







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

Received: 26 August 2024

Accepted: 31 March 2025

Published online: 30 April 2025

 Check for updates

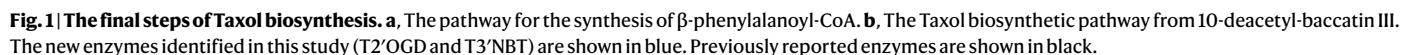
Feiyan Liang¹ , Yuman Xie¹ , Chi Zhang, Yong Zhao¹ ,
Mohammed S. Motawia²  & Sotirios C. Kampranis¹  

Taxol (paclitaxel) is a widely used anti-cancer drug with a complex biosynthetic pathway that has puzzled biochemists for decades. Owing to inefficient chemical synthesis, Taxol supply depends on costly semi-synthesis. Elucidating the Taxol biosynthesis will solve a long-standing question in biochemistry and enable cost-effective production using biotechnological methods. While recent advances have improved our understanding of the steps leading up to the intermediate baccatin III, the final steps of the pathway remain elusive. Here we use gene co-expression analysis, chemically synthesized intermediates and a stepwise learning-by-building approach to reveal the enzymes that catalyse the final two modifications, that is, C2'α hydroxylation and 3'-N benzoylation, which are essential for Taxol's bioactivity. To replace the current semi-synthetic method of Taxol production, we reconstruct the late pathway in yeast and synthesize Taxol from the readily available intermediate baccatin III. This work provides a complete understanding of Taxol biosynthesis and establishes a foundation for its biotechnological production.

Taxol is among the most commonly used chemotherapeutic drugs, effective in the treatment of breast, ovarian, cervical, nasopharyngeal and non-small cell lung cancer¹. It exerts its anti-mitotic activity by preventing the disassembly of microtubules². Initially isolated from the inner bark of the mature Pacific yew tree (*Taxus brevifolia*)^{3,4}, sourcing Taxol from yew trees is not viable owing to its remarkably low content in the bark. Chemical synthesis of Taxol is also inefficient because of its complex chemical structure^{5–7}. Consequently, pharmaceutical Taxol is currently produced either from *Taxus* cell culture or by semi-synthesis from more available plant-derived taxoids. Both methods are costly, making Taxol and its derivatives (docetaxel and cabazitaxel) among the highest-priced small molecule active pharmaceutical ingredients⁸. To improve Taxol supply, biotechnological production presents a promising alternative. However, progress has been hindered by incomplete knowledge of the biosynthetic pathway.

The remarkably complex structure of Taxol has prompted intense research into its biosynthesis for over 30 years. The pathway is believed to involve at least 18 biosynthetic steps. Early investigations in the 1990s established the first step as the cyclization of geranylgeranyl diphosphate to the basic diterpene scaffold, taxadiene, by the enzyme taxadiene synthase^{9,10}. Subsequent studies revealed several cytochrome P450 (CYP) enzymes that oxidize the taxadiene structure^{11–16} and acyltransferases that further decorate the oxygenated taxoids^{17–20}. However, further progress was hampered by the lack of comprehensive genomic and transcriptomic information. The recent genome sequencing of two Taxol-producing yew species^{21,22} has rekindled efforts to elucidate the Taxol pathway. Several studies have since contributed to understanding the early steps leading to the key intermediate baccatin III^{23–27}. These studies revisited earlier proposed biosynthetic schemes, discovering missing oxidase enzymes^{23–27}, accessory proteins essential

Biochemical Engineering Group, Plant Biochemistry Section, Department of Plant and Environment Sciences, Faculty of Science, University of Copenhagen, Frederiksberg C, Denmark. ✉ e-mail: soka@plen.ku.dk



Despite this progress, our knowledge of the final steps from baccatin III to Taxol remains incomplete. There are five putative enzymatic steps in the conversion of baccatin III to Taxol: the isomerization of α -phenylalanine to β -phenylalanine, the activation of β -phenylalanine by coenzyme A (CoA), the addition of the β -phenylalanine moiety to the C13 position of baccatin III, the oxidation of the β -phenylalaninyl side chain at C2' and the addition of a

benzoyl group at its 3'-N position (Fig. 1). Although at least one putative enzyme for each step has been proposed^{18,23,28–33}, the characterization of some steps was based on the indirect observation of downstream products^{23,32}, making their functional assignment inconclusive. One critical such step is the oxidation of the β -phenylalanyl side chain at C2', for which a CYP enzyme, TB506, has been proposed. However, its proposed activity has not been supported by experiments directly demonstrating the oxidation of the β -phenylalanyl side chain and only implied on the basis of the observation of Taxol production in plant cells or cell extracts^{23,32}, which may be influenced by the presence of host plant enzymes. Furthermore, the identification of 3'-N-debenzoyl-2'-deoxy-taxol N-benzoyltransferase (DBTNBT), the benzoyltransferase proposed to be responsible for the final step

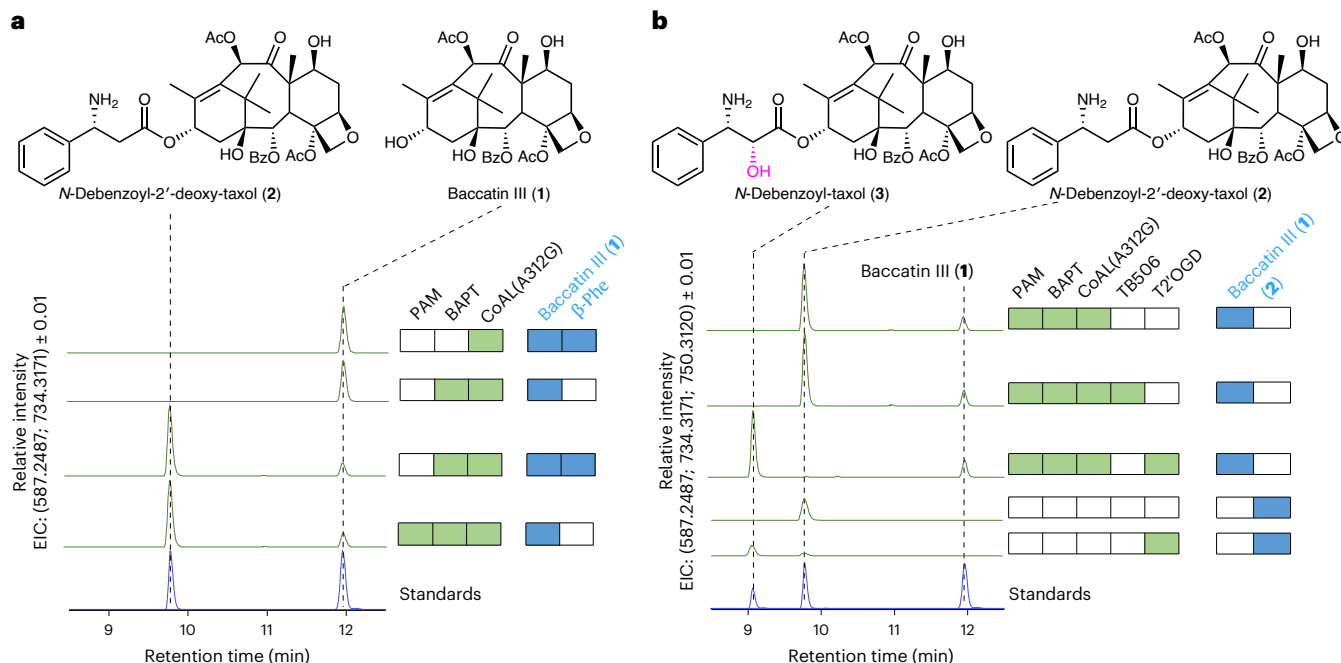


Fig. 2 | Biosynthesis of the Taxol intermediate *N*-debenzoyl-taxol (3) from *N*-debenzoyl-2'-deoxy-taxol (2) is catalysed by taxoid-2'-oxoglutarate-dependent dioxygenase. **a, Establishing the biosynthesis of *N*-debenzoyl-2'-deoxy-taxol (2) in tobacco. Extracted ion chromatograms (EICs, positive) of extracts of tobacco leaves show the levels of baccatin III (1, $C_{31}H_{39}O_{11}^+$, $m/z = 587.2487 \pm 0.01$) and *N*-debenzoyl-2'-deoxy-taxol (2, $C_{40}H_{48}NO_{12}^+$, $m/z = 734.3171 \pm 0.01$). For each chromatogram, the corresponding transiently produced enzymes (PAM, BAPT and/or CoAL(A312G)) are shown as solid green squares. Solid blue squares indicate the supply of exogenous substrate (200 mg l⁻¹ β -phenylalanine (β -phe) and/or 200 mg l⁻¹ baccatin III). Open squares**

indicate the absence of the corresponding enzyme or substrate. The formation of β -phenylalanine by PAM is sufficient to drive efficient biosynthesis of *N*-debenzoyl-2'-deoxy-taxol (2) since the supply of exogenous β -phenylalanine does not increase the *N*-debenzoyl-2'-deoxy-taxol (2) levels. **b**, The biosynthesis of *N*-debenzoyl-taxol (3) in tobacco. EICs (positive) show the levels of baccatin III (1), *N*-debenzoyl-2'-deoxy-taxol (2) and *N*-debenzoyl-taxol (3, $C_{40}H_{48}NO_{13}^+$, $m/z = 750.3120 \pm 0.01$). TB506, a CYP enzyme previously proposed to hydroxylate 2, cannot synthesize 3. The newly identified T2'OGD can efficiently synthesize 3 when 2 is present (either provided by the coordinated action of PAM, BAPT and CoAL(A312G) or supplied exogenously at 50 mg l⁻¹).

of the pathway, was based on its ability to benzoylate a surrogate substrate, *N*-debenzoyl-2'-deoxy-taxol, and not its native substrate *N*-debenzoyl-taxol³¹. Thus, elucidating these final steps is essential to completing the Taxol pathway and advancing its biotechnological production.

Here we address these gaps by systematically identifying and validating the enzymes responsible for the final steps of Taxol biosynthesis. By integrating gene co-expression analysis, chemical synthesis of intermediates and functional characterization in both plant and microbial systems, we provide experimental evidence for the enzymes catalysing the critical C2' hydroxylation and 3'-N benzoylation steps. These findings complete the Taxol biosynthetic pathway and lay the foundation for scalable and sustainable Taxol production in engineered microbial or plant hosts.

Results and discussion

Establishing the biosynthesis of *N*-debenzoyl-2'-deoxy-taxol

We set out to validate the previously proposed activities using a learning-by-building approach, reconstructing the pathway step by step. Given that transient heterologous gene expression in tobacco leaves is an efficient and reliable method to characterize plant enzyme activity^{34,35}, we opted to use *Nicotiana benthamiana* (hereafter tobacco) for pathway reconstruction. Additionally, as previous efforts in the elucidation of Taxol biosynthesis were hampered by the lack of isolated or chemically synthesized pathway intermediates, we chemically synthesized five putative pathway intermediates to use as substrates or standards: *N*-debenzoyl-2'-deoxy-taxol (2), *N*-debenzoyl-taxol (3), 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6), 10-deacetyl-*N*-debenzoyl-taxol (7) and 2'-deoxy-taxol (9) (Fig. 1 and Supplementary Figs. 1–15).

To reconstruct the pathway in tobacco, each gene, driven by the M24 transcript promoter of the *Mirabilis mosaic virus*, shown to support high-level gene expression³⁶ was introduced into tobacco through *Agrobacterium tumefaciens*-mediated (AGL1) infiltration (agroinfiltration). First, we studied the function of *Taxus baccata* phenylalanine aminomutase (PAM), which was reported to convert α -phenylalanine to β -phenylalanine²⁸. We tested PAM in tobacco and confirmed β -phenylalanine production by ultra-high-performance liquid chromatography–quadrupole time of flight–mass spectrometry (UPLC–qTOF–MS) analysis (Supplementary Fig. 16), verifying the reported function of PAM.

The next step in the pathway is the synthesis of β -phenylalanoyl-CoA, for which two candidates have been reported: β -phenylalanine- and 4-coumarate-CoA ligase (TBPCCL) from *T. baccata*²⁹ and acyl-activating enzyme 16 (AAE16) from *Taxus chinensis*²³. Since β -phenylalanoyl-CoA is unstable and can decompose during extraction, we combined its synthesis with the subsequent step of the pathway: the condensation of β -phenylalanoyl-CoA with baccatin III into *N*-debenzoyl-2'-deoxy-taxol (2; Fig. 1). This reaction is catalysed by the enzyme baccatin III 13-*O*-(3-amino-3-phenylpropanoyl) transferase (BAPT)³³. When TBPCCL or AAE16 were co-expressed with *Taxus cuspidata* BAPT in tobacco, a peak with the same retention time and mass spectrum as chemically synthesized *N*-debenzoyl-2'-deoxy-taxol (2) was observed (Supplementary Fig. 17). However, the levels of *N*-debenzoyl-2'-deoxy-taxol (2) obtained were low, hindering further reconstruction of the pathway. Therefore, we examined a variant of a CoA ligase from *Penicillium chrysogenum*, CoAL(A312G), which is reported to catalyse the same reaction^{23,37}. The combination of CoAL(A312G) with BAPT and PAM enabled robust synthesis of *N*-debenzoyl-2'-deoxy-taxol (2), converting over 60% of exogenous baccatin III (Fig. 2a; Supplementary Fig. 17).

Elucidating the C2'α hydroxylation step

From *N*-debenzoyl-2'-deoxy-taxol (**2**), the next step is postulated to be the hydroxylation of C2' of the β-phenylalaninyl side chain, leading to the formation of *N*-debenzoyl-taxol (**3**; Fig. 1). A CYP enzyme, TB506, has been proposed to catalyse this step using a *Pisum sativum* protoplast system³². However, no direct production of *N*-debenzoyl-taxol (**3**) from TB506 has been demonstrated. To validate the proposed function of TB506, we tested its ability to directly convert *N*-debenzoyl-2'-deoxy-taxol (**2**) into *N*-debenzoyl-taxol (**3**) in tobacco cells expressing PAM, CoAL(A312G) and BAPT. Although Western blot analysis confirmed the production of TB506 protein in the agro-infiltrated leaves (Supplementary Fig. 18), we were unable to obtain a product that matched the retention time and mass spectrum of chemically synthesized *N*-debenzoyl-taxol (**3**; Fig. 2b). When exogenous *N*-debenzoyl-2'-deoxy-taxol (**2**) was directly supplied to TB506-expressing tobacco leaves, we were again unable to detect a peak corresponding to the expected product. Therefore, we concluded that TB506 is probably not the correct enzyme for this step.

This prompted us to search for new CYP candidates for the enzyme catalysing C2'α hydroxylation. We utilized available genomic and transcriptomic data of *Taxus wallichiana*²¹, *Taxus chinensis* var. *mairei*²² and *Taxus yunnanensis*³⁸ to perform a gene co-expression analysis across different cell types (leaf, root, stem and strobilus) and under diverse conditions (for example, methyl-jasmonate treatment). Known Taxol biosynthetic genes were used as baits (Supplementary Table 2). The analysis revealed that the known genes form two separate co-expression groups. One group includes taxadiene synthase^{9,10}, taxoid-5α-hydroxylase¹⁶, taxoid-10β-hydroxylase¹¹, taxoid acetyltransferase 19¹⁷, benzoyl-CoA:taxane 2α-*O*-benzoyltransferase¹⁹, 10-deacetyl-baccatin III-10-*O*-acetyltransferase (DBAT)¹⁸ and BAPT³³. The second group consists of PAM²⁸, taxoid-2-hydroxylase¹³, taxoid-13α-hydroxylase¹² and taxoid-7β-hydroxylase¹⁵ (Supplementary Fig. 19). To identify new candidates for C2'α hydroxylation, we selected CYPs whose expression pattern correlated with at least one of the two groups (Supplementary Fig. 20). This led to the selection of 19 CYPs from *T. chinensis* for further characterization. Additionally, we included 16 CYPs from *T. chinensis* previously reported as differentially expressed in Taxol-producing cultured cells but not functionally characterized (Supplementary Table 3)³⁹. We also examined all CYPs previously reported to act on the taxadiene skeleton to explore the possibility of promiscuous activity. However, none of the selected candidates was able to produce *N*-debenzoyl-taxol (**3**; Supplementary Figs. 21 and 22).

Having screened over 40 different CYP candidates without success, we began to suspect that this reaction may not be catalysed by a CYP. Therefore, we turned our attention to another large oxidase family, the 2-oxoglutarate-dependent oxygenases (OGDs). OGDs employ 2-oxoglutarate and O₂ as co-substrates and are known to catalyse a wide range of oxidative transformations on small molecules and proteins^{40,41}. Notably, members of this enzyme family have been shown to hydroxylate aspartyl or asparaginyl residues⁴², a reaction that closely resembles the oxidation at the 2'α position of *N*-debenzoyl-2'-deoxy-taxol (Supplementary Fig. 23). This functional similarity highlighted OGDs as strong candidates for further investigation. We re-analysed the co-expression matrix and found nine *T. chinensis* OGD genes that showed strong expression correlation with at least one of the two aforementioned expression groups (Supplementary Fig. 24). These nine candidates were further studied in tobacco plants producing *N*-debenzoyl-2'-deoxy-taxol (**2**) through the co-expression of PAM, CoAL(A312G) and BAPT. One of these candidates produced a new peak confirmed to be *N*-debenzoyl-taxol (**3**) by comparison of its retention time and mass spectrum with an authentic standard (Fig. 2b and Supplementary Figs. 25 and 26). The identified OGD demonstrated high efficiency, converting nearly all of the available *N*-debenzoyl-2'-deoxy-taxol (**2**) to *N*-debenzoyl-taxol (**3**; Fig. 2b). The observed activity was further validated by feeding chemically synthesized *N*-debenzoyl-2'-deoxy-taxol (**2**) to tobacco leaves expressing only the new enzyme

(Fig. 2b). We named this enzyme taxoid-2'-oxoglutarate-dependent dioxygenase (T2'OGD; GenBank accession no. [PQ015324](#)).

Benzoylation of the C13 side chain

The final step of the pathway involves the benzoylation of *N*-debenzoyl-taxol (**3**). The enzyme DBTNBT has been proposed to catalyse this step³¹. To validate the activity of DBTNBT, we transiently expressed it in tobacco plants that produced *N*-debenzoyl-taxol (**3**), the proposed substrate of DBTNBT, and we detected low amounts of Taxol (Fig. 3a). However, similar Taxol levels were also detected in samples where DBTNBT was absent (Fig. 3a), suggesting that the Taxol observed was probably produced by tobacco enzymes. To ensure that the inability to detect DBTNBT-mediated Taxol production was not due to the enzyme being non-functional in our system, we assessed its reported ability to benzoylate *N*-debenzoyl-2'-deoxy-taxol (**2**) to 2'-deoxy-taxol (**9**)^{30,31} by co-expressing it with PAM, CoAL(A312G) and BAPT. Indeed, DBTNBT successfully converted *N*-debenzoyl-2'-deoxy-taxol (**2**) to 2'-deoxy-taxol (**9**; Supplementary Fig. 27), confirming that, although it is functionally produced in tobacco, it cannot efficiently catalyse Taxol synthesis. Therefore, we concluded that previous reports of Taxol production in tobacco mediated by DBTNBT²³ were probably influenced by the low levels of Taxol produced by endogenous tobacco enzymes.

Therefore, we embarked on identifying the correct 3'-*N*-benzoyltransferase using gene co-expression analysis as a guide for candidate selection. In total, 25 BAHD family (BAHD stands for the first four characterized enzymes in this family: benzylalcohol *O*-acetyltransferase, anthocyanin *O*-hydroxycinnamoyltransferase, anthranilate *N*-hydroxycinnamoyl/benzoyltransferase and deacetyl-vindoline 4-*O*-acetyltransferase) candidates from *T. chinensis* were chosen (Supplementary Fig. 28) and individually tested in tobacco plants producing *N*-debenzoyl-taxol (**3**) through co-expression of PAM, CoAL(A312G), BAPT and T2'OGD. One candidate was found to consume *N*-debenzoyl-taxol (**3**) and produce Taxol, confirmed by comparison with an authentic Taxol standard (Fig. 3a and Supplementary Figs. 29 and 30). This newly identified enzyme, named here as taxoid-3'-*N*-benzoyltransferase (T3'NBT, GenBank accession no. [PQ015327](#)), exhibited high efficiency, converting 75% of the available *N*-debenzoyl-taxol (**3**) to Taxol. Compared with the background level of Taxol produced by tobacco in the absence or presence of DBTNBT, Taxol derived from T3'NBT was 280 times higher (Fig. 3b), clearly suggesting that T3'NBT is the preferred enzyme for the 3'-*N*-benzoylation of taxoids. Under these conditions, Taxol production reached 1.99 ± 0.19 μg per gram of leaf wet weight.

Sequence of biosynthetic events

Having identified T2'OGD and T3'NBT as the enzymes catalysing the last two steps of the Taxol pathway, we investigated the sequence of these two biosynthetic events. First, we assessed the ability of T2'OGD to hydroxylate 2'-deoxy-taxol (**9**) by supplying T2'OGD-expressing tobacco cells with chemically synthesized 2'-deoxy-taxol (**9**). We found that 2'-deoxy-taxol (**9**) is not a substrate for T2'OGD (Supplementary Fig. 31), establishing that C2'α hydroxylation by T2'OGD precedes benzoylation by T3'NBT (Fig. 1). Subsequently, we expressed T3'NBT in tobacco leaves synthesizing *N*-debenzoyl-2'-deoxy-taxol (**2**). T3'NBT was unable to convert *N*-debenzoyl-2'-deoxy-taxol (**2**) to 2'-deoxy-taxol (**9**), confirming the essential role of T2'OGD-mediated C2'α hydroxylation for benzoylation by T3'NBT (Supplementary Fig. 32). These findings were further supported by analysing the metabolites present in *T. baccata* bark extracts, where it was evident that, although *N*-debenzoyl-2'-deoxy-taxol (**2**) and *N*-debenzoyl-taxol (**3**) were present, 2'-deoxy-taxol (**9**) could not be detected (Supplementary Fig. 33).

Metabolic grid leading to Taxol

Baccatin III is synthesized from 10-deacetyl-baccatin III through the action of the enzyme DBAT¹⁸ (Fig. 1). It has been reported that DBAT

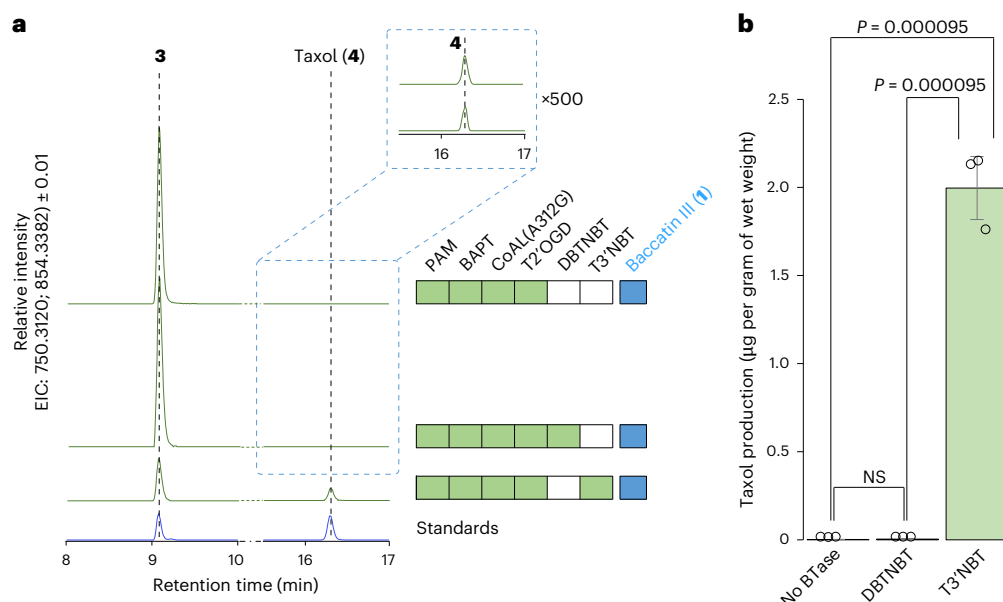


Fig. 3 | T3'NBT catalyses the biosynthesis of Taxol from *N*-debenzoyl-taxol (3). **a**, EICs (positive) of extracts of tobacco leaves transiently expressing PAM, BAPT, CoAL(A312G), T2'OGD and DBTNBT (the previously proposed enzyme for the 3'-N-benzoylation reaction) or T3'NBT (the newly identified enzyme for the 3'-N-benzoylation reaction). Inset: 500× magnifications of the EICs show low levels of Taxol (4, $C_{47}H_{52}NO_{14}^+$, $m/z = 854.3382 \pm 0.01$) can be produced by the conversion of *N*-debenzoyl-taxol (3, $C_{40}H_{48}NO_{13}^+$, $m/z = 750.3120 \pm 0.01$) by endogenous tobacco enzymes. When DBTNBT is present, Taxol levels do not increase above the background. Taxol is produced from *N*-debenzoyl-taxol (3) above background levels only when PAM, BAPT, CoAL(A312G), T2'OGD and

T3'NBT are present. Transiently produced enzymes are shown in solid green squares. Baccatin III was supplied at 200 mg l^{-1} at 2 days after agro-infiltration. **b**, Taxol levels in tobacco expressing PAM, BAPT, CoAL(A312G) and T2'OGD in the absence of a benzoyltransferase (No BTase), in the presence of DBTNBT and in the presence of T3'NBT. Data are shown as mean \pm s.d. from $n = 3$ independent biological replicates. The one-way analysis of variance (ANOVA) P value ($P = 0.000095$) indicates a statistically significant difference between T3'NBT and the other two samples. There is no statistically significant difference (NS) between the DBTNBT sample and the sample without a benzoyltransferase. Individual data points are indicated by open circles.

also acetylates the C10 hydroxyl group of 10-deacetyl-taxol (8) to form Taxol⁴³. This raises the question of whether baccatin III is the exclusive entry point into the final steps of Taxol biosynthesis or if these steps function as a metabolic grid, allowing 10-deacetyl-baccatin III to serve as an alternative entry point. To explore this, we investigated whether BAPT, T2'OGD or T3'NBT could use taxoids with the 10-deacetyl-baccatin III core as substrates.

Feeding 10-deacetyl-baccatin III to tobacco leaves expressing PAM, CoAL(A312G) and BAPT, we confirmed that BAPT adds the phenylalanine side chain to C13 of 10-deacetyl-baccatin III, forming 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6; Fig. 4a and Supplementary Fig. 34). Furthermore, T2'OGD oxidizes 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6) to form 10-deacetyl-*N*-debenzoyl-taxol (7), and T3'NBT benzoylates 10-deacetyl-*N*-debenzoyl-taxol (7) to form 10-deacetyl-taxol (8; Fig. 4a and Supplementary Fig. 34). Notably, T3'NBT does not benzoylate 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6; Supplementary Fig. 35), confirming its strict requirement for the presence of the C2'α hydroxyl group.

These findings suggest that the final steps of Taxol biosynthesis probably operate as a metabolic grid. To further investigate, we examined whether DBAT acetylates intermediates between 10-deacetyl-baccatin III and 10-deacetyl-taxol, namely 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6) and 10-deacetyl-*N*-debenzoyl-taxol (7). Tobacco feeding experiments revealed that DBAT indeed acetylates these intermediates, producing *N*-debenzoyl-2'-deoxy-taxol (2) and *N*-debenzoyl-taxol (3), respectively (Fig. 4b). Furthermore, analysis of *T. baccata* bark extracts confirmed the presence of 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6), 10-deacetyl-*N*-debenzoyl-taxol (7), *N*-debenzoyl-2'-deoxy-taxol (2), *N*-debenzoyl-taxol (3) and 10-deacetyl-taxol (8), supporting the grid structure of the pathway (Supplementary Fig. 33).

Biotechnological production of Taxol in yeast

To further confirm the function of the identified enzymes and to establish the basis for the biotechnological production of Taxol, we reconstructed the pathway from baccatin III to Taxol in a stepwise manner in baker's yeast (*Saccharomyces cerevisiae*). As in tobacco, we started by establishing efficient production of *N*-debenzoyl-2'-deoxy-taxol (2) by UPLC-qTOF-MS analysis of yeast culture extracts. As baccatin III import or active export could be a limiting factor in this step, we examined the levels of baccatin III inside the yeast cells and in the medium. We observed similar intracellular and extracellular levels of baccatin III, suggesting that another obstacle, such as poor expression or activity of BAPT or CoAL(A312G), could be responsible. Thus, we proceeded to evaluate the protein levels of these two enzymes by Western blotting and found that the levels of BAPT were considerably lower than those of CoAL(A312G) (Supplementary Fig. 36). Suspecting that the low BAPT levels could be due to poor stability or solubility, we lowered the yeast cultivation temperature from 30 to 20 °C. This improved BAPT protein levels (Supplementary Fig. 37) enabled detectable conversion of baccatin III to *N*-debenzoyl-2'-deoxy-taxol (2; Supplementary Fig. 38). As observed in tobacco, the two previously proposed CoA-ligases, TBPCCL and AAE16, were markedly less efficient than CoAL(A312G) in supporting *N*-debenzoyl-2'-deoxy-taxol synthesis in yeast (Supplementary Fig. 38).

Despite this improvement, only ~0.001% of baccatin III was converted to *N*-debenzoyl-2'-α-deoxy-taxol under these conditions. Thus, we focused on improving the *N*-debenzoyl-2'-deoxy-taxol (2) production by enhancing the performance of BAPT. First, we utilized

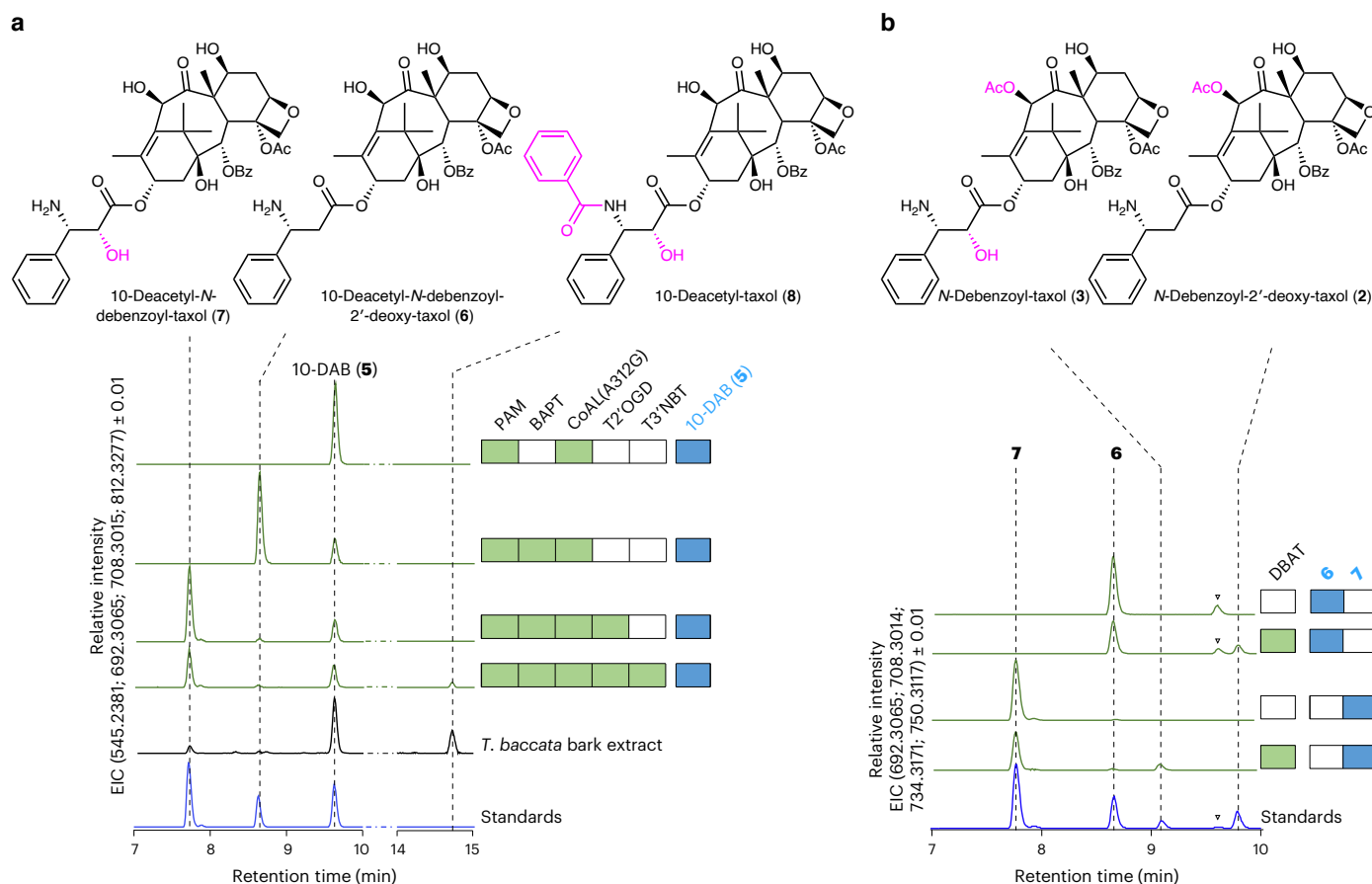


Fig. 4 | 10-Deacetyl-baccatin III serves as an alternative entry point into the late Taxol biosynthetic pathway. a, Biosynthesis of 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6), 10-deacetyl-*N*-debenzoyl-taxol (7) and 10-deacetyl-taxol (8) from exogenously supplied 10-deacetyl-baccatin III (10-DAB, 5) in tobacco leaves transiently expressing PAM, BAPT, CoAL(A312G), T2'OGD and T3'NBT. EICs (positive) of 10-DAB (5, $C_{29}H_{37}O_{10}^+$, $m/z = 545.2381 \pm 0.01$), 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6, $C_{38}H_{46}NO_{11}^+$, $m/z = 692.3065 \pm 0.01$), 10-deacetyl-*N*-debenzoyl-taxol (7, $C_{38}H_{45}NO_{12}^+$, $m/z = 708.3015 \pm 0.01$) and 10-deacetyl-taxol (8, $C_{45}H_{50}NO_{13}^+$, $m/z = 812.3277 \pm 0.01$). **b**, DBAT is a promiscuous enzyme that accepts both 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6) and 10-deacetyl-*N*-debenzoyl-taxol (7) as substrates and produces *N*-debenzoyl-

2'-deoxy-taxol (2) and *N*-debenzoyl-taxol (3), respectively. EICs (positive) of 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6), 10-deacetyl-*N*-debenzoyl-taxol (7), *N*-debenzoyl-2'-deoxy-taxol (2, $C_{40}H_{48}NO_{12}^+$, $m/z = 734.3171 \pm 0.01$) and *N*-debenzoyl-taxol (3, $C_{40}H_{48}NO_{13}^+$, $m/z = 750.3120 \pm 0.01$) are shown. Transiently produced enzymes are shown in solid green squares. Solid blue squares indicate the exogenous supply of the corresponding substrate (200 mg l⁻¹ 10-DAB (5), 50 mg l⁻¹ 10-deacetyl-*N*-debenzoyl-2'-deoxy-taxol (6) or 50 mg l⁻¹ 10-deacetyl-*N*-debenzoyl-taxol (7)) 2 days after the agro-infiltration event. Inverted triangles indicate an uncharacterized byproduct of the chemical synthesis of compound 6 that has the same exact mass as 6 (Supplementary Fig. 34d).

computational tools to assess the BAPT solubility and guide the design of improved variants. We searched for BAPT homologs from other *Taxus* species (Supplementary Table 5) and aligned their amino acid sequences (Supplementary Fig. 39) to identify conserved or nearly conserved residues probably critical for enzyme folding and activity. Thirty-six residues were identified as probably non-essential and suitable for mutagenesis to enhance solubility. Next, we employed the Protein-Sol predictive algorithm⁴⁶ to evaluate the impact of substituting each of these residues on the solubility of *T. cuspidata* BAPT. This analysis identified nine positions predicted to positively affect solubility, which were combined into a new variant, BAPtm (Supplementary Fig. 39). BAPtm exhibited higher predicted solubility than any individual BAPT enzyme identified in our search (Supplementary Table 5). Introducing BAPtm into yeast increased the *N*-debenzoyl-2'-deoxy-taxol (2) production by 27% compared with wild-type BAPT (Fig. 5a).

Fusing proteins with maltose binding protein (MBP) is a general strategy to increase protein stability and solubility^{47,48}. Therefore, to further improve the performance of BAPtm, we constructed fusions with MBP at either the N- or C-terminus of BAPtm using a four-amino

acid linker (IGGG). While appending MBP to the C-terminus reduced the *N*-debenzoyl-2'-deoxy-taxol (2) production compared with unfused BAPtm, fusing MBP at the N-terminus (MBPig3BAPtm) improved the performance (Supplementary Fig. 37). Consequently, yeast strain LT03, carrying MBPig3BAPtm and CoAL(A312G), exhibited a 77% increase in *N*-debenzoyl-2'-deoxy-taxol (2) production compared with LT01 (Fig. 5a).

We proceeded to incorporate the next enzyme, T2'OGD, into the yeast cells expressing CoAL(A312G) and MBPig3BAPtm, resulting in strain LT04 (Supplementary Table 4). T2'OGD was efficient in producing *N*-debenzoyl-taxol in yeast, consuming 71 ± 1% of the available *N*-debenzoyl-2'-deoxy-taxol (2, Fig. 5b), further confirming its role in Taxol biosynthesis.

To complete the pathway, we introduced CoAL(A312G), MBPig3BAPtm, T2'OGD and T3'NBT into EGY48, resulting in strain LT05 (Supplementary Table 4). Upon feeding baccatin III and β-phenylalanine, Taxol production was clearly observed (Fig. 5b and Supplementary Figs. 40 and 41). Under these conditions, *N*-debenzoyl-taxol (3) was efficiently consumed by T3'NBT, further confirming its function as a taxoid-3'-*N*-benzoyltransferase. The Taxol titer was 0.59 ± 0.03 μg l⁻¹.

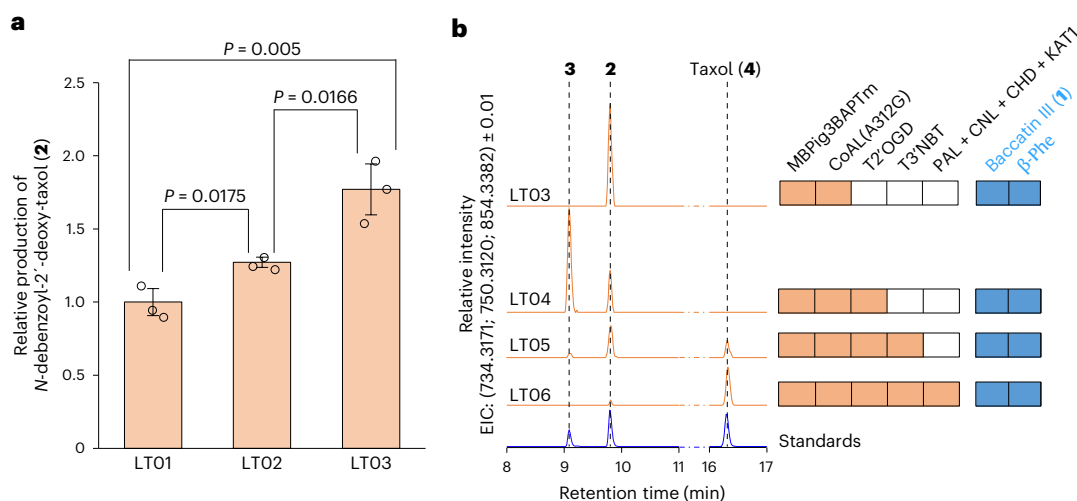


Fig. 5 | Reconstruction of the late Taxol biosynthetic pathway in *S. cerevisiae*. **a**, BAPT optimization to increase the production of *N*-debenzoyl-2'-deoxy-taxol (2; $C_{40}H_{48}NO_{12}^+$, $m/z = 734.3171 \pm 0.01$) in *S. cerevisiae*. All three strains express CoAL(A312G) from episomal vectors. Additionally, LT01 expresses BAPT, LT02 expresses BAPTm and LT03 expresses BAPTm fused with maltose binding protein (MBPig3BAPTm). Data are shown as mean \pm s.d. from $n = 3$ biological replicates. One-way ANOVA *P* values indicate significant statistical differences. Production

of *N*-debenzoyl-2'-deoxy-taxol (2) in LT01 is set as 1. **b**, Biosynthesis of Taxol (4) in yeast. EICs (positive) showing the levels of *N*-debenzoyl-2'-deoxy-taxol (2), *N*-debenzoyl-taxol (3, $C_{40}H_{48}NO_{13}^+$, $m/z = 750.3120 \pm 0.01$) and Taxol (4, $C_{47}H_{52}NO_{14}^+$, $m/z = 854.3382 \pm 0.01$) in ethyl acetate extracts of yeast cultures supplemented with 100 mg l⁻¹ baccatin III (1) and 100 mg l⁻¹ β-phenylalanine. Solid orange squares indicate the produced proteins.

Engineering a benzoyl-CoA-supplying module

Benzoylation by T3'NBT requires benzoyl-CoA as a co-substrate. However, benzoyl-CoA levels in yeast may be insufficient for efficient Taxol production. To enhance Taxol synthesis, we set out to introduce a benzoyl-CoA biosynthetic module to boost the yeast benzoyl-CoA pool. In *Petunia hybrida*, a beta-oxidation pathway is responsible for producing benzoyl-CoA for the biosynthesis of benzenoids. This pathway involves four core steps that convert cinnamic acid to benzoyl-CoA. In the first step, cinnamic acid ligase (CNL) attaches CoA to cinnamic acid to form cinnamoyl-CoA⁴⁹. Next, the bifunctional enzyme cinnamoyl-CoA hydratase-dehydrogenase (CDH) catalyses two reactions: conversion of cinnamoyl-CoA to 3-hydroxy-3-phenylpropanoyl-CoA (3H3PP-CoA) and then reduction of 3H3PP-CoA to 3-oxo-3-phenylpropanoyl-CoA (3O3PP-CoA)⁵⁰. Finally, 3O3PP-CoA is cleaved by 3-ketoacyl-CoA thiolase 1 (KAT1) to produce benzoyl-CoA⁵¹ (Fig. 6, Supplementary Table 1). This pathway normally operates in the peroxisome of *P. hybrida* cells. Thus, to establish a benzoyl-CoA-producing module in the yeast cytosol, we removed the peroxisomal targeting signals from CNL, CDH and KAT1 and introduced these genes into strain LT05 via chromosomal integration. Furthermore, to provide the core steps with cinnamic acid, we introduced the enzyme phenylalanine ammonia-lyase 2 (PAL) from *Arabidopsis thaliana*, which converts the amino acid phenylalanine to cinnamic acid⁵² (resulting in strain LT06; Supplementary Table 4).

Introducing the benzoyl-CoA module had a positive impact on Taxol production. The improved availability of benzoyl-CoA supported a strong pull by T3'NBT in the last step of the reconstructed pathway, resulting in the almost complete conversion of the intermediates *N*-debenzoyl-2'-deoxy-taxol (2) and *N*-debenzoyl-taxol (3; Fig. 5b). As a result, a 1.6-fold increase in the Taxol titer was obtained, reaching $0.97 \pm 0.05 \mu\text{g l}^{-1}$.

Conclusions

In this Article, we employed a systematic approach that included the functional validation of key enzymes and the identification of missing activities to complete the elusive final steps of Taxol biosynthesis. Using a step-by-step approach, we discovered that the final two modifications (C2'α hydroxylation and 3'-N benzoylation), which

are critical for Taxol's anti-mitotic activity⁵³, had been incorrectly assigned. To accurately identify the correct enzymes, we conducted a co-expression analysis across three different *Taxus* species, leading to the identification of T2'OGD and T3'NBT as the correct enzymes for these steps. We validated the functions of T2'OGD and T3'NBT in both tobacco and yeast systems, where they showed high activity and efficiency in converting their substrates. These findings complete the understanding of a biosynthetic pathway that has mystified biochemists for decades.

Building on this knowledge, we leveraged the benefits of microbial systems to develop a sustainable and scalable method for Taxol production, establishing a yeast cell factory for the bioconversion of baccatin III to Taxol. Currently, Taxol is primarily produced through chemical semi-synthesis from other taxoids, such as baccatin III⁵⁴. These taxoids can be sustainably sourced from the needles of cultivated *Taxus* species without causing irreversible damage to the plant, making them relatively scalable and cost-efficient starting materials. However, the chemical semi-synthesis of Taxol is limited in efficiency and depends on highly pure starting materials as well as environmentally harmful catalysts and solvents^{54,55}. The sustainable biotechnological method developed here offers substantial advantages over semi-synthesis. It reduces the environmental impact of Taxol production by eliminating the reliance on petrochemicals and the use of catalysts and solvents. Furthermore, while chemical synthesis demands relatively pure starting materials, biotechnological production can utilize crude *Taxus* needle extracts, further lowering costs. Importantly, this method is non-destructive to the starting materials, allowing unused taxoids to be recycled through subsequent bioconversion rounds.

At this stage, the efficiency of baccatin III bioconversion to Taxol in *S. cerevisiae* remains low. Analysing the levels of pathway intermediates revealed that the steps downstream of C13 side-chain addition are not limiting, as introducing the benzoyl-CoA module in strain LT06 resulted in near-complete consumption of intermediates. Instead, a main bottleneck was identified in the early steps of the pathway leading to *N*-debenzoyl-2'-deoxy-taxol, making these steps critical targets for further optimization.

Among the early pathway enzymes, BAPT was identified as a major obstacle. While BAPT achieves 60% conversion of baccatin III

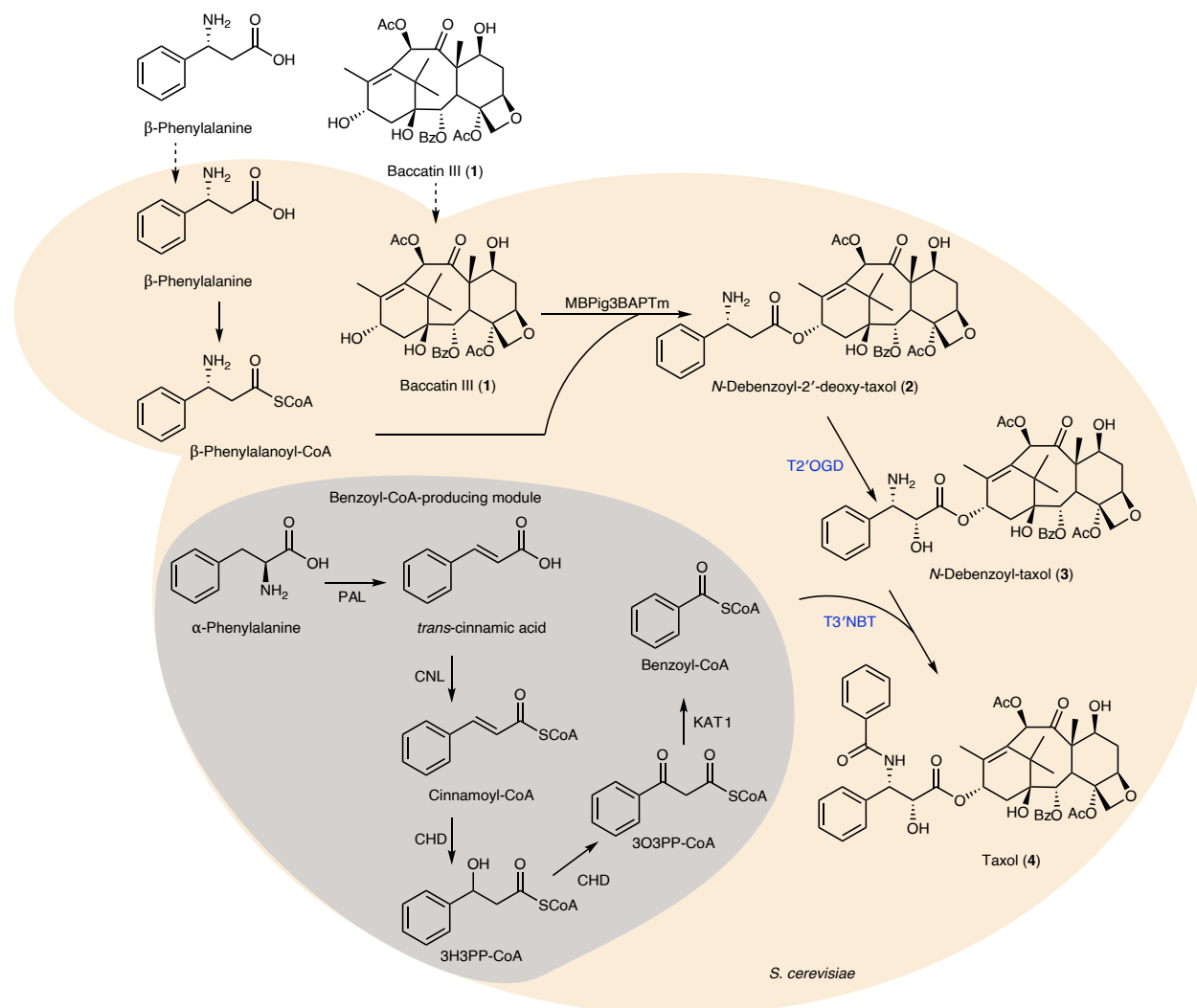


Fig. 6 | Biotechnological production of Taxol from baccatin III in yeast.

The final steps of Taxol biosynthesis were introduced in the base strain EGY48, and a five-step 'benzoyl-CoA production module' was incorporated to facilitate Taxol production. This module consisted of the enzymes *A. thaliana* PAL,

P. hybrida CNL, *P. hybrida* CHD and *P. hybrida* KAT1. MBPig3BAPTm indicates a fusion protein with a MBP fused at the N-terminus of the mutant form of BAPT. Substrates β-phenylalanine and baccatin III were supplied at 100 mg l⁻¹ each.

in tobacco, its performance in yeast is suboptimal. Enhancing BAPT stability and activity in yeast could involve exploring homologous enzymes from other *Taxus* species, employing alternative yeast or bacterial hosts that offer a more suitable environment for BAPT activity or applying structure-guided protein engineering. Another contributor to inefficiency may be the availability of β-phenylalanoyl-CoA. Owing to the poor efficiency of the two previously proposed CoA-ligases (TBPCCL and AAE16) in tobacco and yeast, we employed CoAL(A312G) as a surrogate enzyme. Although efficient in tobacco, CoAL(A312G) may lack the performance required for robust synthesis of *N*-debenzoyl-2'-deoxy-taxol in yeast. Thus, identifying an efficient enzyme from *Taxus* may be required for optimal pathway reconstruction in yeast. At the late stages of this revision, a pre-print reported a new candidate CoA-ligase²⁷. It would be valuable to test this new candidate in yeast and evaluate whether it can improve the critical C13 side-chain synthesis step.

In the later steps of the pathway, the reactions catalysed by T2'OGD and T3'NBT did not present bottlenecks at this stage. However, using yeast codon-optimized versions of these genes could further enhance the overall pathway efficiency. Additionally, benzoyl-CoA production was localized to the cytosol to couple *N*-debenzoyl-taxol synthesis with subsequent benzoylation. This approach avoids reliance on

peroxisomal export of benzoyl-CoA and has proven effective, providing sufficient benzoyl-CoA levels to support near-complete conversion of *N*-debenzoyl-taxol. However, as early pathway steps are improved, new limitations in benzoyl-CoA supply could emerge. In this case, alternative strategies, such as enzyme engineering, bacterial-derived benzoyl-CoA pathways or re-localizing key steps to the peroxisome, could be explored.

Overall, our results demonstrate the feasibility of producing Taxol by bioconversion of pathway intermediates. With further optimization of the limiting steps and the conditions of the bioconversion process, this biotechnological platform holds strong potential for achieving industrial-scale Taxol production.

Methods

Functional re-annotation of the *T. chinensis* var. *mairei* genome

The putative encoded proteins were retrieved from the genome of *T. chinensis* var. *mairei*²². Two paralleled functional re-annotations were implemented by reverse position-specific BLAST (RPS-BLAST) against the Conserved Domain Database (CDD; <https://www.ncbi.nlm.nih.gov/cdd/>) with e-value e⁻¹⁰ and HMMsearch⁵⁶ against Pfam⁵⁷ database with e-value e⁻⁵, respectively.

RNA-seq data analysis and co-expression network construction

Three sets of transcriptomes were retrieved from three independent previous studies, *Taxus chinensis* var. *mairei*²², *Taxus yunnanensis*³⁸ and *Taxus wallichiana*²¹, respectively, ending with 59 data sets in total. The chromosome-level genome of *Taxus chinensis* var. *mairei* was used as the reference genome. RNA-sequencing reads were mapped to the reference genome using HISAT v2.2.1⁵⁸, and the fragments per kilobase of transcript per million mapped reads value was calculated by using StringTie v2.2.1⁵⁹. Eleven characterized taxol biosynthetic genes (Supplementary Table 2) were used as baits, and 26,407 genes with Spearman correlation coefficient >0.6 were identified and extracted. The co-expression network of the 26,407 genes was later created using Cytoscape 3.10.0⁶⁰.

Transient gene expression in *N. benthamiana*

Vector pLIFE33n³⁵, which contains a USER⁶¹ cloning site, was used for transient gene expression in tobacco. Expression vectors were constructed by inserting the M24³⁶ promoter and gene of interest into pLIFE33n through USER cloning, utilizing the enzymes AsiI and Nb.BsmI. The candidate genes were synthesized by Twist Bioscience according to the cDNA sequence of *T. chinensis* var. *mairei* (Genebank [GCA_019776745.2](https://www.ncbi.nlm.nih.gov/nuclot/GCA_019776745.2)). USER cloning primers were synthesized from TAG Copenhagen A/S, Denmark, and their sequences are listed in Supplementary Data 1. Expression vectors were individually electro-transformed into *A. tumefaciens* strain AGL-1-GV3850. Agro-infiltration was carried out by following the previously published protocol, using different combinations of agrobacterium strains carrying the target gene of interest⁶².

Tobacco plants were cultivated in a greenhouse with a 16/8 h light/dark cycle for 4–6 weeks until they had five or six leaves. Two young leaves, the fourth one and the fifth one, of each plant were agro-infiltrated. When exogenous substrates were added, substrates were fed 48 h after agro-infiltration. Substrates were dissolved in 5% methanol/H₂O at 200 mg l⁻¹ each. Tobacco was grown for another 5 days before metabolite extraction from the agro-infiltrated leaves.

Gene expression and bioconversion in *S. cerevisiae*

EGY48^{44,45} yeast strain was used as the parental strain. Yeast codon-optimized synthetic genes for CoAL (A312G), BAPT, PAL, CNL, CHD and KAT1 were obtained from Thermo Fisher. The genes encoding the four enzymes of the benzoyl-CoA module, PAL, CNL, CHD and KAT1, were integrated into the XI-2 locus⁶³ of the yeast genome. Yeast transformation was carried out using a lithium acetate protocol⁶⁴. All the other genes were expressed from episomal vectors. For taxane production, seed cultures were inoculated into yeast synthetic media with the relevant autotrophic selections and cultivated overnight at 30 °C. The next day, overnight cultures were washed and re-suspended in galactose synthetic media (20 g l⁻¹ galactose and 10 g l⁻¹ raffinose) at an optical density of 600 nm (OD₆₀₀) of 0.5 and grown overnight at 20 °C. On the third day, the cultures were spun down and re-suspended in galactose-containing synthetic media, buffered with 200 mM 2-(*N*-morpholino)ethane sulfonic acid (pH 7) containing 100 mg l⁻¹ β-phenylalanine and 100 mg l⁻¹ baccatin III, and cultivated for another 3 days at 20 °C before metabolite extraction. The OD₆₀₀ of each engineered strain was monitored every day.

Metabolite extraction

Methanol was used to extract metabolites from tobacco leaves. Before methanol extraction, leaves were quickly frozen in liquid nitrogen and ground. Specifically, a leaf disc with diameter of 2 cm was extracted with 450 μl methanol. The extraction mixture was sonicated for 30 min in a water bath and spun down, and the supernatant was filtered through a 0.22-μm polyvinylidene difluoride filter before UPLC–qTOF–MS analysis. The same extraction method was used for preparing the *T. baccata* bark samples.

Ethyl acetate was used to extract metabolites from yeast cultures. Yeast pellets were collected by centrifugation and re-suspended in 1 ml MilliQ H₂O before the addition of 1 ml ethyl acetate. An aliquot of 200 mg 0.5 mm acid-washed glass beads (Mini-BeadBeater Glass Mill Beads, Cole-Parmer) was added to each sample to facilitate cell disruption through vortexing. The upper layer was collected for vacuum evaporation. After evaporation, the residues were re-suspended in methanol. Methanol-dissolved extracts were filtered through a 0.22-μm polyvinylidene difluoride filter before UPLC–qTOF–MS analysis.

UPLC–qTOF–MS analysis

Ultra-high-performance liquid chromatography with quadrupole time-of-flight high-resolution mass spectrometry analysis was carried out in a Dionex Ultimate 3000 quaternary rapid separation UPLC focused system (Thermo Fisher Scientific) connected to a Bruker Daltonics compact qTOF mass spectrometer equipped with an electrospray ionization interface (Bruker Daltonics). Electrospray ionization–MS settings were as follows: capillary voltage of 4,000 V, end-plate offset of –500 V, nebulizer pressure of 2.0 bar, drying gas flow of 8 l min⁻¹ and drying temperature of 220 °C. A Phenomenex Kinetex 1.7 μXB-C18 LC column (150 × 2.1 mm) was used for separation. Solvents A (H₂O acidified with 0.05% formic acid) and B (acetonitrile acidified with 0.05% formic acid) were used as mobile phases. The taxoids were detected in positive mode using the following separation programme: 0 min, 10% B; 27 min, 100% B; 32 min, 100% B; 33 min, 10% B; 38 min, 10% B. For α/β-phenylalanine and β-phenylalanine-CoA detection, MS data were acquired in negative mode. In this case, the separations were performed using the following gradient profile: 0 min, 2% B; 15 min, 100% B; 18 min, 100% B; 19 min, 2% B; 26 min, 10% B. A sodium formate solution was injected at the beginning of each chromatographic run as an internal standard. HyStar 3.2 software (Bruker Daltonics) was used for ultra-high-performance liquid chromatography–high-resolution mass spectrometry data acquisition, and Bruker Compass DataAnalysis 4.20 software was used for mass spectra and quantification.

Data analysis and illustrations

ChemDraw Professional 15.1 (PerkinElmer) was used to draw chemical structures. Microsoft Excel 2016 was used for bar charts, and Microsoft PowerPoint 2016 was used for the preparation of illustrations.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data necessary to interpret, verify and extend the research presented in the article are provided within the paper and in the Supplementary Information and Source data. The nucleotide sequences of T2'OGD and T3'NBT have been deposited in the National Center for Biotechnology Information GenBank under accessions [PQ015324](https://www.ncbi.nlm.nih.gov/nuclot/PQ015324) and [PQ015327](https://www.ncbi.nlm.nih.gov/nuclot/PQ015327), respectively. Databases and software used in genome functional re-annotation are available via the Conserved Domain Database at <https://www.ncbi.nlm.nih.gov/cdd/>, HMMsearch at <http://hmmer.org/> and the Pfam database at <http://pfam.xfam.org/>. Software used in RNA-seq data analysis and co-expression network construction are available at <https://daehwankimlab.github.io/hisat2/> (HISAT v2.2.1), <https://ccb.jhu.edu/software/stringtie/> (StringTie v2.2.1) and <https://cytoscape.org/> (Cytoscape 3.10.0). Source data are provided with this paper.

References

1. Model list of essential medicines: 21st list. World Health Organization <https://iris.who.int/handle/10665/325771> (2019).
2. Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer* **4**, 253–265 (2004).

3. Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & Mcphail, A. T. Plant antitumor agents. VI. The isolation and structure of Taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **93**, 2325–2327 (1971).
4. Wani, M. C. & Horwitz, S. B. Nature as a remarkable chemist: a personal story of the discovery and development of Taxol®. *Anticancer Drugs* **25**, 482 (2014).
5. Nicolaou, K. C. et al. Total synthesis of Taxol. *Nature* **367**, 630–634 (1994).
6. Holton, R. A. et al. First total synthesis of Taxol. 1. Functionalization of the B ring. *J. Am. Chem. Soc.* **116**, 1597–1598 (1994).
7. Flam, F. Race to synthesize Taxol ends in a tie. *Science* **263**, 911 (1994).
8. *Active Pharmaceutical Ingredients*. (PHARMA COMPASS, access 12 April 2025); <https://www.pharmacompass.com/active-pharmaceutical-ingredients>
9. Wildung, M. R. & Croteau, R. A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. *J. Biol. Chem.* **271**, 9201–9204 (1996).
10. Köksal, M., Jin, Y., Coates, R. M., Croteau, R. & Christianson, D. W. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **469**, 116–120 (2011).
11. Schoendorf, A., Rithner, C. D., Williams, R. M. & Croteau, R. B. Molecular cloning of a cytochrome P450 taxane 10 β -hydroxylase cDNA from *Taxus* and functional expression in yeast. *Proc. Natl Acad. Sci. USA* **98**, 1501–1506 (2001).
12. Jennewein, S., Rithner, C. D., Williams, R. M. & Croteau, R. B. Taxol biosynthesis: taxane 13 α -hydroxylase is a cytochrome P450-dependent monooxygenase. *Proc. Natl Acad. Sci. USA* **98**, 13595–13600 (2001).
13. Chau, M. D. & Croteau, R. Molecular cloning and characterization of a cytochrome P450 taxoid 2 α -hydroxylase involved in Taxol biosynthesis. *Arch. Biochem. Biophys.* **427**, 48–57 (2004).
14. Jennewein, S., Rithner, C. D., Williams, R. M. & Croteau, R. Taxoid metabolism: taxoid 14 β -hydroxylase is a cytochrome P450-dependent monooxygenase. *Arch. Biochem. Biophys.* **413**, 262–270 (2003).
15. Chau, M. D., Jennewein, S., Walker, K. & Croteau, R. Taxol biosynthesis: molecular cloning and characterization of a cytochrome P450 taxoid 7 β -hydroxylase. *Chem. Biol.* **11**, 663–672 (2004).
16. Hefner, J. et al. Cytochrome P450-catalyzed hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5 α -ol: the first oxygenation step in taxol biosynthesis. *Chem. Biol.* **3**, 479–489 (1996).
17. Chau, M., Walker, K., Long, R. & Croteau, R. Regioselectivity of taxoid-O-acetyltransferases: heterologous expression and characterization of a new taxadien-5 α -ol-O-acetyltransferase. *Arch. Biochem. Biophys.* **430**, 237–246 (2004).
18. Walker, K. & Croteau, R. Molecular cloning of a 10-deacetyl baccatin III-10-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97**, 583–587 (2000).
19. Walker, K. & Croteau, R. Taxol biosynthesis: molecular cloning of a benzoyl-CoA: taxane 2 α -O-benzoyltransferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97**, 13591–13596 (2000).
20. Walker, K. et al. Partial purification and characterization of acetyl coenzyme A: taxa-4(20),11(12)-dien-5 α -ol-O-acetyl transferase that catalyzes the first acylation step of taxol biosynthesis. *Arch. Biochem. Biophys.* **364**, 273–279 (1999).
21. Cheng, J. et al. Chromosome-level genome of Himalayan yew provides insights into the origin and evolution of the paclitaxel biosynthetic pathway. *Mol. Plant* **14**, 1199–1209 (2021).
22. Xiong, X. et al. The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nat. Plants* **7**, 1026–1036 (2021).
23. Zhang, Y. et al. Synthetic biology identifies the minimal gene set required for paclitaxel biosynthesis in a plant chassis. *Mol. Plant* **16**, 1951–1961 (2023).
24. Jiang, B. et al. Characterization and heterologous reconstitution of *Taxus* biosynthetic enzymes leading to baccatin III. *Science* **383**, 622–629 (2024).
25. Zhao, Y. et al. Oxetane ring formation in taxol biosynthesis is catalyzed by a bifunctional cytochrome P450 enzyme. *J. Am. Chem. Soc.* **146**, 801–810 (2024).
26. Yang, C. et al. Biosynthesis of the highly oxygenated tetracyclic core skeleton of Taxol. *Nat. Commun.* **15**, 2339 (2024).
27. McClune, C. J. et al. Multiplexed perturbation of yew reveals cryptic proteins that enable a total biosynthesis of baccatin III and Taxol precursors. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.11.06.622305> (2024).
28. Walker, K. D., Klettke, K., Akiyama, T. & Croteau, R. Heterologous expression, and characterization of a phenylalanine aminomutase involved in Taxol biosynthesis. *J. Biol. Chem.* **279**, 53947–53954 (2004).
29. Ramírez-Estrada, K. et al. Transcript profiling of jasmonate-elicited *Taxus* cells reveals a β -phenylalanine-CoA ligase. *Plant Biotechnol. J.* **14**, 85–96 (2016).
30. Long, R. M., Lagisetti, C., Coates, R. M., Croteau, R. B. & Croteau, R. Specificity of the *N*-benzoyl transferase responsible for the last step of Taxol biosynthesis. *Arch. Biochem. Biophys.* **477**, 384–389 (2008).
31. Walker, K., Long, R. & Croteau, R. The final acylation step in Taxol biosynthesis: cloning of the taxoid C13-side-chain *N*-benzoyltransferase from *Taxus*. *Proc. Natl Acad. Sci. USA* **99**, 9166–9171 (2002).
32. Sanchez-Muñoz, R. et al. A novel hydroxylation step in the taxane biosynthetic pathway: a new approach to paclitaxel production by synthetic biology. *Front. Bioeng. Biotechnol.* **8**, 410 (2020).
33. Walker, K., Fujisaki, S., Long, R. & Croteau, R. Molecular cloning and heterologous expression of the C13 phenylpropanoid side chain-CoA acyltransferase that functions in Taxol biosynthesis. *Proc. Natl Acad. Sci. USA* **99**, 12715–12720 (2002).
34. Zhao, Y. et al. Biosynthesis and biotechnological production of the anti-obesity agent celastrol. *Nat. Chem.* **15**, 1236–1246 (2023).
35. Forman, V. et al. A gene cluster in *Ginkgo biloba* encodes unique multifunctional cytochrome P450s that initiate ginkgolide biosynthesis. *Nat. Commun.* **13**, 5143 (2022).
36. Dey, N. & Maiti, I. B. Structure and promoter/leader deletion analysis of mirabilis mosaic virus (MMV) full-length transcript promoter in transgenic plants. *Plant Mol. Biol.* **40**, 771–782 (1999).
37. Koetsier, M. J., Jekel, P. A., Wijma, H. J., Bovenberg, R. A. L. & Janssen, D. B. Aminoacyl-coenzyme A synthesis catalyzed by a CoA ligase from *Penicillium chrysogenum*. *FEBS Lett.* **585**, 893–898 (2011).
38. He, C. T. et al. Transcriptome profiling reveals specific patterns of paclitaxel synthesis in a new *Taxus yunnanensis* cultivar. *Plant Physiol. Biochem.* **122**, 10–18 (2018).
39. Liao, W. et al. Transcriptome assembly and systematic identification of novel cytochrome P450s in *Taxus chinensis*. *Front. Plant Sci.* **8**, 1468 (2017).
40. Kawai, Y., Ono, E. & Mizutani, M. Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* **78**, 328–343 (2014).
41. Martinez, S. & Hausinger, R. P. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).
42. Markolovic, S., Wilkins, S. E. & Schofield, C. J. Protein hydroxylation catalyzed by 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **290**, 20712–20722 (2015).

43. Li, B. J. et al. Improving 10-deacetylbaaccatin III-10- β -O-acetyltransferase catalytic fitness for Taxol production. *Nat. Commun.* **8**, 15544 (2017).
44. Gyuris, J., Golemis, E., Chertkov, H. & Brent, R. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* **75**, 791–803 (1993).
45. Golemis, E. A. & Khazak, V. Alternative yeast two-hybrid systems. The interaction trap and interaction mating. *Methods Mol. Biol.* **63**, 197–218 (1997).
46. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**, 3098–3100 (2017).
47. Nallamsetty, S. & Waugh, D. S. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **45**, 175–182 (2006).
48. Fox, J. D., Kapust, R. B. & Waugh, D. S. Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins. *Protein Sci.* **10**, 622–630 (2001).
49. Klempien, A. et al. Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. *Plant Cell* **24**, 2015–2030 (2012).
50. Qualley, A. V., Widhalm, J. R., Adebesein, F., Kish, C. M. & Dudareva, N. Completion of the core β -oxidative pathway of benzoic acid biosynthesis in plants. *Proc. Natl Acad. Sci. USA* **109**, 16383–16388 (2012).
51. Van Moerkercke, A., Schauvinhold, I., Pichersky, E., Haring, M. A. & Schuurink, R. C. A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. *Plant J.* **60**, 292–302 (2009).
52. Wanner, L. A., Li, G., Ware, D., Somssich, I. E. & Davis, K. R. The phenylalanine ammonia-lyase gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* **27**, 327–338 (1995).
53. Kingston, D. G. I. The shape of things to come: structural and synthetic studies of taxol and related compounds. *Phytochemistry* **68**, 1844–1854 (2007).
54. Zhang, S. et al. Research advances in clinical applications, anticancer mechanism, total chemical synthesis, semi-synthesis and biosynthesis of paclitaxel. *Molecules* **28**, 7517 (2023).
55. Mountford, P. G. The Taxol® Story—Development of a Green Synthesis via Plant Cell Fermentation. in *Green Chemistry in the Pharmaceutical Industry*. (eds Dunn, P. J., et al) Ch. 7 (Wiley-VCH, 2010).
56. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, 1002195 (2011).
57. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
58. Zhang, Y., Park, C., Bennett, C., Thornton, M. & Kim, D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.* **31**, 1290–1295 (2021).
59. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
60. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498 (2003).
61. Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T. & Halkier, B. A. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* **35**, 55 (2007).
62. Bach, S. S. et al. High-throughput testing of terpenoid biosynthesis candidate genes using transient expression in *Nicotiana benthamiana*. *Methods Mol. Biol.* **1153**, 245–255 (2014).
63. Flagfeldt, D. B., Siewers, V., Huang, L. & Nielsen, J. Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast* **26**, 545–551 (2009).
64. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).

Acknowledgements

We are grateful to F. Geu-Flores and I. Pateraki (University of Copenhagen) for critically reading the manuscript. We thank D. Staerk and L. Kjærulff for their support in running nuclear magnetic resonance analyses. We thank J. Olsen and M. Ramirez (University of Copenhagen) for their assistance in running analytical instruments. This work was financially supported by Independent Research Fund Denmark grant no. 0136-00410B (to S.C.K.) and Novo Nordisk Foundation grants NNF19OC0055204 (to S.C.K.) and NNF20OC0064680 (to F.L.).

Author contributions

S.C.K. conceived the study. F.L., Y.Z. and S.C.K. designed experiments. F.L., Y.X. and Y.Z. conducted all cloning, engineering of *S. cerevisiae* and chromatographic data analysis. F.L., Y.X. and Y.Z. performed *N. benthamiana* and yeast cultivation and extraction. F.L., Y.X. and Y.Z. conducted the isolation of taxoids from yeast cultures and *N. benthamiana* and performed the nuclear magnetic resonance analysis. M.S.M. performed the chemical synthesis of taxoids used as standards in this study. C.Z. performed the genome re-annotation and co-expression analysis. F.L., Y.Z. and S.C.K. wrote the paper. All authors read and commented on the final version of the paper.

Competing interests

F.L., Y.X., C.Z., Y.Z. and S.C.K. are co-inventors on two patent applications disclosing the use of T2'OGD and T3'NBT for the biotechnological production of paclitaxel (PCT/EP2024/088095 and EP24386083.0). M.S.M. declares no commercial or non-commercial interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44160-025-00800-z>.

Correspondence and requests for materials should be addressed to Sotirios C. Kampranis.

Peer review information *Nature Synthesis* thanks Vincent Courdavault, Leonardo Rios-Solis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Thomas West, in collaboration with the *Nature Synthesis* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Topspin ver. 4.0 (Bruker Biospin, Karlsruhe) for NMR data acquisition; HyStar 3.2 (Bruker) for UPLC-HRMS data acquisition.
Data analysis	Bruker Compass DataAnalysis 4.20 was used for mass spectra and quantification analysis; ChemDraw Professional 15.1 (PerkinWlmer) was used to draw chemical structures; Microsoft Excel 2016 was used for bar charts and graphs and Microsoft PowerPoint 2016 was used for the preparation of illustrations; Topspin ver.4.0.6 was used for NMR spectra analyses. For RNA-seq data analysis and co-expression network construction the following software was used: HISAT2 (https://daehwankimlab.github.io/hisat2/), StringTie (https://ccb.jhu.edu/software/stringtie/), and Cytoscape (https://cytoscape.org/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data necessary to interpret, verify, and extend the research presented in the article are provided within the paper and in the Supplementary Information and Datasets. The nucleotide sequences of T2'OGD and T3'NBT have been deposited in the NCBI GenBank under accessions PQ015324 and PQ015327, respectively. A reporting summary for this article is available as a Supplementary Information file. Source data are provided as a Source Data file. Databases and softwares used in functional reannotation of genome can be found through the following links: Conserved Domain Database (CDD, <https://www.ncbi.nlm.nih.gov/cdd/>), HMMsearch(<http://hmmer.org/>), and Pfam database (<http://pfam.xfam.org/>). Software used in RNA-seq data analysis and co-expression network construction can be found through the following links: HISAT v2.2.1 (<https://daehwankimlab.github.io/hisat2/>), StringTie v2.2.1 (<https://ccb.jhu.edu/software/stringtie/>), and Cytoscape 3.10.0 (<https://cytoscape.org/>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Predetermination of sample size was not relevant to this study. Based on previous experience with the yeast assays performed, replication consisting of three biological replicates was sufficient to provide statistically reliable data (e.g., Ignea, C. et al. Orthogonal monoterpene biosynthesis in yeast constructed on an isomeric substrate. Nat. Commun. 10, 3799 (2019)).
Data exclusions	No data exclusion.
Replication	All determinations were repeated in triplicates. All attempts at replication of the experiments using re-establishing of yeast strains or re-infiltration of tobacco plants and reanalysis of the products have been successful.
Randomization	For each of the constructed yeast strains containing high copy plasmids (each biological replicate), 20-30 colonies from the same plate were randomly selected, pooled together, and analyzed. For yeast strains engineered via chromosome integration, 3 different colonies were selected randomly and analyzed. No further allocation into experimental groups was carried out and control of covariates was not relevant.
Blinding	Blinding was not relevant for data collection because the type of analysis carried out could not be biased by lack of blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input type="checkbox"/>	<input checked="" type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	anti-c-myc (9E10), catalog number: sc-40 (Santa Cruz Biotechnology)
Validation	Validation details can be found at: https://www.scbt.com/p/c-myc-antibody-9e10

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Public health
<input checked="" type="checkbox"/>	<input type="checkbox"/> National security
<input checked="" type="checkbox"/>	<input type="checkbox"/> Crops and/or livestock
<input checked="" type="checkbox"/>	<input type="checkbox"/> Ecosystems
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Demonstrate how to render a vaccine ineffective
<input checked="" type="checkbox"/>	<input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent
<input checked="" type="checkbox"/>	<input type="checkbox"/> Increase transmissibility of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Alter the host range of a pathogen
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable evasion of diagnostic/detection modalities
<input checked="" type="checkbox"/>	<input type="checkbox"/> Enable the weaponization of a biological agent or toxin
<input checked="" type="checkbox"/>	<input type="checkbox"/> Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a