2016.08.003

12. Mock, M. B., Zhang, S., Pakulski, K., Hutchison, C., Kapperman, M., Dreischarf, T., & Summers, R. M, Production of 1-methylxanthine via the biodegradation of theophylline by an optimized Escherichia coli strain. *Journal of Biotechnology*, *379*, 25–32, (2024).

    DOI: https://doi.org/10.1016/j.jbiotec.2023.11.005

13. Kim, J. H., Kim, B. H., Brooks, S., Kang, S. Y., Summers, R. M., & Song, H. K, Structural and Mechanistic Insights into Caffeine Degradation by the Bacterial N-Demethylase Complex. *Journal of Molecular Biology*, *431*(19), 3647–3661, (2019).

    DOI: https://doi.org/10.1016/j.jmb.2019.08.004

14. Mock, M. B., Mills, S. B., Cyrus, A., Campo, H., Dreischarf, T., Strock, S., & Summers, R. M, Biocatalytic production and purification of the high-value biochemical paraxanthine. *Biotechnology and Bioprocess Engineering*, *27*(4), 640–651, (2022).

    DOI: https://doi.org/10.1007/s12257-021-0301-0

15. Mills, S. B., Mock, M. B., & Summers, R. M. Rational Protein Engineering of Bacterial N-demethylases to Create Biocatalysts for the Production of Methylxanthines. *bioRxiv*, (2021).

DOI: https://doi.org/10.1101/2021.12.17.472166

16. Chen, J., Wang, Y., Zheng, P., & Sun, J, Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends in Biotechnology*, *40*(7), 773–776, (2022).

    DOI: https://doi.org/10.1016/j.tibtech.2022.01.010

17. Romero, P. A., & Arnold, F. H., Exploring protein fitness landscapes by directed evolution. *Nature Reviews. Molecular Cell Biology,* 10(12), 866–876, (2009).

    DOI: https://doi.org/10.1038/nrm2805

18. Williams, T. C., Pretorius, I. S., & Paulsen, I. T., Synthetic Evolution of Metabolic Productivity Using Biosensors. *Trends in Biotechnology,* 34(5), 371–381,(2016).

    DOI: https://doi.org/10.1016/j.tibtech.2016.02.002

19. Summers, R. M., Louie, T. M., Yu, C., Gakhar, L., Louie, K. C., & Subramanian, M, Novel, highly specific N -Demethylases enable bacteria to live on caffeine and related purine alkaloids. *Journal of Bacteriology*, *194*(8), 2041–2049, (2012).

    DOI: https://doi.org/10.1128/jb.06637-11

20. Wrist, A., Sun, W., & Summers, R. M., The Theophylline Aptamer: 25 Years as an Important Tool in Cellular Engineering Research. *ACS Synthetic Biology,* 9(4), 682–697, (2020).

    DOI: https://doi.org/10.1021/acssynbio.9b00475

21. Guedez, A., Sherman, M., & Ryu, Y., Dual genetic selection of the theophylline riboswitch with altered aptamer specificity for caffeine. *Biochemical and Biophysical Research Communications,* 579, 105–109, (2021).

DOI: https://doi.org/10.1016/j.bbrc.2021.09.058

22. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., & Arnold, F. H, Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences*, *116*(18), 8852–8858, (2019).

DOI: https://doi.org/10.1073/pnas.1901979116

23. Liu, M, Faculty Opinions recommendation of Abundant associations with gene expression complicate GWAS follow-up. *Nature Genetics*, 51(5), 768–769, (2019).

DOI: https://doi.org/10.3410/f.735645309.793561318

24. Raman, S., Rogers, J. K., Taylor, N. D., & Church, G. M, Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences*, *111*(50), 17803–17808, (2014).

DOI: https://doi.org/10.1073/pnas.1409523111

25. Ishida, T, Effects of Point Mutation on Enzymatic Activity: Correlation between Protein Electronic Structure and Motion in Chorismate Mutase Reaction. *Journal of the American Chemical Society*, *132*(20), 7104–7118, (2010).

DOI: https://doi.org/10.1021/ja100744h

26. Kukol, A, Lipid models for United-Atom molecular dynamics simulations of proteins. *Journal of Chemical Theory and Computation*, *5*(3), 615–626, (2009).

DOI: https://doi.org/10.1021/ct8003468

27. Adcock, S. A., & McCammon, J. A, Molecular Dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5), 1589–1615, (2006).

DOI: https://doi.org/10.1002/chin.200630297

28. Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., & Wriggers, W, Atomic-Level characterization of the structural dynamics of proteins. *Science*, *330*(6002), 341–346, (2010).

DOI: https://doi.org/10.1126/science.1187409

29. Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J, Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols*, *11*(5), 905–919, (2016).

DOI: https://doi.org/10.1038/nprot.2016.051

30. Zhou, C., Ye, B., Cheng, S., Zhao, L., Liu, Y., Jiang, J., & Yan, X, Promoter engineering enables overproduction of foreign proteins from a single copy expression cassette in Bacillus subtilis. *Microbial Cell Factories*, *18*(1), 111–121, (2019).

DOI: https://doi.org/10.1186/s12934-019-1159-0

31. Alper, H., Fischer, C., Nevoigt, E., & Stephanopoulos, G., Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences,* 102(36), 12678–12683, (2005).

DOI: https://doi.org/10.1073/pnas.0504604102

32. Wittmann, B. J., Johnston, K. E., Wu, Z., & Arnold, F. H, Advances in machine learning for directed evolution. *Current Opinion in Structural Biology,* 69, 11–18, (2021).

DOI: https://doi.org/10.1016/j.sbi.2021.01.008

33. Zhao, L., Wei, J., Hu, Y., Pi, D., Jiang, M., & Lang, T., Caffeine Synthesis and Its Mechanism and Application by Microbial Degradation, A Review. *Foods,* 12(14), 2721–2721, (2023).

DOI: https://doi.org/10.3390/foods12142721

34. Radivojević, T., Costello, Z., Workman, K., & Martin, H. G, A machine learning Automated Recommendation Tool for synthetic biology. *Nature Communications*, *11*(1), 4879, (2020).

DOI: https://doi.org/10.1038/s41467-020-18008-4

35. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., & Arnold, F. H, Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences*, *116*(18), 8852–8858, (2019b).

DOI: https://doi.org/10.1073/pnas.1901979116

36. Xia, Y., Li, K., Li, J., Wang, T., Gu, L., & Xun, L, T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Research*, *47*(3), e15, (2018).

DOI: https://doi.org/10.1093/nar/gky1169

37. Thomason, L., Court, D. L., Bubunenko, M., Costantino, N., Wilson, H., Datta, S., & Oppenheim, A, Recombineering: Genetic engineering in bacteria using homologous recombination. *Current Protocols in Molecular Biology*, 78(1), 1.16.1–1.16.39, (2007).

DOI: https://doi.org/10.1002/0471142727.mb0116s78

38. Engler, C., & Marillonnet, S, Golden Gate cloning. *Methods in Molecular Biology*, 1116, 119–131, (2013).

DOI: https://doi.org/10.1007/978-1-62703-764-8_9

39. Froger, A., & Hall, J. E, Transformation of Plasmid DNA into *E. coli* Using the Heat Shock Method. *Journal of Visualized Experiment*s, 6(6), 253, (2007).

DOI: https://doi.org/10.3791/253

40. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Structure Function and Bioinformatics*, *78*(8), 1950–1958, (2010).

DOI: https://doi.org/10.1002/prot.22711

41. Slonczewski, J. L., Rosen, B. P., Alger, J. R., & Macnab, R. M, pH homeostasis in Escherichia coli: measurement by 31P nuclear magnetic resonance of methylphosphonate and phosphate. *Proceedings of the National Academy of Sciences*, *78*(10), 6271–6275, (1981).

DOI: https://doi.org/10.1073/pnas.78.10.6271

42. Morris, G. M., Huey, R., & Olson, A. J, Using AutoDock for Ligand-Receptor docking. *Current Protocols in Bioinformatics*, *24*(1), (2008).

DOI: https://doi.org/10.1002/0471250953.bi0814s24