

Elucidation of the final steps in Taxol biosynthesis and its biotechnological production

In the format provided by the
authors and unedited

Contents

Supplementary methods.....	3
Chemicals, enzymes, and plant materials	3
Chemical synthesis of taxoids	3
NMR experiments	4
HRMS and NMR data of chemically synthesized taxoids	4
<i>N</i> -debenzoyl-2'-deoxy-taxol (compound 2).....	5
<i>N</i> -debenzoyl-taxol (compound 3).....	5
10-deacetyl- <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 6).....	6
10-deacetyl- <i>N</i> -debenzoyl-taxol (compound 7).....	6
(3'R/3'S)-2'-deoxy-taxol.....	7
Protein extraction from tobacco leaves.....	7
Yeast media.....	8
BAPT optimization	8
Protein extraction from yeast.....	8
Supplementary Figures	10
Fig. S1. Chemical synthesis of <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 2).	10
Fig. S2. Chemical synthesis of <i>N</i> -debenzoyl-taxol (compound 3).	11
Fig. S3. Chemical synthesis of 10-deacetyl- <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 6).....	12
Fig. S4. Chemical synthesis of 10-deacetyl- <i>N</i> -debenzoyl-taxol (compound 7).	13
Fig. S5. Chemical synthesis of 2'-deoxy-taxol (compound 9)	14
Fig. S6. ¹ H-NMR spectrum of <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 2, 600 MHz, DMSO- <i>d</i> ₆). 15	
Fig. S7. ¹³ C-NMR spectrum of <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 2, 125 MHz, DMSO- <i>d</i> ₆).	15
Fig. S8. ¹ H-NMR spectrum of <i>N</i> -debenzoyl-taxol (compound 3, 600 MHz, DMSO- <i>d</i> ₆).....	16
Fig. S9. ¹³ C-NMR spectrum of <i>N</i> -debenzoyl-taxol (compound 3, 125 MHz, DMSO- <i>d</i> ₆).....	16
Fig. S10. ¹ H-NMR spectrum of 10-deacetyl- <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 6, 600 MHz, DMSO- <i>d</i> ₆).	17
Fig. S11. ¹³ C-NMR spectrum of 10-deacetyl- <i>N</i> -debenzoyl-2'-deoxy-taxol (compound 6, 125 MHz, DMSO- <i>d</i> ₆).	17
Fig. S12. ¹ H-NMR spectrum of 10-deacetyl- <i>N</i> -debenzoyl-taxol (compound 7, 600 MHz, DMSO- <i>d</i> ₆).	18
Fig. S13. ¹³ C-NMR spectrum of 10-deacetyl- <i>N</i> -debenzoyl-taxol (compound 7, 125 MHz, DMSO- <i>d</i> ₆).	18
Fig. S14. ¹ H-NMR spectrum of (3'R/3'S)-2'-deoxy-taxol (600 MHz, DMSO- <i>d</i> ₆).	19
Fig. S15. ¹³ C-NMR spectrum of (3'R/3'S)-2'-deoxy-taxol (125 MHz, DMSO- <i>d</i> ₆).	19
Fig. S16. Transient expression of <i>T. baccata</i> phenylalanine aminomutase (PAM) in tobacco leaves via agro-infiltration	20
Fig. S17. The CoA ligase from <i>P. chrysogenum</i> , CoAL(A312G), is more efficient than TBPCCL or AAE16	21

Fig. S18. Western blot of TB506 from tobacco leaves.....	23
Fig. S19. Correlation heatmap of co-expression patterns between the known Taxol pathway genes	24
Fig. S20. Correlation heatmap of co-expression patterns between selected <i>T. chinensis</i> CYP candidates and baits in this study	25
Fig. S21. Analysis of 19 candidate CYPs (Fig. S20) obtained from our co-expression analysis. None of the candidates was capable of	26
Fig. S22. Analysis of 21 candidate CYPs, listed in Tables S2 and S3. None of the candidates was capable of	28
Fig. S23. Scheme depicting the similarity between C2'α hydroxylation of the C13 side chain of Taxol and hydroxylation of asparaginy and aspartyl residues of peptides.	30
Fig. S24. Correlation heatmap of co-expression patterns between the selected <i>Taxus chinensis</i> OGD candidates and baits in this study	31
Fig. S25. Analysis of nine candidate OGDs obtained from our co-expression analysis (Fig. S24) revealed one candidate, T2'OGD, with potent C2'α hydroxylation activity.....	32
Fig. S26. Mass spectra of N-debenzoyl-taxol (3) in positive ionization mode.....	33
Fig. S27. <i>Taxus canadensis</i> DBTNBT benzoylates N-debenzoyl-2'-deoxy-taxol (2) to form 2'-deoxy-taxol (9).....	34
Fig. S28. Correlation heatmap of co-expression patterns between selected <i>T. chinensis</i> acyltransferase candidates and baits in this study	35
Fig. S29. Analysis of 25 candidate BAHD family acyltransferases obtained from our co-expression analysis (Fig. S28) revealed one candidate, T3'NBT with potent 3'-N benzoylation activity. Compound.....	36
Fig. S30. Mass spectra of Taxol (4) in positive ionization mode	38
Fig. S31. 2'-deoxy-taxol (9) is not a substrate of T2'OGD.	39
Fig. S32. N-debenzoyl-2'-deoxy-taxol (2) is not a substrate of T3'NBT.....	40
Fig. S33. Taxoids in <i>T. baccata</i> bark extracts	41
Fig. S34. Mass spectra of: (a) 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (6), (b) 10-deacetyl-N-debenzoyl-taxol (7) and (c) 10-deacetyl-taxol (8) in positive ionization mode ([M+H] ⁺), and (d) chromatograms of chemically synthesized 6 and 7	42
Figure S35. T3'NBT does not use 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (6) as a substrate when produced heterologously in tobacco leaves	43
Fig. S36. Western blot of CoAL(A312G) and BAPT extracted from engineered yeast cells cultivated at 30°C.....	44
Fig. S37. Western blot of BAPT variants extracted from engineered yeast cells cultivated at different temperatures.	45
Fig. S38. N-debenzoyl-2'-deoxy-taxol (2) production in <i>S. cerevisiae</i>	46
Fig. S39. Amino acid sequence alignment of BAPTs (Table. S5).....	47
Fig. S40. Mass spectra of Taxol (4) in positive ionization mode ([M+H] ⁺).....	48
Fig. S41. Growth of engineered yeast strains (see main text and Table S4 for strain description). .	49
Supplementary Tables.....	50
Table S1. Enzymes used for pathway reconstruction.....	50

Table S2. <i>Taxus</i> genes used as baits in co-expression analysis.....	51
Table S3. CYPs of the CYP725 family obtained from the literature that were tested in this work for C2' α hydroxylation activity.	52
Table S4. <i>S. cerevisiae</i> strains constructed in this study.	53
Table. S5. NCBI access numbers and predicted scaled solubility of BAPT sequences used in BAPT optimization.	54
Supplementary References.....	55
Additional Supplementary Figures	56
Additional Supplementary Fig 1. Unprocessed Gels and Blots for Supplementary Figure 17b. ...	59
Additional Supplementary Fig. 2. Unprocessed Gels and Blots for Supplementary Figure 18. ...	63
Additional Supplementary Fig. 3. Unprocessed Gels and Blots for Supplementary Figure 36. ...	67
Additional Supplementary Fig. 4. Unprocessed Gels and Blots for Supplementary Figure 37. ...	71

Supplementary methods

Chemicals, enzymes, and plant materials

Taxol (Product no. PHL89806) and β -phenylalanine (Product No.: 159492) were purchased from Merck. α -phenylalanine was purchased from Carbosynth (product no.: FP15788). 10-deacetyl-baccatin III (**5**), and baccatin III (**1**) were purchased from Biosynth (Product no.s.: FB11737 and FB18122, respectively). D-(+)-glucose monohydrate (16301) and D-(+)-galactose (G0625), were purchased from Sigma-Aldrich (USA). Raffinose pentahydrate (R1030,) and Yeast Nitrogen Base w/o AA (Y2025) were obtained from US Biological (USA). DNA restriction enzymes, USER[®] enzyme, and Q5U[®] Hot Start High-Fidelity DNA Polymerase for PCR amplifications were obtained from New England Biolabs (Herlev, Denmark). *Taxus baccata* materials were obtained from the botanical garden of Copenhagen University, Denmark.

c-Myc (9E10) primary antibody was purchased from Santa Cruz Biotechnology, Inc. Polyclonal Rabbit Anti-mouse Immunoglobulins/HRP was purchased from Agilent Technologies (P0260). K₂CO₃ (209619), Urea (U5128), Sodium dodecyl sulfate (SDS, 436143), 1,4-Dithiothreitol (DTT,1114740001, Tris (GE17-1321-01), 2-Mercaptoethanol (M6250), and cOmplete[™] EDTA-free Protease Inhibitor Cocktail (11873580001) were purchased from Merck-Sigma-Aldrich.

Chemical synthesis of taxoids

Melting points (mp) were determined on a Kofler apparatus and are uncorrected. Optical rotations were measured at 21 $\pm$ 2 °C with an Optical Activity Ltd AA-1000 Polarimeter. Unless otherwise indicated, commercially available starting materials, reagents, solvents and dry solvents were purchased from Sigma-Aldrich. All reactions were carried out under an atmosphere of argon in oven-dried glassware with magnetic stirring and monitored by thin layer chromatography (TLC) on aluminum sheets coated with silica gel 60F254 (0.2 mm thickness, E. Merck, Darmstadt, Germany). UV light was used as the visualizing agent and the components present were detected by charring with 10% H₂SO₄ in MeOH. Column chromatography was carried out using silica gel 60 (particle size 0.040-0.063 mm, 230-400 mesh ASTM, E. Merck). Solvent extracts were dried with anhydrous MgSO₄ unless otherwise specified.

The starting materials for synthesizing the different taxoids, including D-3-(tert-

butoxycarbonylamino)-3-phenylpropanoic acid (**12**)¹, (4S,5R)-3-(tert-butoxycarbonyl)-2,2-dimethyl-4-phenyloxazolidine-5-carboxylic acid (**16**) (commercially available from Sigma Aldrich), DL-3-(tert-butoxycarbonylamino)-3-phenylpropanoic acid (**21**)^{1,2}, 7,10-di(2,2,2-Trichloroethoxycarbonyl)-10-deacetyl baccatin III (**11**)^{3,4} and 7-(2,2,2-Trichloroethoxycarbonyl)baccatin III (**14**)^{3,4}, were prepared according to the corresponding literature procedures.

The taxoids, including 10-deacetyl-N-debenzoyl-2deoxy-taxol (**6**), N-debenzoyl-2'-deoxy-taxol (**2**), N-debenzoyl-taxol (**3**), 2'-deoxy-taxol (**9**), and 10-deacetyl-N-debenzoyl-taxol (**7**), were chemically synthesized from the aforementioned starting materials using Steglich esterification (DCC and DMAP) as indicated in the charts (Figs. S1-S5) according to literature procedures⁵⁻¹¹. The subsequent steps involve the selective de-protection of the Troc (2,2,2-trichloroethyloxycarbonyl) ether, as illustrated in the charts. The general protocol for synthesis of each of the taxoids is as follows. DCC (83.0 mg, 0.40 mmol, 2.0 eq.) was added to a solution of **11** (180.0 mg, 0.20 mmol, 1.0 eq.) and/or **14** (150.0 mg, 0.20 mmol, 1.0 eq.), **12** (58.4 mg, 0.22 mmol, 1.1 eq.), **16** (70.7 mg, 0.22 mmol, 1.1 eq.) and/or **21** (58.4 mg, 0.22 mmol, 1.1 eq.), and DMAP (2.50 mg, 0.02 mmol, 10.0 % eq.) in dry CH₂Cl₂ as shown in Figs S1-S5. Each reaction mixture was stirred for 2 h at room temperature. After stirring, the byproduct DCU was filtered off, and the solution was concentrated under vacuum to get the corresponding products, which were further used for the appropriate de-protection to afford the desired taxoid **2** (99.80 mg, 0.136 mmol, 68%), **3** (99.50 mg, 0.106 mmol, 53%), **6** (95.50 mg, 0.138 mmol, 69%), **7** (76.40 mg, 0.108 mmol, 54%), and **9** (58.60 mg, 0.070 mmol, 35%).

NMR experiments

The NMR data were acquired on a 600 MHz Bruker Avance III instrument equipped with a cryogenically cooled 1.7-mm TCI probe head and a Bruker SampleJet sample changer (Bruker Biospin, Karlsruhe, Germany). ¹H-NMR and ¹³C-NMR spectra were acquired with 30° pulses and 64k data points. IconNMR ver. 4.2 (Bruker Biospin, Karlsruhe, Germany) was used for controlling automated sample change and acquisition of NMR data. Topspin ver. 4.0.6 (Bruker Biospin, Karlsruhe, Germany) was used for acquisition and processing of NMR data.

HRMS and NMR data of chemically synthesized taxoids

The structures of the five chemically synthesized taxoids were verified by NMR, as shown in **Figs. S6-15**. The ¹H NMR and ¹³C NMR δ -values were calibrated using residual NMR solvent

peaks; coupling constants (J) are given in Hz. The multiplicities are reported as follows: singlet (s), doublet (d), doublet of doublets (dd), doublet of doublet of doublets (ddd), multiplet (m) and triplet (t).

N-debenzoyl-2'-deoxy-taxol (compound 2):

White solid; mp = 148-150 °C; $[\alpha]_D -10.8^\circ$ ($c = 0.66$, CH₃OH). HR-ESI-MS (positive mode) m/z 734.3176 $[M+H]^+$ (calculated for C₄₀H₄₈NO₁₂⁺ 734.3171, $\Delta M = -0.6$ ppm). ¹H NMR (600 MHz, DMSO-*d*₆): δ 1.05 (3H, s, H-17), 1.07 (3H, s, H-16), 1.54 (3H, s, H-19), 1.66-1.68 (2H, m, H-6a, H-14a), 1.86 (3H, s, H-18), 2.08 (1H, m, H-14b), 2.12 (3H, s, 4-OAc), 2.14 (3H, s, 10-OAc), 2.37 (1H, m, H-6b), 2.76 (2H, m, H-2'), 3.70 (1H, d, $J = 7.0$ Hz, H-3), 4.03 (2H, m, H-20), 4.15 (1H, dd, $J = 10.6, 6.7$ Hz, H-7), 4.30 (1H, t, $J = 6.9$ Hz, H-3'), 4.92 (1H, d, $J = 10.0$ Hz, H-5), 5.47 (1H, d, $J = 7.1$ Hz, H-2), 5.97 (1H, t, $J = 8.2$ Hz, H-13), 6.36 (1H, s, H-10), 7.24 (1H, m, H-*p*-Ph), 7.34 (2H, m, H-*m*-Ph), 7.40 (2H, d, $J = 7.0$ Hz, H-*o*-Ph), 7.57 (2H, m, H-*m*-Bz), 7.67 (1H, m, H-*p*-Bz), 7.98 (2H, dd, $J = 7.7, 1.3$ Hz, H-*o*-Bz). ¹³C NMR (125 MHz, DMSO-*d*₆): δ 9.79 (C-19), 14.35 (C-18), 20.75 (C-17), 21.09 (CH₃ of 10-OAc), 22.34 (CH₃ of 4-OAc), 26.38 (C-16), 35.37 (C-14), 36.68 (C-6), 43.01 (C-15), 44.24 (C-2'), 46.37 (C-3), 52.67 (C-3'), 57.66 (C-8), 69.23 (C-13), 70.63 (C-7), 74.46 (C-2), 75.03 (C-10), 75.47 (C-20), 76.77 (C-1), 80.47 (C-4), 83.70 (C-5), 126.39 (C-*m*-Ph), 126.89 (C-*p*-Ph), 128.28 (C-*o*-Ph), 128.77 (C-*m*-Bz), 129.63 (C-*o*-Bz), 130.02 (C-1-Bz), 133.30 (C-11), 133.45 (C-*p*-Bz), 139.70 (C-12), 145.54 (C-1-Ph), 165.26 (C=O of Bz), 168.92 (C=O of 10-OAc), 169.75 (C=O of 4-OAc), 171.32 (C-1'), 202.67 (C-9).

N-debenzoyl-taxol (compound 3):

White solid; mp = 190-192 °C; $[\alpha]_D -19.5^\circ$ ($c = 0.76$, CHCl₃). HR-ESI-MS (positive mode) m/z 750.3130 $[M+H]^+$ (calculated for C₄₀H₄₈NO₁₃⁺ 750.3120, $\Delta M = -1.3$ ppm). ¹H NMR (600 MHz, DMSO-*d*₆): δ 1.02 (3H, s, H-17), 1.03 (3H, s, H-16), 1.50 (3H, s, H-19), 1.63-1.69 (2H, m, H-6a, H-14a), 1.79 (1H, m, H-14b), 1.80 (3H, s, H-18), 2.11 (3H, s, 4-OAc), 2.11 (3H, s, 10-OAc), 2.32 (1H, m, H-6b), 3.58 (1H, d, $J = 7.3$ Hz, H-3), 3.99 (2H, m, H-20), 4.10 (1H, dd, $J = 10.3, 6.7$ Hz, H-7), 4.16 (1H, d, $J = 7.8$ Hz, H-3'), 4.20 (1H, d, $J = 7.8$ Hz, H-2'), 4.88 (1H, d, $J = 10.4$ Hz, H-5), 5.42 (1H, d, $J = 7.3$ Hz, H-2), 5.86 (1H, t, $J = 8.8$ Hz, H-13), 6.29 (1H, s, H-10), 7.22 (1H, m, H-*p*-Ph), 7.38-7.39 (4H, m, H-*m*-Ph, H-*o*-Ph), 7.63 (2H, m, H-*m*-Bz), 7.71 (1H, m, H-*p*-Bz), 7.94 (2H, d, $J = 7.2, 1.2$ Hz, H-*o*-Bz). ¹³C NMR (125 MHz, DMSO-*d*₆): δ 9.92 (C-19), 14.05 (C-18), 20.83 (C-17), 21.47 (CH₃ of 10-OAc), 22.54 (CH₃ of 4-OAc), 26.47 (C-16), 34.85 (C-14), 36.63 (C-6), 43.13 (C-15), 46.25 (C-3), 57.62 (C-8), 58.37 (C-3'), 69.76

(C-13), 70.65 (C-7), 74.62 (C-2), 74.98 (C-10), 77.52 (C-2'), 75.55 (C-20), 76.93 (C-1), 80.43 (C-4), 83.79 (C-5), 127.68 (C-*m*-Ph), 127.98 (C-*p*-Ph), 128.50 (C-*o*-Ph), 128.88 (C-*m*-Bz), 129.70 (C-*o*-Bz), 130.07 (C-1, Bz), 133.44 (C-12), 133.70 (C-*p*-Bz), 139.30 (C-12), 139.59 (C-1, Ph), 165.40 (C=O of Bz), 169.06 (C=O of 10-OAc), 170.01 (C=O of 4-OAc), 172.81 (C-1'), 202.69 (C-9).

10-deacetyl-N-debenzoyl-2'-deoxy-taxol (compound 6):

White solid; mp = 160-162 °C; $[\alpha]_D -13.9^\circ$ (c = 0.85, CH₃OH). HR-ESI-MS (positive mode) m/z 692.3070 [M+H]⁺ (calculated for C₃₈H₄₆NO₁₁⁺ 692.3065, $\Delta M = -0.7$ ppm). ¹H NMR (600 MHz, DMSO-*d*₆): δ 1.01 (3H, s, H-17), 1.04 (3H, s, H-16H), 1.55 (3H, s, H-19), 1.64 (1H, overlapped, H-6a), 1.68 (2H, m, H-14a), 1.78 (3H, s, H-18), 2.04 (1H, m, H-14b), 2.12 (3H, s, OAc), 2.29 (1H, m, H-6b), 2.76 (2H, m, H-2'), 3.74 (1H, d, $J = 7.0$ Hz, H-3), 4.03-4.08 (3H, overlapped, H-20a, H-20b, H-7), 4.31 (1H, t, $J = 6.4$ Hz, H-3'), 4.91 (1H, d, $J = 9.5$ Hz, H-5), 5.14 (1H, s, H-10), 5.45 (1H, d, $J = 7.0$ Hz, H-2), 5.96 (1H, t, $J = 8.5$ Hz, H-13), 7.25 (1H, m, H-*p*-Ph), 7.34 (2H, m, H-*m*-Ph), 7.40 (2H, m, H-*o*-Ph), 7.57 (2H, m, H-*m*-Bz), 7.67 (1H, t, $J = 7.4$ Hz, H-*p*-Bz), 7.98 (2H, d, $J = 7.4$ Hz, H-*o*-Bz). ¹³C NMR (125 MHz, DMSO-*d*₆): δ 9.82 (C-19), 14.21 (C-18), 20.59 (C-17), 22.36 (CH₃ of Ac), 26.60 (C-16), 35.69 (C-14), 36.60 (C-6), 42.94 (C-15), 44.23 (C-2'), 46.21 (C-3), 52.64 (C-3'), 57.23 (C-8), 69.49 (C-13), 70.96 (C-7), 74.00 (C-10), 74.75 (C-2), 75.58 (C-20), 76.91 (C-1), 80.48 (C-4), 83.84 (C-5), 126.42 (C-*m*-Ph), 126.93 (C-*p*-Ph), 128.30 (C-*o*-Ph), 128.77 (C-*m*-Bz), 129.62 (C-*o*-Bz), 130.12 (C-1, Bz), 133.42 (C-*p*-Bz), 136.23 (C-12), 136.81 (C-11), 145.44 (C-1-Ph), 165.31 (C=O of Bz), 169.67 (C=O of OAc), 171.37 (C-1'), 209.65 (C-9).

10-deacetyl-N-debenzoyl-taxol (compound 7):

White Solid; mp = 174-176 °C (dec.); $[\alpha]_D -30.7^\circ$ (c = 0.55, CH₃OH). HR-ESI-MS (positive mode) m/z 708.3020 [M+H]⁺ (calculated for C₃₈H₄₆NO₁₂⁺ 708.3015, $\Delta M = -0.8$ ppm). ¹H NMR (600 MHz, DMSO-*d*₆): δ 0.99 (3H, s, H-17), 1.03 (3H, s, H-16H), 1.53 (3H, s, H-19), 1.64-1.68 (2H, overlapped, H-6a, H-14a), 1.75 (3H, s, H-18), 1.83 (1H, m, H-14b), 2.11 (3H, s, OAc), 2.28 (1H, m, H-6b), 3.65 (1H, d, $J = 4.9$ Hz, H-3), 4.00 (1H, m, H-20a), 4.02-4.07 (4H, overlapped, H-2', H-3', H-20b, H-7), 4.52 (1H, br s, OH-10), 4.89 (1H, d, $J = 8.2$ Hz, H-5), 5.11 (1H, s, H-10), 5.42 (1H, d, $J = 4.7$ Hz, H-2), 5.89 (1H, t, $J = 7.8$ Hz, H-13), 7.19 (1H, m, H-*p*-Ph), 7.35 (2H, m, H-*m*-Ph), 7.35 (2H, m, H-*o*-Ph), 7.62 (2H, m, H-*m*-Bz), 7.71 (1H, m, H-*p*-Bz), 7.96 (2H, d, $J = 6.0$ Hz, H-*o*-Bz). ¹³C NMR (125 MHz, DMSO-*d*₆): δ 9.93 (C-19), 13.86 (C-18), 20.96 (C-17), 22.44 (CH₃ of OAc), 26.66 (C-16), 35.19 (C-14), 36.57 (C-6),

43.04 (C-15), 46.08 (C-3), 57.15 (C-8), 58.82 (C-3'), 69.78 (C-13), 70.94 (C-7), 73.94 (C-10), 74.91 (C-2), 75.60 (C-20), 76.60 (C-1), 77.06 (C-2'), 80.42 (C-4), 83.90 (C-5), 127.33 (C-*p*-Ph), 127.42 (C-*m*-Ph), 128.23 (C-*o*-Ph), 128.81 (C-*m*-Bz), 129.68 (C-*o*-Bz), 130.16 (C-1-Bz), 133.58 (C-*p*-Bz), 136.33 (C-12), 136.79 (C-11), 141.74 (C-1-Ph), 165.42 (C=O of Bz), 169.86 (C=O of OAc), 173.38 (C1'), 209.62 (C-9). The NMR data are in agreement with those reported in the literature⁷.

(3'R/3'S)-2'-deoxy-taxol (referred to as 2'-deoxy-taxol (**9**) in the main text for simplicity):

White solid; mp = 214-216 °C (dec.); $[\alpha]_D -20.8^\circ$ (c = 0.75, CH₃OH). HR-ESI-MS (positive mode) m/z 838.3447 [M+H]⁺ (calculated for C₄₇H₅₂NO₁₃⁺ 838.3433, $\Delta M = -1.6$ ppm). ¹H NMR (600 MHz, DMSO-*d*₆): δ 1.00/1.01 (3H, s, H-17), 1.03 (3H, s, H-16H), 1.52/1.57 (3H, s, H-19), 1.84 (3H, s, H-18), 2.08/2.09 (3H, s, 10-OAc), 2.17 (3H, s, 4-OAc), 2.35 (1H, m, H-6b), 3.05/3.08 (1H, m, H-2'b), 3.17/3.18 (1H, m, H-2'a), 3.69 (1H, m, H-3), 4.02/4.03 (1H, m, H-20a), 4.93/4.97 (1H, m, H-5), 5.46 (1H, m, H-2), 5.52/5.56 (1H, m, H-3'), 5.96 (1H, m, H-13), 6.29/6.33 (1H, s, H-10), 7.25 (1H, m, H-*p*-Ph), 7.35 (2H, m, H-*m*-Ph), 7.35 (2H, m, H-*o*-Ph), 7.45 (2H, m, H-*m*-NBz), 7.51 (1H, m, H-*p*-NBz), 7.54 (2H, m, H-*m*-OBz), 7.65 (1H, m, H-*p*-OBz), 7.85 (2H, m, H-*o*-NBz), 7.96 (2H, m, H-*o*-OBz). ¹³C NMR (125 MHz, DMSO-*d*₆): δ 9.98/10.00 (C-19), 14.08/14.47 (C-18), 20.59/21.18 (C-17), 22.51/22.74 (CH₃ of Ac), 26.54/26.57 (C-16), 35.52 (C-14), 36.85/36.88 (C-6), 40.56/40.79 (C-2'), 43.20/43.17 (C-15), 46.55/46.62 (C-3), 50.08/50.52 (C-3'), 57.85/57.87 (C-8), 69.45/69.59 (C-13), 70.82 (C-7), 74.60/74.67 (C-10), 75.20 (C-2), 75.67/75.72 (C-20), 76.83/77.0 (C-1), 80.61/80.65 (C-4), 83.93 (C-5), 126.88/127.12 (C-*m*-Ph), 127.55/127.57 (C-*p*-Ph), 127.55 (C-*o*-NBz), 128.60 (C-*o*-Ph), 128.61 (C-*m*-NBz), 128.60 (C-*m*-NBz), 129.02 (C-*m*-OBz), 129.83 (C-*o*-OBz), 130.15 (C-1-OBz), 131.68 (C-1-NBz), 133.52/133.71 (C-*p*-OBz), 134.49/134.54 (C-11), 139.79/139.83 (C-12), 142.22/142.52 (C-1-Ph), 165.52 (C=O of OBz), 166.00/166.07 (C=O of NBz), 169.16/169.17 (C=O of 10-OAc), 170.01/170.05 (C=O of 4-OAc), 170.50/170.77 (C-1'), 202.94/202.96 (C-9). The NMR data are in agreement with those reported in the literature¹¹.

Protein extraction from tobacco leaves

Protein was extracted from tobacco leaves four days after agro-infiltration using TCA/acetone precipitation with modified solubilization solution¹². After 2-mercaptoethanol/acetone rinsing and centrifugation, the pellet was re-suspended in solubilization solution (8M Urea, 5 mM K₂CO₃, 2% SDS, 5% DTT, 5% Trion X-100, and 1× Protease Inhibitor cocktail in 100 mM Tris-

HCl (pH 7.5)) using 100 μ L solubilization solution for 100 mg leaves (wet weight). Western blot was performed following a published protocol¹³.

Yeast media

Yeast glucose media: 2% (w/v) glucose, 0.13% (w/v) Yeast Synthetic Dropout Medium Supplements (all essential amino acids), 0.67% (w/v) Yeast Nitrogen Base w/o AA. Yeast galactose/raffinose medium: 2% (w/v) galactose, 1% (w/v) raffinose, 0.13% (w/v) Yeast Synthetic Dropout Medium Supplements (all essential amino acids), 0.67% (w/v) Yeast Nitrogen Base w/o AA (Y2025, US Biologicals).

BAPT optimization

The BAPT engineering was based on solubility prediction. Six BAPT homologs from other *Taxus* species were retrieved from the NCBI database. We aligned the amino acid sequences of *Taxus cuspidata* BAPT (the enzyme used in this study) with these homologs to identify residues that varied among the sequences, considering that these would likely not be essential for BAPT function or folding. We identified 36 such amino acid residues and then applied the ProteinSol algorithm to calculate the predicted scaled solubility of *Taxus cuspidata* BAPT sequences in which each of these residues was substituted individually. Through this analysis, we identified nine substitutions that were predicted to significantly enhance solubility. These nine changes were introduced into the original *T. cuspidata* BAPT sequence via site-directed mutagenesis, generating a variant we termed BAPT_m. BAPT_m displayed a predicted scaled solubility of 0.357, compared to 0.236 for the wild-type *T. cuspidata* BAPT. The predicted solubility scores of the other homologous BAPTs were similar to that of wild-type *T. cuspidata* BAPT (**Table S5**).

Protein extraction from yeast

For protein extraction, seed cultures were inoculated into yeast synthetic media with the relevant autotrophic selection and cultivated overnight at 30 °C. The next day, overnight cultures were washed and resuspended in galactose synthetic media (20 g/L galactose and 10 g/L raffinose) at an OD₆₀₀ = 0.5 and grew overnight at 20 °C or 30°C as required. On the third day, the cultures were spun down and resuspended in galactose synthetic media, buffered with 200 mM MES (2-(N-morpholino) ethane sulfonic acid) (pH=7), containing 100 mg/L β -phenylalanine and 100 mg/L baccatin III, and cultivated for another three days at 20 °C or one day at 30 °C before protein extraction.

A 1 mL aliquot of $OD_{600}=7$ yeast culture was centrifuged and the pellet was collected for total protein extraction. The cell pellet was treated with 200 μ L NaOH for 3 min at room temperature. Subsequently, 50 μ L lysis buffer (50 mM Tris-HCl (pH 6.8), 50 mM DTT, 10% glycerol, 2% SDS, and 4% 2-mercaptonethanol) were added together with 50 μ L of acid washed glass beads. The mixture was vortexed (Votex GeniE2, Scientific Industries) for 10 rounds of 1 min vortexing and 1 min resting on ice. Following centrifugation (17,115g for 10 min), the supernatant was collected for SDS-PAGE analysis.

Supplementary Figures

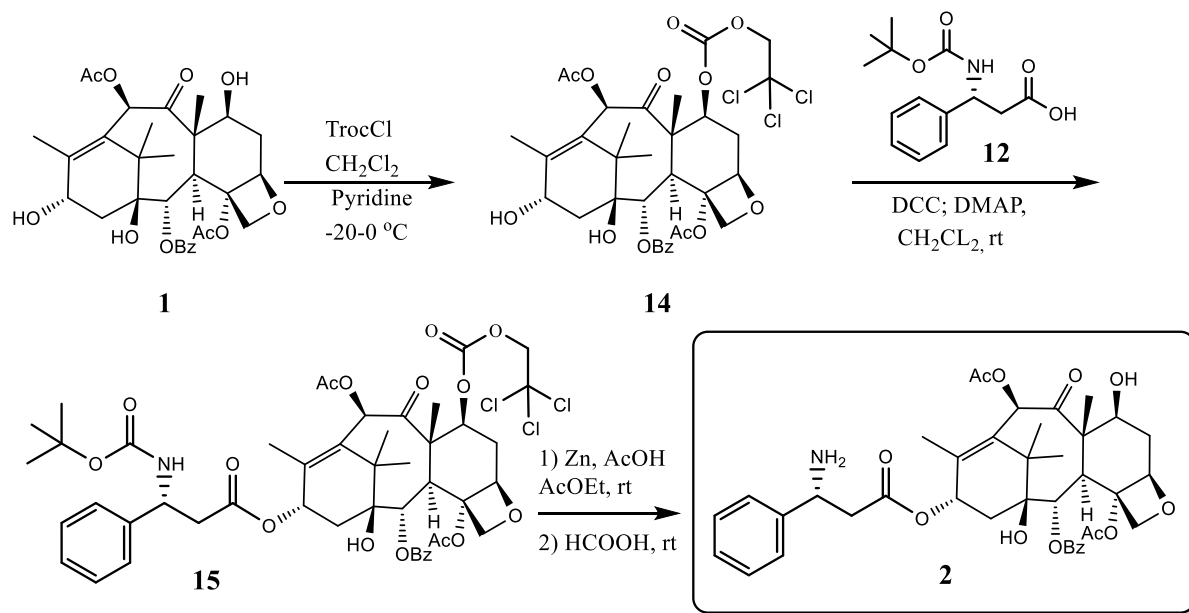


Fig. S1. Chemical synthesis of N-debenzoyl-2'-deoxy-taxol (compound **2**).

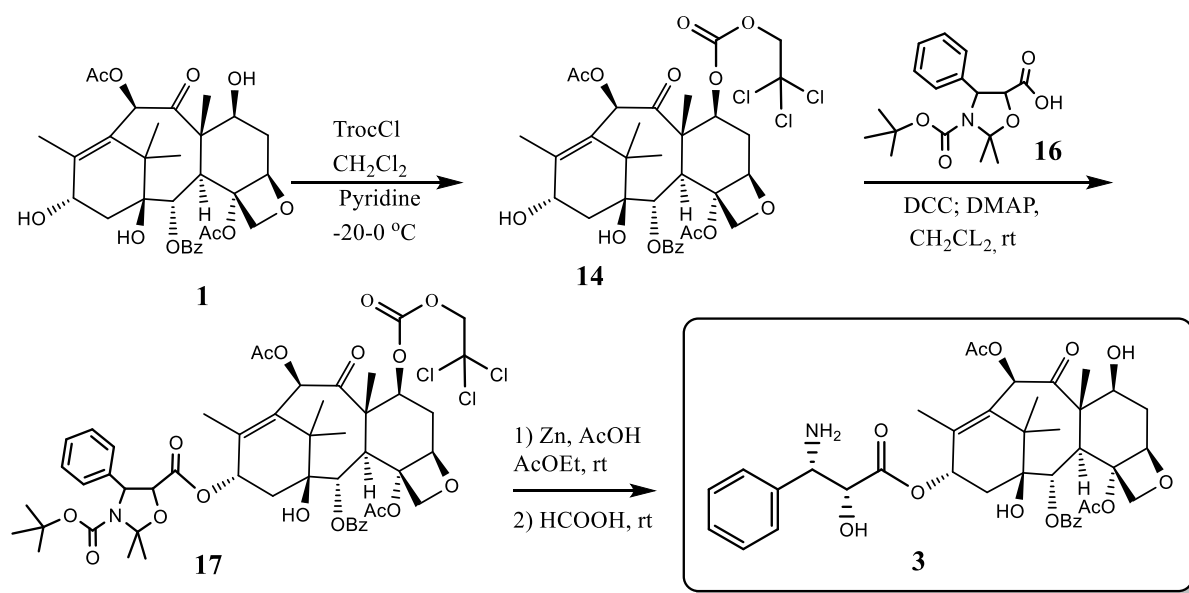


Fig. S2. Chemical synthesis of N-debenzoyl-taxol (compound **3**).

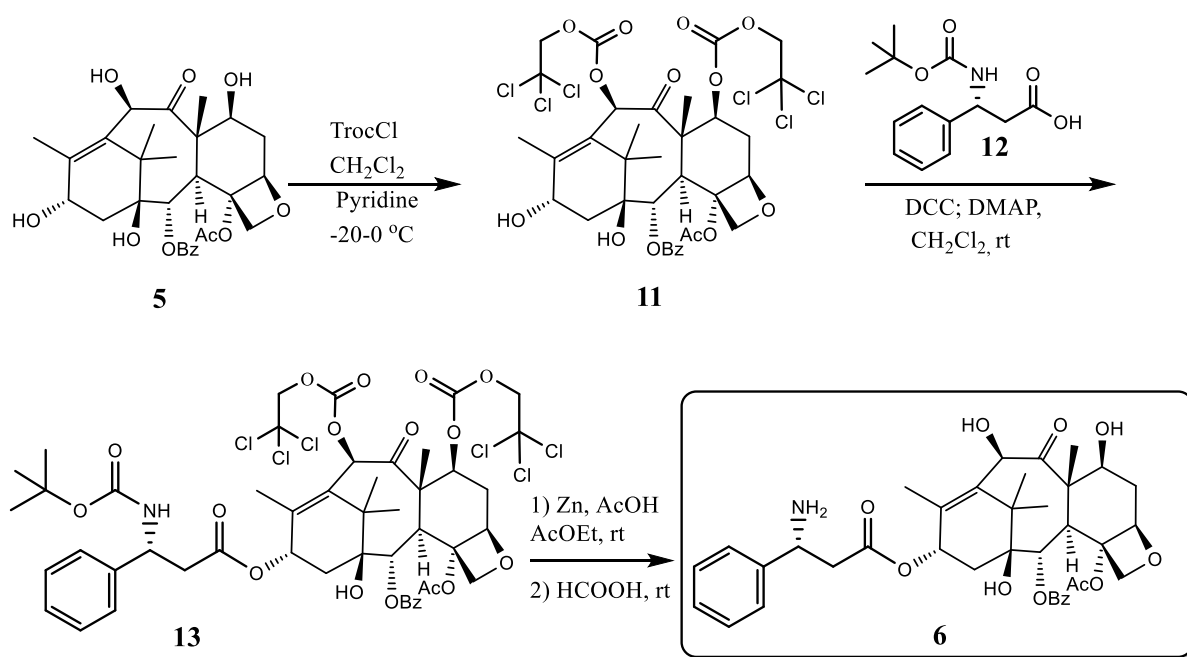


Fig. S3. Chemical synthesis of 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (compound **6**).

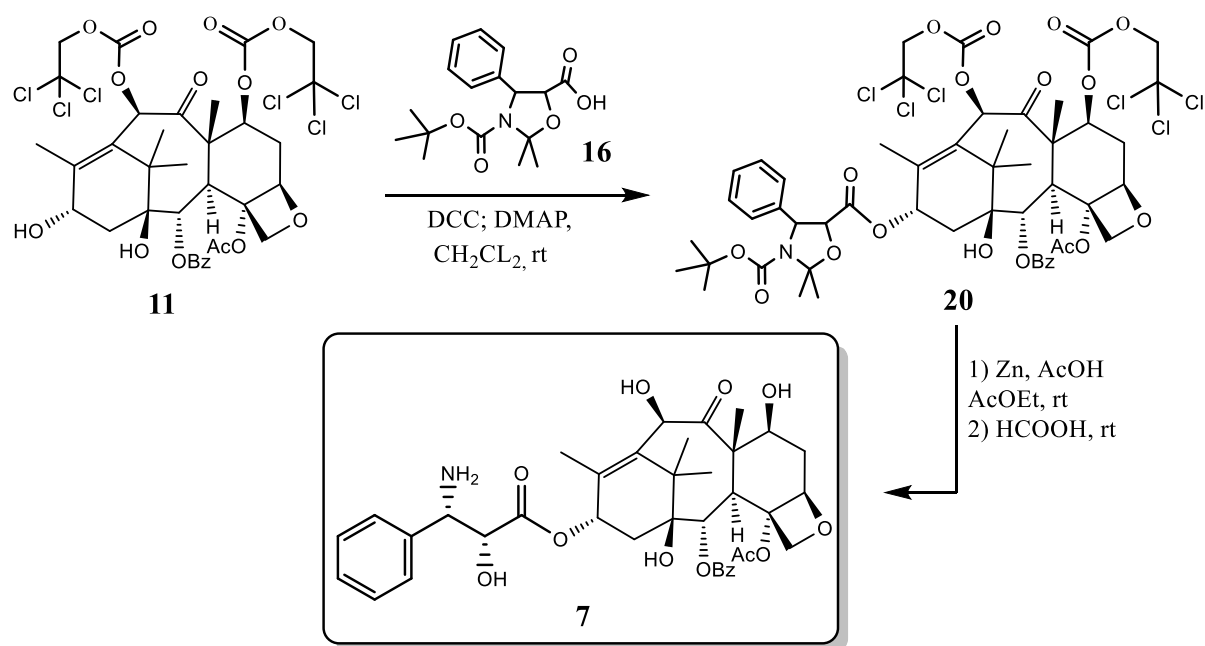


Fig. S4. Chemical synthesis of 10-deacetyl-N-debenzoyl-taxol (compound 7).

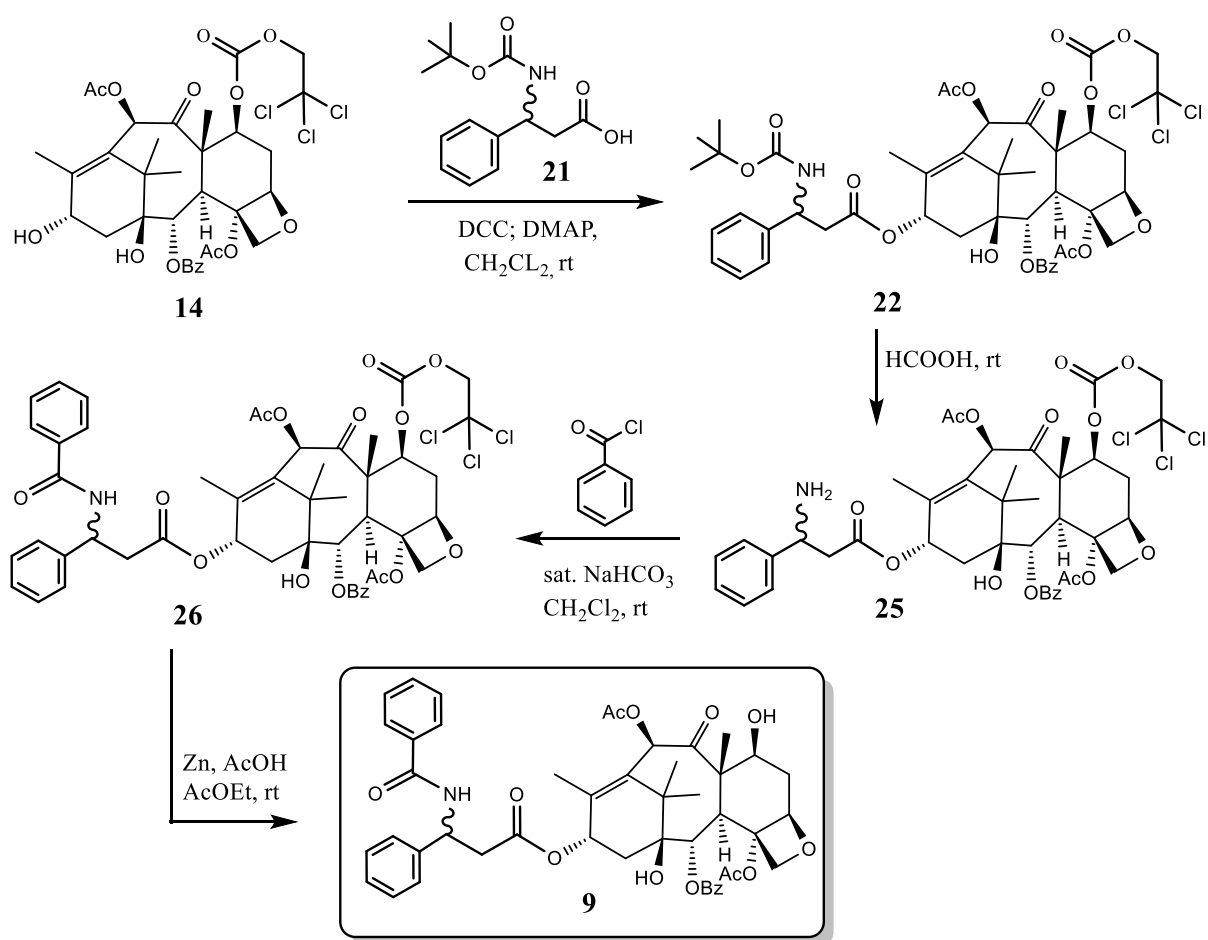
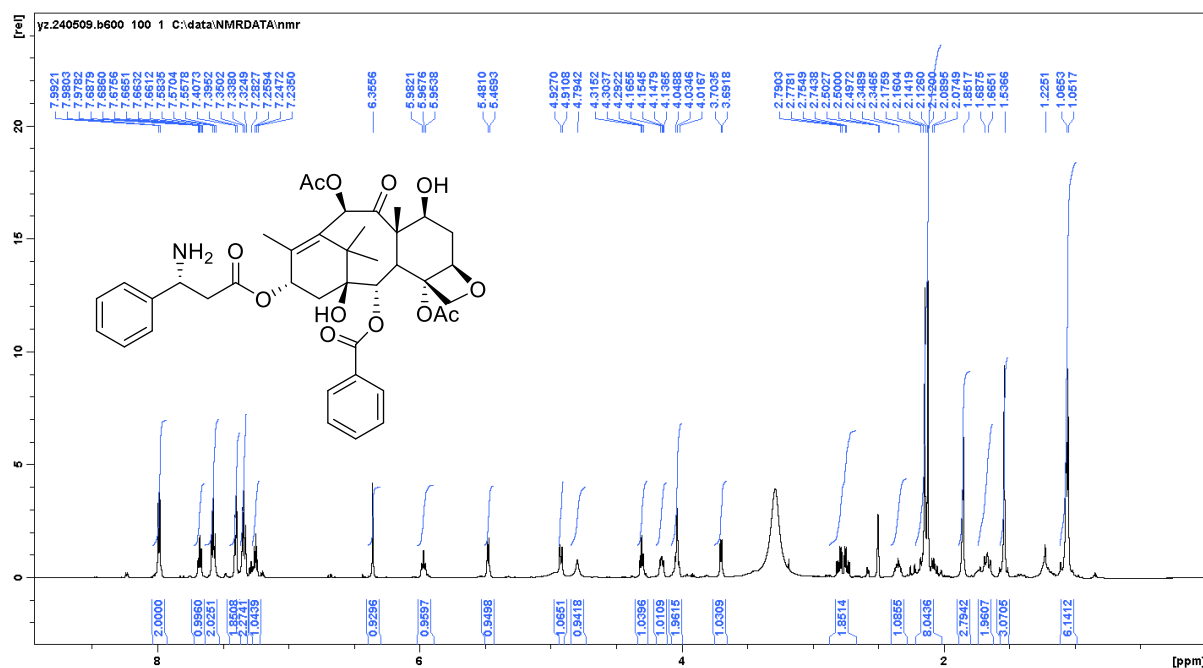


Fig. S5. Chemical synthesis of 2'-deoxy-taxol (compound **9**)



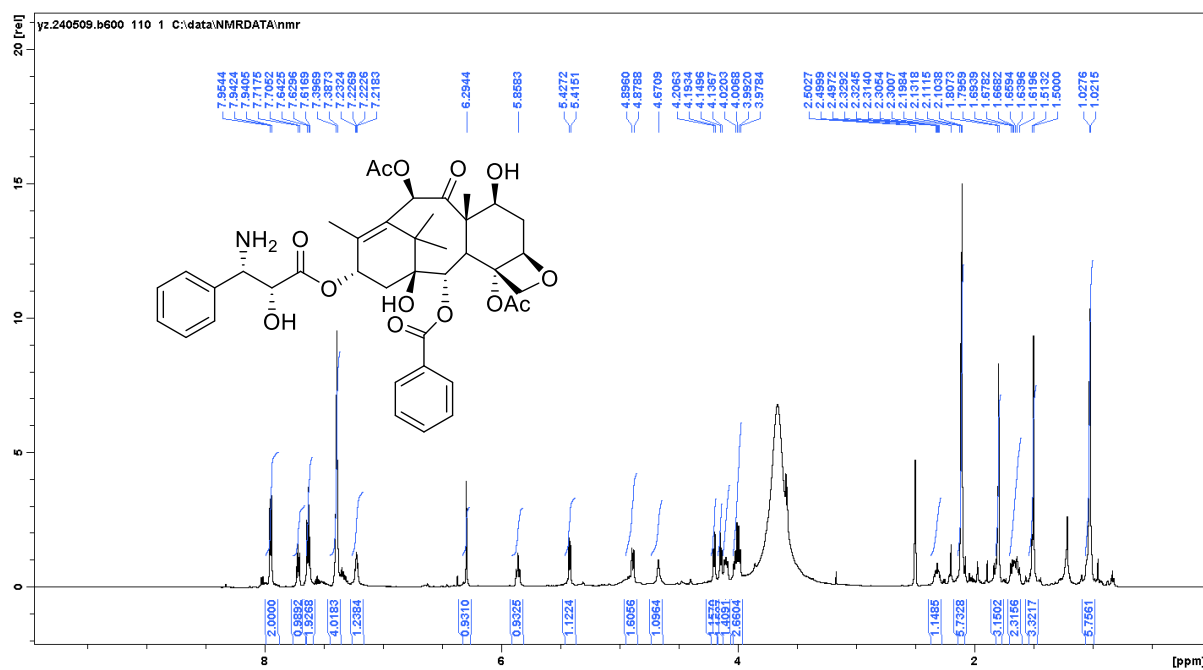


Fig. S8. ^1H -NMR spectrum of N-debenzoyl-taxol (compound 3, 600 MHz, $\text{DMSO}-d_6$).

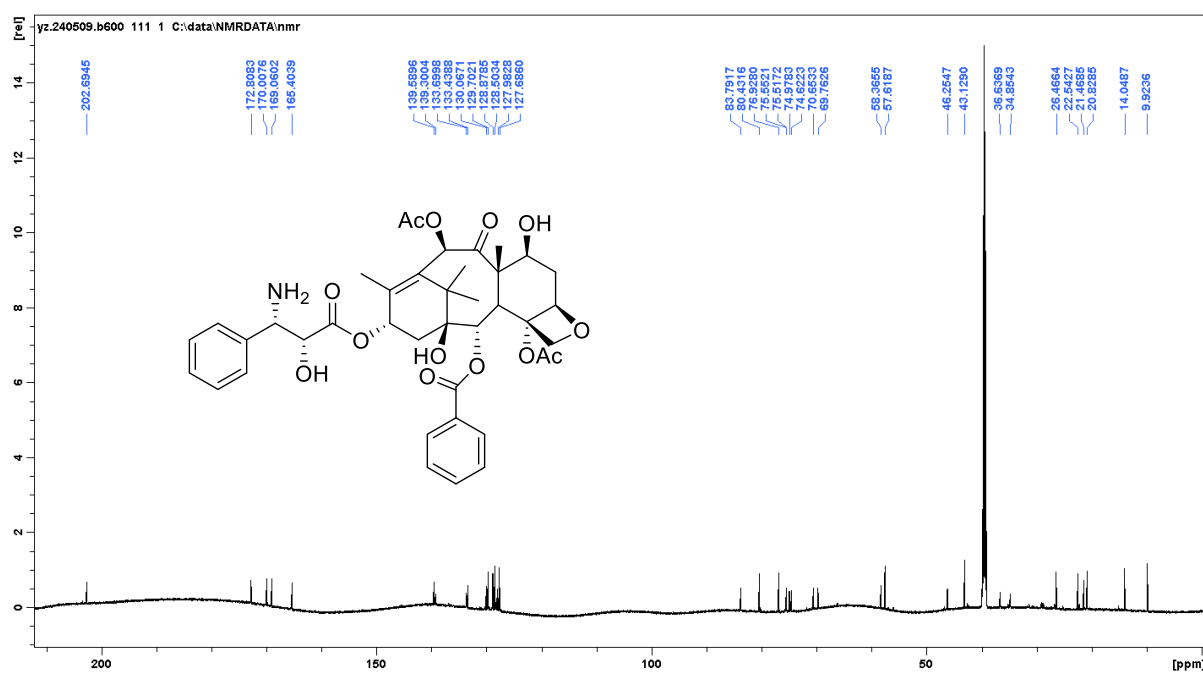


Fig. S9. ^{13}C -NMR spectrum of N-debenzoyl-taxol (compound 3, 125 MHz, $\text{DMSO}-d_6$).

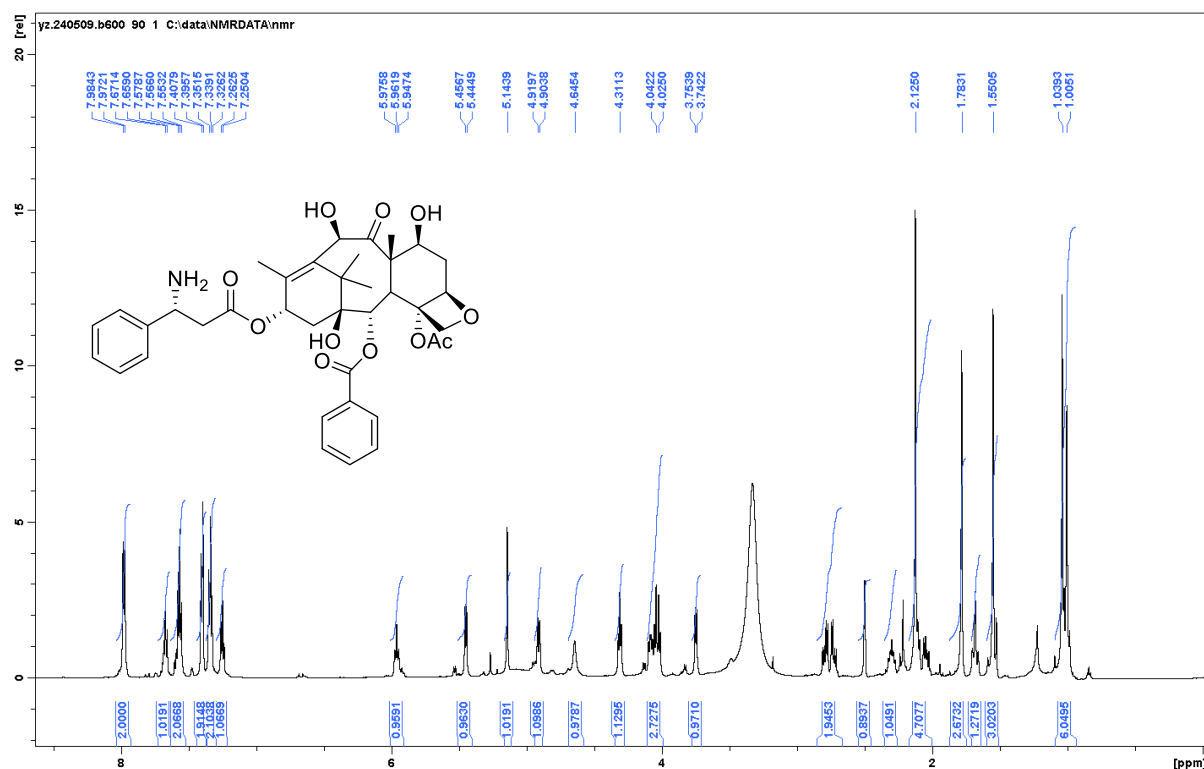


Fig. S10. ¹H-NMR spectrum of 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (compound **6**, 600 MHz, DMSO-*d*₆).

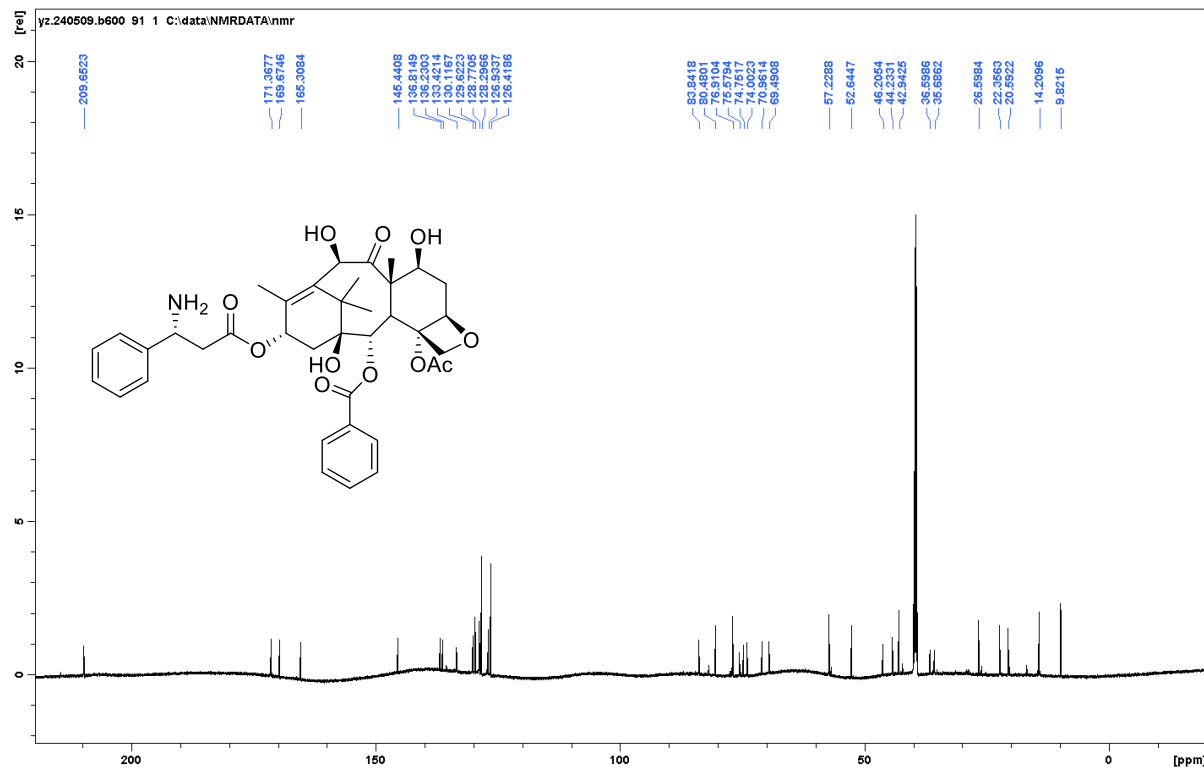


Fig. S11. ¹³C-NMR spectrum of 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (compound **6**, 125 MHz, DMSO-*d*₆).

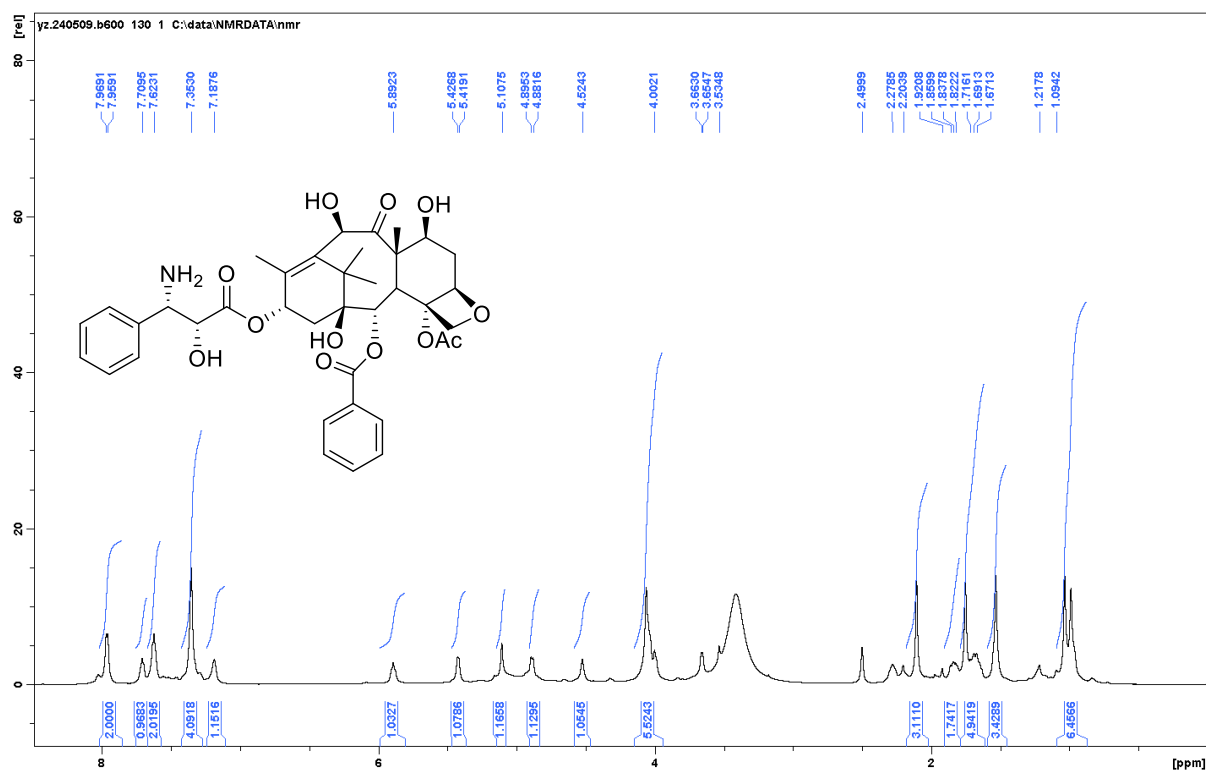


Fig. S12. ^1H -NMR spectrum of 10-deacetyl-N-debenzoyl-taxol (compound 7, 600 MHz, $\text{DMSO}-d_6$).

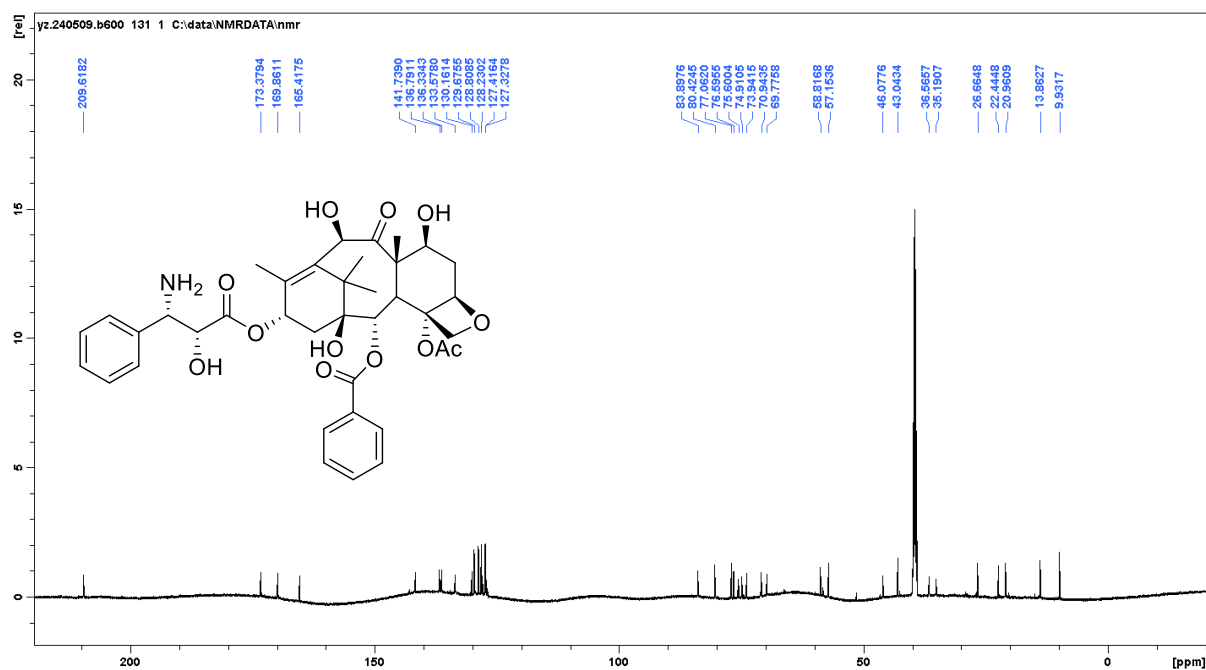


Fig. S13. ^{13}C -NMR spectrum of 10-deacetyl-N-debenzoyl-taxol (compound 7, 125 MHz, $\text{DMSO}-d_6$).

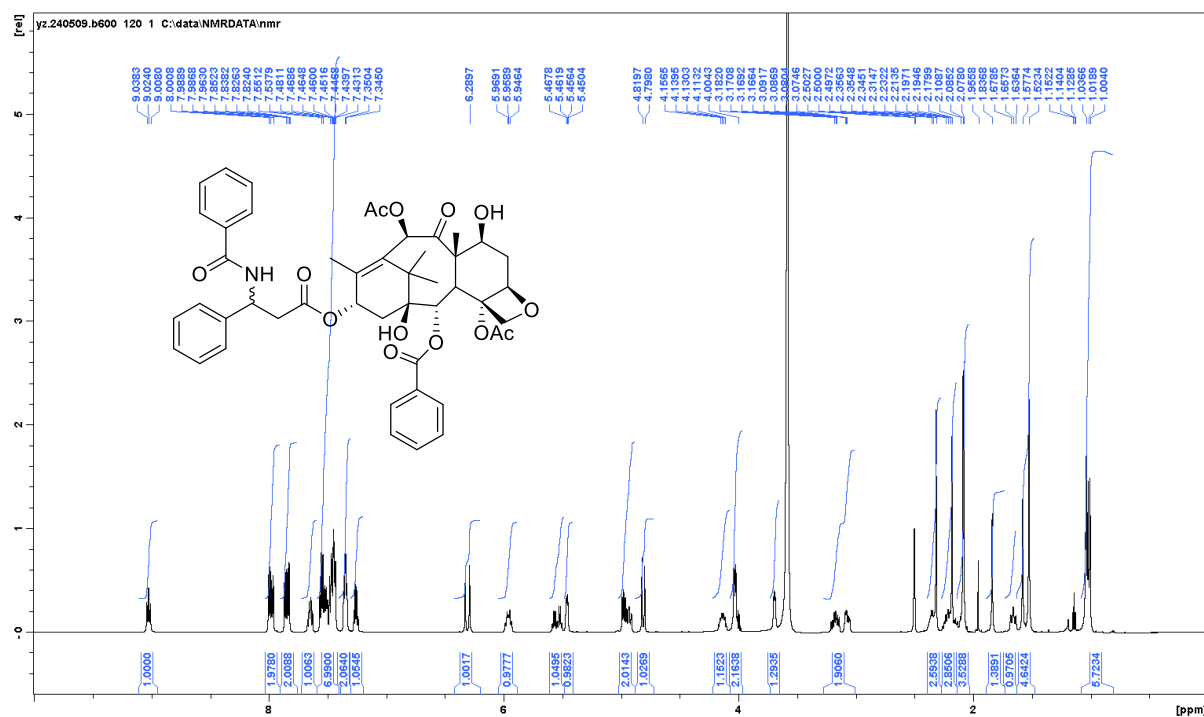


Fig. S14. ^1H -NMR spectrum of (3'R/3'S)-2'-deoxy-taxol (600 MHz, $\text{DMSO}-d_6$).

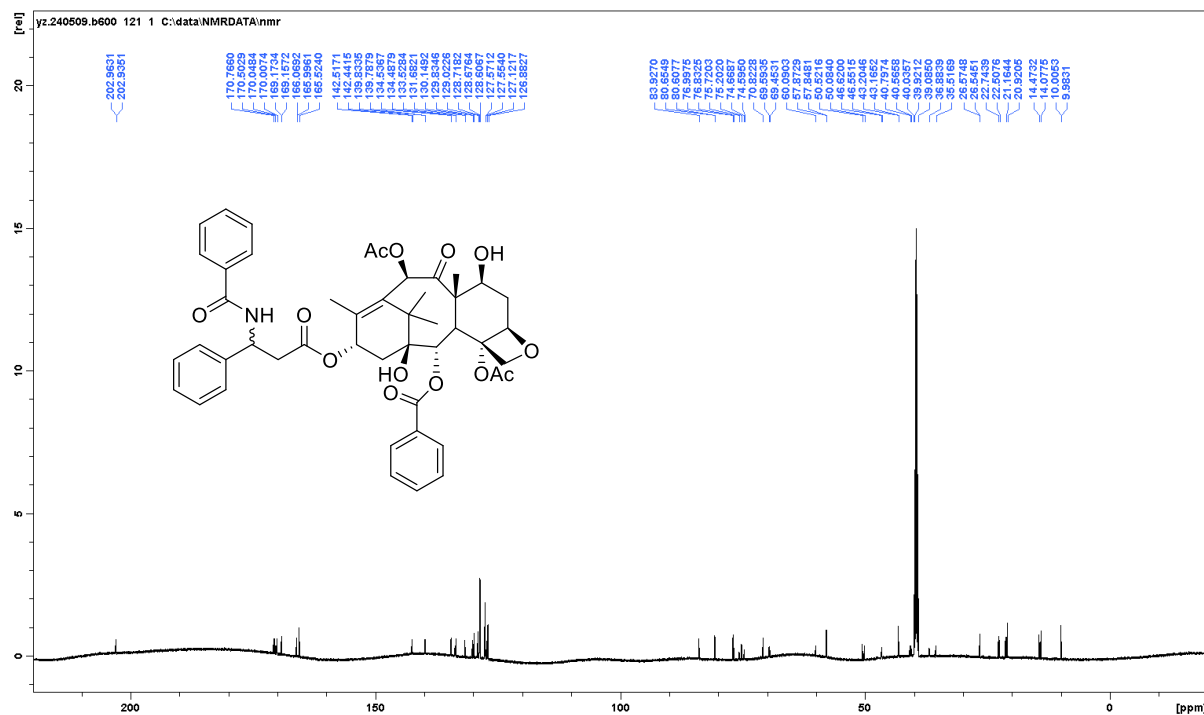


Fig. S15. ^{13}C -NMR spectrum of (3'R/3'S)-2'-deoxy-taxol (125 MHz, $\text{DMSO}-d_6$).

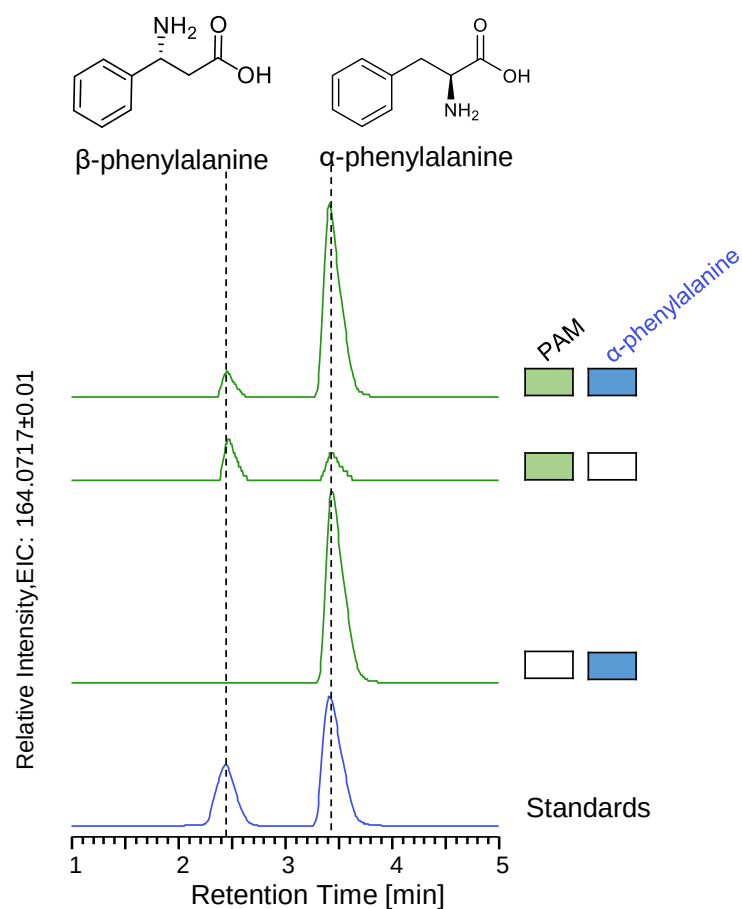


Fig. S16. Transient expression of *T. baccata* phenylalanine aminomutase (PAM) in tobacco leaves via agro-infiltration resulted in the formation of β -phenylalanine ($C_9H_{11}NO_2$) when endogenous α -phenylalanine ($C_9H_{11}NO_2$) was used as substrate. Feeding α -phenylalanine (400 mg/L) to tobacco leaves did not increase the production of β -phenylalanine. This indicates that the levels of endogenous α -phenylalanine in tobacco leaves are sufficient to saturate the catalytic capacity of PAM. Green squares indicate the genes used for agro-infiltration of each sample, while blue squares indicate the exogenously supplied substrate. A filled square indicates the presence, and an empty square indicates the absence, of the corresponding gene or substrate. $[M-H]^-$, $m/z = 164.0717 \pm 0.01$ extracted ion chromatogram.

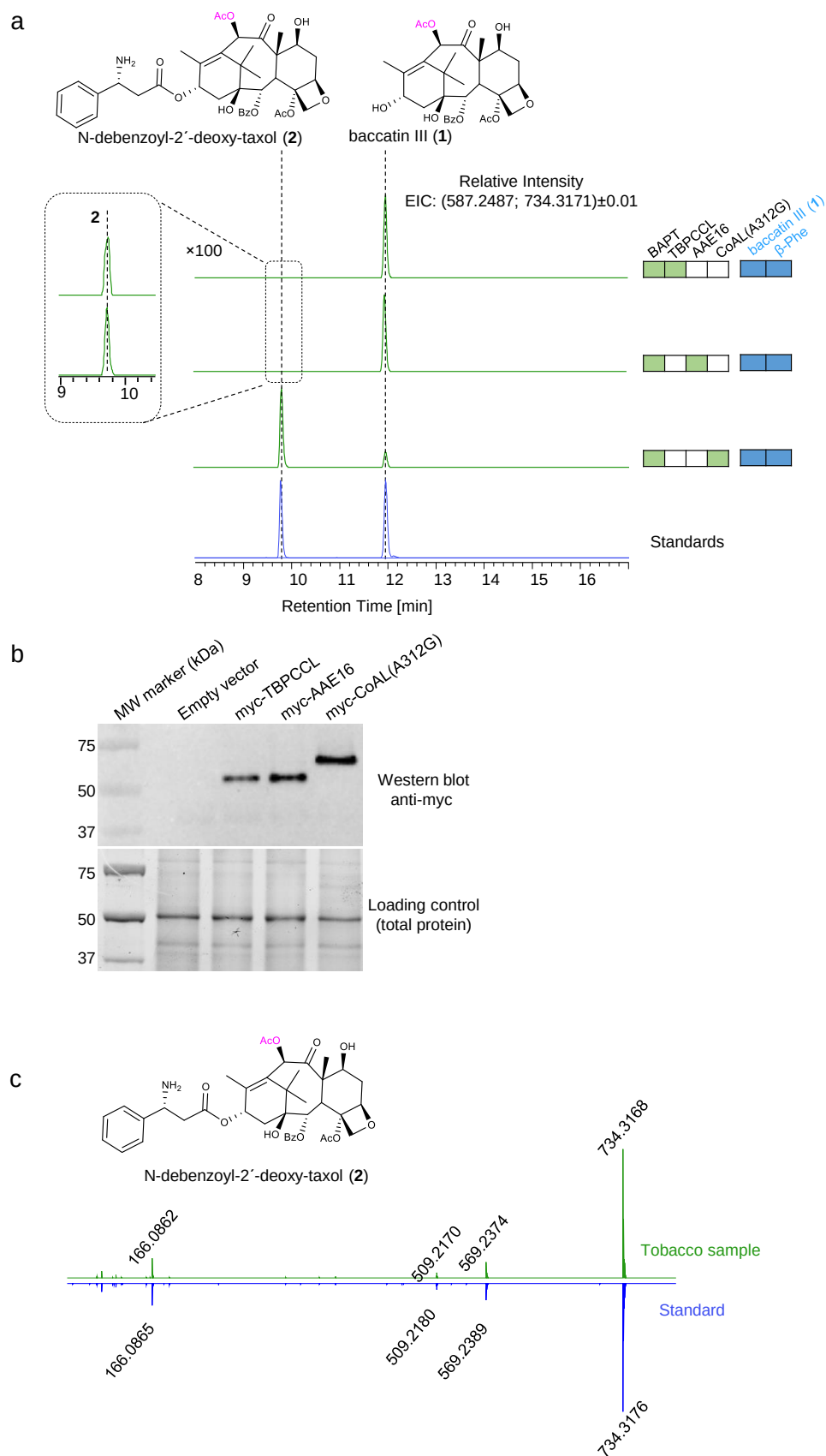


Fig. S17. The CoA ligase from *P. chrysogenum*, CoAL(A312G), is more efficient than TBPCCL or AAE16 in converting β -phenylalanine to β -phenylalanyl-CoA in tobacco, as

indicated by the levels of N-debenzoyl-2'-deoxy-taxol (**2**) produced in tobacco leaves. **a.** Extracted ion chromatograms (EIC, positive) of baccatin III (**1**, $C_{31}H_{38}O_{11}^+$, $m/z = 587.2487 \pm 0.01$) and N-debenzoyl-2'-deoxy-taxol (**2**, $C_{40}H_{47}NO_{12}^+$, $m/z = 734.3171 \pm 0.01$) of extracts of tobacco leaves transiently expressing the different CoA ligase genes via agro-infiltration. The corresponding area of the top two chromatograms at x100 magnification is shown in the inset. Green squares indicate the genes used in agro-infiltration of each sample, while blue squares indicate the supplied substrates. A filled square indicates the presence, and an empty square indicates the absence, of the corresponding gene or substrate. **b.** Western blot of the three CoA ligases bearing an N-terminal myc-tag. Total protein was extracted from tobacco leaves four days after agro-infiltration. The myc-tag (EQKLISEEDL) was fused at the N terminus each of the three CoA ligases using an SGSGSG linker. Empty vector indicates extracts of tobacco leaves where only the inhibitor of silencing P19 was expressed. **c.** Mass spectra of N-debenzoyl-2'-deoxy-taxol (**2**) in positive ionization mode ($[M+H]^+$). The green spectrum is from the compound produced in tobacco, while the blue spectrum is from the chemically synthesized standard.

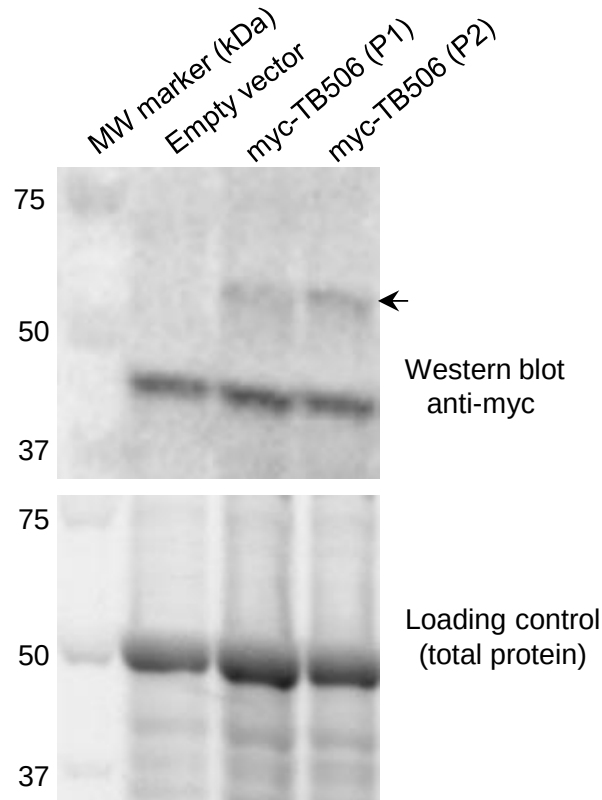


Fig. S18. Western blot of TB506 from tobacco leaves. Total protein was extracted from tobacco leaves four days after agro-infiltration. Two tobacco plants were used in this experiment (P1 and P2). A myc-tag was fused at the N terminus of TB506 through an SGSGSG linker. Empty vector indicates the extract of tobacco leaves where only the inhibitor of silencing P19 was expressed. The arrow points to the TB506 band, which is predicted to be 59 kDa.

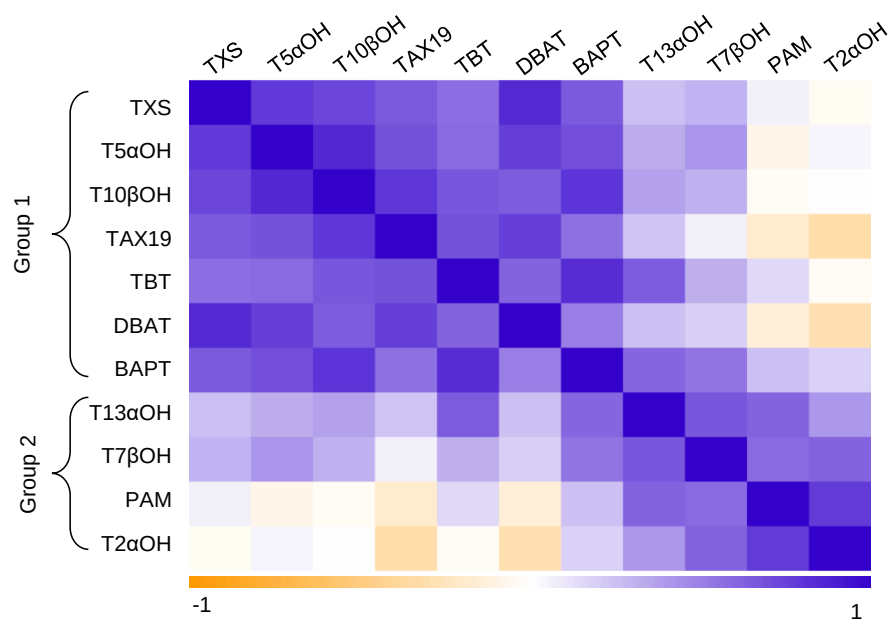


Fig. S19. Correlation heatmap of co-expression patterns between the known Taxol pathway genes used as baits in this study (**Table S2**). The color bar indicates the correlation factor value. Blue indicates a positive correlation, while orange indicates a negative correlation.

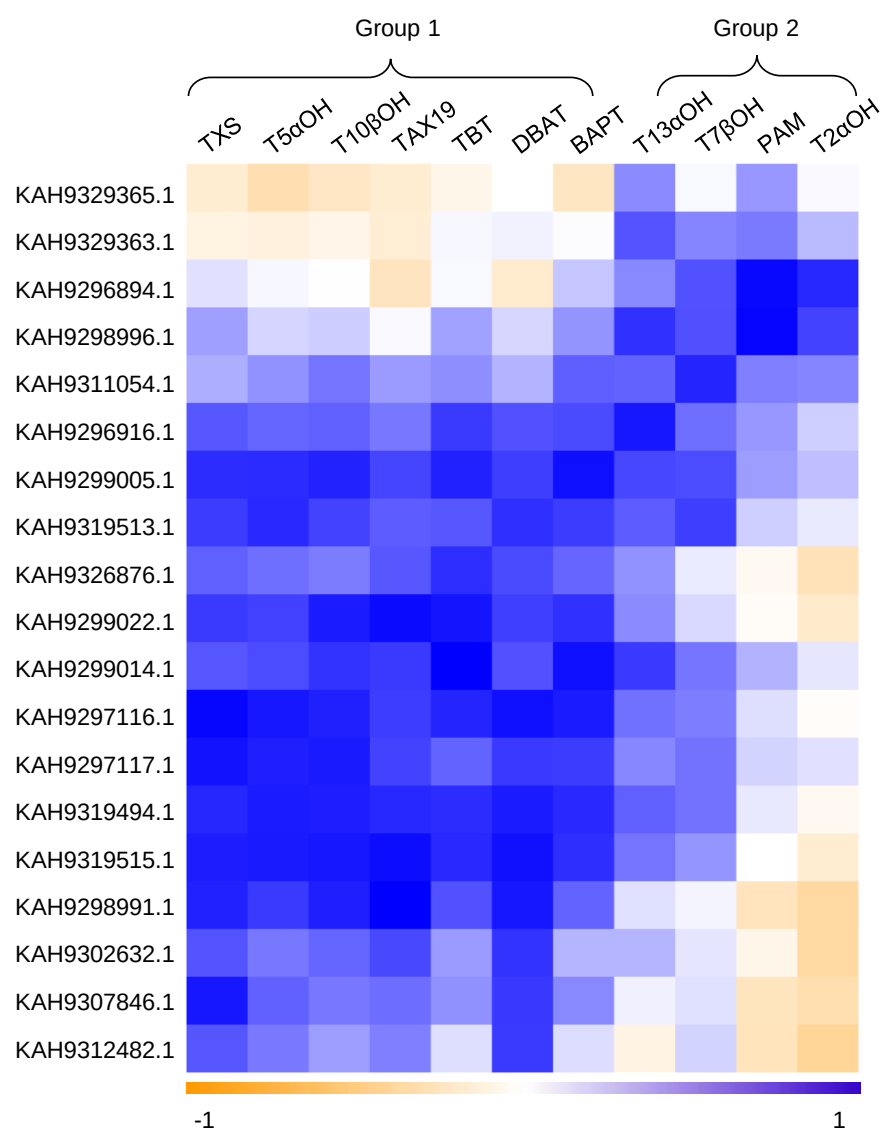


Fig. S20. Correlation heatmap of co-expression patterns between selected *T. chinensis* CYP candidates and baits in this study (**Table S2**). The color bar indicates the correlation factor value. Blue indicates a positive correlation, while orange indicates a negative correlation.

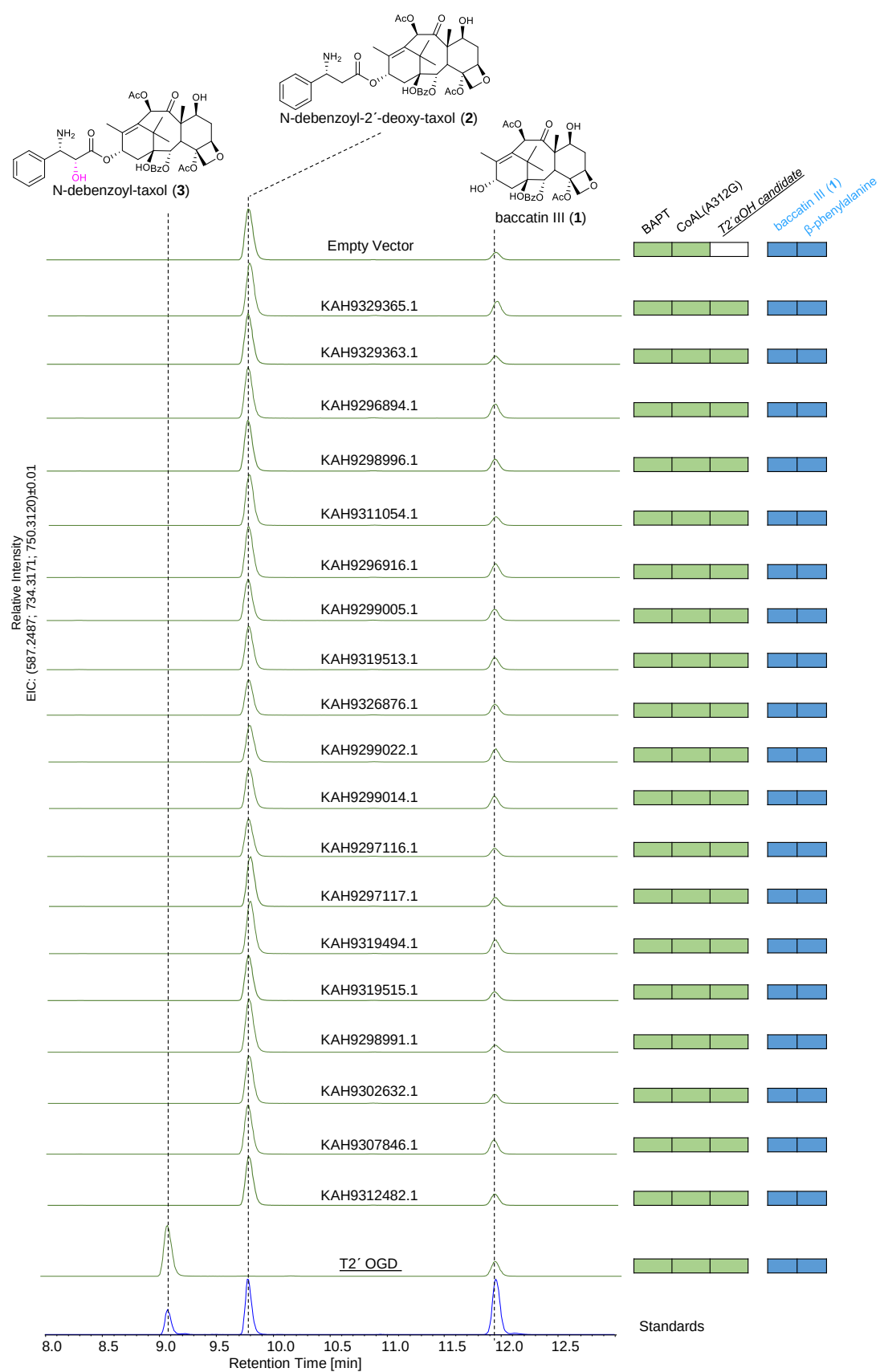


Fig. S21. Analysis of 19 candidate CYPs (**Fig. S20**) obtained from our co-expression analysis.

None of the candidates was capable of synthesizing compound **3** ($\text{C}_{40}\text{H}_{48}\text{NO}_{13}^+$, $m/z =$

750.3120±0.01) using compound **2** ($C_{40}H_{47}NO_{12}^+$, $m/z = 734.3171 \pm 0.01$) as substrate. C2'α hydroxylation by T2'OGD is shown here as positive control. Compound **2** was synthesized by transient expression of BAPT and CoAL(A312G) and feeding baccatin III and β-phenylalanine. Green squares indicate the genes used in agro-infiltration of each sample, while blue squares indicate the exogenously supplied substrates. A filled square indicates the presence, and an empty square indicates the absence, of the corresponding gene or substrate.

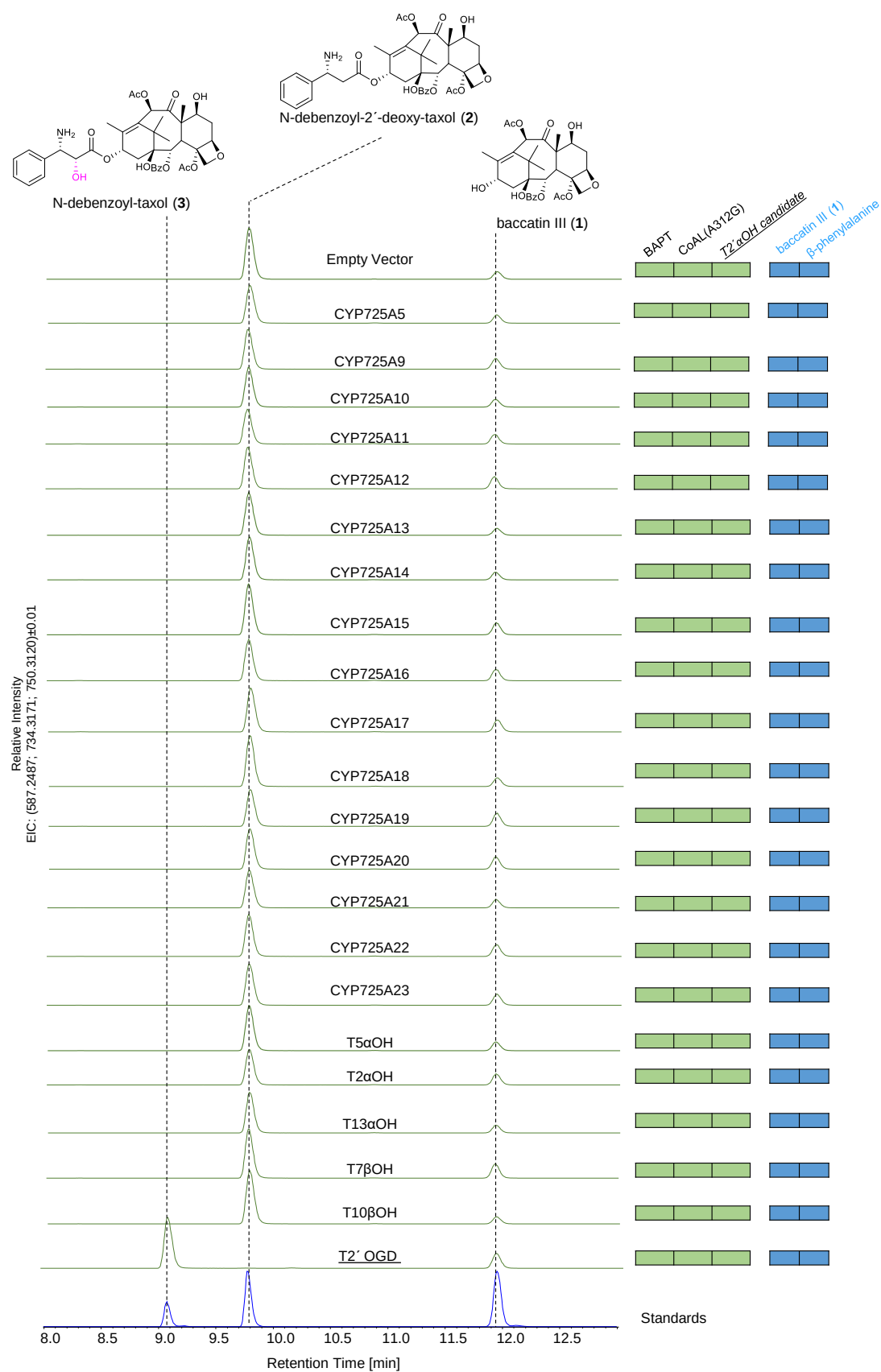


Fig. S22. Analysis of 21 candidate CYPs, listed in Tables S2 and S3. None of the candidates was capable of synthesizing compound **3** ($C_{40}H_{48}NO_{13}^+$, $m/z = 750.3120 \pm 0.01$) using

compound **2** ($\text{C}_{40}\text{H}_{47}\text{NO}_{12}^+$, $m/z = 734.3171 \pm 0.01$) as substrate. C2' α hydroxylation by T2'OGD is shown here as positive control. Compound **2** was synthesized by transient expression of BAPT and CoAL(A312G) and feeding baccatin III and β -phenylalanine. Green squares indicate the genes used in agro-infiltration of each sample, while blue squares indicate the exogenously supplied substrates. A filled square indicates the presence, and an empty square indicates the absence, of the corresponding gene or substrate.

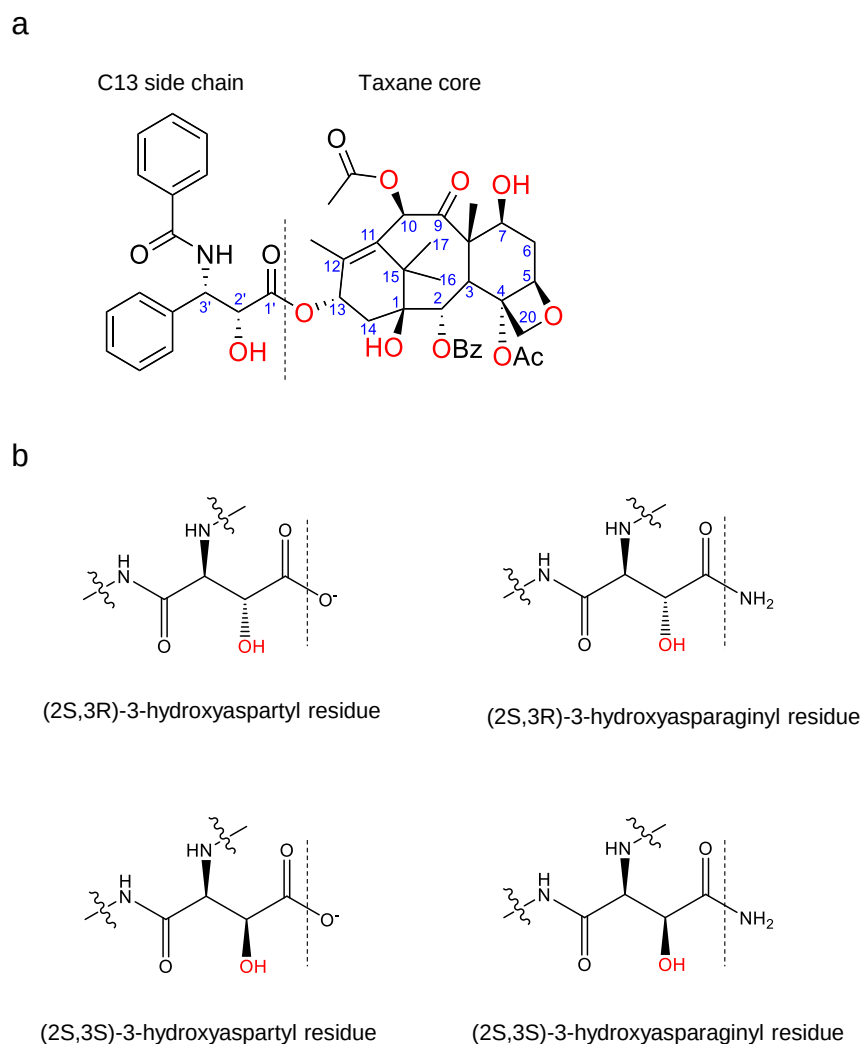


Fig. S23. Scheme depicting the similarity between C2'α hydroxylation of the C13 side chain of Taxol and hydroxylation of asparaginyl and aspartyl residues of peptides. **a.** The molecular structure of Taxol. The dashed vertical line is shown to facilitate substructure comparison. **b.** 2-oxoglutarate-dependent-oxidases (OGDs) catalyze the hydroxylation of asparaginyl and aspartyl residues during post-translational modification of proteins. The OGD-added hydroxy groups are shown in red.

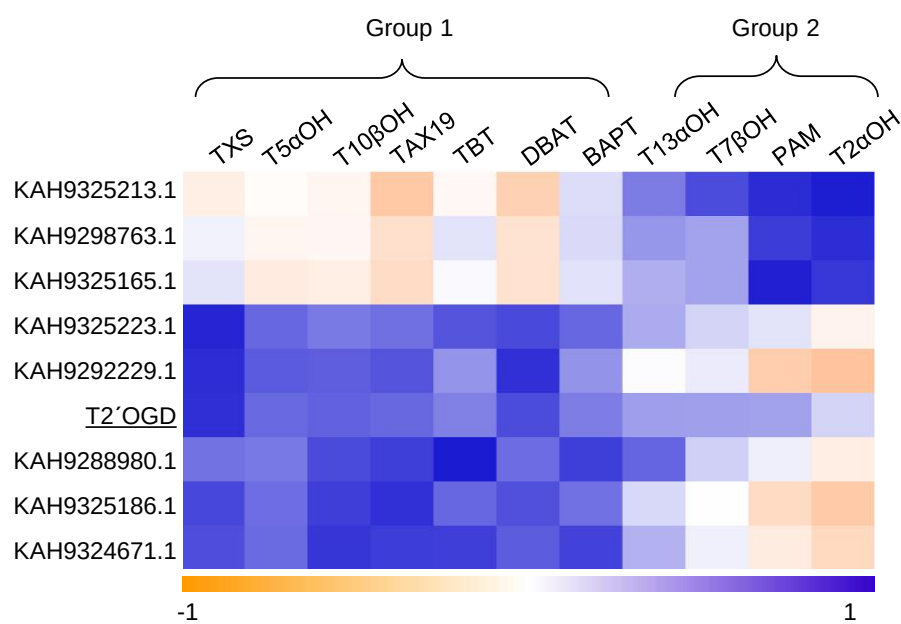


Fig. S24. Correlation heatmap of co-expression patterns between the selected *Taxus chinensis* OGD candidates and baits in this study (**Table S2**). The color bar indicates the correlation factor value. Blue indicates a positive correlation, while orange indicates a negative correlation.

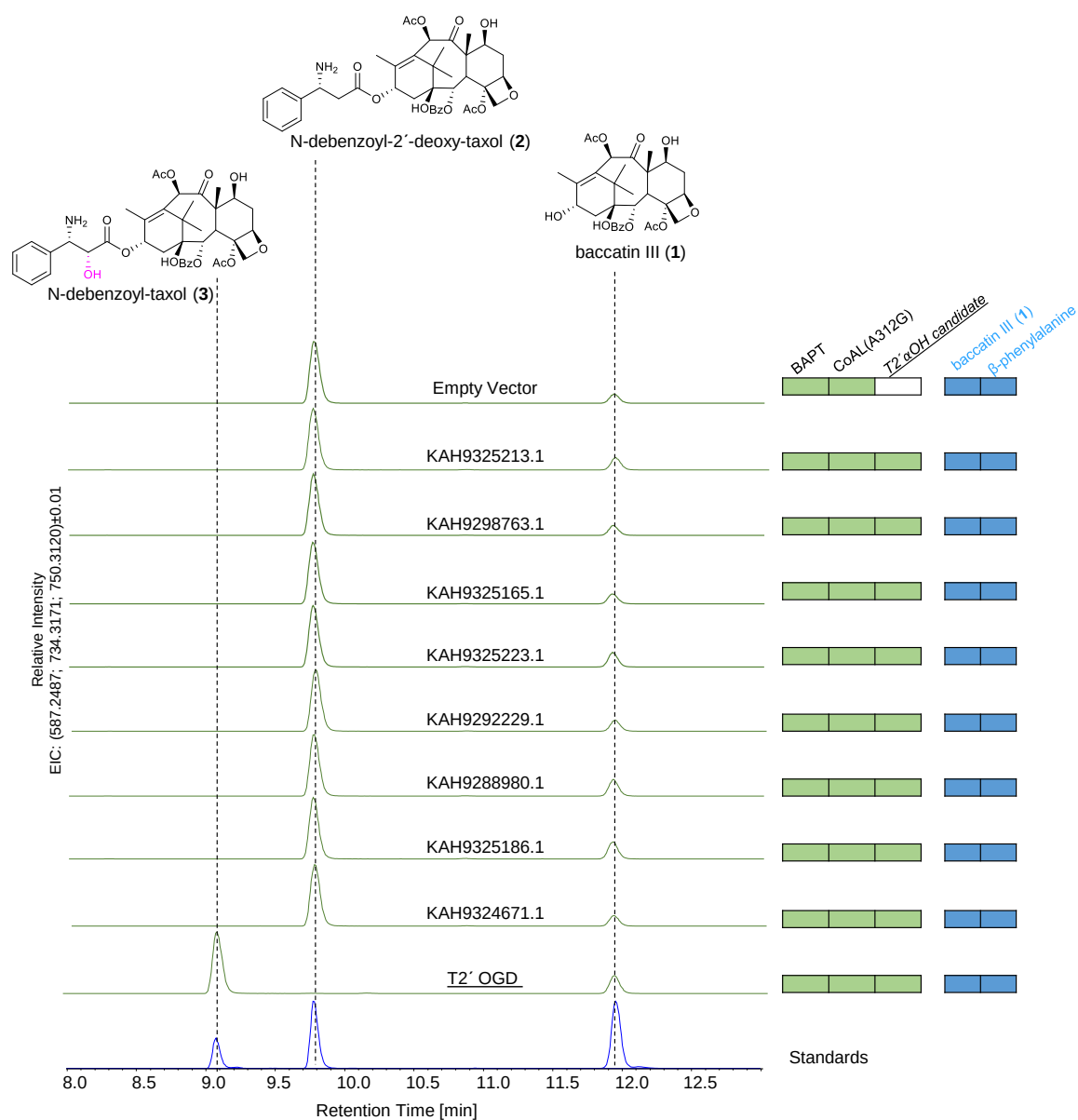


Fig. S25. Analysis of nine candidate OGDs obtained from our co-expression analysis (Fig. S24) revealed one candidate, T2'OGD, with potent C2' α hydroxylation activity. Compound 2 was synthesized in situ by transient expressing BAPT and CoAL(A312G) and feeding baccatin III and β -phenylalanine. Green squares indicate the genes used in agro-infiltration of each sample, while blue squares indicate the exogenously fed substrate. A filled square indicates the presence, and an empty square indicates the absence, of the corresponding gene or substrate.

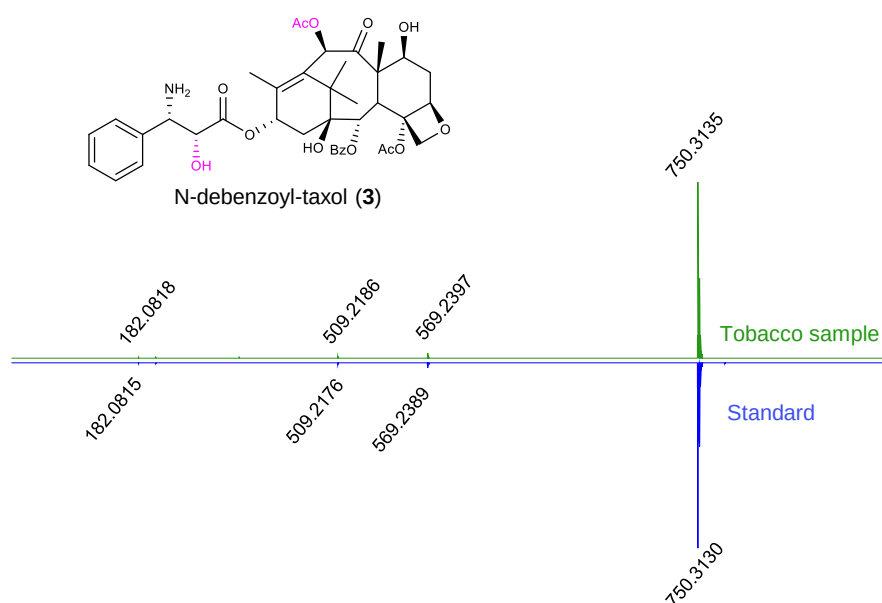


Fig. S26. Mass spectra of N-debenzoyl-taxol (**3**) in positive ionization mode ($[M+H]^+$). The green spectrum is from the compound produced in tobacco, while the blue spectrum is from the chemically synthesized standard.

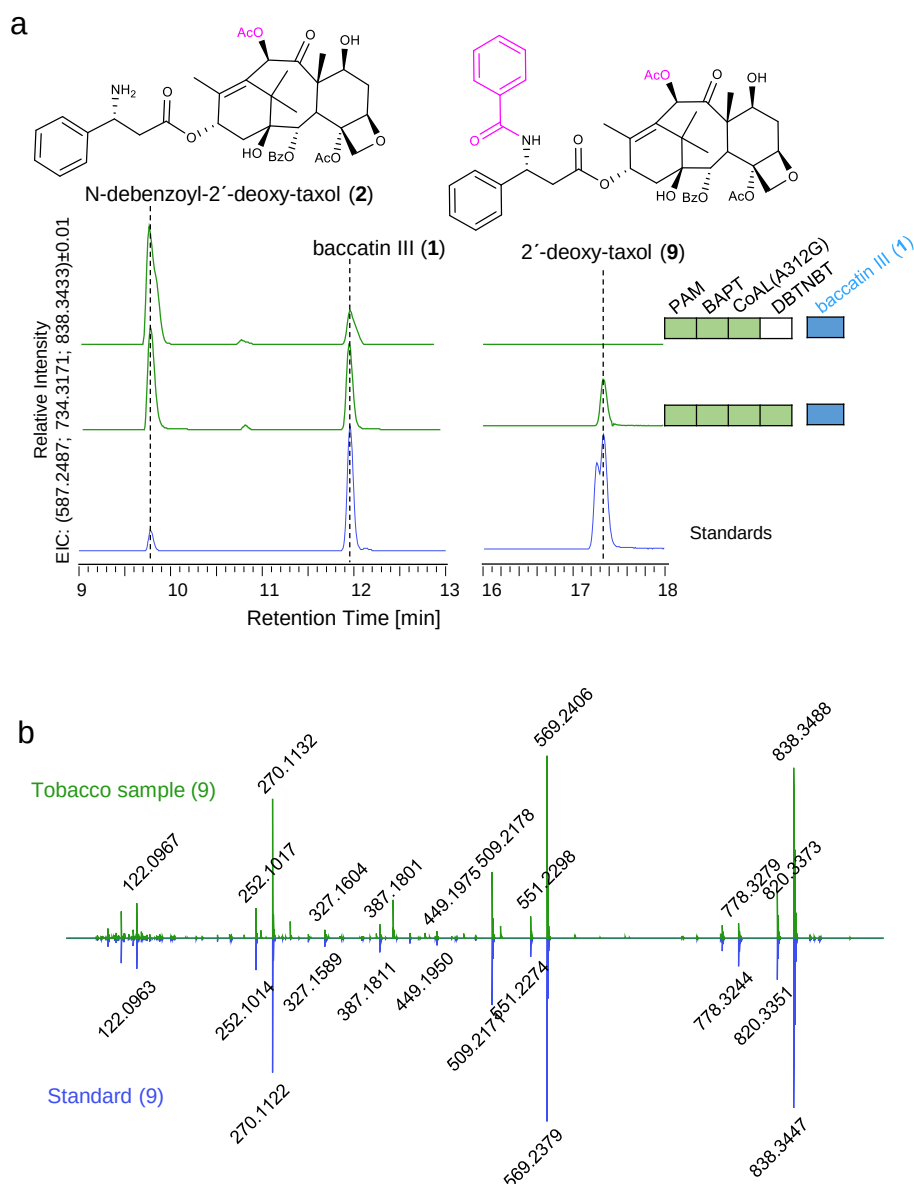


Fig. S27. *Taxus canadensis* DBTNBT benzoylates N-debenzoyl-2'-deoxy-taxol (**2**) to form 2'-deoxy-taxol (**9**). **a.** Transient expression of DBTNBT in tobacco leaves producing compound **2** results in the production of a new peak with $[M+H]^+$, $m/z = 838.3433 \pm 0.01$. The peak has the same retention time as the authentic standard of 2'-deoxy-taxol (**9**). Green squares indicate the enzymes used in agro-infiltration of each sample, while blue squares indicate the exogenously fed substrate. A filled square indicates the presence and an empty square indicates the absence of the corresponding gene or substrate. **b.** Mass spectra of 2'-deoxy-taxol (**9**) in positive ionization mode ($[M+H]^+$). The green spectrum is from the 2'-deoxy-taxol (**9**) produced in tobacco. The blue spectrum is from the chemically synthesized standard.

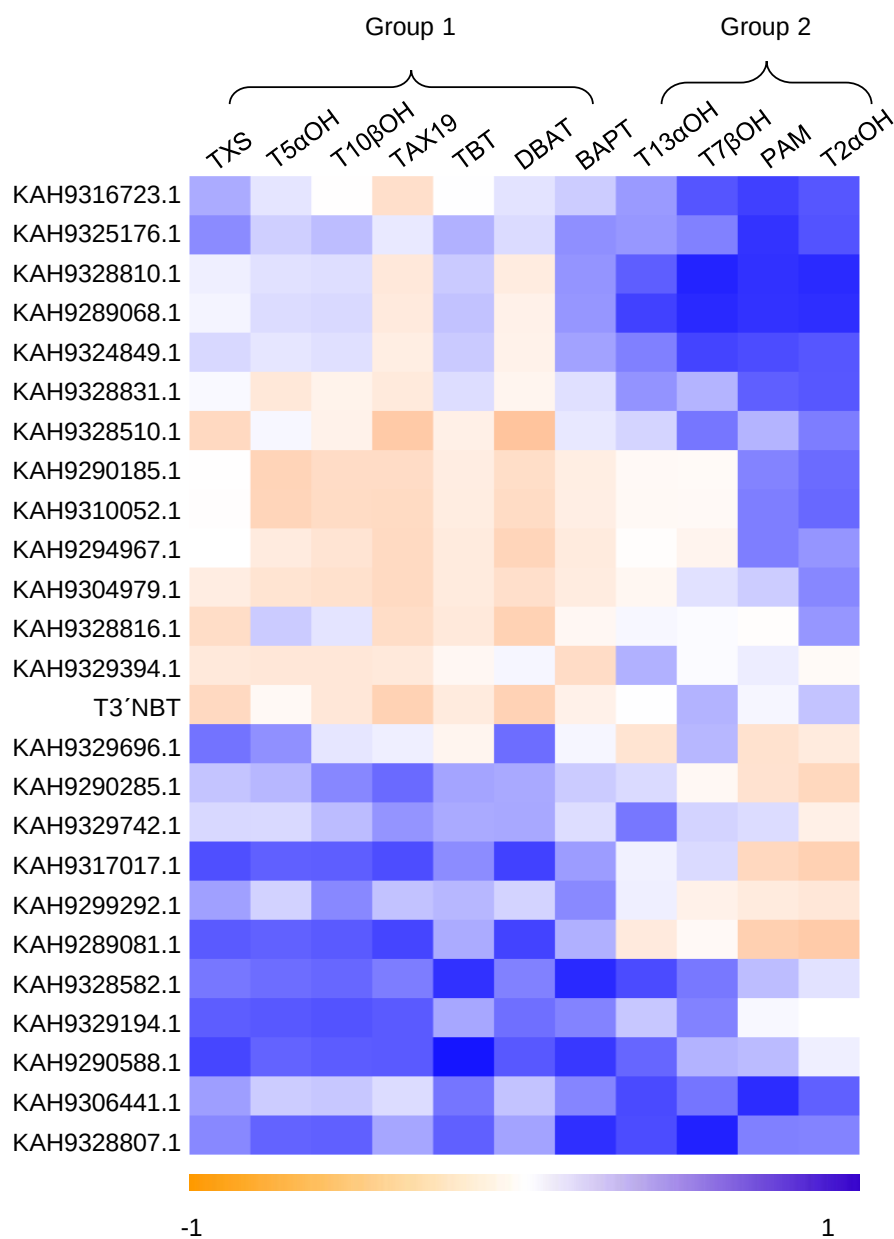


Fig. S28. Correlation heatmap of co-expression patterns between selected *T. chinensis* acyltransferase candidates and baits in this study (**Table S2**). The color bar indicates the correlation factor value. Blue indicates a positive correlation, while orange indicates a negative correlation.

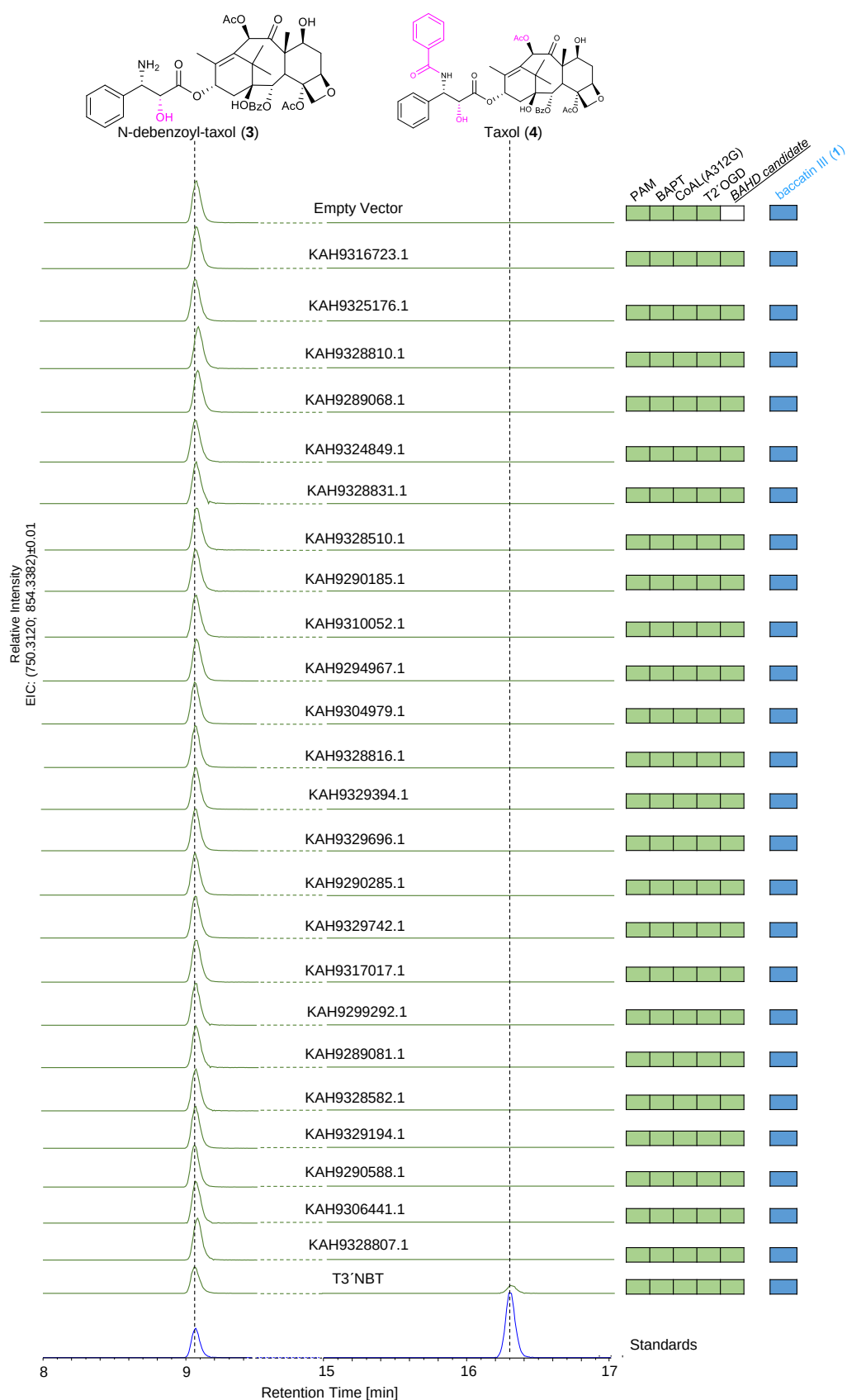


Fig. S29. Analysis of 25 candidate BAHD family acyltransferases obtained from our co-expression analysis (Fig. S28) revealed one candidate, T3'NBT with potent 3'-N benzoylation

activity. Compound **3** was synthesized by transient expression of PAM, BAPT, CoAL(A312G) and T2'OGD and feeding baccatin III. Green squares indicate the genes used in agro-infiltration of each sample, while blue squares indicate the exogenously fed substrate. A filled square indicates the presence, and an empty square indicates the absence, of the corresponding gene or substrate.

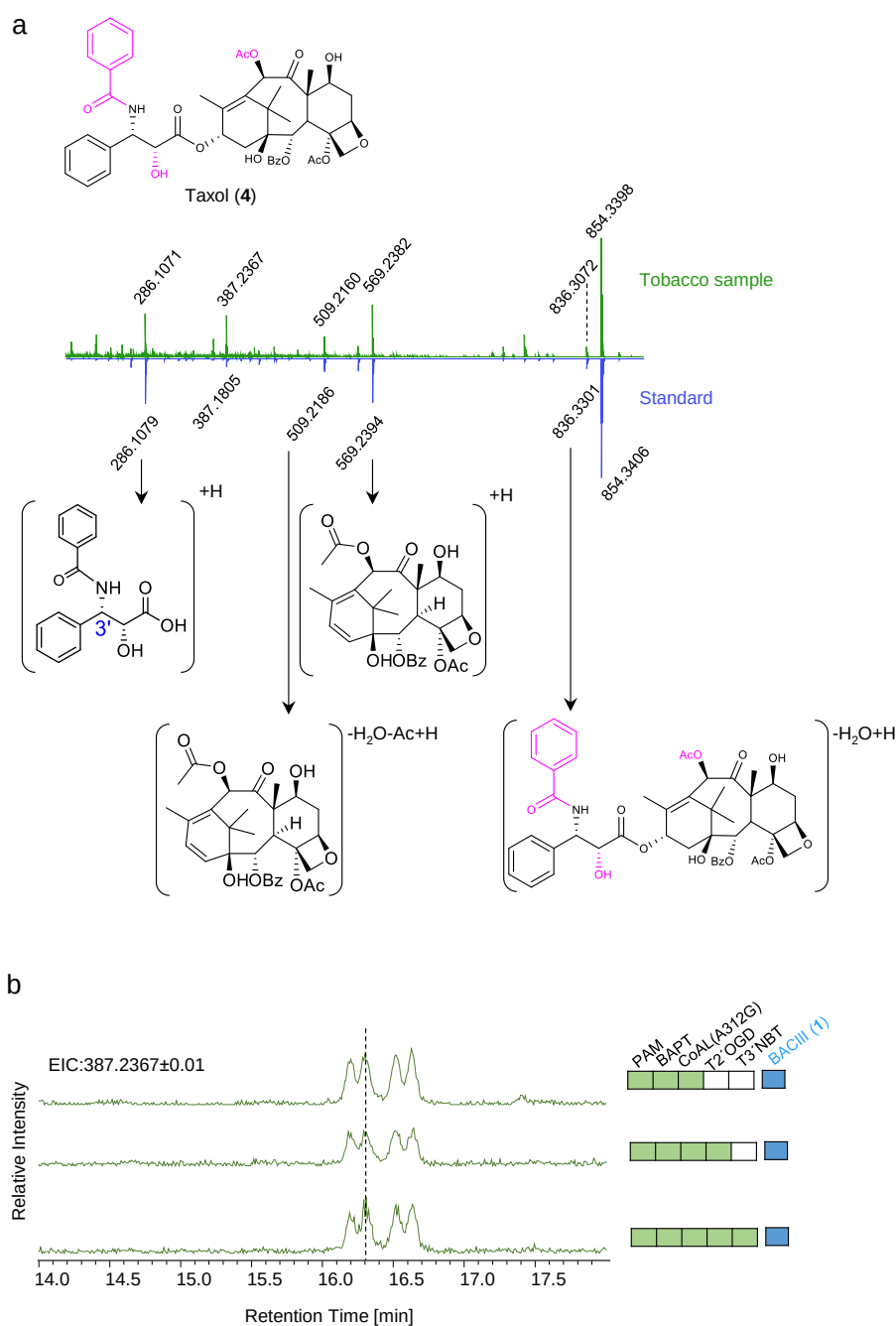


Fig. S30. Mass spectra of Taxol (**4**) in positive ionization mode ($[M+H]^+$). **a.** The green spectrum is from the compound produced in tobacco while the blue spectrum is from the commercially available authentic standard. Proposed structures of each fragment are indicated. **b.** Extracted ion chromatogram (EIC 387.2367) of extracts of tobacco leaves producing **2**, **3** and Taxol. Comparison between the three chromatograms reveals that the 387.2367 fragment (indicated by a dashed line) is present in all samples, suggesting that it originates from a co-eluting compound present also in the control samples.

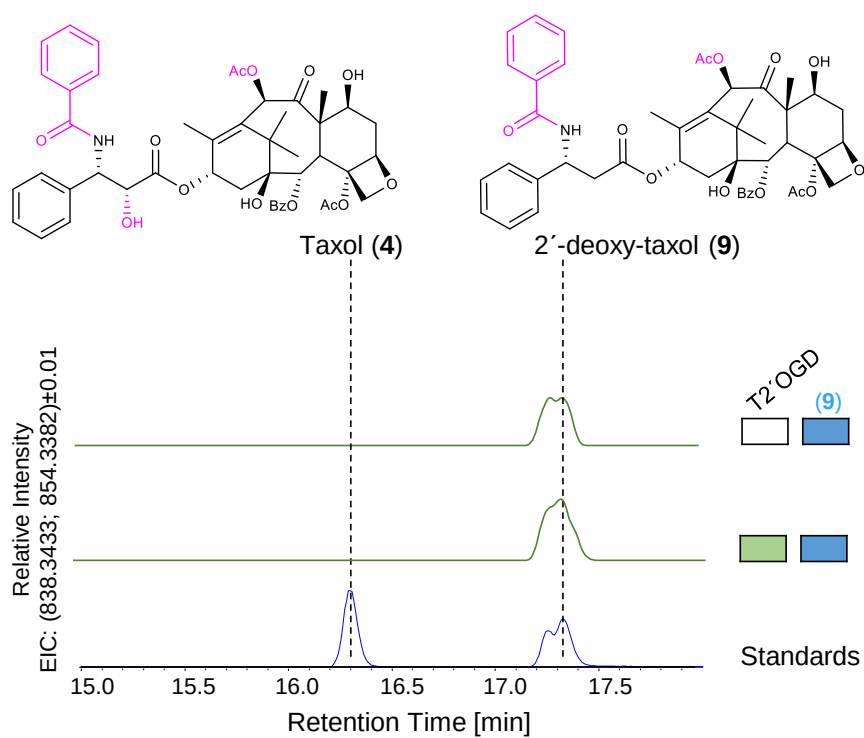


Fig. S31. 2'-deoxy-taxol (**9**) is not a substrate of T2'OGD. Transient expression of T2'OGD in tobacco leaves supplied exogenously with **9** does not lead to the production of Taxol (**4**).

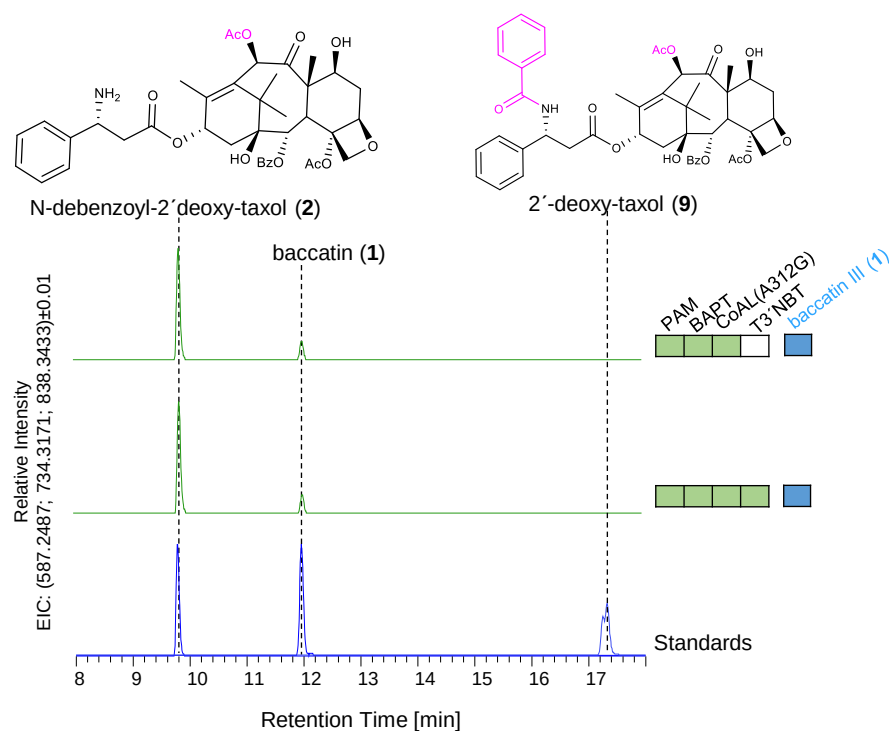


Fig. S32. N-debenzoyl-2'-deoxy-taxol (**2**) is not a substrate of T3'NBT. Transient expression of PAM, BAPT, and CoAL(A312G) in tobacco leaves via agro-infiltration leads to the production of **2**. However, when T3'NBT is additionally expressed, no synthesis of 2'-deoxy-taxol (**9**) is observed. Green squares indicate the genes used in agro-infiltration, while blue squares indicate the supplied substrate.

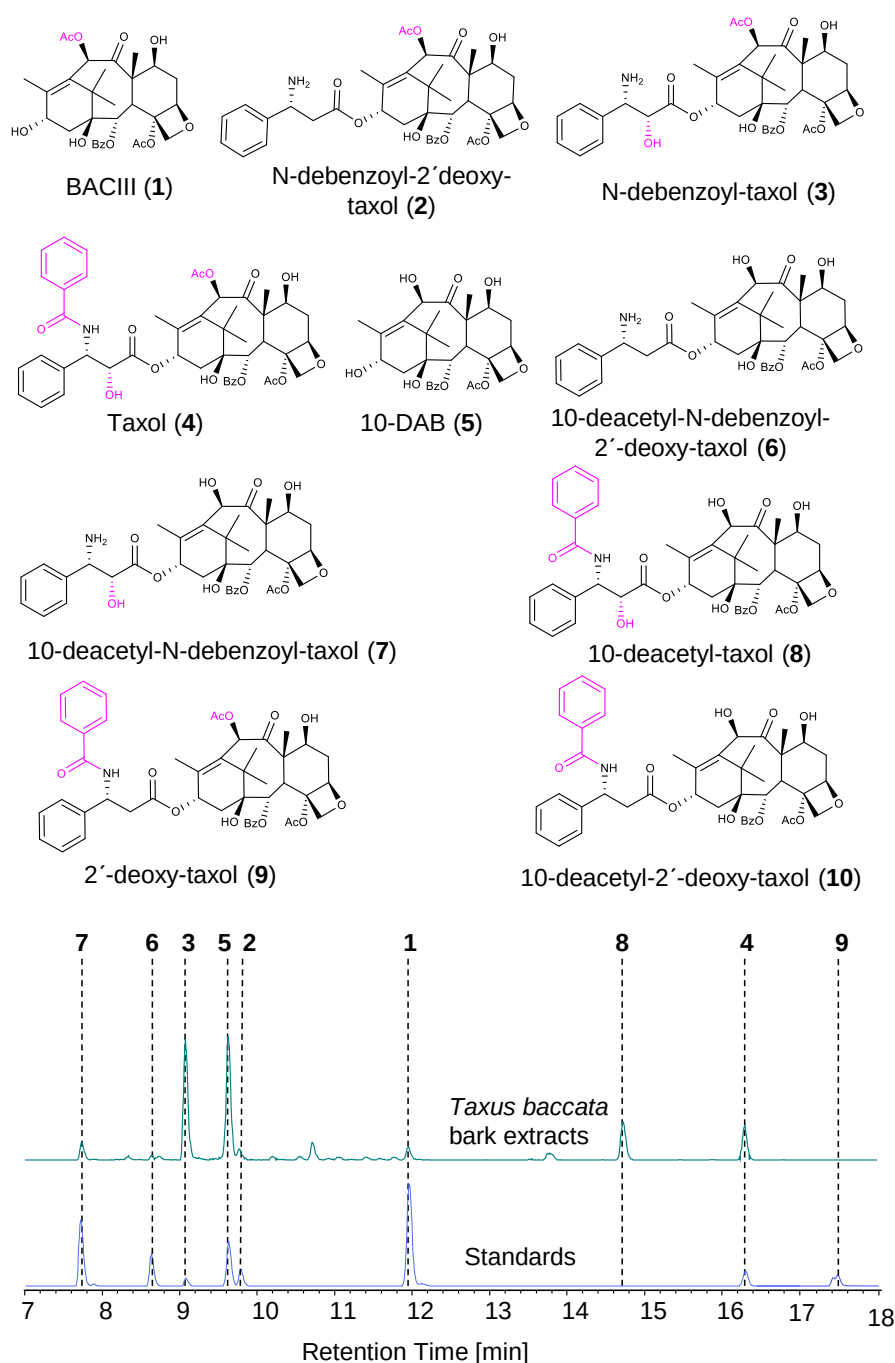


Fig. S33. Taxoids in *T. baccata* bark extracts. Compounds **1-8** can be found in the extracts, as shown by the black chromatogram where the corresponding ions ($[M+H]^+$; $m/z = (545.2381; 692.3065; 708.3015; 796.3328; 812.3277; 587.2487; 734.3171; 750.3120; 838.3433; 854.3382) \pm 0.01$) have been extracted. The identity of these taxoids was further confirmed by comparison with authentic standards (blue chromatogram). No peaks corresponding to compound **9** ($[M+H]^+$ $m/z = 838.3433$) or **10** ($[M+H]^+$ $m/z = 796.3328$) could be extracted.

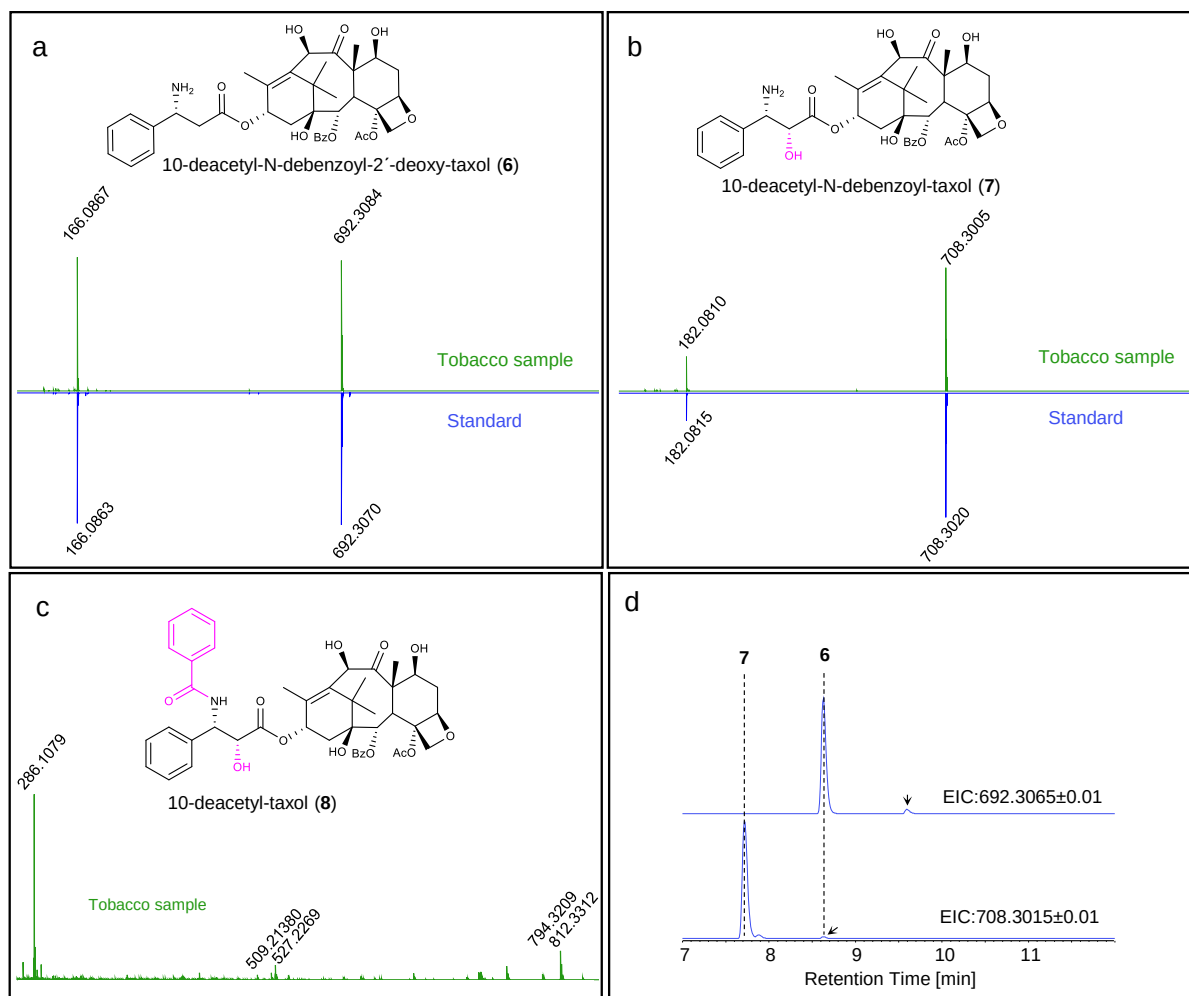


Fig. S34. Mass spectra of: (a) 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (**6**), (b) 10-deacetyl-N-debenzoyl-taxol (**7**) and (c) 10-deacetyl-taxol (**8**) in positive ionization mode ($[M+H]^+$), and (d) chromatograms of chemically synthesized **6** and **7**. The green spectra are from compounds produced in tobacco while the blue are from the chemically synthesized authentic standards. The exact mass of **8** ($[M+H]^+$, $m/z = 812.3312 \pm 0.01$) indicates that **8** is a benzoylated product of **7**. Compound **8** breaks into fragments $[M+H]^+$, $m/z = 527.2269 \pm 0.01$ and $[M+H]^+$, $m/z = 286.1079 \pm 0.01$, suggesting that benzoylation happens on the C13 side chain instead of the taxane core. The fragmentation pattern of **8** is similar to Taxol (**4**). Therefore, **8** is predicted to be 10-deacetyl-taxol. Chemically synthesized **6** and **7**, each contain minor amounts of synthesis by-products that share the same exact mass with **6** and **7**, respectively. These by-products are indicated by arrows.

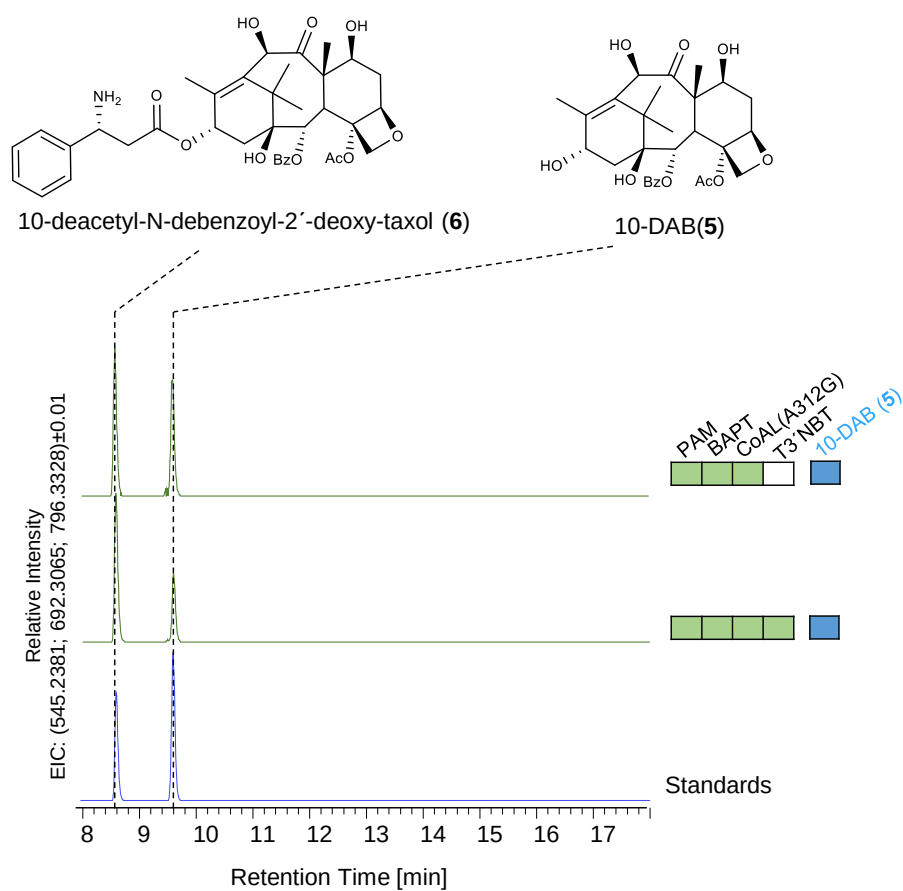


Figure S35. T3'NBT does not use 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (**6**) as a substrate when produced heterologously in tobacco leaves. BAPT, CoAL(A312G), and PAM catalyze the formation of 10-deacetyl-N-debenzoyl-2'-deoxy-taxol (**6**). However, no conversion of **6** and no new peaks are observed when T3'NBT is expressed in tobacco via agro-infiltration. Green squares indicate the genes used in agro-infiltration, while blue squares indicate the exogenously supplied substrate.

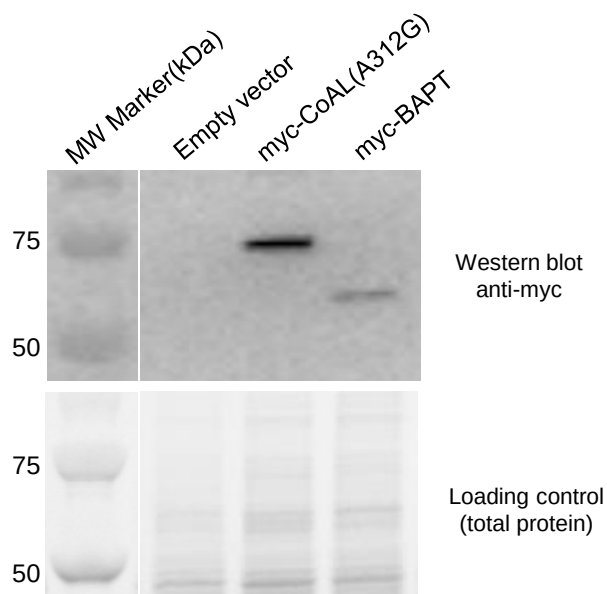


Fig. S36. Western blot of CoAL(A312G) and BAPT extracted from engineered yeast cells cultivated at 30°C. A myc-tag was fused at the N terminus of CoAL(A312G) and BAPT through an SGSGSG linker. The empty vector control corresponds to the extract of yeast cells carrying an empty vector.

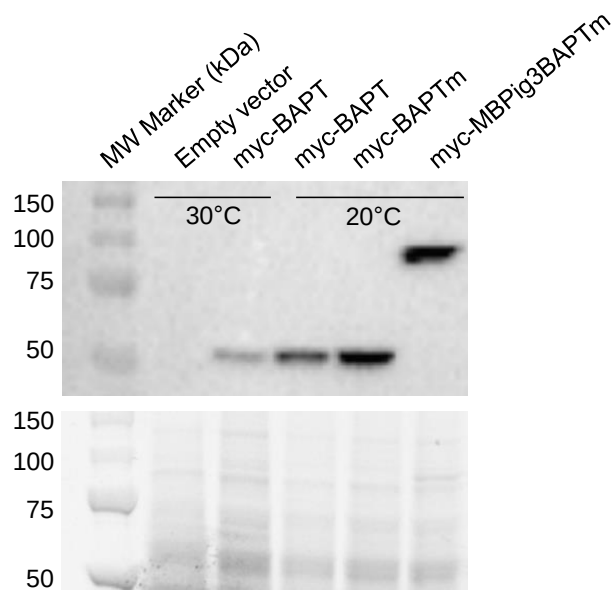


Fig. S37. Western blot of BAPT variants extracted from engineered yeast cells cultivated at different temperatures. A myc-tag (EQKLISEEDL) was fused at the N terminus of each BAPT variant. The empty vector control corresponds to the extract of yeast cells carrying an empty vector.

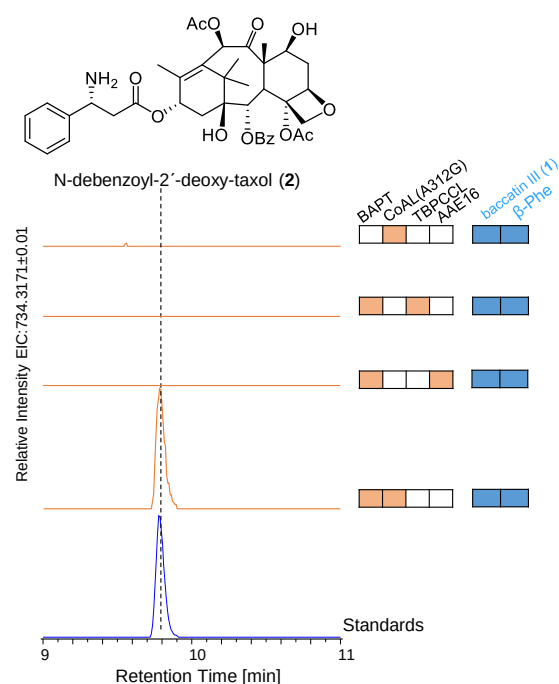


Fig. S38. N-debenzoyl-2'-deoxy-taxol (2) production in *S. cerevisiae* using BAPT and CoAL(A312G) and feeding baccatin III (100 mg/L) and β -phenylalanine (100 mg/L). The two CoA ligases, TBPCCL and AAE16, from *Taxus* cannot produce detectable amounts of compound 2. Orange squares indicate the genes expressed in yeast strains via episomal vectors, while blue squares indicate the substrates supplied.

<i>T.cu</i> BAPT	MKKTGSFAEF	HVNMIERVMV	RPCLPSPKTI	LPLSAIDNMA	RAFSNVLLVY	AANMDRVSD	60
ACN62085.1	MKKTGSFAEF	HVNMIERVMV	RPCMLSPKTI	LPLSAIDNMA	RAFSNVLLVY	TADMDRVSD	60
AAT73200.1	MKKTGSLAEF	HVNMIERVMV	RPCLPSPKTI	LPLSAIDNMA	RAFSNVLLVY	AANMDRVSD	60
AEJ84001.1	MKKTGSLAEF	HVNMIERVMV	RPCLPSPKTI	LPLSAIDNMA	RAFSNVLLVY	AANMDRVSD	60
AFD32416.1	MKKTGSLAEF	HVNMIERVMV	RPCPPSPKTI	LPLSAIDNMA	RAFSNVLLVY	AANMDRVSD	60
AGT51232.1	MKKTGSFAEF	HVNMIERVMV	RPCLPSPKTI	LPLSAIDNMA	RAFSNVLLVY	TANMDRVSD	60
QDM39187.1	MKKTGSLAEF	HVNMIERVMV	RPCLPSPKTI	LPLSAIDNMA	RAFSNVLLVY	AANMDRVSD	60
BAPTm	MKKTGSFAEF	HVNMIERVMV	RPCLPSPKTI	LPLSAIDNMA	RAFSNVLLVY	AADMDRVSD	60
<i>T.cu</i> BAPT	PAKVIREALS	KVLVYYYPPFA	GRLRNKNGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
ACN62085.1	PAKVIREALS	KVLVYYYPPFA	GRLRNKEDGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
AAT73200.1	PAKVIREALS	KVLVYYYPPFA	GRLRNKNGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
AEJ84001.1	PAKVIREALS	KVLVYYYPPFA	GRLRNKNGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
AFD32416.1	PAKVIREALS	KVLVYYYPPFA	GRLRNKNGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
AGT51232.1	PAKVIREALS	KVLVYYYPPFA	GRLRNKNGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
QDM39187.1	PAKVIREALS	KVLVYYYPPFA	GRLRNKNGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
BAPTm	PAKVIREALS	KVLVYYYPPFA	GRLRNKEDGE	LEVECTGQGV	LFLEAMADSD	LSVLTDLDDY	120
<i>T.cu</i> BAPT	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGAN	VYGSACDAKG	FGQFLQSGMAE	180
ACN62085.1	KPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGAN	VYGSVCDACKG	FGQFLQSGMAE	180
AAT73200.1	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGAN	VYGSTCDAKG	FGQFLQSGMAE	180
AEJ84001.1	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGTN	VYGSVCDACKG	FGQFLQSGMAE	180
AFD32416.1	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGAN	VYGSTCDAKG	FGQFLQSGMAE	180
AGT51232.1	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGTN	VYGSVCDACKG	FGQFLQSGMAE	180
QDM39187.1	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTHF	TCGGFVVGAN	VYGSACDAKG	FGQFLQSGMAE	180
BAPTm	NPSFQQLIFS	LPQDTDIEDL	HLLIVQVTRF	TCGGFVVGAN	VYGSVCDACKG	FGQFLQSGMAE	180
<i>T.cu</i> BAPT	MARGEVKPSI	EPIWNRELVK	LEHCPMFRMS	HLQIIHAPVI	EEKFVQTSLV	INFEIINHIR	240
ACN62085.1	MARGEVKPSI	EPIWNRELVN	PEDCMSFRMS	HLQIIHAPLI	EEKFVQTSLV	INFEIINHIR	240
AAT73200.1	MARGEVKPSI	EPIWNRELVK	LEDCEMFRMS	HLQIIHAPVI	EEKFVQTSLV	INFEIINHIR	240
AEJ84001.1	MARGEVKPSI	EPIWNRELVK	PEDCEMFRMS	HLQIIHAPLI	EEKFVQTSLV	INFEIINHIR	240
AFD32416.1	MARGEVKPSI	EPIWNRELVK	LEDCEMFRMS	HLQIIHAPVI	EEKFVQTSLV	INFEIINHIR	240
AGT51232.1	MARGEVKPSI	EPIWNRELVK	LEDCEMFRMS	HLQIIHAPLI	EEKFVQTSLV	INFEIINHIR	240
QDM39187.1	MARGEVKPSI	EPIWNRELVK	LEDCEMFRMS	HLQIIHAPVI	EEKFVQTSLV	INFEIINHIR	240
BAPTm	MARGEVKPSI	EPIWNRELVN	PEDCEMFRMS	HLQIIHAPVI	EEKFVQTSLV	INFEIINHIR	240
<i>T.cu</i> BAPT	RRIMEERKES	LSSFEIVAAL	VWLAKIKAFQ	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
ACN62085.1	QRIMEECKES	FSSFEIVAAL	VWLAKIKAFQ	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
AAT73200.1	RRIMEECKES	LSSFEIVAAL	VWLAKIKAFQ	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
AEJ84001.1	RRIMEERKES	LSSFEIVAAL	VWLAKIKAFR	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
AFD32416.1	RRIMEECKES	LSSFEIVAAL	VWLAKIKAFQ	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
AGT51232.1	RRIMEECKES	LSSFEIVAAL	VWLAKIKAFR	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
QDM39187.1	RRIMEECKES	LSSFEIVAAL	VWLAKIKAFQ	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
BAPTm	RRIMEERKES	FSSFEIVAAL	VWLAKIKAFQ	IPHSENVKLL	FAMDLRSSFN	PPLPHGYGYN	300
<i>T.cu</i> BAPT	AFGIACAMDN	VHDLLSGSLL	RTIMIIKKSK	FSLHKELNSK	TVMSSSVVDV	NTKFEDVVS	360
ACN62085.1	AFGIACAMDN	VHDLLSGSLL	RAIMIIKKAK	FSLHNELNSK	TVMSPSVVYV	NTKFEDVVS	360
AAT73200.1	AFGIACAMDN	VHDLLSGSLL	RTIMIIKKSK	FSLHKELNSK	TVMSSSVVDV	NTKFEDVVS	360
AEJ84001.1	AFGIACAMDN	VHDLLSGSLL	RAIMIIKKSK	FSLHKELNSK	TVMSPSVVDV	NMKFEDVVS	360
AFD32416.1	AFGIACAMDN	VHDLLSGSLL	RTIMIIKKSK	FSLHKELNSK	TVMSSSVVDV	NTKFEDVVS	360
AGT51232.1	AFGIACAMDN	VHDLLSGSLL	RAIMIIKKSK	FSLHKELNSK	TVMSSSVVDV	NTKFEDVVS	360
QDM39187.1	AFGIACAMDN	VHDLLSGSLL	RAIMIIKKSK	FSLHKELNSK	TVMSPSVVDV	NTKFEDVVS	360
BAPTm	AFGIACAMDN	VHDLLSGSLL	RTIMIIKKSK	FSLHKELNSK	TVMSSSVVDV	NTKFEDVVS	360
<i>T.cu</i> BAPT	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQSTKNMPDG	IKMLMFMPPS	420
ACN62085.1	SDWRHSIYYE	VDFGWGDAMN	VSTILQQQGH	EKSLPTYFSF	LQPTKNMPDG	IKMLMFMPPS	420
AAT73200.1	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQSTKNMPDG	IKMLMFMPPS	420
AEJ84001.1	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQPTKNMPDG	IKMLMFMPPS	420
AFD32416.1	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQSTKNMPDG	IKMLMFMPPS	420
AGT51232.1	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQSTKNMPDG	IKMLMFMPPS	420
QDM39187.1	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQSTKNMPDG	IKMLMFMPPS	420
BAPTm	SDWRHSIYYE	VDFGWGDAMN	VSTMILQQQEH	EKSLPTYFSF	LQSTKNMPDG	IKMLMFMPPS	420
<i>T.cu</i> BAPT	KLKTKFKIEIE	AMIKKYVTKV	CPSKL	445			
ACN62085.1	KLKTKFKIEIE	AMINKYVTEV	CPSKL	445			
AAT73200.1	KLKTKFKIEIE	AMINKYVTKV	CPSKL	445			
AEJ84001.1	KLKTKFKIEIE	AMINKYVTKV	CPSKL	445			
AFD32416.1	KLKTKFKIEIE	AMINKYVTKV	CPSKL	445			
AGT51232.1	KLKTKFKIEIE	AMINKYVTKV	CPSKL	445			
QDM39187.1	KLKTKFKIEIE	AMIKKYVTKV	CPSKL	445			
BAPTm	KLKTKFKIEIE	AMIKKYVTEV	CPSKL	445			

Fig. S39. Amino acid sequence alignment of BAPTs (Table. S5). Gray boxes indicate residues different between BAPTs. Blue boxes indicate the nine residues substituted in *Taxus cuspidata* BAPT (AAL92459.1) to construct BAPTm.

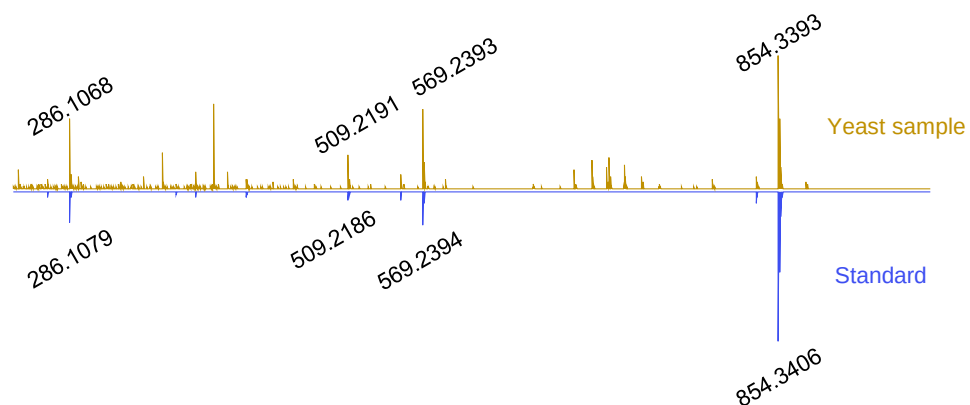


Fig. S40. Mass spectra of Taxol (**4**) in positive ionization mode ($[M+H]^+$). The yellow spectrum is from Taxol produced in yeast, while the blue spectrum is from the commercially available authentic Taxol standard.

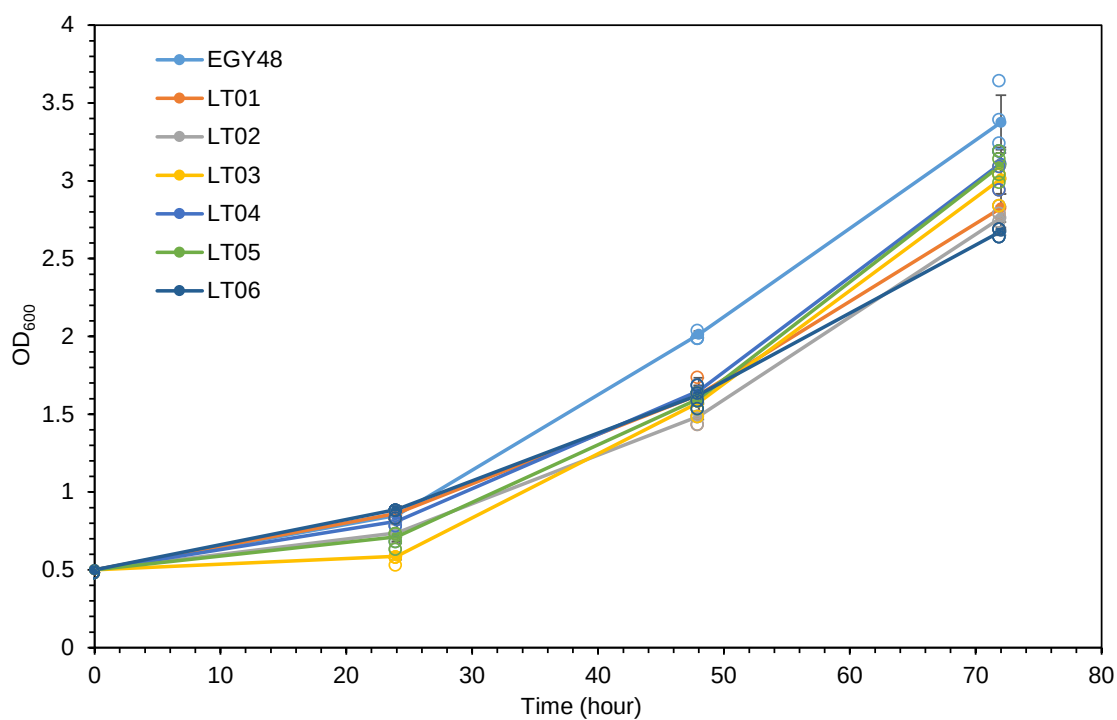


Fig. S41. Growth of engineered yeast strains (see main text and Table S4 for strain description). Control strain EGY48 carries empty vectors instead of vectors expressing pathway genes. All strains are tryptophan, leucine and uracil autotrophic. Error bars correspond to the standard derivation of four biological replicates (n=4).

Supplementary Tables

Table S1. Enzymes used for pathway reconstruction.

ENZYME	ORGANISM OF ORIGIN	NCBI ACCESSION NUMBER (PROTEIN)
PAM	<i>Taxus canadensis</i>	AAT47186.1
TBPCCL	<i>Taxus baccata</i>	AKA59704.1
BAPT	<i>Taxus cuspidata</i>	AAL92459.1
TB506	<i>Taxus cuspidata</i>	QYV42687.1
DBTNBT	<i>Taxus canadensis</i>	Q8LL69.1
AAE16	<i>Taxus chinensis</i>	KAH9307324.1
CoAL(A312G)	<i>Penicillium chrysogenum</i>	Pc21g20650
PAL	<i>Arabidopsis thaliana</i>	NP_190894.1
CNL	<i>Petunia x hybrida</i>	AEO52693.1
CHD	<i>Petunia x hybrida</i>	AFS41246.1
KAT1	<i>Petunia x hybrida</i>	ACV70032.1

Table S2. *Taxus* genes used as baits in co-expression analysis.

ENZYME	NCBI ACCESSION NUMBER
TXS	KAH9297158
T5 α OH	KAH9297121
T7 β OH	KAH9296915
T2 α OH	KAH9296917
T13 α OH	KAH9296885
T10 β OH	KAH9296919
TAX19	KAH9291366
DBAT	KAH9297248
BAPT	KAH9331117
PAM	KAH9313992
TBT	KAH9328583

Table S3. CYPs of the CYP725 family obtained from the literature that were tested in this work for C2' α hydroxylation activity.

ENZYME	NCBI ACCESSION NUMBER
CYP725A5	AAR21106.1
CYP725A9	ATG29988.1
CYP725A10	ATG29941.1
CYP725A11	ATG29896.1
CYP725A12	ATG29938.1
CYP725A13	ATG29940.1
CYP725A14	ATG29937.1
CYP725A15	ATG29939.1
CYP725A16	ATG29989.1
CYP725A17	ATG29897.1
CYP725A18	ATG29898.1
CYP725A19	ATG29942.1
CYP725A20	ATG29959.1
CYP725A21	ATG29965.1
CYP725A22	ATG29967.1
CYP725A23	ATG29968.1

Table S4. *S. cerevisiae* strains constructed in this study.

<i>S. cerevisiae</i> strain	Genotype
EGY48	<i>MATalpha ura3 his3 trp1 6xLexAop::LEU2</i>
LT01	EGY48; pHUS- <i>P_{GAL1}</i> -BAPT- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CoAL(A312G)- <i>T_{CYC1}</i> ;
LT02	EGY48; pHUS- <i>P_{GAL1}</i> -BAPTm- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CoAL(A312G)- <i>T_{CYC1}</i> ;
LT03	EGY48; pHUS- <i>P_{GAL1}</i> -MBPig3BAPTm- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CoAL(A312G)- <i>T_{CYC1}</i> ;
LT04	EGY48; pHUS- <i>P_{GAL1}</i> -MBPig3BAPTm- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CoAL(A312G)- <i>T_{CYC1}</i> ; pWUS- <i>P_{GAL1}</i> -T2'OGD- <i>T_{AHD1}</i>
LT05	EGY48; pHUS- <i>P_{GAL1}</i> -MBPig3BAPTm- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CoAL(A312G)- <i>T_{CYC1}</i> ; pWUS- <i>P_{GAL1}</i> -T2'OGD- <i>T_{AHD1}</i> ; pLUS- <i>P_{GAL1}</i> -T3'NBT- <i>T_{AHD1}</i>
LT06	EGY48; pHUS- <i>P_{GAL1}</i> -MBPig3BAPTm- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CoAL(A312G)- <i>T_{CYC1}</i> ; pWUS- <i>P_{GAL1}</i> -T2'OGD- <i>T_{AHD1}</i> ; pLUS- <i>P_{GAL1}</i> -T3'NBT- <i>T_{AHD1}</i> ; XI-2:: <i>P_{GAL1}</i> -PAL- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -CHD- <i>T_{CYC1}</i> & <i>P_{GAL1}</i> -CNL- <i>T_{AHD1}</i> - <i>P_{GAL10}</i> -KAT1- <i>T_{CYC1}</i>

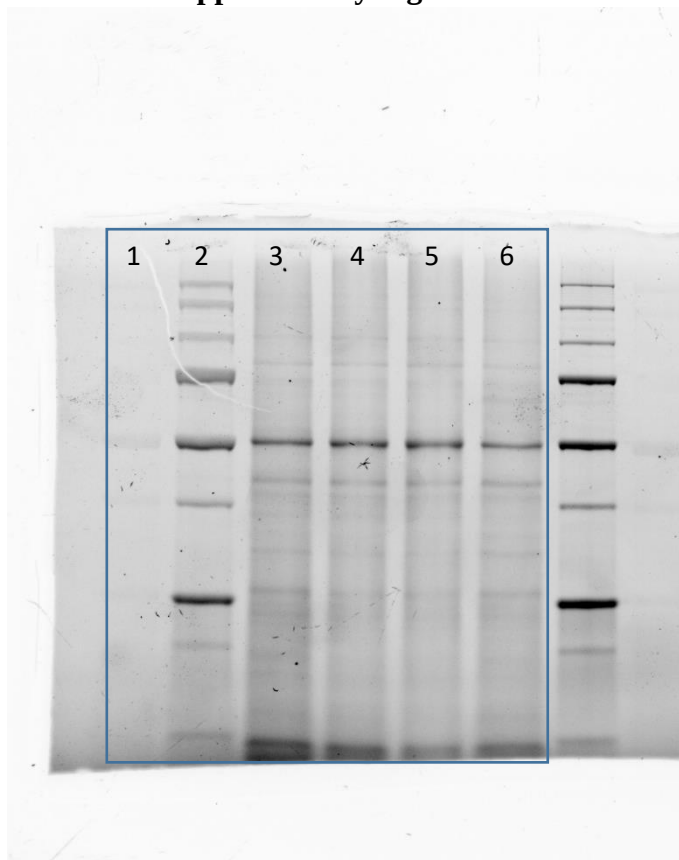
Table.S5. NCBI access numbers and predicted scaled solubility of BAPT sequences used in BAPT optimization.

Accession number	Source	Predicted scaled solubility using Protein-Sol
AAL92459.1 ^a	<i>Taxus cuspidata</i>	0.236
ACN62085.1	<i>Taxus sumatrana</i>	0.283
AAT73200.1	<i>Taxus media</i>	0.244
AEJ84001.1	<i>Taxus mairei</i>	0.244
AFD32416.1	<i>Taxus media</i>	0.254
AGT51232.1	<i>Taxus cuspidata</i>	0.257
QDM39187.1	<i>Taxus media</i>	0.226
BAPTm	Engineered variant	0.357

Supplementary References

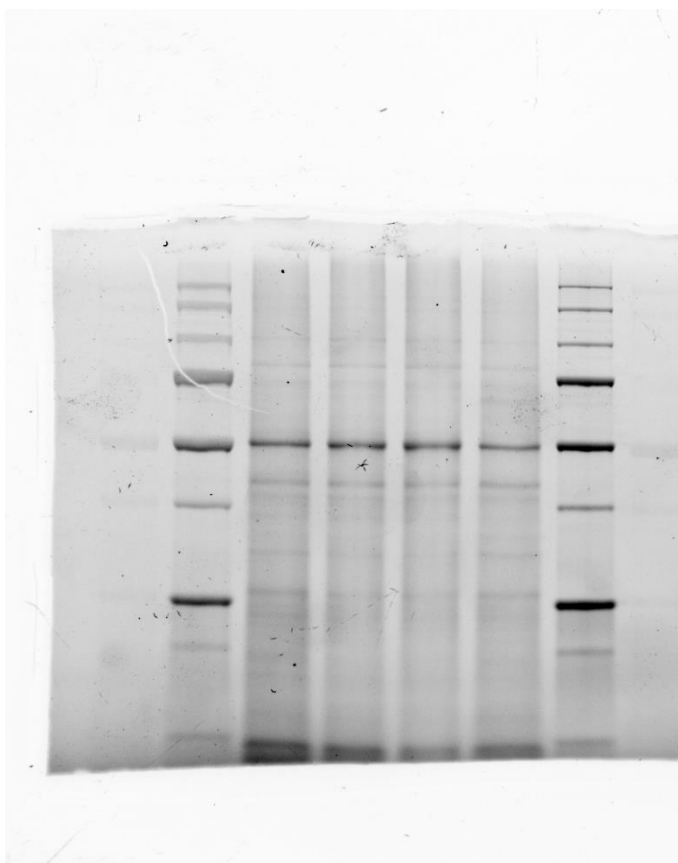
1. Thornburg, C. K., Walter, T. & Walker, K. D. Biocatalysis of a Paclitaxel Analogue: Conversion of Baccatin III to *N*-Debenzoyl- *N*-(2-furoyl)paclitaxel and Characterization of an Amino Phenylpropanoyl CoA Transferase. *Biochemistry* **56**, 5920–5930 (2017).
2. Zhang, J. *et al.* Design, synthesis and biological evaluation of peptidyl epoxyketone proteasome inhibitors composed of β -amino acids. *Chem. Biol. Drug Des.* **84**, 497–504 (2014).
3. Afon'Kin, A. A. *et al.* Regio- and stereoselective methods for the conversion of (2S,3R)- β -phenylglycidic acid esters to taxoids and other enantiopure (2R,3S)-phenylisoserine esters. *Russ. Chem. Bull.* **61**, 2149–2162 (2012).
4. Guéritte-Voegelein, F., Sénilh, V., David, B., Guénard, D. & Potier, P. Chemical studies of 10-deacetyl baccatin III: Hemisynthesis of taxol derivatives. *Tetrahedron* **42**, 4451–4460 (1986).
5. Chen, M., Liu, J., Tian, Z., Liu, X. & Zhang, S. Synthesis, cytotoxic activity and binding model analysis of novel isoxazole-docetaxel analogues with C3'-N modification. *Med. Chem. Res.* **27**, 1355–1365 (2018).
6. Chen, M., Chen, H., Ma, J., Liu, X. & Zhang, S. Synthesis and anticancer activity of novel quinoline-docetaxel analogues. *Bioorg. Med. Chem. Lett.* **24**, 2867–2870 (2014).
7. Sekhar, N. M., Vishweshwar, P., Acharyulu, P. V. R. & Anjaneyulu, Y. Alternative synthesis and the determination of absolute configuration of docetaxel, an anticancer drug. *Synth. Commun.* **42**, 3482–3492 (2012).
8. Lu, H. F., Sun, X., Xu, L., Lou, L. G. & Lin, G. Q. Design, synthesis and biological evaluation of novel fluorinated docetaxel analogues. *Eur. J. Med. Chem.* **44**, 482–491 (2009).
9. Qi, C. M., Wang, Y. F. & Yang, L. C. A novel method to synthesize docetaxel and its isomer with high yields. *J. Heterocycl. Chem.* **42**, 679–684 (2005).
10. Hayashi, Y. *et al.* A novel approach of water-soluble paclitaxel prodrug with no auxiliary and no byproduct: Design and synthesis of isotaxel. *J. Med. Chem.* **46**, 3782–3784 (2003).
11. Moyna, G. *et al.* Conformational Studies of Paclitaxel Analogs Modified at the C-2' Position in Hydrophobic and Hydrophilic Solvent Systems. *J. Med. Chem.* **40**, 3305–3311 (1997).
12. Méchin, V., Damerval, C. & Zivy, M. Total protein extraction with TCA-acetone. *Methods Mol. Biol.* **355**, 1–8 (2007).
13. Kaur, J. & Bachhawat, A. K. A modified Western blot protocol for enhanced sensitivity in the detection of a membrane protein. *Anal. Biochem.* **384**, 348–349 (2009).

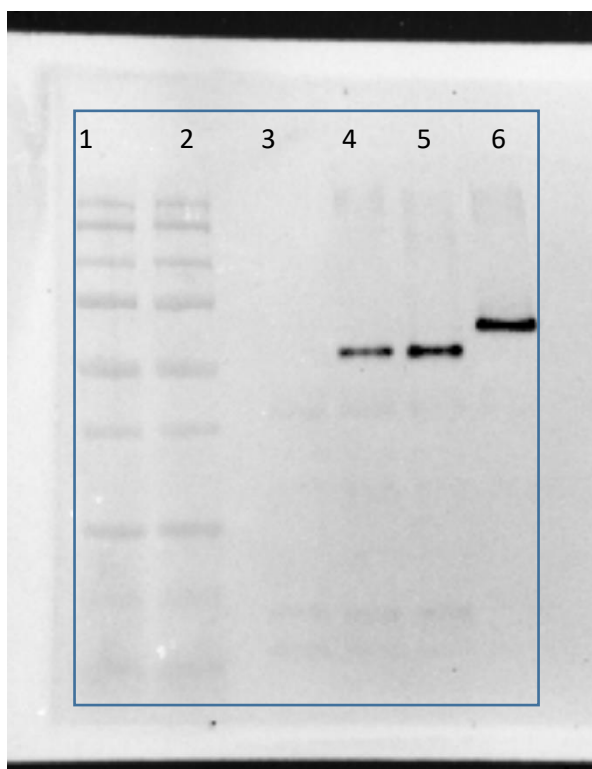
Additional Supplementary Figures



SDS-PAGE for total protein used as loading control

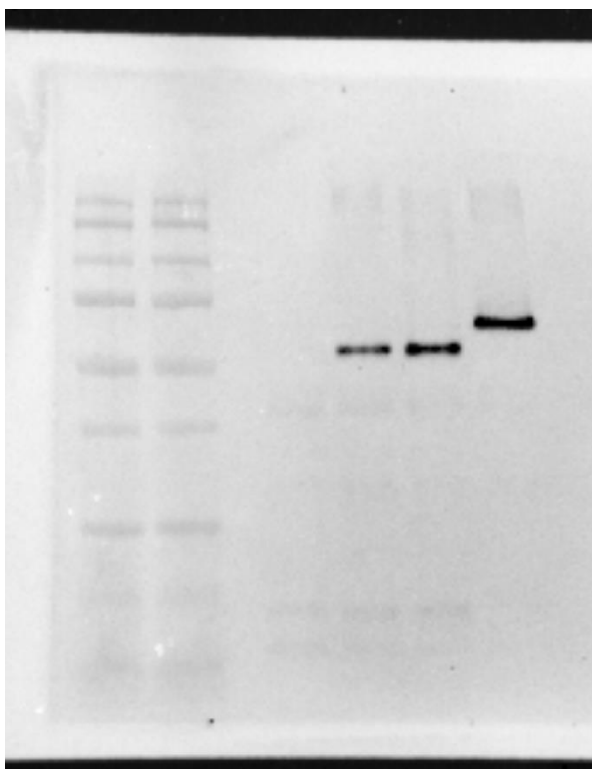
lane number	samples
1	MW marker (pre-stained)
2	MW marker
3	empty vector
4	myc-TBPCCL
5	myc-AAE16
6	myc-CoAL(A312G)



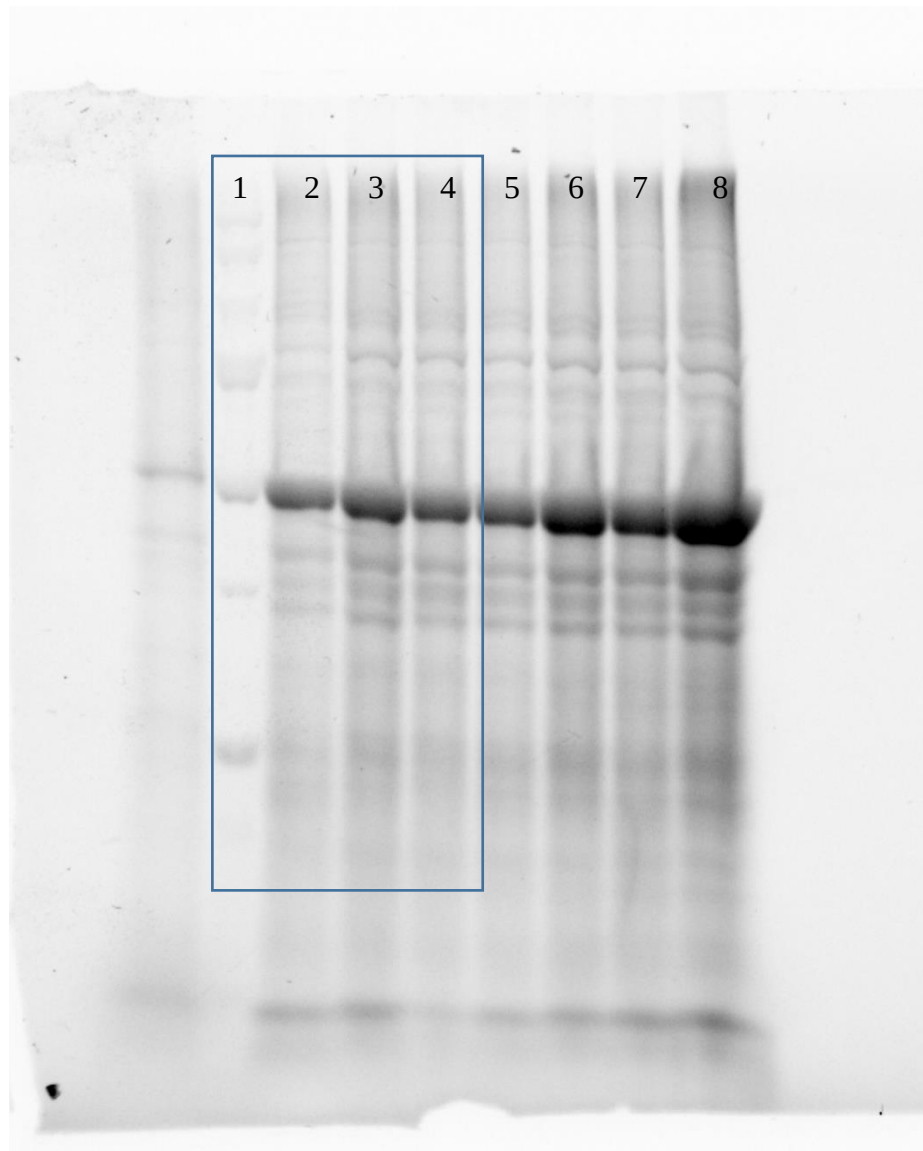


Western blot developed with anti-myc

lane number	samples
1	MW marker (pre-stained)
2	MW marker
3	empty vector
4	myc-TBPCCL
5	myc-AAE16
6	myc-CoAL(A312G)

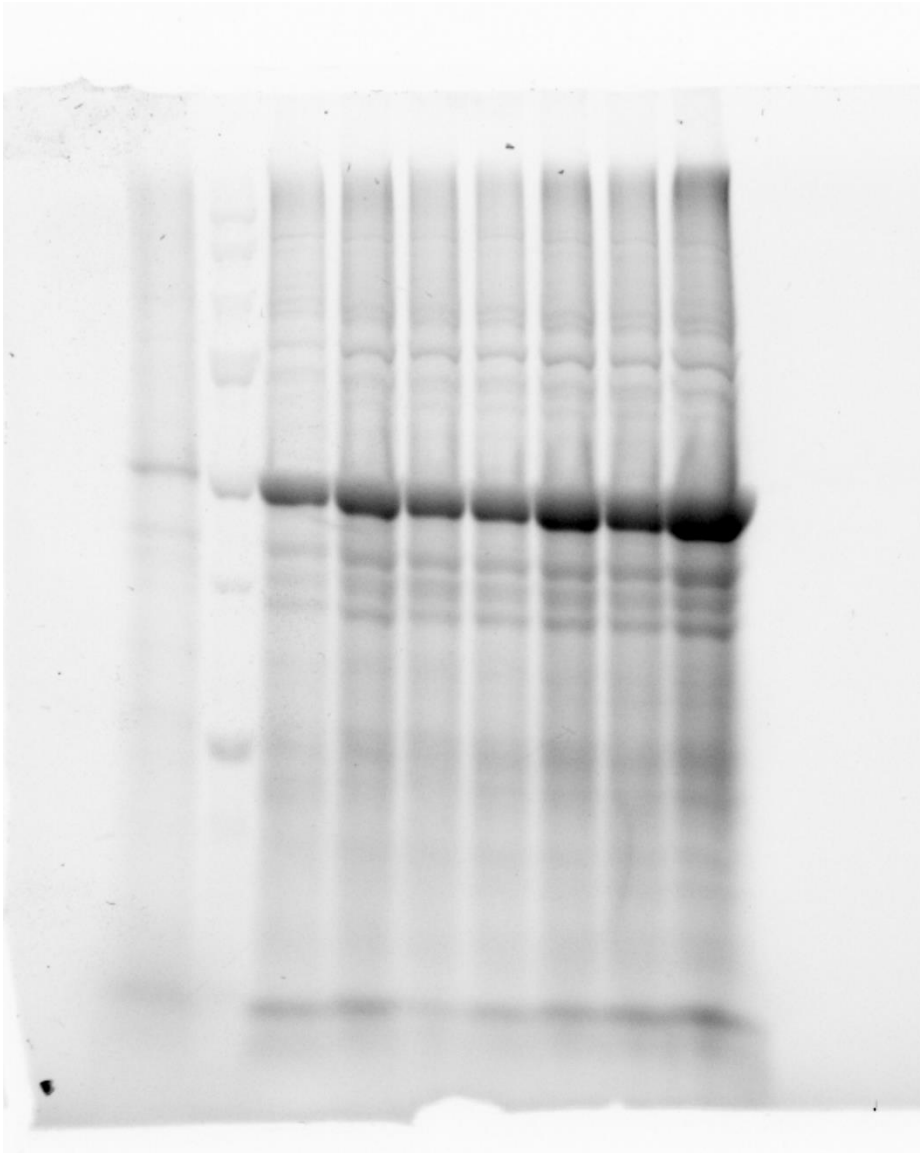


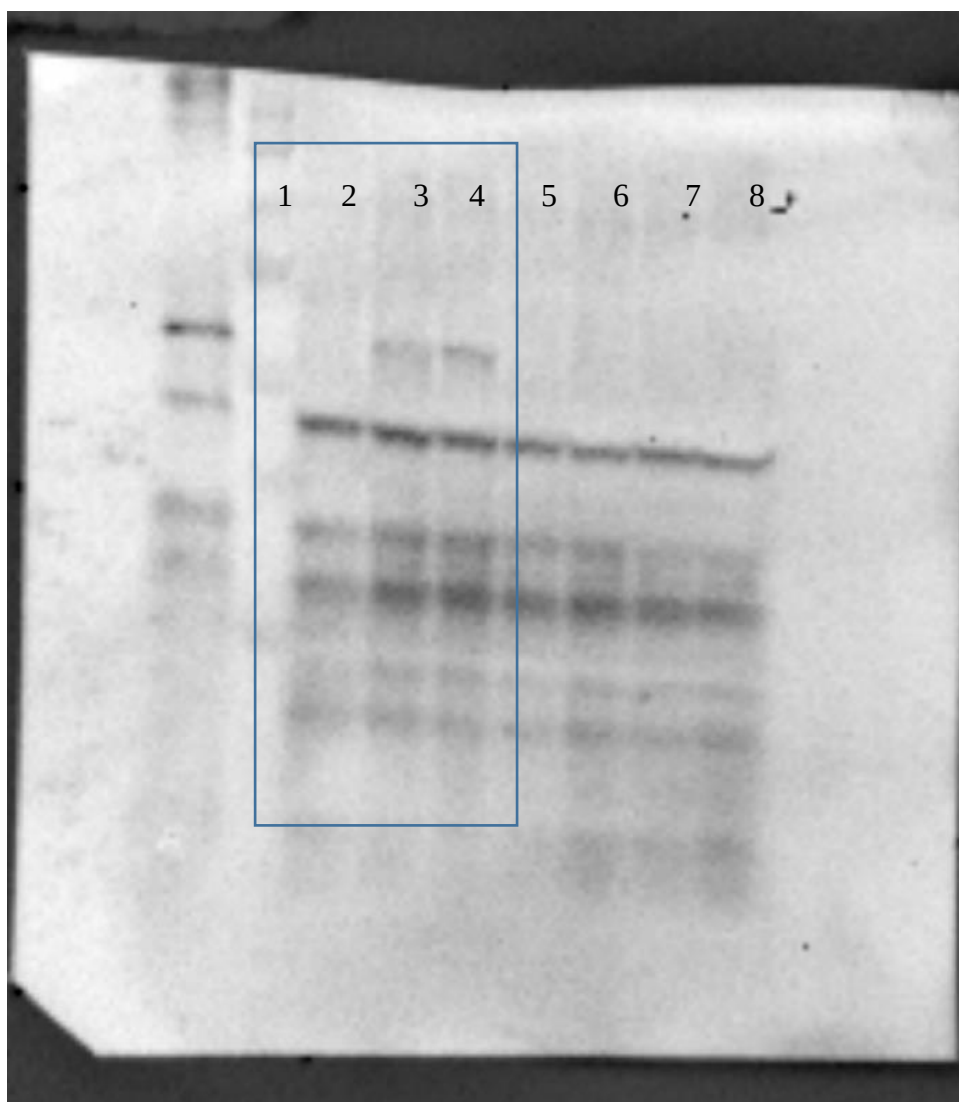
Additional Supplementary Fig 1. Unprocessed Gels and Blots for Supplementary Figure 17b.



SDS-PAGE for total protein used as loading control

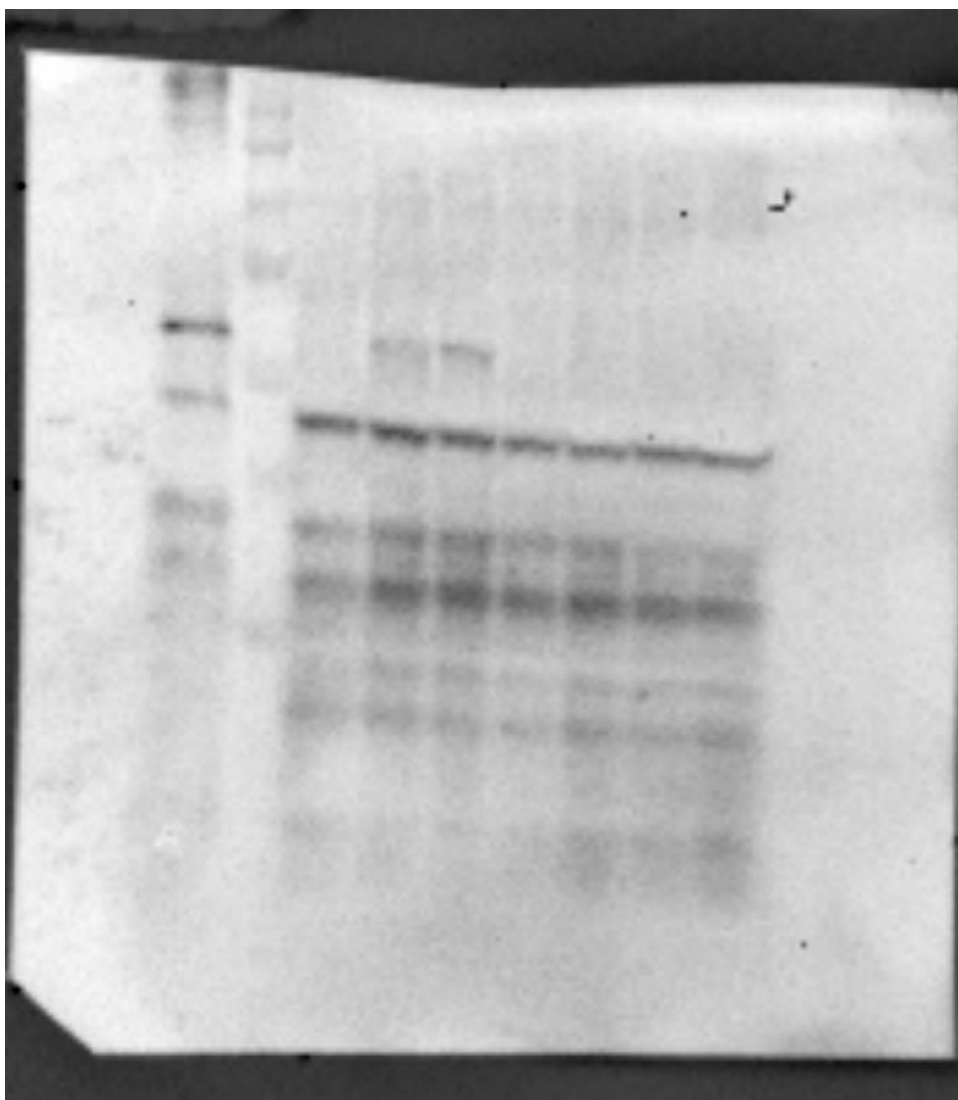
lane number	samples
1	MW marker
2	empty vector
3	myc-TB506
4	myc-TB506
5	TB506
6	TB506
7	strep-TB506
8	strep-TB506



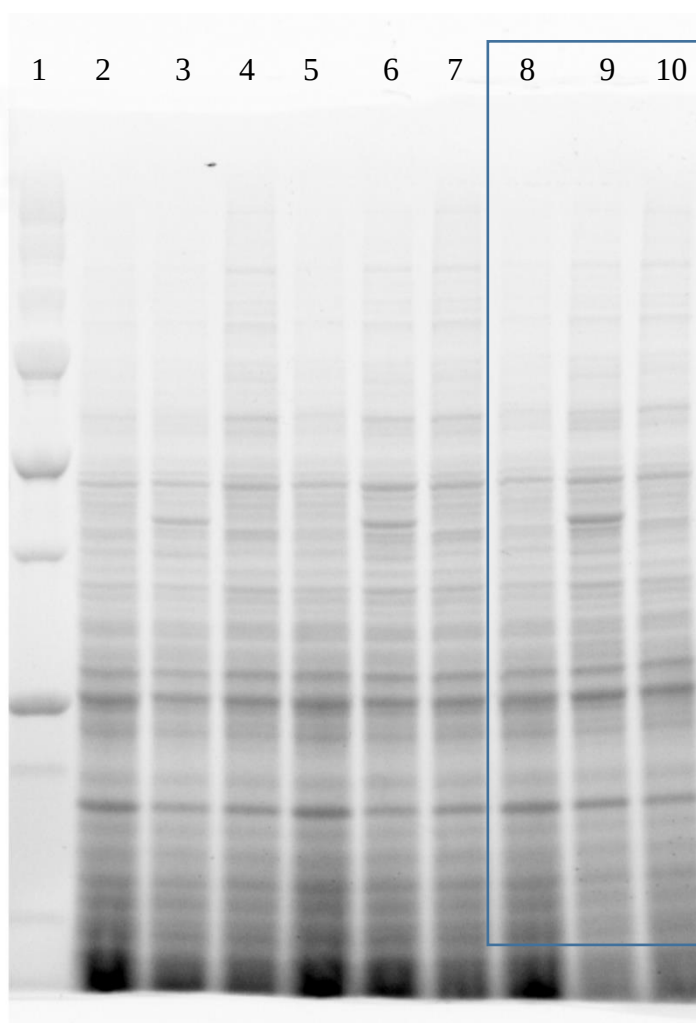


Western blot developed with anti-myc

lane number	samples
1	MW marker
2	empty vector
3	myc-TB506
4	myc-TB506
5	TB506
6	TB506
7	strep-TB506
8	strep-TB506

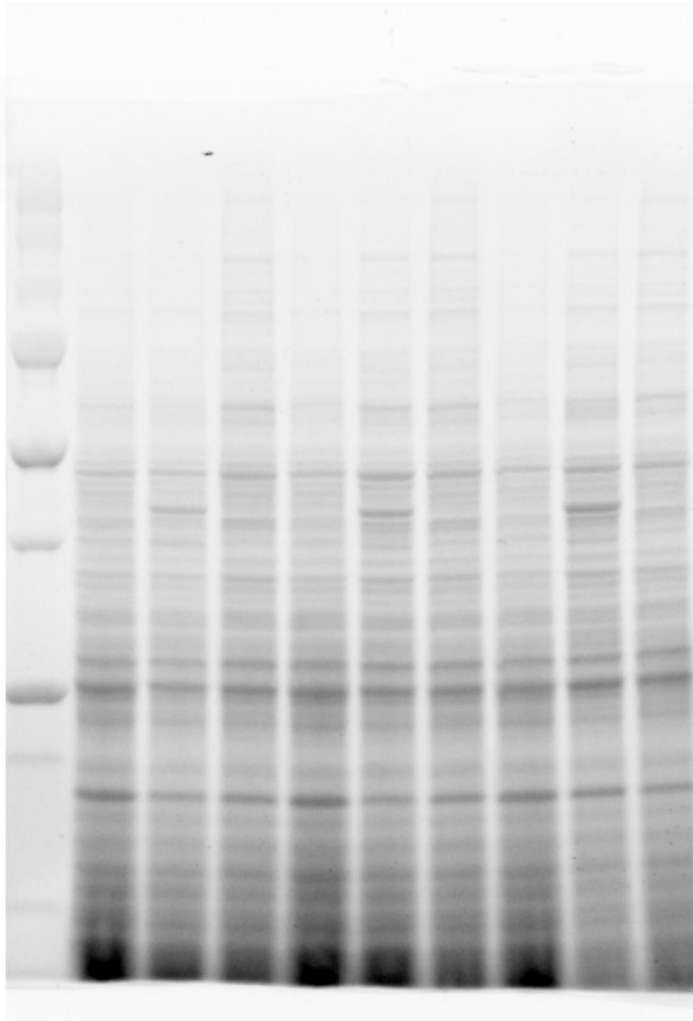


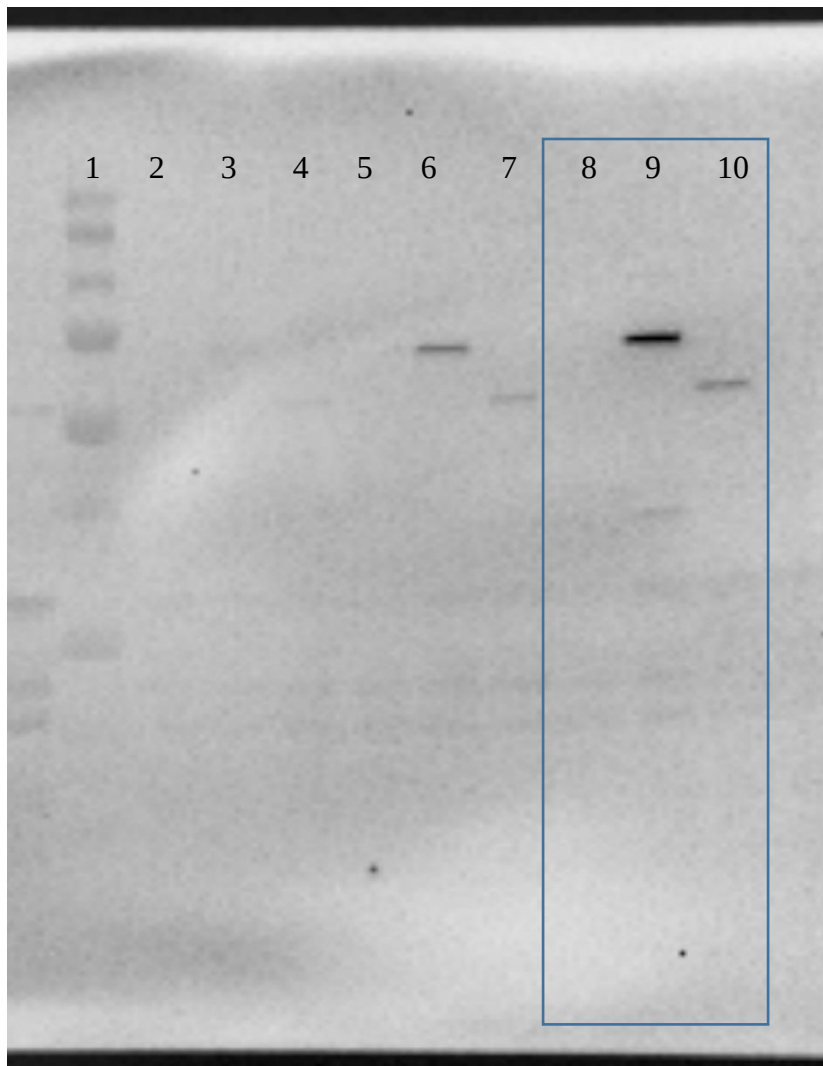
Additional Supplementary Fig. 2. Unprocessed Gels and Blots for Supplementary Figure 18.



SDS-PAGE for total protein used as loading control

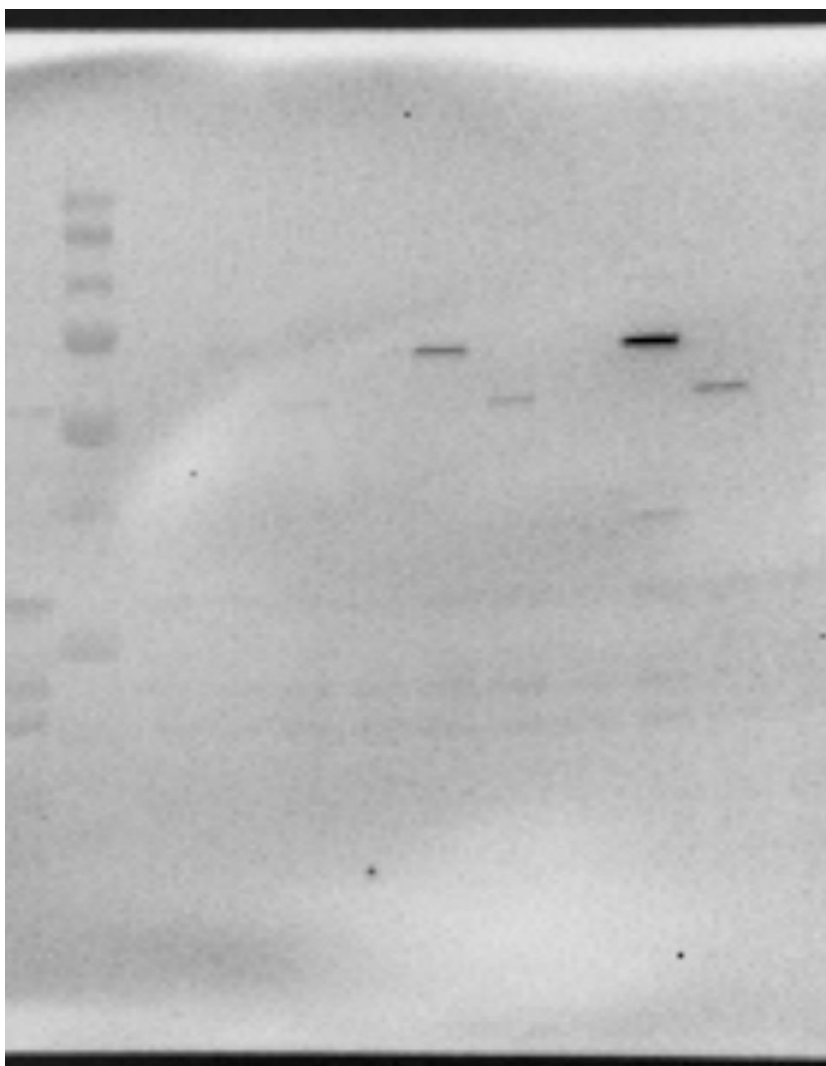
lane number	samples
1	MW marker
2	empty vector
3	CoAL(A312G) without induction
4	BAPT without induction
5	empty vector
6	CoAL(A312G) batch 1
7	BAPT batch 1
8	empty vector
9	CoAL(A312G) batch 2
10	BAPT batch 2



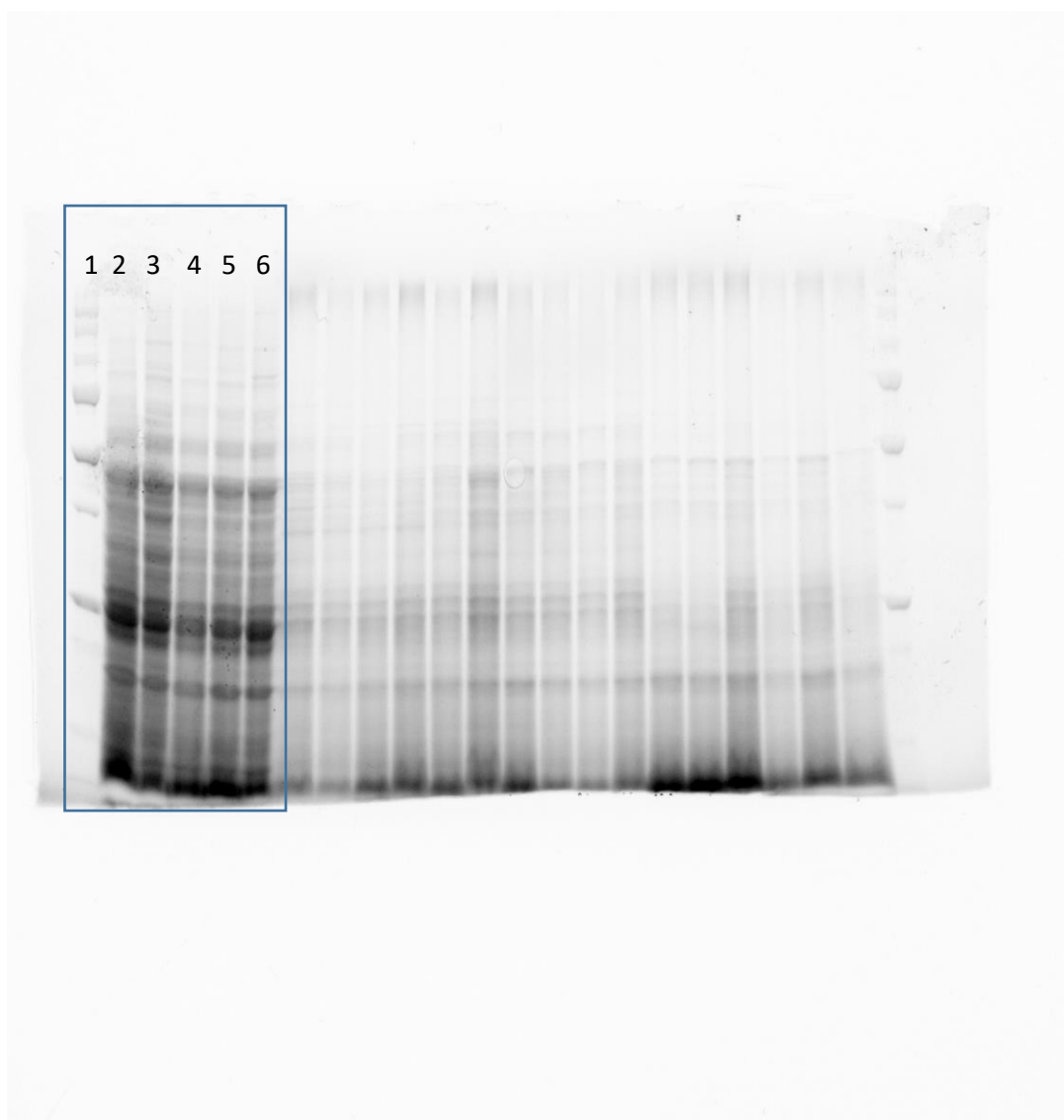


Western blot developed with anti-myc

lane number	samples
1	MW marker
2	empty vector
3	CoAL(A312G) without induction
4	BAPT without induction
5	empty vector
6	CoAL(A312G) batch 1
7	BAPT batch 1
8	empty vector
9	CoAL(A312G) batch 2
10	BAPT batch 2

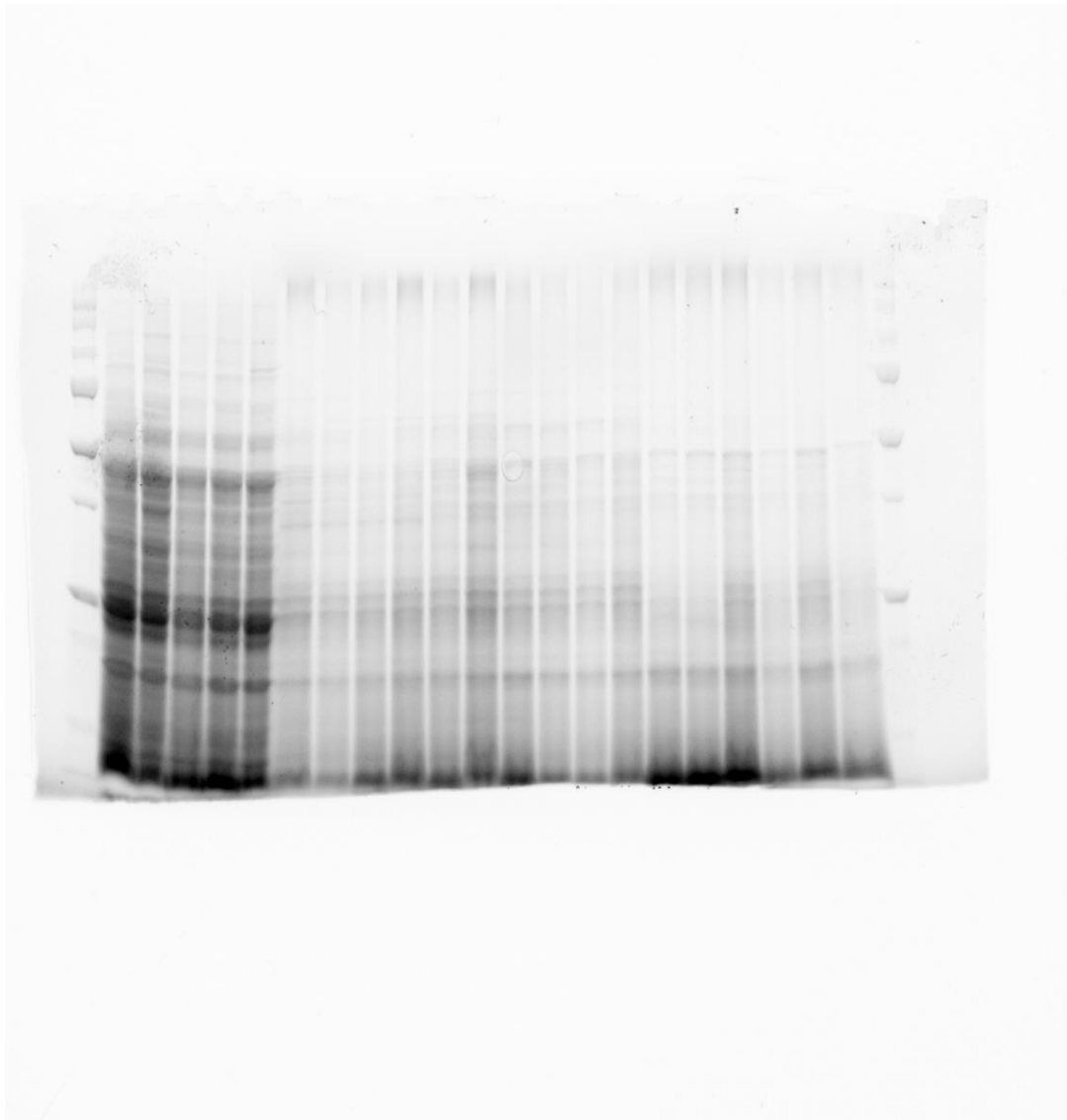


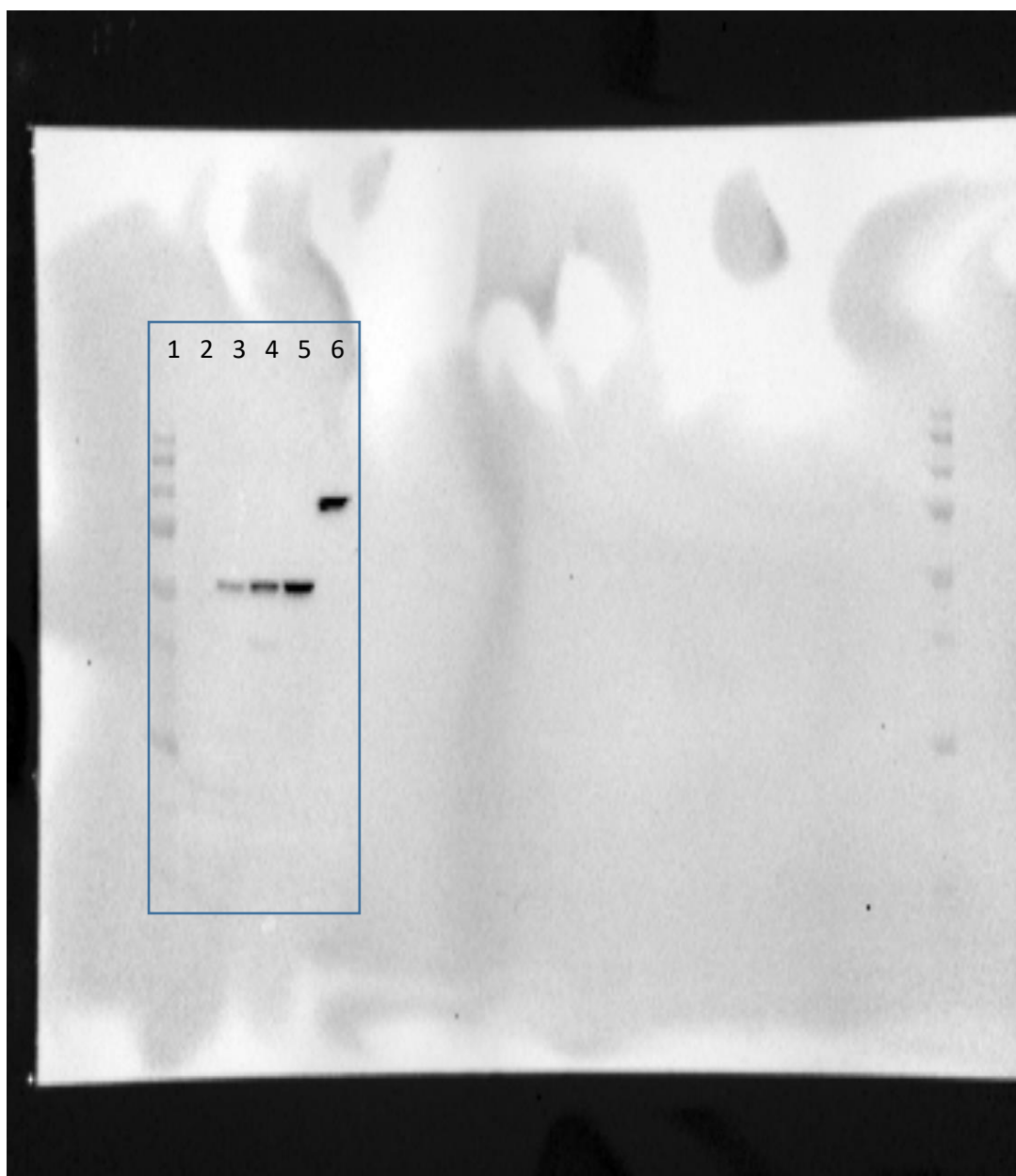
Additional Supplementary Fig. 3. Unprocessed Gels and Blots for Supplementary Figure 36.



SDS-PAGE for total protein used as loading control

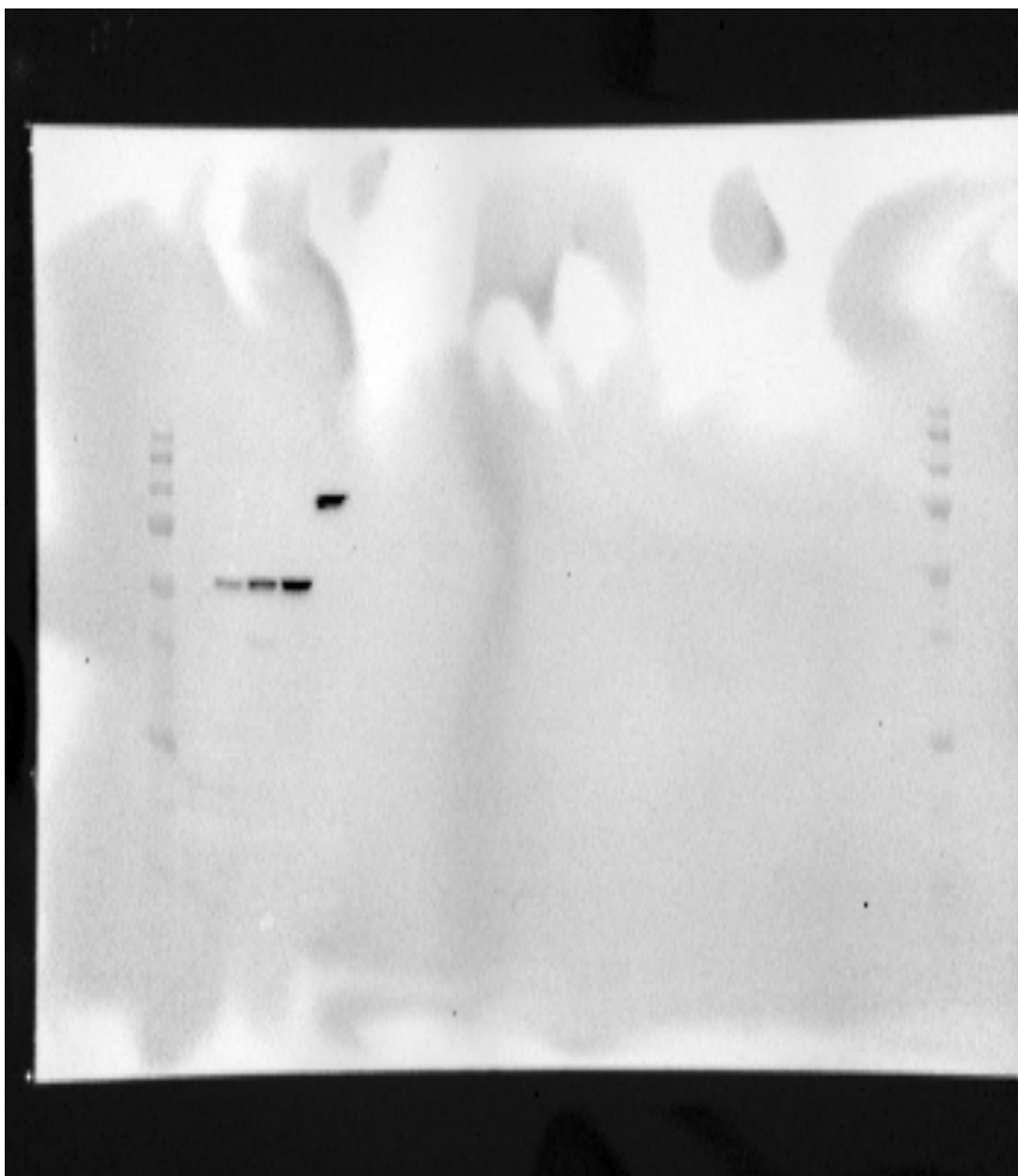
lane number	samples
1	MW marker
2	empty vector, 30°C
3	BAPT, 30°
4	BAPT, 20°C
5	BAPTm, 20°C
6	MBPig3BAPTm, 20°C





Western blot developed with anti-myc

lane number	samples
1	MW marker
2	empty vector, 30°C
3	BAPT, 30°
4	BAPT, 20°C
5	BAPTm, 20°C
6	MBPig3BAPTm, 20°C



Additional Supplementary Fig. 4. Unprocessed Gels and Blots for Supplementary Figure 37.