

Supplementary Information

Methods

Experimental Validation

Strains, media and chemicals. The sources, identifiers of bacterial strain, and the media and chemicals used in this study are provided in Supplementary Table 11.

Gene Synthesis, Enzyme Expression, and Purification. Representative genes from TnaA Clusters 3–6 were synthesized with codon optimization for *Escherichia coli* expression and cloned into pET-28a (+) via NcoI/XhoI sites. Recombinant plasmids were transformed into *E. coli* BL21(DE3), validated by colony PCR and Sanger sequencing (Cowin Biotech, Jiangsu, China). For protein expression, cells were grown in Luria Broth medium supplemented with kanamycin (50 µg/mL) at 37°C with shaking at 215 rpm until OD₆₀₀ reached 0.5, measured using a spectrophotometer (UV-1800, Shimadzu Corporation, Kyoto, Japan). The culture temperature was then reduced to 18°C and growth continued until OD₆₀₀ reached 0.8, then induced with 100 µM isopropyl-1-thio-D-galactopyranoside (IPTG) and incubated overnight. Following IPTG induction at an OD₆₀₀ of 0.6–0.8, cultures were shifted from 37°C to 18°C and expression was continued for 16 h. This low-temperature induction strategy slows the rate of protein synthesis relative to folding and reduces inclusion body formation, thereby enhancing soluble yield for PLP-dependent enzymes in *E. coli* (1). Cells were harvested by centrifugation at 4,500×g for 20 min at 4°C using a TGL21M high-speed refrigerated centrifuge (Xiangzhi Centrifuge Instrument, Changsha, China). Cell pellets were resuspended in Buffer A (100,000 µM potassium phosphate pH 7.5, 50,000 µM imidazole, 50 µM pyridoxal-5'-phosphate), supplemented with DNase, lysozyme, and a complete EDTA-free protease inhibitor tablet. Cells were lysed by sonication and clarified by centrifugation at 12,000×g for 45 min using a TGL21M high-speed refrigerated centrifuge (Xiangzhi Centrifuge Instrument, Changsha, China). The cleared lysate was applied to a Ni-NTA HP column (GE Healthcare, Uppsala, Sweden) on an AKTA Protein Purification System (GE Healthcare) and washed with Buffer A. Target proteins were eluted with Buffer B (100,000 µM potassium phosphate pH 7.5, 350,000 µM imidazole, 50 µM pyridoxal-5'-phosphate). Eluted fractions were combined and subjected to size exclusion chromatography using a Superose 6 Increase 10/300GL column (GE Healthcare) equilibrated with Buffer C (100,000 µM potassium phosphate pH 7.5, 50 µM pyridoxal-5'-phosphate, 5,000 µM 2-mercaptoethanol).

Fractions were analyzed by SDS-PAGE, and those with >95% purity were pooled, concentrated, and stored at -80°C.

Circular Dichroism Spectroscopy for Structural Verification. To confirm correct folding of recombinant TnaA variants beyond SDS-PAGE purity, far-UV circular dichroism (CD) spectra were recorded for all four variants (Clusters 3-6) using a CD spectropolarimeter (Jasco 1500, Jasco Corp., Tokyo, Japan), with heat-inactivated enzymes (incubated at 95 °C for 10 min) as the negative controls. Measurements were performed in Buffer D (10 mM potassium phosphate pH 7.5, 50 µM pyridoxal-5'-phosphate, 50 mM NaCl, 5% glycerol, 1 mM DTT, ddH₂O was added to a final volume of 50 mL) at 25°C using a quartz cuvette with a 1-mm path length. Protein concentration was determined using a Bradford Protein Assay Kit (Beyotime Biotechnology, Shanghai, China). Only samples with protein concentrations ≥ 0.2 mg/mL were used for subsequent CD measurements. Spectra were recorded from 190 to 260 nm at a scan rate of 50 nm/min, with three accumulations per spectrum for signal averaging. Buffer baselines were subtracted. Secondary structure content was estimated using the CONTIN algorithm implemented in the CDPro software package.

Enzyme Activity Assay. Reaction mixtures (100 µL) contained 50 µL assay buffer (100,000 µM potassium phosphate, pH 7.5), 10 µL purified TnaA enzyme (final concentration 1.9 µM), and 40 µL L-Trp (final concentrations 2,500 µM). Negative controls reactions included: (1) Heat inactivated TnaA. The purified TnaA was incubated at 95 °C for 10 min to abolish catalytic activity and cooled on ice prior to use. Complete inactivation was confirmed prior to use. (2) Reactions without the PLP cofactor to confirm that all measured activity is PLP-dependent. (3) Bovine serum albumin (BSA, 1.9 µM) at equivalent protein concentration to exclude non-specific protein effects on the Kovács's colorimetric reaction. Reactions were incubated at 37°C and quenched at six time points (0, 1, 5, 10, 15 and 20 min) with 100 µL of 1,000,000 µM hydrochloric acid. Samples were incubated on ice for 10 min, then centrifuged at 13,000×g for 10 min. Supernatant (100 µL) was mixed with 150 µL Kovacs's reagent and incubated at ambient temperature for 15 min. Absorbance was measured at 540 nm using a Tecan Spark Multimode Microplate Reader (Tecan Group Ltd., Switzerland). Indole concentrations were determined using a standard curve prepared with authentic indole (0-1000 µM).

Enzyme Kinetic Analysis. Initial reaction velocities (v_0) were determined from the linear portion of indole production time courses. Indole formation was monitored at 9

substrate concentrations (0.1-2,100 μM) over 20 min, with measurements at 0, 1, 5, 10, 15 and 20 min. Kinetic experiments were performed with four independently prepared enzyme batches for each TnaA variant (Clusters 3-6), with duplicate measurements at each substrate concentration.

Enzyme concentration ($[E]_{\text{total}}$) was determined by active-site titration with S-(o-nitrophenyl)-L-cysteine (SOPC) at 370 nm (2). Specifically, purified TnaA (10 $\mu\text{g/mL}$) was incubated with SOPC concentrations (2,570 μM) in assay buffer at 25 $^{\circ}\text{C}$. Absorbance at 370 nm was monitored over 5 min with measurements at 10 s intervals. $[E]_{\text{total}}$ was calculated as $\Delta A_{370}/(\epsilon_{370} \times b)$, where $\epsilon_{370} = 7,800 \text{ M}^{-1}\text{cm}^{-1}$ and $b = 1 \text{ cm}$ (cuvette path length), assuming 1:1 stoichiometry between SOPC and active site. Heat-inactivated TnaA served as negative control. All determinations were performed in triplicate with relative standard deviation (RSD) $< 5\%$.

Kinetic parameters (K_M and V_{max}) were estimated by nonlinear least-squares fitting to the Michaelis-Menten equation using GraphPad Prism v10.2.3 (GraphPad Software, San Diego, CA, USA). For each independent enzyme preparation, 95% confidence intervals (CIs) were calculated from the regression results. All Michaelis-Menten curve fits achieved $R^2 > 0.90$. The turnover number (k_{cat}) was calculated for each preparation as $V_{\text{max}}/[E]_{\text{total}}$. The 95% CI of k_{cat} was calculated by dividing the V_{max} confidence limits by $[E]_{\text{total}}$. The catalytic efficiency (k_{cat}/K_M) was determined for each enzyme preparation. Its 95% CI was computed by Monte Carlo simulation (10,000 iterations) based on the fitting-derived uncertainties of K_M and V_{max} , with $[E]_{\text{total}}$ held constant. For statistical comparison, kinetic parameters among the four variants were compared using one-way analysis of variance (ANOVA) followed by Tukey's honestly significant difference (HSD) post-hoc test for pairwise comparisons. Statistical significance was defined as $P < 0.05$.

Isolation and Identification of Target Bacteria. Bacterial strains were isolated from rumen fluid of HG-fed cattle using L-Trp-enriched medium using an anaerobic workstation (DG250, Don Whitley Scientific, Shipley, UK) and indole content was determined using high-performance liquid chromatography (HPLC) on a Shimadzu LC10ADvp system equipped with an auto-injector (SIL-10ADvp) and fluorescence detector (RF-10Axl). Genomic DNA was extracted using the Bacterial DNA Kit (D3350, Solarbio Science & Technology, Beijing, China), and the 16S rRNA gene was amplified with primers 27F/1492R. Reference 16S rRNA sequences from target indole-producing MAGs (*Prevotella* SOTU346 and Lachnospiraceae SOTU1618) were

103 extracted using cmscan in Infernal v1.1.5 (3) against the Rfam bacterial small subunit
104 rRNA model (RF00177) (4). Sequences with E-values < 0.01 were retrieved using
105 bedtools v2.31.0 (5) for downstream strain identification. PCR amplicons were
106 sequenced and compared against the extracted 16S rRNA sequences from *Prevotella*
107 SOTU346 and Lachnospiraceae SOTU1618. Isolates exhibiting ≥ 95 % sequence
108 identity were selected for whole-genome sequencing (6). Indole concentrations were
109 quantified using HPLC. The chromatography was performed isocratically with a mobile
110 phase consisting of 70% acetic acid solution and 30% isopropanol at a flow rate of 1
111 mL/min, with an injection volume of 5 μ L. The chromatographic separation was carried
112 out on a reverse-phase platinum C18 column (150 x 4.6 mm). Fluorescence excitation
113 was set to 285 nm and emission to 350 nm for the detection of indole. Data acquisition
114 and peak processing were performed using Shimadzu Class-VP software version 5.032
115 (7).

116 **Cross-validation of indole quantification methods.** Kovác's reagent assay was
117 used for high-throughput screening of enzymatic activity in 96-well plates (semi-
118 quantitative), while HPLC provided precise quantification in bacterial culture
119 supernatants. To ensure consistency between methods, we performed a cross-validation
120 by measuring identical samples ($n = 16$, spanning 0-500 μ M indole) using both methods
121 in parallel. Linear regression between Kovác's absorbance (OD₅₄₀) and HPLC-
122 determined indole concentration yielded $R^2 = 0.851$ ($P < 0.001$, Supplementary Fig. 5),
123 confirming strong agreement across the quantification range. All IC₅₀ values reported
124 in the manuscript were determined using Kovác's reagent under standardized conditions.
125 All physiological indole measurements were determined by HPLC. The two methods
126 were not mixed within any single dataset or comparison.

127 **Characterization of Indole Production.** Isolated strains were cultured in L-Trp-
128 enriched medium at 39 °C under anaerobic conditions. Two negative control strains,
129 *Butyrivibrio hungatei* LGM 122 and *Eubacterium* sp. I LGM 112, both from our
130 previous study (8) and confirmed to lack the *tnaA* gene and indole production capability,
131 were included to validate the indole production phenotype of test strains. Effects of
132 glucose concentration (0-2 g/L) and pH (5.6, 6.3, 7.0) on growth and indole production
133 were assessed over 24 hours in triplicate. Growth was monitored by OD₆₀₀
134 measurements, and indole production was quantified by HPLC analysis at 2-hour
135 intervals.

Relationship of Indole Production and LPS production. *Prevotella* SOTU346 was cultured in L-Trp-enriched medium containing low-, middle- and high-glucose concentrations (0.5, 1 and 2 g/L glucose, pH 6.3) under anaerobic conditions. Growth was monitored by measuring OD₆₀₀. Culture samples were collected at 2-hour intervals for temporal analysis. Indole production was quantified as described previously. LPS was prepared by the phenol-water method (9) LPS content was determined using a 96-well microplate and the phenol-sulfuric acid method. The specific procedure was as follows: the LPS sample was mixed with 5% phenol, followed by the addition of concentrated sulfuric acid (98%). After incubation at room temperature for 30 min for color development, the absorbance was measured at 490 nm. Quantitative analysis was performed using a standard curve ($R^2 = 0.99$).

Development of Targeted Inhibitors

Molecular Docking. A total of 12,250 plant-derived compounds from the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP) (10) were prepared using Open Babel v3.1.1 (11). The TnaA structure predicted by AlphaFold3 was prepared with AutoDockTools 1.5.7 (12). Docking was performed using AutoDock Vina v1.2.5 (13) with an exhaustiveness of 32. The binding site was defined based on the *Proteus vulgaris* TnaA-L-Trp complex (PDB: 8V6P). A cubic search space of $20 \times 20 \times 20 \text{ \AA}^3$ was centered on the active site. Docking poses with binding affinity scores ≤ -7.0 kcal/mol were selected as potential inhibitors for experimental validation.

Inhibitor Effects on Enzyme Activity. Potential inhibitors were purchased as solids and dissolved in DMSO and assay buffer (100,000 μM potassium phosphate, pH 7.5; 1:1) to a final concentration of 100,000 μM , while SOPC was dissolved in DMSO at a concentration of 2,570 μM , aliquoted and stored at -80°C until further use. Serial dilutions of inhibitor stocks were prepared to achieve final inhibitor concentrations ranging from 0 to 100 μM in the assay mixture. Reactions were conducted in 96-well, black, clear-bottom assay plates (Costar, Corning Inc., Corning, NY, USA) at 37°C . Assay mixtures contained 5 μL purified TnaA enzymes from Clusters 3 and 6 (varying final concentrations between 1.4-40 μM), 50 μL assay buffer, and 5 μL test compound or vehicle control, with a final DMSO concentration of 2% (v/v). To optimize signal-to-background ratio and generate robust dose-response curves, enzyme concentrations were tested across the range of 1.4-40 μM to determine the optimal concentration for each cluster that yielded maximal inhibition discrimination. Vehicle control reactions

containing 2% DMSO without inhibitor were used to establish baseline enzyme activity. Negative controls for IC₅₀ determinations included: (1) heat-inactivated TnaA (prepared as described above) exposed to luteolin across the full concentration range (0-100 µM), confirming that the inhibition curve reflects active TnaA catalysis rather than compound interference with the detection system; (2) BSA (varying final concentrations between 1.4-40 µM) treated with luteolin across the full concentration range, confirming that luteolin does not non-specifically react with the Kovács's reagent; (3) Luteolin in substrate-containing reaction buffer without enzyme, confirming the absence of spontaneous indole generation; and (4) 2% DMSO vehicle controls to confirm that DMSO alone does not inhibit enzyme activity.

Following pre-incubation at 37°C for 5 minutes, 45 µL of 1,100 µM SOPC was added to initiate the reaction. Substrate degradation was measured via absorbance at 370 nm using a Tecan Spark Multimode Microplate Reader (Tecan Group Ltd., Switzerland). Indole product concentrations were determined using a standard curve (0-900 µM). Initial velocity values were normalized to the 2% DMSO vehicle control baseline. The initial velocity curves were fit using linear regression in Excel, inhibitor concentrations were log transformed, and inhibition curves were fitted using non-linear regression in GraphPad Prism. IC₅₀ values were converted into K_i values using the Cheng-Prusoff equation.

Statistical Analysis. The kinetic parameters of all TnaA variants were determined from four independently prepared enzyme batches, with duplicate measurements at each substrate concentration and 95% CIs were calculated for each parameter. Pairwise comparisons among the four TnaA variants were assessed using Tukey's HSD post-hoc test following one-way ANOVA, with statistical significance defined as $P < 0.05$. For differential enrichment analyses of TnaA variants across diets and co-occurrence network analyses, Benjamini-Hochberg FDR correction was applied to control for multiple testing, with statistical significance defined as $FDR < 0.05$.

Reference

1. Vera A, González-Montalbán N, Arís A, Villaverde A. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol Bioeng.* 2006;96(6):1101-6.
2. Sundararaju B, Antson AA, Phillips RS, Demidkina TV, Barbolina MV, Gollnick P, et al. The Crystal Structure of *Citrobacter freundii* Tyrosine Phenol-lyase Complexed with 3-(4'-Hydroxyphenyl)propionic Acid, Together with Site-

- Directed Mutagenesis and Kinetic Analysis, Demonstrates That Arginine 381 Is Required for Substrate Specificity. *Biochemistry*. 1997;36(21):6502-10.
3. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29(22):2933-5.
4. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res*. 2021;49(D1):D192-D200.
5. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-2.
6. Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*. 2014;64(Pt 2):346-51.
7. Schreurs N, Tavendale M, Lane G, Barry T, Marotti D, McNabb W. Postprandial indole and skatole formation in the rumen when feeding white clover, perennial ryegrass and *Lotus corniculatus*. *Proc N Z Soc Anim Prod*. 1999.
8. Zhang H, Mao W, Lai Z, Zhu Y, Dong J, Li L, et al. Integrated metatranscriptomics identifies *Lachnospiraceae* as keystone taxa regulating rumen biohydrogenation and milk ω -6/ ω -3 polyunsaturated fatty acids ratio in dairy cows. *Microbiome*. 2026;14(1):1.
9. Shen X, Yang Y-B, Gao Y, Wang S, Wang H, Sun M, et al. Lipid A-modified *Escherichia coli* can produce porcine parvovirus virus-like particles with high immunogenicity and minimal endotoxin activity. *Microb Cell Fact*. 2024;23(1):222.
10. Ru J, Li P, Wang J, Zhou W, Li B, Huang C, et al. TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J Cheminform*. 2014;6:13.
11. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform*. 2011;3:33.
12. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785-91.
13. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model*. 2021;61(8):3891-8.