Supporting Information

Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C–C Bond Formation in cyclo-1-Tyr-1-Tyr and Its Coupling with Nucleobases

J. Liu, L. Harken, Y. Yang, X. Xie, S.-M. Li*
Widely Distributed Bifunctional Bacterial Cytochrome P450 Enzymes Catalyze both Intramolecular C-C Bond Formation in cyclo-L-Tyr-L-Tyr and Its Coupling with Nucleobases

Jing Liu,[a, c] Lauritz Harken,[a, c] Yiling Yang,[a] Xiulan Xie,[b] and Shu-Ming Li*[a]

[a] Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany;

[b] Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Straße 4, 35032 Marburg, Germany;

[c] These authors contributed equally: Jing Liu, Lauritz Harken.

* Email: shuming.li@staff.uni-marburg.de
# Table of Contents

**Methods** .......................................................................................................................... 4
- Computer-assisted sequence analysis .................................................................................. 4
- Bacterial strains, plasmids, and growth conditions .............................................................. 4
- Genetic manipulation, PCR amplification, and gene cloning .............................................. 4
- Heterologous gene expression in *S. albus* J1074 and cultivation for secondary metabolite production ........................................................................................................... 4
- LC-MS analysis .................................................................................................................. 5
- Overproduction in *E. coli* and purification of the recombinant GymB₁ – GymB₆ ............. 5
- *In vitro* assays with GymB₁ – GymB₆ .............................................................................. 5
- Large-scale fermentation, extraction, and isolation of secondary metabolites ................. 6
- Determination of production yields ................................................................................... 6
- NMR analysis and structural elucidation ........................................................................... 6
- Cultivation of the *S. albus* transformant in media containing ^1⁵N-labeled salt .............. 7

**Supplementary Tables** .................................................................................................. 8
- Table S1. Comparison of gym genes with known entries .................................................... 8
- Table S2. Bacterial strains used in this study ..................................................................... 9
- Table S3. Key residues for the substrate binding pockets of CDPSs .................................. 10
- Table S4. Cloning and expression constructs used in this study ....................................... 11
- Table S5. Product yields of the *S. albus* transformants harboring guatyromycine and mycocyclosin clusters .......................................................... 12
- Table S6. ^1H NMR data of cYY (1), cYF (2), and mycocyclosin (3) in DMSO-d₆ ............ 13
- Table S7. NMR data of guatyromycine A (4) in DMSO-d₆ .............................................. 14
- Table S8. NMR data of guatyromycine B (5) in DMSO-d₆ .............................................. 15

**Supplementary Figures** ............................................................................................. 16
- Figure S1. Phylogenetic tree of CDPSs associated with cytochrome P450 enzymes ........ 16
- Figure S2. Phylogenetic tree of CDPS-associated cytochrome P450 enzymes ................ 17
- Figure S3. Alignments of GymA₁ – GymA₆ with AlbC .................................................. 18
- Figure S4. HPLC analysis of *E. coli* extracts harboring gymA₁ – gymA₆ ....................... 19
- Figure S5. UV spectra of 1 – 5 ..................................................................................... 20
- Figure S6. LC-MS analysis of extracts of *S. albus* J1074 harboring gym(AB)₁ – gym(AB)₆ and rv2275/rv2276 by UV detection and EICs for 3, 4, and 5 ......................... 21
Figure S7. LC-MS analysis of extracts of *S. albus* J1074 harboring *gym(AB)1* — *gym(AB)6* and *rv2275/rv2276* by UV detection and EICs for potential cYF derivatives .......................................................... 22

Figure S8. $^1$H NMR spectrum of cYY (1) in DMSO-$d_6$. ........................................................................................................ 23

Figure S9. $^1$H NMR spectrum of cYF (2) in DMSO-$d_6$. ........................................................................................................ 24

Figure S10. $^1$H NMR spectrum of mycocyclosin (3) in DMSO-$d_6$. ...................................................................................... 25

Figure S11. $^1$H NMR spectrum of guatyromycine A (4) in DMSO-$d_6$. .................................................................................. 26

Figure S12. $^{13}$C NMR spectrum of guatyromycine A (4) in DMSO-$d_6$. ................................................................................. 27

Figure S13. $^1$H-$^1$H COSY spectrum of guatyromycine A (4) in DMSO-$d_6$. .......................................................... 28

Figure S14. HSQC spectrum of guatyromycine A (4) in DMSO-$d_6$. ..................................................................................... 29

Figure S15. HMBC spectrum of guatyromycine A (4) in DMSO-$d_6$. ..................................................................................... 30

Figure S16. $^1$H NMR spectrum of guatyromycine B (5) in DMSO-$d_6$. .................................................................................. 31

Figure S17. $^{13}$C NMR spectrum of guatyromycine B (5) in DMSO-$d_6$. .................................................................................. 32

Figure S18. $^1$H-$^1$H COSY spectrum of guatyromycine B (5) in DMSO-$d_6$. ............................................................... 33

Figure S19. HSQC spectrum of guatyromycine B (5) in DMSO-$d_6$. .................................................................................... 34

Figure S20. HMBC spectrum of guatyromycine B (5) in DMSO-$d_6$. .................................................................................... 35

Figure S21. MS analysis of extracts of *S. albus* J1074 harboring *gymA6B6* cultivated with and without $^{15}$NH$_4$Cl .......................................................................................................................... 36

Figure S22. SDS-PAGE analysis of the purified P450s ................................................................................................................. 37

Figure S23. LC-MS analysis of cYY assays with GymB$_1$ — GymB$_6$ without nucleobases. ........................................... 38

References .............................................................................................................................................................................. 39
Methods

Computer-assisted sequence analysis

Nucleotide and amino acid sequences used in this study (Table S1) were obtained from NCBI databases (http://www.ncbi.nlm.nih.gov). Comparison of protein sequences was carried out by using BLASTP program (http://blast.ncbi.nlm.nih.gov/). The phylogenetic tree of CDPSs and P450s (Figures S1 and S2) were created by MEGA version X (http://www.megasoftware.net) using the maximum-likelihood method. Multiple sequence alignments were carried out with the program ClustalW and visualized with ESPript 3.0 (http://esprit.ibcp.fr/ESPrcript/cgi-bin/ESPrcript.cgi) to identify strictly conserved amino acid residues (Figure S3).

Bacterial strains, plasmids, and growth conditions

Strains and plasmids used and generated in this study are listed in Tables S2 and S4, respectively. Recombinant E. coli strains were cultivated in liquid or on solid Luria-Bertani (LB) with 100 µg/mL ampicillin, 50 µg/mL kanamycin, 50 µg/mL apramycin or 25 µg/mL chloramphenicol when necessary.

S. flavidovirens DSM 40150 and S. indicus DSM 42001 were purchased from German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). Streptomyces sp. NRRL S-646, Streptomyces sp. NRRL B-3648, S. katrae NRRL ISP-5550 as well as S. aureus NRRL B-2808 were obtained from the ARS Culture Collection (NRRL). S. albus J1074[1] was kindly gifted by Prof. Dr. Andriy Luzhetskyy (Saarland University). S. albus J1074 and the generated ex-conjugants were maintained on MS plates (mannitol 20.0 g/L, soy flour 20.0 g/L, agar 20.0 g/L) at 28 °C for sporulation. For secondary metabolite production, S. albus J1074 transformants were cultivated in liquid modified R5 medium (Table S2) at 28 °C for 7 days.

Genetic manipulation, PCR amplification, and gene cloning

Genetic manipulation in E. coli was performed according to the protocol by Green and Sambrook.[2] Genomic DNA isolation from Streptomyces was performed as described in the literature.[3]

Genes were amplified by PCR from the genomic DNA of the native strains by using primers listed in Table S4 with Phusion® High-Fidelity DNA Polymerase from New England Biolabs. The generated PCR fragments were cloned into pGEM-T Easy vector (Promega) and the sequence integrity was confirmed by sequencing. Subsequently, the fragments were released with restriction endonucleases from pGEM-T Easy and ligated into pPWW50A[4] for overexpression in S. albus J1074, or pET28a(+) for overexpression in E. coli BL21(DE3), which were both linearized with the same endonucleases. The plasmid pLH11 harboring rv2275 and rv2276 from M. tuberculosis was cloned directly into pPWW50A, linearized with restriction endonucleases, via homologous recombination in E. coli DH5α. The generated constructs are listed in Table S4.

Heterologous gene expression in S. albus J1074 and cultivation for secondary metabolite production

The constructed plasmids harboring different genes/gene clusters were firstly transferred into non-methylated E. coli ET12567/pUZ8002, then conjugated with S. albus J1074. The positive
conjugants were firstly selected by the phenotype showing apramycin resistance and further confirmed by PCR. The spores of the S. albus J1074 transformants were inoculated into 50 mL modified R5 liquid media supplied with 50 µg/mL of apramycin in 250 mL baffled flasks and cultured at 28 °C and 200 rpm for 7 days. 500 µL of such cultures were extracted with the same volume of EtOAc for three times. After that, the organic phases were combined, evaporated, and the dried residues were dissolved in 250 µL of methanol with 30 µL DMSO. 5 µL of such samples were subjected to LC-MS for analysis.

**LC-MS analysis**

LC-MS analysis was performed on an Agilent HPLC 1260 series system equipped with a photo diode array detector and a microTOF QIII mass spectrometer (Bruker, Bremen, Germany) by using a Multospher 120 RP-18 column (250 × 2 mm, 5 µm, CS-Chromatographie Service GmbH). For secondary metabolite analysis, a linear gradient of 5 — 100% ACN in H2O, both containing 0.1% HCOOH, in 40 min and a flow rate at 0.25 mL/min were used. The column was then washed with 100% ACN containing 0.1% HCOOH for 5 min and equilibrated with 5% ACN in H2O for 5 min. The parameters of the mass spectrometer were set as following: electrospray positive ion mode for ionization, capillary voltage with 4.5 kV, collision energy with 8.0 eV.

For analysis of the enzyme assays, samples were measured on the same LC-MS system as described above, but with a VDSpher PUR 100 C18-M-SE column (150 × 2 mm, 3 µm, VDS optilab Chromatographie Technik GmbH). A linear gradient of 5 — 100% ACN in H2O, both containing 0.1% HCOOH, in 30 min and a flow rate at 0.3 mL/min were used. The column was then washed and equilibrated as the described above.

**Overproduction in E. coli and purification of the recombinant GymB₁ – GymB₆**

The recombinant E. coli cells were cultivated for 16 h in 50 mL LB media supplied with 50 µg/mL kanamycin as preculture. 5 mL of the preculture were transferred into 500 mL LB media (with 50 µg/mL kanamycin) in 2 L-Erlenmeyer flasks and grown at 37 °C and 230 rpm to an absorption of 0.6 at 600 nm. Gene expression was induced with 0.1 mM IPTG at 16 °C for 20 h. The bacterial cultures were harvested by centrifugation (4,500 rpm, 20 min, 4 °C) and the cells were resuspended in lysis buffer (50 mM Tris-HCl, 10 mM imidazole, 300 mM NaCl, pH 8.0) with 2 – 5 mL/g wet weight. Lysozyme from the chicken egg white was added to the mixture at a final concentration of 1 mg/mL, which was incubated on ice for 30 min. The cells were then lysed by sonication on ice. Cell debris was removed by centrifugation at 13,000 rpm and 4 °C for 30 min. One-step purification of the recombinant His₆-tagged protein was performed by using Ni-NTA agarose (Macherey-Nagel, Düren, Germany) according to the manufacturer’s instructions. The storage buffer was changed to 50 mM Tris-HCl (pH 8.0) containing 15% (v/v) glycerol through a PD MiniTrap G-25 column (Cytiva, Freiburg, Germany), which had been equilibrated with the same buffer. After proof of their purity on 12% (w/v) SDS-PAGE (Figure S22), the recombinant GymB₁ – GymB₆ were stored frozen at -80 °C.

**In vitro assays with GymB₁ – GymB₆**

The enzyme assays of GymB₁ – GymB₆ were carried out with 100 µM cYY (1), 10 µM each of the six P450s, 8 mM NADPH, 25 µM spinach ferredoxin (Sigma-Aldrich), 0.4 unit/mL spinach
ferredoxin-NADP⁺ reductase (Sigma-Aldrich) with or without 100 µM guanine and 100 µM hypoxanthine, and 50 mM Tris-HCl buffer (pH 8.0) in a total volume of 50 µL. The reactions were performed at 30 °C for 16 h and then quenched with 50 µL ice-cold MeOH. After centrifugation at 13,000 rpm for 5 min, 20 µL of the supernatants were subjected to LC-MS analysis. Assays with heat-inactivated P450s were used as negative controls.

Large-scale fermentation, extraction, and isolation of secondary metabolites

For structural elucidation of the accumulated products, \textit{S. albus} J1074 transformants harboring \textit{gymA}_6\textit{B}_6 were cultivated in modified R5 medium on large scales (12 L) at 28 °C for 7 days. The culture supernatants were collected and extracted with equal volume of EtOAc for three times. The EtOAc phases were combined, evaporated to dryness, dissolved in MeOH, and applied to chromatography on a silica gel column and eluted with a gradient of CH₂Cl₂: MeOH in ratios of 100:2, 100:3, 100:5, 100:10, 100:20 and 100:100. The fractions containing the target products were further purified on an Agilent HPLC 1260 series by using a semi-preparative Agilent ZORBAX Eclipse XDB C18 HPLC column (250 × 9.4 mm, 5 µm). The flow rate was set to 2.0 mL/min.

Determination of production yields

For quantification, UV absorptions at 280 nm were used for 4, and 5 and at 294 nm for 3. 0.5 mL culture of \textit{S. albus} J1074 transformants harboring \((\text{gymAB})_1 — (\text{gymAB})_6\) were extracted with same volume of EtOAc for three times. The organic phases were combined and evaporated to dryness. The residues were dissolved in 250 µL MeOH and 30 µL DMSO. 5 µL were analyzed on LC-MS described above. The isolated products were used as authentic standards.

NMR analysis and structural elucidation

All the purified compounds were dissolved in DMSO-\textit{d}_6 for NMR analysis. The NMR spectra of the purified compounds 1 — 3 were recorded on a JOEL ECA-500 MHz spectrometer (JEOL, Tokyo, Japan). For compound 4 and 5, NMR experiments were performed on a Bruker AVIII 500 spectrometer equipped with a 5 mm BBO cryo probe Prodigy with z-gradient. All spectra were processed with MestReNova 5.2.2 (Mestrelab Research, Santiago de Compostella, Spain). NMR data and spectra of the identified compounds are provided in Tables S6 – S8 and Figures S8 – S20, respectively.

By comparison of their \textsuperscript{1}HNMR data with those of authentic standards, compounds 1 and 2 were identified to be \textit{cyclo-L-Tyr-L-Tyr} (cYY) and \textit{cyclo-L-Tyr-L-Phy} (cYF), respectively. Compound 3 was identified as mycocyclosin by comparison of its \textsuperscript{1}HNMR data with those published previously.\textsuperscript{[8]}

Inspection of the NMR data of guatyromycine A (4), including \textsuperscript{1}H, \textsuperscript{13}C, \textsuperscript{1}H—\textsuperscript{1}H COSY, HSQC, and HMBC (Table S7, Figures S11 – S15), suggested a cYY derivative with a guaninyl residue. Five signals at δ\textsubscript{c} 156.2 (C-2'), 155.2 (C-4'), 115.1 (C-5'), 159.7 (C-6'), and 152.8 ppm (C-8') can be assigned to the guaninyl unit. Two sets of AA'BB' systems at δ\textsubscript{H} 6.71 and 6.90, and at δ\textsubscript{H} 7.02 and 7.17 ppm in the \textsuperscript{1}H NMR spectrum and at δ\textsubscript{c} 115.2 and 131.1, and at δ\textsubscript{c} 130.7 and 119.1 ppm in the \textsuperscript{13}C NMR spectrum indicated the presence of two different 1,4-disubstituted benzene rings. All these data supported that the guanine residue in 4 is attached to a hydroxyl group of
cYY (Figure 1B). The loss of the characteristic singlet of H-8’ at approximate 8 ppm supported the linkage between C-8’ and O-21.

Structure elucidation of guatyromycine B (5) confirmed our assumption that a hypoxanthine is transferred onto cYY. The five $^{13}$C signals of the hypoxanthinyl residue were found at $\delta_C$ 153.2 (C-2’), 155.8 (C-4’), 114.9 (C-5’), 156.1 (C-6’), and 143.4 ppm (C-8’). The signal of C-2’ is de-shielded compared to that of a hypoxanthin-cWW conjugate.\cite{6} Furthermore, the signal of H-9 in the tyrosyl moiety appears as a singlet and that of H-10 is disappeared. Thus, hypoxanthine is attached via its C-2’ to C-10 of the cYY residue (Figure 1B), which is unequivocally confirmed by the correlation from H-9 to C-2’ in the HMBC spectrum (Table S8, Figures S16 – S20).

**Cultivation of the S. albus transformant in media containing $^{15}$N-labeled salt**

Cultivation of *S. albus* transformant harboring gymA$_6$B$_6$ in 50 mL medium containing 100 mg $^{15}$NH$_4$Cl and subsequent LC-MS analysis of the cultural extract demonstrated that at least two $^{15}$N atoms are incorporated into in 3, seven in 4, and six in 5 (Figure S21). These data provide additional evidence for the structures of 3, 4, and 5 depicted in Figure 1.
### Supplementary Tables

**Table S1.** Comparison of *gym* genes with known entries

| Streptomyces strains | Protein | Accession No. | Length (aa) | Sequence identity (%) to |
|----------------------|---------|---------------|-------------|--------------------------|
| *S. flavidovirens* DSM 40150 | GymA<sub>1</sub> | WP_078606819.1 | 236 | Rv2275 (63) |
|                        | GymB<sub>1</sub> | WP_028811244.1 | 395 | Rv2276 (60) |
| *S. indicus* DSM 42001 | GymA<sub>2</sub> | WP_093612344.1 | 241 | Rv2275 (65) |
|                        | GymB<sub>2</sub> | WP_093612346.1 | 402 | Rv2276 (63) |
| *Streptomyces* sp. NRRL S-646 | GymA<sub>3</sub> | WP_199931956.1 | 243 | Rv2275 (63) |
|                        | GymB<sub>3</sub> | WP_030942274.1 | 395 | Rv2276 (62) |
| *S. aureus* NRRL B-2808 | GymA<sub>4</sub> | WP_079042211.1 | 237 | Rv2275 (60) |
|                        | GymB<sub>4</sub> | WP_055602838.1 | 397 | Rv2276 (60) |
| *S. katrae* NRRL ISP-5550 | GymA<sub>5</sub> | KJY31545.1 | 236 | Rv2275 (62) |
|                        | GymB<sub>5</sub> | KJY31546.1 | 393 | Rv2276 (57) |
| *Streptomyces* sp. NRRL B-3648 | GymA<sub>6</sub> | WP_078946746.1 | 229 | Rv2275 (64) |
|                        | GymB<sub>6</sub> | WP_053711053.1 | 403<sup>a</sup> | Rv2276 (59) |

<sup>a</sup> Due to insolubility of the recombinant protein with its original coding region, a 186 bp genomic sequence upstream *gymB<sub>6</sub>* was added for expression in *E. coli* BL21(DE3).
### Table S2. Bacterial strains used in this study

| Strain                        | Source              | Cultivation media                      |
|-------------------------------|---------------------|----------------------------------------|
| *E. coli* DH5α                | Invitrogen          | LB                                     |
| *E. coli* ET12567/pUZ8002     | {7}                 | LB                                     |
| *S. albus* J1074              | DSMZ                | MS (solid), modified R5 (liquid)       |
| *S. flavidovirens* DSM 40150  | DSMZ                | modified R5, GYM                       |
| *S. indicus* DSM 42001        | DSMZ                | modified R5, GYM                       |
| *Streptomyces* sp. NRRL S-646 | NRRL                | modified R5, GYM                       |
| *S. katrae* NRRL ISP-5550     | NRRL                | modified R5, GYM                       |
| *Streptomyces* sp. NRRL B-3648| NRRL                | modified R5, GYM                       |
| *S. aureus* NRRL B-2808       | NRRL                | modified R5, GYM                       |
| *M. tuberculosis* DSM 43990 a | DSMZ                | ---                                    |

*DNA was ordered instead of the bacterium*

**NRRL**: ARS Culture Collection

**DSMZ**: German Collection of Microorganisms and Cell Cultures GmbH

**LB medium**: tryptone 10.0 g/L, yeast extract 5.0 g/L, NaCl 10.0 g/L.

**MS medium**: mannitol 20.0 g/L, soy flour 20.0 g/L, agar 20.0 g/L.

**Modified R5 medium**: sucrose 103.0 g/L, glucose 10.0 g/L, yeast extract 5.0 g/L, MgCl$_2$·6H$_2$O 10.12 g/L, K$_2$SO$_4$ 0.25 g/L, Difco casaminoacids 0.1 g/L, MOPS 21.0 g/L, trace element solution 2 mL/L, pH 7.2

**GYM medium**: glucose 4.0 g/L, yeast extract 4.0 g/L, malt extract 10.0 g/L, pH 7.2
Table S3. Key residues for the substrate binding pockets of the investigated CDPSs compared to AlbC[^8]

| Protein | amino acid residues of pocket P1 | amino acid residues of pocket P2 | Predicted product |
|---------|----------------------------------|----------------------------------|------------------|
| GymA1   | V G V T M L F N                  | L T F T Q L P                   | cYY              |
| GymA2   | V G V T M L F N                  | L T F D Q L P                   | cYY              |
| GymA3   | V G V T M L F N                  | L S F T Q L P                   | cYY              |
| GymA4   | V G V T M L F N                  | L I F G Q L P                   | cYY              |
| GymA5   | V G I T M L F N                  | L I F S Q L P                   | cYY              |
| GymA6   | V G V T M L F N                  | L R F A Q L P                   | cYY              |

Prediction of CDPS products were carried out based on the previous reported methods.[^9;10]
## Table S4. Cloning and expression constructs used in this study

| Gene          | Primer sequences (5'-3')                                                                 | Cloning construct | Expression vector | Cloning sites | Expression constructs |
|---------------|-----------------------------------------------------------------------------------------|-------------------|-------------------|---------------|-----------------------|
| gymA1-D40150  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pLH01            | pET28a(+)         | Ndel/BamH1    | pLH06                 |
| gym(AB)1-D40150 | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pLH02            | pPWW50A           | Ndel/BamH1    | pLH07                 |
| gymA2-D42001  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL85            | pET28a(+)         | Ndel/BamH1    | pJL95                 |
| gym(AB)2-D42001 | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL86            | pPWW50A           | Ndel/BamH1    | pJL96                 |
| gymA3-S-646   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL87            | pET28a(+)         | Ndel/BamH1    | pJL97                 |
| gym(AB)3-S-646 | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL88            | pPWW50A           | Ndel/BamH1    | pJL98                 |
| gymA4-B2808   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL93            | pET28a(+)         | Ndel/BamH1    | pJL103                |
| gym(AB)4-B2808 | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL94            | pPWW50A           | Ndel/BamH1    | pJL104                |
| gymA5-ISP550  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL95            | pET28a(+)         | Ndel/BamH1    | pJL100                |
| gym(AB)5-ISP550 | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL96            | pPWW50A           | Ndel/BamH1    | pJL101 & pJL117       |
| gymB1-D40150  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL97            | pET28a(+)         | Ndel/BamH1    | pJL102                |
| gymB2-D42001  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL98            | pPWW50A           | Ndel/BamH1    | pJL11                 |
| gymB3-S-646   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL99            | pET28a(+)         | Ndel/BamH1    | pJL11                 |
| gymB4-B2808   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL100           | pPWW50A           | Ndel/BamH1    | pJL11                 |
| gymB5-ISP550  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL101           | pPWW50A           | Ndel/BamH1    | pJL11                 |
| gymB6-B3648   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL102           | pPWW50A           | Ndel/BamH1    | pJL11                 |
| rv2275rv2276  | CACACGGGAGCCATATCAGAGGTTGCTCATAGTGCGAGCC                                              | ---              | pPWW50A           | Ndel/BamH1a   | pLH11                 |
| M.tuberc..    | CACACGGGAGCCATATCAGAGGTTGCTCATAGTGCGAGCC                                              | ---              | pPWW50A           | Ndel/BamH1a   | pLH11                 |
| gymB15-D40150 | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL103           | pPWW50A           | Ndel/BamH1    | pJL111                |
| gymB2-D42001  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL104           | pPWW50A           | Ndel/BamH1    | pJL112                |
| gymB3-S-646   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL105           | pPWW50A           | Ndel/BamH1    | pJL113                |
| gymB4-B2808   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL106           | pPWW50A           | Ndel/BamH1    | pJL114                |
| gymB5-ISP550  | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL107           | pPWW50A           | Ndel/BamH1    | pJL115                |
| gymB6-B3648   | CATATGACCACACGACTGAGAATCTGAGGCTCATGGGCG                                               | pJL108           | pPWW50A           | Ndel/BamH1    | pJL116                |

Restriction sites for cloning are underlined in the primer sequences. Cloning constructs are based on pGEM T® EASY vector.

a pLH11 was cloned via homologous recombination directly into pPWW50A
Table S5. Product yields of the *S. albus* transformants harboring guatyromycine and mycocyclisin clusters

| *S. albus* with | Yields [mg/L] |
|----------------|---------------|
|                | mycocyclisin (3) | guatyromycine A (4) | guatyromycine B (5) |
| gymA1B1        | 2.3 ± 0.1       | 3.4 ± 0.2            | 2.3 ± 0.2            |
| gymA2B2        | 1.6 ± 0.2       | 2.6 ± 0.2            | 2.4 ± 0.1            |
| gymA3B3        | 5.6 ± 0.2       | 5.0 ± 0.7            | 5.4 ± 0.5            |
| gymA4B4        | 9.4 ± 0.3       | 5.5 ± 0.3            | 5.4 ± 0.1            |
| gymA5B5        | 10.1 ± 0.1      | 4.1 ± 0.0            | 3.3 ± 0.1            |
| gymA6B6        | 24.0 ± 1.3      | 5.5 ± 0.1            | 3.6 ± 0.1            |
| rv2275/ rv2276 | 53.9 ± 2.8      | Not detected         | Not detected         |

The data are mean values and the deviation are given as ± values (n = two or three independent experiments)
| Position | δ<sub>H</sub>, multi. (J in Hz) | δ<sub>H</sub>, multi. (J in Hz) | δ<sub>H</sub>, multi. (J in Hz) |
|----------|-------------------------------|-------------------------------|-------------------------------|
| 1        | 7.73, d (1.5)                 | 7.82, s                       | 7.97, s                       |
| 2        | 3.85, t (6.1)                 | 3.88, t (6.0)                 | 4.33, d (5.5)                 |
| 4        | 7.73, d (1.5)                 | 7.82, s                       | 7.97, s                       |
| 5        | 3.85, t (6.1)                 | 3.95, t (6.0)                 | 4.33, d (5.5)                 |
| 7a       | 2.51, dd (13.7, 6.1)          | 2.59<sup>b</sup>, dd (13.7, 6.0) | 3.46, d (15.6)               |
| 7b       | 2.12, dd (13.7, 6.1)          | 2.18<sup>a</sup>, m           | 2.65, dd (15.6, 5.5)          |
| 9        | 6.84, d (8.4)                 | 6.84, d (8.3)                 | 6.58, d (2.4)                 |
| 10       | 6.67, d (8.4)                 | 6.67, d (8.3)                 | -                             |
| 11       | 9.19 (OH), s                  | 9.20 (OH), s                  | 9.04 (OH), br s               |
| 12       | 6.67, d (8.4)                 | 6.67, d (8.3)                 | 6.62, d (8.1)                 |
| 13       | 6.84, d (8.4)                 | 6.84, d (8.3)                 | 6.84, dd, (8.3, 2.3)          |
| 14a      | 2.51, dd (13.7, 6.1)          | 2.54<sup>b</sup>, m           | 3.46, d (15.6)               |
| 14b      | 2.12, dd (13.7, 6.1)          | 2.18<sup>a</sup>, m           | 2.65, dd (15.6, 5.5)          |
| 16       | 6.84, d (8.4)                 | 7.04, d (7.2)                 | 6.58, d (2.4)                 |
| 17       | 6.67, d (8.4)                 | 7.27, t (7.2)                 | -                             |
| 18       | 9.19 (OH), s                  | 7.20, t (7.2)                 | 9.04 (OH), br s               |
| 19       | 6.67, d (8.4)                 | 7.27, t (7.2)                 | 6.62, d (8.1)                 |
| 20       | 6.84, d (8.4)                 | 7.04, d (7.2)                 | 6.84, dd, (8.3, 2.3)          |

<sup>a</sup> signal overlapping with solvent peak, <sup>b</sup> signals overlapping with each other.

The spectra of 1 and 2 correspond well to those of the authentic standards and 3 to those published previously.\[5\]
Table S7. NMR data of guatyromycine A (4) in DMSO-$d_6$

![Guatyromycine A (4)]

| Position | $\delta$C | $\delta$H, multi. ($J$ in Hz) | COSY | HMBC |
|----------|----------|-------------------------------|------|------|
| 1        | -        | 7.87, d (2.1)                | H-2  | C-2, 3, 6, 15 |
| 2        | 55.5     | 3.87, m                       | H-1, H-14 | C-3, 6, 15 |
| 3        | 166.3    | -                             | -    | -    |
| 4        | -        | 8.00, d (1.7)                 | H-5  | C-5, 6 |
| 5        | 55.7     | 3.99, m                       | H-4, H-7 | C-3, 7, 8 |
| 6        | 166.4    | -                             | -    | -    |
| 7a       | 38.3     | 2.63, dd, (13.7, 5.0)        | -    | C-5, 6, 8, 9, 13 |
| 7b       | 2.51     | m                             | -    | C-5, 6, 8, 9, 13 |
| 8        | 126.3    | -                             | -    | -    |
| 9        | 131.1    | 6.90, d (8.4)                 | H-10 | C-5, 7, 10, 11, 13 |
| 10       | 115.2    | 6.71, d (8.4)                 | H-9  | C-9, 11 |
| 11       | 156.3    | -                             | -    | -    |
| 12       | 115.2    | 6.71, d (8.4)                 | H-13 | C-11, 13 |
| 13       | 131.1    | 6.90, d (8.4)                 | H-12 | C-5, 7, 9, 11, 12 |
| 14a      | 39.8     | 2.51, m                       | -    | C-2, 3, 15, 16, 20 |
| 14b      | 1.93     | 1.93, dd (13.6, 7.3)         | -    | C-2, 3, 15, 16, 20 |
| 15       | 133.1    | -                             | -    | -    |
| 16       | 130.7    | 7.02, d (8.5)                 | H-17 | C-14, 17, 18, 20 |
| 17       | 119.1    | 7.17, d (8.5)                 | H-16 | C-15, 16, 18, 19 |
| 18       | 153.2    | -                             | -    | -    |
| 19       | 119.1    | 7.17, d (8.5)                 | H-20 | C-15, 17, 18, 20 |
| 20       | 130.7    | 7.02, d (8.5)                 | H-19 | C-14, 16, 18, 19 |
| 1'       | -        | 10.80, s                      | -    | -    |
| 2'       | 156.2    | -                             | -    | -    |
| 4'       | 155.2    | -                             | -    | -    |
| 5'       | 115.1    | -                             | -    | -    |
| 6'       | 159.7    | -                             | -    | -    |
| 8'       | 152.8    | -                             | -    | -    |
| 9'       | -        | not observed                  | -    | -    |
| 10'      | -        | 6.35, br s                    | -    | -    |
### Table S8. NMR data of guatyromycine B (5) in DMSO-\(d_6\)

![Diagram of guatyromycine B (5)]

| Position | \(\delta_C\) | \(\delta_H,\) multi., \((J\text{ in Hz})\) | COSY | HMBC |
|-----------|---------------|----------------------------------------|------|------|
| 1         | -             | 7.84, s                                | H-2  | C-2, 6 |
| 2         | 55.5          | 3.93b, m                               | H-14 | C-3, 14, 15 |
| 3         | 166.5         | -                                       | -    | - |
| 4         | -             | 7.76, s                                | H-5  | C-3, 5, 7 |
| 5         | 55.9          | 3.94b, m                               | H-7  | C-6, 7, 8 |
| 6         | 166.3         | -                                       | -    | - |
| 7a        | 39.6 \(^a\)  | 2.63\(^c\), dd (13.7, 4.3)            | H-5, 7 | C-5, 6, 8, 13 |
| 7b        | 1.98         | 1.98, dd (13.7, 7.4)                   | H-5, 7 | C-5, 6, 8, 13 |
| 8         | 126.5         | -                                       | -    | - |
| 9         | 127.4         | 7.80, s                                | -    | C-7, 13, 2\(^{'}\) |
| 10        | 115.1         | -                                       | -    | - |
| 11        | 156.4         | -                                       | -    | - |
| 12        | 116.4         | 6.85\(^d\), br d (9.3)                 | H-13 | C-8, 10, 11 |
| 13        | 131.3         | 6.87, br d (9.3)                       | H-9, 12 | C-9, 11 |
| 14a       | 38.7          | 2.61\(^c\), dd (13.0, 4.2)            | H-2, 14 | C-2, 3, 15, 16, 20 |
| 14b       | 2.39          | 2.39, dd (13.0, 5.2)                   | H-2, 14 | C-2, 3, 15, 16, 20 |
| 15        | 126.3         | -                                       | -    | - |
| 16        | 130.9         | 6.84\(^d\), d (8.4)                   | H-17, 19 | C-17, 18, 19 |
| 17        | 115.3         | 6.64, d (8.4)                          | H-16, 20 | C-15, 16, 18, 20 |
| 18        | 156.3         | -                                       | -    | - |
| 19        | 115.1         | 6.64, d (8.4)                          | H-16, 20 | C-15, 16, 18, 20 |
| 20        | 130.8         | 6.84\(^d\), d (8.4)                   | H-17, 19 | C-17, 18, 19 |
| 1\(^{'}\) | -             | not observed                           | -    | - |
| 2\(^{'}\) | 153.2         | -                                       | -    | - |
| 4\(^{'}\) | 155.8         | -                                       | -    | - |
| 5\(^{'}\) | 114.9         | -                                       | -    | - |
| 6\(^{'}\) | 156.1         | -                                       | -    | - |
| 8\(^{'}\) | 143.4         | 7.88, s                                | -    | C-4\(^{'}\) |
| 9\(^{'}\) | -             | not observed                           | -    | - |

\(^a\) signal overlapping with solvent peak  
\(^b, c, d\) signals with same letter overlapping with each other
Figure S1. Phylogenetic tree of CDPSs associated with cytochrome P450 enzymes.
CDPSs investigated in this study (see Table S1 for details) are depicted in red, Rv2275 in green,[5] CDPSs associated to nucleobase transferases in blue,[6;11-13] those associated to dimerases in pink,[14-16] BcmA in brown,[17] and YvmC in light blue.[18]
Figure S2. Phylogenetic tree of CDPS-associated cytochrome P450 enzymes with accession numbers from databases. P450s investigated in this study (see Table S1 for details) are depicted in red, CYP121 in green,[5] known nucleobase transferases in blue,[6,11-13] known CDP dimerases in pink,[14-16] BcmD in brown,[17] and CYP134 in light blue.[18]
Figure S3. Alignments of GymA₁ – GymA₆ with AlbC[8] to determine the key residues for the substrate binding pockets P1 and P2 (Table S3) as reported previously.[9;10]
Figure S4. HPLC analysis of E. coli extracts harboring gymA₁ – gymA₆. The integrity of 1 (cYY) and 2 (cYF) was confirmed by ¹H NMR analysis after isolation (Table S6).
Figure S5. UV spectra of 1 – 5
Figure S6. LC-MS analysis of extracts of *S. albus* J1074 harboring *gym(AB)*, and *rv2275/rv2276*, detected by UV absorptions at 280 – 300 nm and extracted ion chromatography (EICs) for 3, 4, and 5.
Figure S7. LC-MS analysis of extracts of S. albus J1074 harboring gym(AB)₁ – gym(AB)₆ and rv2275/rv2276, detected by UV absorptions at 280 – 300 nm and extracted ion chromatograms (EICs) for potential cYF derivatives, e.g. cYF - 2H, cYF + guanine (6), and cYF + hypoxanthine
Figure S8. $^1$H NMR spectrum of cYY (1) in DMSO-$d_6$. 
Figure S9. $^1$H NMR spectrum of cYF (2) in DMSO-$d_6$. 
Figure S10. $^1$H NMR spectrum of mycocyclosin (3) in DMSO-$d_6$. 
Figure S11. $^1$H NMR spectrum of guatyromycine A (4) in DMSO-$d_6$. 
Figure S12. $^{13}$C NMR spectrum of guatyromycine A (4) in DMSO-$d_6$. 
Figure S13. $^1$H-$^1$H COSY spectrum of guatyromycin A (4) in DMSO-$d_6$. 
Figure S14. HSQC spectrum of guatromycine A (4) in DMSO-
$\text{d}_6$. 

S29
**Figure S15.** HMBC spectrum of guatyromicine A (4) in DMSO-$d_6$. 
Figure S16. $^1$H NMR spectrum of guatyromycine B (5) in DMSO-$d_6$. 
Figure S17. $^{13}$C NMR spectrum of guatyromycin B (5) in DMSO-$d_6$. 
Figure S18. $^1$H-$^1$H COSY spectrum of guatyromycine B (5) in DMSO-$d_6$. 
Figure S19. HSQC spectrum of guatyromycine B (5) in DMSO-\textit{d}_6.
Figure S20. HMBC spectrum of guatyromycine B (5) in DMSO-$d_6$. 
Figure S21. MS analysis of extracts of *S. albus* J1074 harboring *gymA*B*6* cultivated in MR5 medium with and without ^15^N NH₄Cl.
Figure S22. SDS-PAGE analysis of the purified P450s

Due to insolubility of the actual GymB₆, its sequence was enlarged to 465 amino acids instead of 402 by addition of upstream sequences. Therefore, its protein mass is significantly larger than those of other GymBs.
Figure 23. LC-MS analysis of cYY assays with GymB₁ – GymB₆ without nucleobases. EIC of 3 refers the [M + H]⁺ ion at m/z 325.118 with a tolerance range of ± 0.005. x: unknown mycocyclosin isomer.
References

[1] N. Zaburannyi, M. Rabyk, B. Ostash, V. Fedorenko, A. Luzhetskyy, *Bmc Genomics* 2014, 15, 97.

[2] M. R. Green, J. Sambrook, *Molecular cloning: a laboratory manual*, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2012.

[3] T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood, *Practical Streptomyces Genetics*, 2nd ed. John Innes Foundation, Norwich, UK 2000.

[4] Y. Zhu, P. Fu, Q. Lin, G. Zhang, H. Zhang, S. Li, J. Ju, W. Zhu, C. Zhang, *Org Lett.* 2012, 14, 2666-2669.

[5] P. Belin, M. H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J. B. Charbonnier, A. Lecoq, R. Thai, M. Courcon, C. Masson, C. Dugave, R. Genet, J. L. Pernodet, M. Gondry, *Proc. Natl. Acad. Sci. U. S. A* 2009, 106, 7426-7431.

[6] H. Yu, X. Xie, S.-M. Li, *Org. Lett.* 2019, 21, 9104-9108.

[7] B. Gust, G. L. Challis, K. Fowler, T. Kieser, K. F. Chater, *Proc. Natl. Acad. Sci. U. S. A* 2003, 100, 1541-1546.

[8] S. Lautru, M. Gondry, R. Genet, J. L. Pernodet, *Chem Biol.* 2002, 9, 1355-1364.

[9] I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J. L. Pernodet, M. Gondry, P. Belin, *Nat. Chem Biol.* 2015, 11, 721-727.

[10] M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J. L. Pernodet, M. Moutiez, *Front Microbiol.* 2018, 9, 46.

[11] H. Yu, X. Xie, S.-M. Li, *Org. Lett.* 2018, 20, 4921-4925.

[12] J. Liu, X. Xie, S.-M. Li, *Angew. Chem. Int. Ed. Engl.* 2019, 58, 11534-11540.

[13] J. Shi, X. Xu, E. J. Zhao, B. Zhang, W. Li, Y. Zhao, R. H. Jiao, R. X. Tan, H. M. Ge, *Org. Lett.* 2019, 21, 6825-6829.

[14] V. V. Shende, Y. Khatri, S. A. Newmister, J. N. Sanders, P. Lindovska, F. Yu, T. J. Doyon, J. Kim, K. N. Houk, M. Movassaghi, D. H. Sherman, *J. Am. Chem. Soc.* 2020, 142, 17413-17424.

[15] H. Yu, S.-M. Li, *Org. Lett.* 2019, 21, 7094-7098.

[16] J. Liu, X. Xie, S.-M. Li, *Chem. Commun.* 2020, 56, 11042-11045.

[17] S. Meng, W. Han, J. Zhao, X. H. Jian, H. X. Pan, G. L. Tang, *Angew. Chem Int. Ed. Engl.* 2018, 57, 719-723.

[18] M. J. Cryle, S. G. Bell, I. Schlichting, *Biochemistry* 2010, 49, 7282-7296.