

Biosynthesis of the highly oxygenated tetracyclic core skeleton of Taxol

Received: 13 December 2023

Accepted: 3 March 2024

Published online: 15 March 2024

Chengshuai Yang^{1,6}, Yan Wang^{1,6}, Zhen Su^{1,2,6}, Lunyi Xiong^{1,2}, Pingping Wang¹, Wen Lei³, Xing Yan¹✉, Dawei Ma⁴✉, Guoping Zhao^{1,5}✉ & Zhihua Zhou¹✉

Taxol is a widely-applied anticancer drug that inhibits microtubule dynamics in actively replicating cells. Although a minimum 19-step biosynthetic pathway has been proposed and 16 enzymes likely involved have been characterized, stepwise biosynthetic reactions from the well-characterized di-oxygenated taxoids to Taxol tetracyclic core skeleton are yet to be elucidated. Here, we uncover the biosynthetic pathways for a few tri-oxygenated taxoids via confirming the critical reaction order of the second and third hydroxylation steps, unearth a taxoid 9 α -hydroxylase catalyzing the fourth hydroxylation, and identify CYP725A55 catalyzing the oxetane ester formation via a cascade oxidation-concerted acyl rearrangement mechanism. After identifying a acetyltransferase catalyzing the formation of C7-OAc, the pathway producing the highly-oxygenated 1 β -dehydroxybaccatin VI with the Taxol tetracyclic core skeleton is elucidated and its complete biosynthesis from taxa-4(20),11(12)-diene-5 α -ol is achieved in an engineered yeast. These systematic studies lay the foundation for the complete elucidation of the biosynthetic pathway of Taxol.

Taxol (paclitaxel), belonging to the diterpenoid alkaloid, was first obtained from the bark of *Taxus brevifolia*¹ and, so far, has been one of the most potent and widely applied anticancer drugs². During the past two decades, enormous efforts including chemical total synthesis^{3–7}, biosynthesis using cell suspension cultures of *Taxus* species⁸, and semisynthetic route using baccatin III or 10-deacetyl baccatin III isolated from needles of *Taxus* species trees as starting material⁹, have been made to relieve the short supply of Taxol due to its low accumulation in *Taxus* plants (~0.01% dry weight)¹⁰. Meanwhile, due to the complexity of biogenesis for its special core skeleton expansion and various modifications including oxidation and acylation, elucidation of Taxol biosynthetic pathway and revealing its key enzymatic reaction mechanisms have been an attractive challenge for both academia and industry.

The biosynthesis of Taxol is presumed to originate from the cyclization of geranylgeranyl diphosphate (GGPP) (the common precursor of diterpenoids). By incubating tritium-labeled GGPP with a *Taxus* bark extract, radiochemical taxa-4(5),11(12)-diene (**1**) was produced¹¹. The incubation of tritium-labeled **1** with *Taxus* stem sections resulted in the production of radiochemical 10-deacetylbaccatin III and taxol, demonstrating that **1** should be the key diterpenoid core skeleton of Taxol¹¹. Besides, the feeding tritium-labeled taxa-4(20),11(12)-diene-5 α -ol (**2**) has also been testified to be converted to 10-deacetylbaccatin III and Taxol in *Taxus* plant tissue¹². Based on these above studies, **2** has been popularly proposed as the mono-oxygenated intermediate of Taxol^{13–15}.

Over the years, numerous efforts were also made in employing enzymatic (or chemical) catalysis to synthesize the probable

¹Key Laboratories of Plant Design and Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China. ²University of Chinese Academy of Sciences, Beijing, China. ³Shanghai Research Institute of Chemical Industry, Shanghai, China. ⁴State Key Laboratory of Chemical Biology, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, China. ⁵Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. ⁶These authors contributed equally: Chengshuai Yang, Yan Wang, Zhen Su. ✉e-mail: yanxing@cemps.ac.cn; madw@mail.sioc.ac.cn; gpzhao@sibs.ac.cn; zhouzhihua@cemps.ac.cn

intermediates (or related functional groups) which were naturally identified or proposed to be related to the Taxol biosynthetic pathway. To this date, at least 19 catalyzing steps including minimum nine oxidation reactions mainly mediated by cytochrome P450s (P450s) and five acylation modification steps catalyzed by acyltransferases were proposed to involve in this pathway¹⁶. Of which, the P450s for the hydroxylation of C5, C10, C13, C7, and C2^{17–21}, the acyltransferases for the acylation of C5-OH, C10-OH, C2-OH, and C13-OH^{22–27}, as well as the enzymes for the C2' hydroxylation and C3'-NH₂ acylation of the β -phenylalanoyl-modified side chain^{28–30}, have been cloned and characterized over the years (Fig. 1, Supplementary Table 1, and Supplementary Data 1).

The first breakthrough in rebuilding the de novo biosynthetic pathway of Taxol was to synthesize **1** from GGPP catalyzed by taxa-4(5),11(12)-diene synthase (TS) in a recombinant *E. coli* strain¹⁴. Employing the same strategy of metabolic engineering, the mono-oxygenated taxoid **2** was synthesized from **1** catalyzed by taxoid 5 α -hydroxylase (T5OH)¹⁴. Later, 5 α -acetytaxa-4(20),11(12)-diene-10 β -ol (**4**), a di-oxygenated taxoid, was produced in cell factories with addition of C5 acetylation catalyzed by a taxoid 5-O-acetyltransferase (T5AT) followed by C10 oxidation catalyzed by a taxoid 10 β -hydroxylase (T10OH)³¹. However, no any downstream intermediate from **4** has ever been reported since then. The expected combinatorial modification using T10OH and taxoid 13 α -hydroxylase (T13OH) to produce the hypothetical intermediates, in particular the tri-

oxygenated taxoids with oxidation at C5, C10, and C13, was unsuccessful^{16,19}. On the other hand, although the formation of di-oxygenated taxoid taxa-4(20),11(12)-diene-5 α ,13 α -diol (**5**) from intermediate **2** catalyzed by T13OH was reported as early as in 2001¹⁹, no further oxidized modification of **5** has ever been published. Therefore, biosynthetic pathways even for the tri-oxygenated taxoids need to be further clarified or elucidated.

The biosynthesis of the tetra-oxygenated taxoids is proposed to be initiated via hydroxylation modification at C9 catalyzed by taxoid 9 α -hydroxylase. Although a putative cDNA clone for taxoid 9 α -hydroxylase was claimed¹⁵, neither its sequence nor further characterization has ever been reported. Besides, hydroxylation modifications at C2, C7, and C1 were assumed to happen in the later steps for the synthesis of the highly oxygenated intermediates of Taxol¹⁵. The characterized taxoid 2 α -hydroxylase (T2OH) and 7 β -hydroxylase (T7OH) were known to hydroxylate a tetra-oxygenated taxoid taxusin to produce 2 α - and 7 β -hydroxytaxusin, respectively, and further, reciprocally, catalyze the hydroxylation of the corresponding 7 β - and 2 α -hydroxytaxusin to form the intermediate hexa-oxygenated 2 α ,7 β -dihydroxytaxusin (**13**) (Fig. 1)^{20,21}. However, taxusin and its derivatives (7 β -hydroxytaxusin, 2 α -hydroxytaxusin and 2 α ,7 β -dihydroxytaxusin), have never been regarded as intermediates in the previously proposed biosynthetic pathway of Taxol^{20,21}. Thus, the enzymes responsible for C1 and C9 hydroxylation modifications, and the reaction order of hydroxylation modifications at C2, C7, and C1 in the biosynthesis of

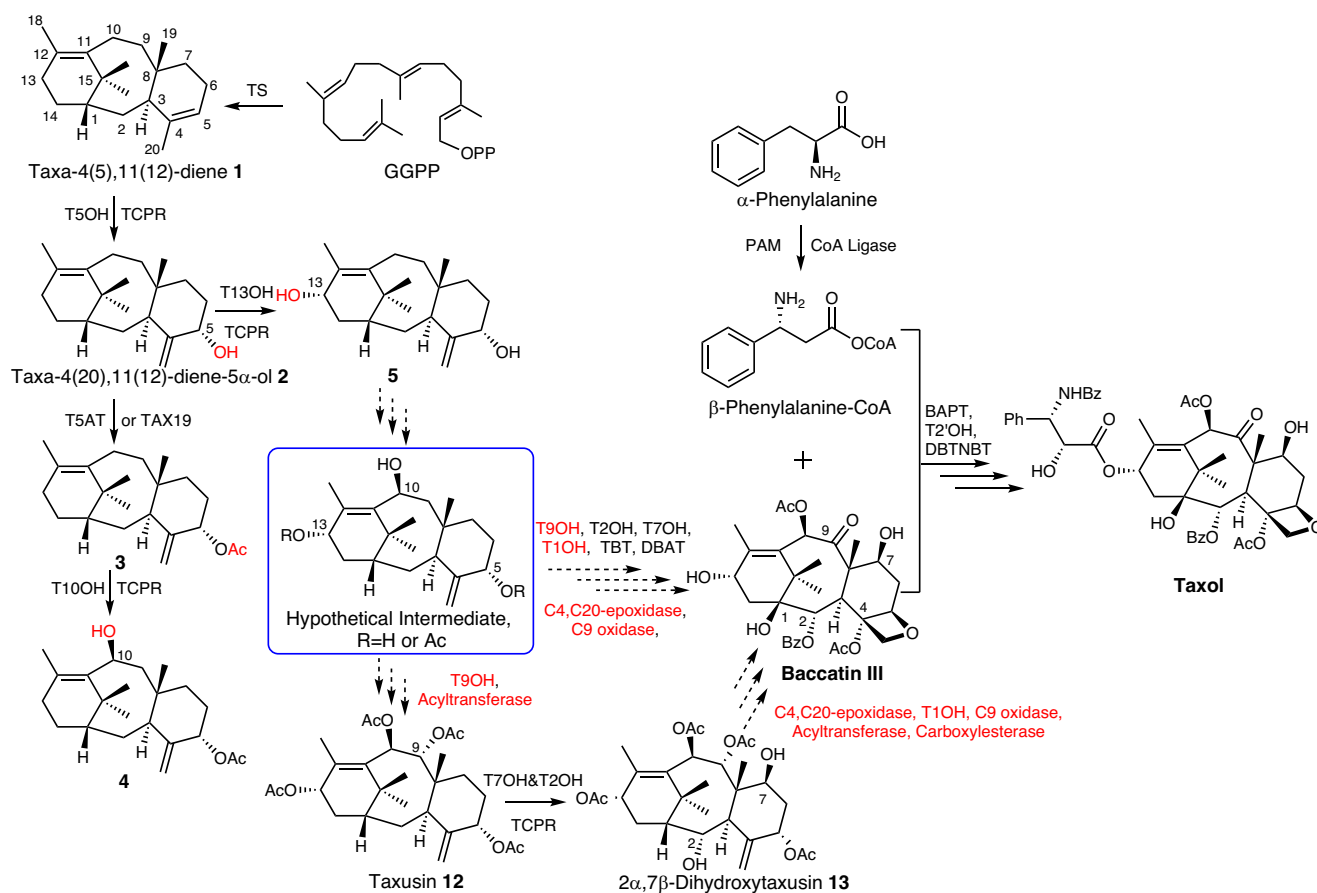


Fig. 1 | Summary of previously reported and supposed reactions involved in the biosynthesis of taxoids. GGPP geranylgeranyl diphosphate, TS taxa-4(5),11(12)-diene synthase, T5OH taxoid 5 α -hydroxylase, TCPR *Taxus* cytochrome P450 reductase, T5AT and TAX19 taxoid 5-O-acetyltransferase, T13OH taxoid 13 α -hydroxylase, T10OH taxoid 10 β -hydroxylase, T9OH taxoid 9 α -hydroxylase, T2OH taxoid 2 α -hydroxylase, T7OH taxoid 7 β -hydroxylase, T1OH taxoid 1 β -hydroxylase, TBT taxoid 2 α -O-benzoyltransferase, DBAT 10-deacetylbaccatin 10 β -O-

acetyltransferase, BAPT 13-O-(3-amino-3-phenylpropanoyl) transferase, T2'OH taxoid 2'-hydroxylase, DBTNBT *N*-debenzoyl-2'-deoxytaxol *N*-benzoyltransferase, PAM phenylalanine aminomutase. Dashed lines represent the proposed steps and unknown enzymes were marked in red. New formed groups in compounds were marked with red. Tri-oxygenated taxoids were speculated as the hypothetical intermediates and hydroxylation modification at C9 catalyzed by T9OH was proposed as the fourth oxidation reaction step in the biosynthetic pathway of Taxol.

Taxol were unclear. In addition, the key step for Taxol biosynthesis, i.e., the formation of Taxol tetracyclic core skeleton with an oxetane ester is unknown although this core skeleton was proposed to form following the formation of an acetylated taxadiene-2,5,7,9,10,13-hexaol intermediate before the formation of baccatin III³², which was employed as the precursor for the industrial semi-synthesis of Taxol⁹.

In this work, we re-characterize the enzymes, the previously reported P450s, i.e., T10OH¹⁸ and T13OH¹⁹, as well as acetyltransferases T5AT²² and TAX19³³, via substrate feeding of *Saccharomyces cerevisiae* cells expressing corresponding genes (in vivo reaction) and substrate incubation with crude protein lysate from the corresponding *S. cerevisiae* cells (in vitro reaction). With further identification of several intermediates including di-oxygenated taxoids 5 α -acetoxytaxa-4(20),11(12)-diene-13 α -ol (**6**) and diacetyl taxoid 5 α ,13 α -diacetoxytaxa-4(20),11(12)-diene (**9**), the first three hydroxylation steps are uncovered, leading to the formation of tri-oxygenated taxoid 5 α ,13 α -diacetoxytaxa-4(20),11(12)-diene-10 β -ol (**10**). Furthermore, from a *Taxus* cDNA library, we discover two CYP450 genes, of which, CYP725A37 encoded taxoid 9 α -hydroxylase catalyzes the fourth hydroxylation step, forming a tetra-oxygenated taxoid (9,10-deacetyl taxusin, **11**). The CYP725A55 catalyzes 2 α -benzoyloxy-7 β -acetoxytaxusin (**15**) to form the Taxol tetracyclic core skeleton of 1 β -dehydroxybaccatin VI (**16**) via a cascade oxidation-concerted acyl rearrangement process. Based on these results, the complete biosynthetic pathway of 1 β -dehydroxybaccatin VI (**16**), the highly oxygenated Taxol intermediate sharing the same Taxol tetracyclic core skeleton as that of baccatin III and Taxol, is elucidated and its biosynthesis from **2** is further confirmed in engineered yeast.

Results

Biosynthesis of tri-oxygenated taxoids

The hydroxylation modifications at C10 and C13 of the mono-oxygenated taxoid taxa-4(20),11(12)-diene-5 α -ol (**2**) were predicted as the second or third oxidation steps in the biosynthetic pathway of Taxol¹⁵. However, the hydroxylation modification of the di-oxygenated taxoid 5 α -acetoxytaxa-4(20),11(12)-diene-10 β -ol (**4**) seems to come to a dead end and the research progress on hydroxylation and acylation modifications of the alternative di-oxygenated taxoid taxa-4(20),11(12)-diene-5 α ,13 α -diol (**5**) was limited. Specifically, there are six potential combined reaction orders (A-F) to synthesize tri-oxygenated taxoids from **2** catalyzed by T5AT, T10OH and T13OH (Supplementary Table 2). Previous studies demonstrated that the efforts to produce hypothetical tri-oxygenated taxoids via these potential reaction orders were either unsuccessful for A, C, D, and F^{16,19} or hard for B^{16,19}, probably due to the conversion rate of T13OH towards 5 α -acetoxytaxa-4(20),11(12)-diene (**3**) was too low to identify the structure of the resulted product. Therefore, we preferentially tried the untested route E.

Co-expression of T13OH and TCPR encoding a cytochrome P450 reductase from *Taxus* in *S. cerevisiae* resulted in the formation of a metabolite in both in vivo and in vitro reactions using **2** as the substrate (Supplementary Fig. 1a–c, e). This metabolite was concluded as **5** with a hydroxyl group at C13 based on spectra analyses including high-resolution electrospray ionization mass spectrometry (HR-ESIMS) and nuclear magnetic resonance (NMR) spectroscopy including ¹H, ¹³C, etc (Supplementary Fig. 1d, f, g). These above results confirmed that T13OH could efficiently install a hydroxyl group at C13 position of the mono-oxygenated taxoid **2** to produce the di-oxygenated taxoid **5**. Next, T5AT catalyzed **5** to produce a metabolite assigned as an acetylated derivative 5 α -acetoxytaxa-4(20),11(12)-diene-13 α -ol (**6**) (Supplementary Fig. 2), indicating that T5AT transferred an acetyl group towards the C5-OH of **5**. Later, the co-expression of T10OH and TCPR in *S. cerevisiae* and feeding the di-oxygenated taxoid **6** resulted in a compound which was confirmed to be a tri-oxygenated taxoid 5 α -acetoxytaxa-4(20),11(12)-diene-10 β ,13 α -diol (**7**) (Supplementary Fig. 3),

suggesting T10OH could install a hydroxyl group at C10 position of **6**. These results confirmed that the reaction order E is a practicable biosynthetic route for converting **2** to **7** through C13 hydroxylation, C5-OH acetylation, and C10 hydroxylation in three sequentially steps. The previous unsuccessful combination F was demonstrated to be accessible when using another acetyltransferase TAX19 instead of T5AT but maybe regarded as minor pathway in the biosynthesis of **7** from **5** because of the low conversion rate of TAX19 towards **8** (Supplementary Fig. 4).

Besides, taking consideration of the fact that TAX19 prefers to add acetyl group towards the east-west pole positions at C5-OH and C13-OH of taxoids³³, we employed **6** as the substrate for both in vivo and in vitro reactions using *S. cerevisiae* strain expressing TAX19. A metabolite was produced and further assigned as a diacetyl taxoid, 5 α ,13 α -diacetoxytaxa-4(20),11(12)-diene (**9**) (Supplementary Fig. 5). This result revealed TAX19 could add an acetyl group towards the C13-OH of **6**. When the di-oxygenated taxoid **9** was employed as the substrate for both in vivo and in vitro reactions with a *S. cerevisiae* strain expressing T10OH and TCPR, the consumption of **9** and formation of another metabolite were observed (Supplementary Fig. 6a, b). This metabolite was further assigned as the tri-oxygenated taxoid 5 α ,13 α -diacetoxytaxa-4(20),11(12)-diene-10 β -ol (**10**), suggesting T10OH introduces a hydroxyl group at C10 position of **9** (Supplementary Fig. 6c–e). Another taxoid 10 β -hydroxylase (T10OH2) exhibited the similar function as T10OH (Supplementary Fig. 7).

We realized the biosynthesis of a few tri-oxygenated taxoids from **2** via re-characterization of the known enzymes and confirming the second and third oxygenated reaction order (Fig. 2a). Besides, to the best of our knowledge, compounds **6**, **7**, and **10** in biosynthetic routes of tri-oxygenated taxoids have never been reported previously (Supplementary Table 3).

Taxoid 9 α hydroxylase for tetra-oxygenated taxoids

The tetra-oxygenated taxoids 9,10-deacetyl taxusin (**11**) and taxusin (**12**) have been isolated from *Taxus* plants previously³⁴. However, their biosynthetic pathways are still unknown. Compared with the structure of **10**, the C9 position was oxidized in both **11** and **12**, which was proposed to be catalyzed by a taxoid 9 α -hydroxylase belonging to CYP725A subfamily¹⁶. In order to characterize the functions of CYP725As, we tried to screen the targeted taxoid 9 α -hydroxylase through in vivo reactions using the *S. cerevisiae* strains harboring a series of uncharacterized CYP725As. A cDNA database based on *Taxus* EST datasets collected from NCBI and RNA-seq datasets generated in this work was established. 26 uncharacterized CYP725As from the cDNAs of *Taxus* twigs were respectively co-expressed with TCPR in *S. cerevisiae*. The co-expression of CYP725A37 and TCPR in *S. cerevisiae* resulted in the formation of a metabolite in both in vivo and in vitro reactions using **10** as the substrate (Fig. 2b and Supplementary Fig. 8a). According to HR-ESIMS and NMR spectroscopy, the structure of this metabolite was elucidated as a tetra-oxygenated taxoid, 9,10-deacetyl taxusin (**11**), which is the C9 hydroxylation product of **10** (Fig. 2c, d and Supplementary Fig. 8b). These results confirmed CYP725A37 was a taxoid 9 α -hydroxylase to install a hydroxyl group at C9 position of taxoids. Based on the identified taxoid 9 α -hydroxylase CYP725A37, the biosynthetic pathway from **2** to **11** could be elucidated involving five sequentially steps including C13 hydroxylation, C5 acetylation, C13 acetylation, C10 hydroxylation and C9 hydroxylation (Fig. 2).

Compared with the structure of **11**, **12** has the same oxidation level but its C9-OH and C10-OH were both acetylated. Thus, the biosynthesis of **12** could be realized via the further acylation of intermediate **11**. T5AT has been demonstrated to preferentially acetylate the “northern” hemisphere hydroxyls at C9 and C10 towards tetra-oxygenated taxoids previously³³. When **11** was employed as the substrate for both in vivo and in vitro reactions with the *S. cerevisiae* strain expressing T5AT, HPLC analysis demonstrated that **12** was formed, accompanied with

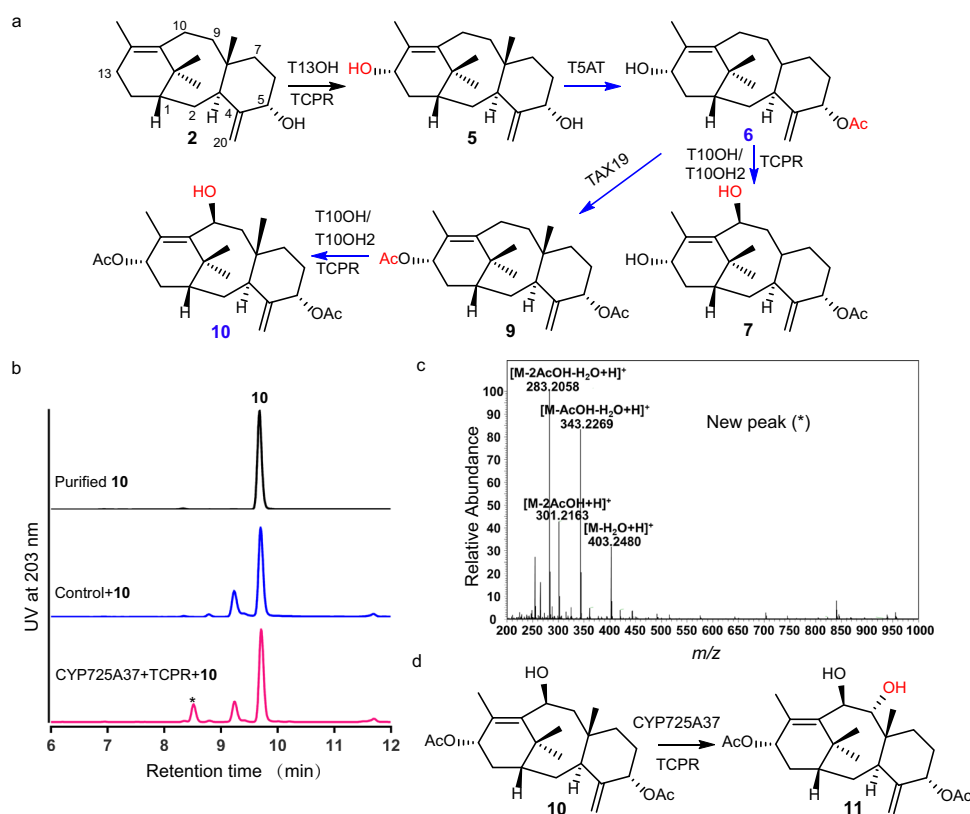


Fig. 2 | The biosynthetic pathways of tri-oxygenated and tetra-oxygenated taxoids and characterization of the taxoid 9 α hydroxylase. **a** The biosynthetic pathways of tri-oxygenated taxoids. Reaction steps and new compounds characterized in this study were marked in blue. New formed groups in compounds were marked with red. **b** HPLC analysis of in vivo reaction for *S. cerevisiae* strain YCYP725A37 expressing CYP725A37 and TCPR using 5 α ,13 α -diacetoxytaxa-

4(20),11(12)-diene-10 β -ol (**10**) as the substrate. Substrate was marked with compound number and product was marked with an asterisk. *S. cerevisiae* strain YTCPR expressing TCPR was used as the control. **c** Mass spectra (ESI) of new peak (*). These experiments were performed three times with similar results each time. **d** The C9 hydroxylation reaction catalyzed by CYP725A37 to produce 9,10-deacetyl taxusin (**11**) from **10**. Source data are provided as a Source Data file.

proposed monoacetylated **11** at C9-OH or C10-OH (Supplementary Fig. 8c–g), confirming T5AT could add acetyl groups towards the C9-OH and C10-OH of **11** to form **12**. Besides, we testified that T7OH and T2OH could convert **12** to yield a hexa-oxygenated taxoid 2 α ,7 β -dihydroxytaxusin (**13**), which has been reported in a previous study²⁰. Thus, we could elucidate a complete biosynthetic pathway of **13** from **2** which involves eight sequentially steps based on the confirmed biosynthetic pathway of tri-oxygenated taxoids, unearthed taxoid 9 α -hydroxylase and re-characterization of the reported enzymes.

Formation of Taxol tetracyclic core skeleton

The hexa-oxygenated intermediate **13** lacks the Taxol tetracyclic core skeleton with an oxetane ester in baccatin III and Taxol as well as their derivatives. P450 was suggested to participate in the formation of oxetane ester in the Taxol tetracyclic core skeleton, especially for the formation of a proposed 4 β , 20-epoxide intermediate in a plausible mechanism for the oxetane ester formation³². However, no enzyme has been reported to catalyze the relative reaction until now. According to the structures of taxoids with the oxetane ester group isolated from *Taxus* plants, we speculated that the preferential substrate for the oxetane ester formation might be 2 α -benzoyloxy-7 β -acetoxytaxusin (**15**) and then semi-synthesized **15** from **13** (Supplementary Fig. 9a, b, g), which was used as the substrate to screen the left uncharacterized CYP725As through in vivo reaction. The co-expression of CYP725A55 and TCPR in *S. cerevisiae* resulted in the formation of two metabolites (Peaks A and B) (Fig. 3a). The structure of one metabolite (peak A) was deduced as 1 β -dehydroxybaccatin VI (**16**) based on its 1D NMR data, being consistent with the related NMR

spectrum of **16** reported in literature³⁵ (Fig. 3a, b, e and Supplementary Fig. 10e). The other metabolite (peak B) with the same molecular formula as **16**, was assigned as a C4-C20 epoxide derivative 2-deacetyl-2 α -benzoylbaccatin I (**17**) based on the HR-ESIMS and NMR spectroscopy (Fig. 3a, c, e and Supplementary Fig. 10f). These results demonstrated that CYP725A55 was responsible for the oxetane ester formation to generate the highly oxygenated Taxol tetracyclic core skeleton in Taxol biosynthesis.

The mechanism of oxetane ester formation

In 1987, Potier and coworkers proposed that the oxetane and tertiary C4 ester of the plant-isolated taxoids was formed via enzyme-catalyzed epoxidation of C4-C20 double bond and subsequently acyl group rearrangement by analyzing the structure of isolated taxoids bearing the oxetane ester motif and its possible allyl alcohol precursors³⁶ (Supplementary Fig. 9c). Decades later, Walker and coworkers reported the 10-deacetyl baccatin 10 β -O-acetyltransferase (DBAT) could catalyze the acylation of the tertiary C4-OH²⁶. Based on the above results, the formation of the oxetane ring and the acetylation of C4-OH catalyzed by P450 and acetyltransferase were proposed as another stepwise pathway to the biosynthesis of oxetane ester. However, none of the aforementioned pathways has been verified because no enzyme-catalyzed oxetane ester formation has been reported to date.

In this study, deuterium labeled compound 5 α -trideuterated acetyl-2 α -benzoyloxy-7 β -acetoxytaxusin (**20**) was synthesized from **5** via the combination of chemosynthetic and biosynthetic strategies and employed as the substrate in an in vivo reaction using *S. cerevisiae* strain co-expressing CYP725A55 and TCPR (Supplementary Fig. 9d–g).

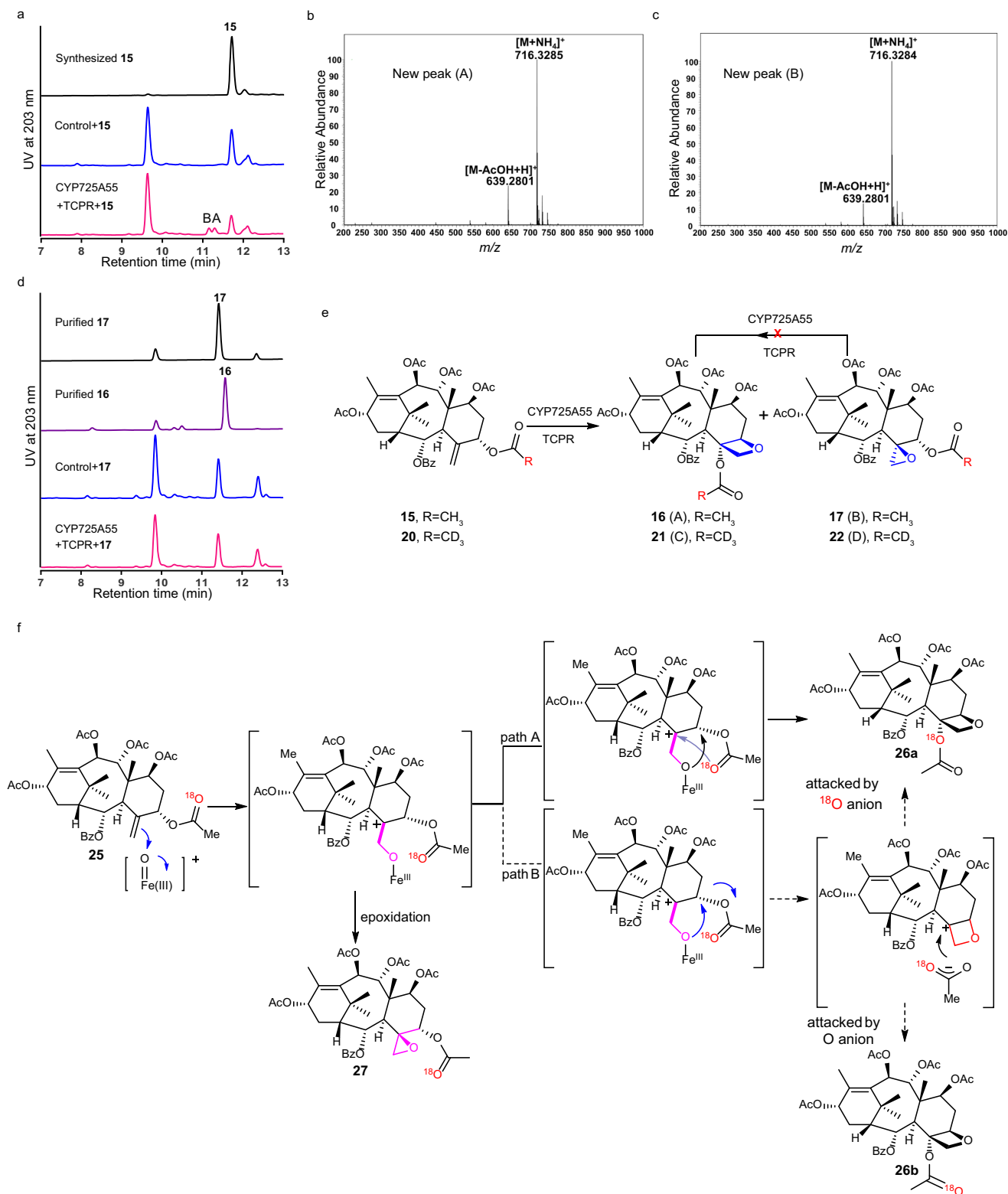


Fig. 3 | The formation of Taxol tetracyclic core skeleton with an oxetane ester.

a HPLC analysis of in vivo reaction for *S. cerevisiae* strain YCYP725A55 expressing CYP725A55 and TCPR using 2 α -benzoyloxy-7 β -acetoxytaxusin (**15**) as the substrate. Substrate was marked with compound number and products were marked with A and B. *S. cerevisiae* strain YTCPR expressing TCPR was used as the control. **b** Mass spectra (ESI) of new peak (A). **c** Mass spectra (ESI) of new peak (B). **d** HPLC analysis of in vivo reaction for *S. cerevisiae* strain YCYP725A55 expressing CYP725A55 and TCPR using 2-deacetyl-2 α -benzoylbaccatin I (**17**) as the substrate. Substrate was

marked with compound number. *S. cerevisiae* strain YTCPR expressing TCPR was used as the control. These experiments were performed three times with similar results each time. **e** Oxidation reaction of hexa-oxygenated taxoid derivatives catalyzed by CYP725A55. Labeled or regular methyl was marked with red. Arrow with a cross indicated this reaction was failed. **f** Proposed process of the formation of oxetane ester. ^{18}O was marked with red. Source data are provided as a Source Data file.

Two compounds were detected (Peaks C and D in Supplementary Fig. 10a). Compounds 1 β -dehydroxy-4 α -trideuterated acetyl-baccatin VI (**21**, peak C) with oxetane ester and 5 α -trideuterated acetyl-2-deacetyl-2-benzoylbaccatin I (**22**, peak D) with the epoxy ring were isolated and characterized by HR-ESIMS and NMR spectra analyses (Supplementary Fig. 10b–f). A significant ^2H signal detected at δ 2.26 ppm for **21** indicates that there was only one deuterated acetyl group. Comparing the ^1H NMR spectra of oxetane ester compound **16** with the deuterated **21**, the signal of methyl group in C4-acetoxy disappeared which indicated that the C4-acetoxy group in the oxetane product was derived from the C5-acetoxy group of 5 α -trideuterated derivative **20** (Supplementary Fig. 10d, e). To further verify whether epoxy ester was involved in the generation of oxetane ester via rearrangement, the epoxy derivatives (**17** and **22**) were employed in the in vivo reaction using *S. cerevisiae* strain co-expressing *CYP725A55* and *TCPR*. However, no oxetane ester product was detected by HPLC analysis (Fig. 3d and Supplementary Fig. 10g, h), implying that the epoxide ester may not be the precursor for the biosynthesis of oxetane ester. Based on these experimental results, the oxetane ester is likely to be formed via a direct oxidation-acyl rearrangement process without the formation of epoxide ester. To further identify whether the acyl rearrangement occurred through concerted or stepwise process or both of them. The C5-acetyl ^{18}O -labeled 2 α -benzoyloxy-7 β -acetoxytaxusin (**25**) was prepared (Supplementary Fig. 11a, b), and employed as substrate in in vivo reaction using the *S. cerevisiae* strain expressing *CYP725A55* and *TCPR*. The ^{18}O -labeled oxetane derivatives **26** (**26a** or the mixture of **26a** and **26b**) could be formed (Supplementary Fig. 11c), which further confirmed the C4 acetoxy group of oxetane ester **26** was derived from the C5 acetoxy group of **25**. All of the ^{18}O in ^{18}O -labeled oxetane derivatives **26** was reserved in the deacetyl product ^{18}O -labeled 4,7,9,10,13-pentadeacetyl-1 β -dehydroxybaccatin VI (**28a**) based on the MS analysis, indicating that the acyl rearrangement overwhelmingly occurred through a concerted process (Fig. 3f and Supplementary Fig. 11d–g).

Biosynthesis of 1 β -dehydroxybaccatin VI

In order to elucidate the complete biosynthetic pathway to 1 β -dehydroxybaccatin VI (**16**) sharing the same Taxol tetracyclic core skeleton as that of baccatin III, Taxol and their derivatives, we need to further find an enzyme responsible for the acylation of C7-OH of taxoid **13** to produce **15**. We cloned and expressed 9 uncharacterized acyltransferases encoding genes from our cDNA database in *E. coli*, and obtained *E. coli* strains YAT1-9. Then we screened these acyltransferases by incubating **13**, benzoyl-CoA, acetyl-CoA, and the combined crude enzymes from the *S. cerevisiae* strain YBT expressing 2 α -O-benzoyl transferase (TBT)²⁴ and each of the *E. coli* strains YAT1-9 in in vitro reaction, respectively. The strain YAT5 expressing *AT5* resulted in the production of **15**, demonstrating acyltransferase *AT5* could acetylate the C7-OH of taxoid **13** (Supplementary Fig. 12a–c). Based on the identified P450s in this work and the acyltransferase *AT5* as well as the re-characterization of several previously reported enzymes^{16,18,19,22,24,33}, biosynthetic pathway of **16** from **2** was elucidated.

Next, we constructed a *S. cerevisiae* strain named as YDBVI to biosynthesize **16** via co-expressing 12 genes from *Taxus* plant, including *T13OH*, *TSAT*, *TAX19*, *T10OH*, *CYP725A37*, *T2OH*, *T7OH*, *AAE4* (encoding enzyme *AAE4* catalyzing benzoic acid to produce benzoyl-CoA)³⁷, *TBT*, *AT5*, *CYP725A55*, and *TCPR* (Fig. 4a). Our target compound **16** could be detected by liquid chromatography-mass spectrometry (LC–MS) after feeding **2** to the strain YDBVI (Fig. 4b). Interestingly, although compound **16** is the predominant product in the culture of YDBVI using **2** as substrate (Fig. 4b), **16** appeared in a 1:1 ratio with **17** in the feeding studies using **15** as the substrate (Fig. 3). We speculate that these seemingly contradictory results might due to the complexity of the in vivo experimental system, i.e., other pathways for **16** from **2** in the YDBVI strain or the complex redox environment in the engineered strains expressing 12 genes.

Almost all of the supposed intermediates in this elucidated biosynthetic pathway were detected in the sequential co-expressing experiment (Supplementary Fig. 12d). And only the compound **11** detected from *Taxus* plants previously³⁴ was missed in these intermediates produced by these engineered yeasts. This result confirmed our elucidated biosynthetic pathway of **16** in this study (Fig. 4a) and illustrated that we have successfully achieved the biosynthesis of **16** from **2**.

Discussion

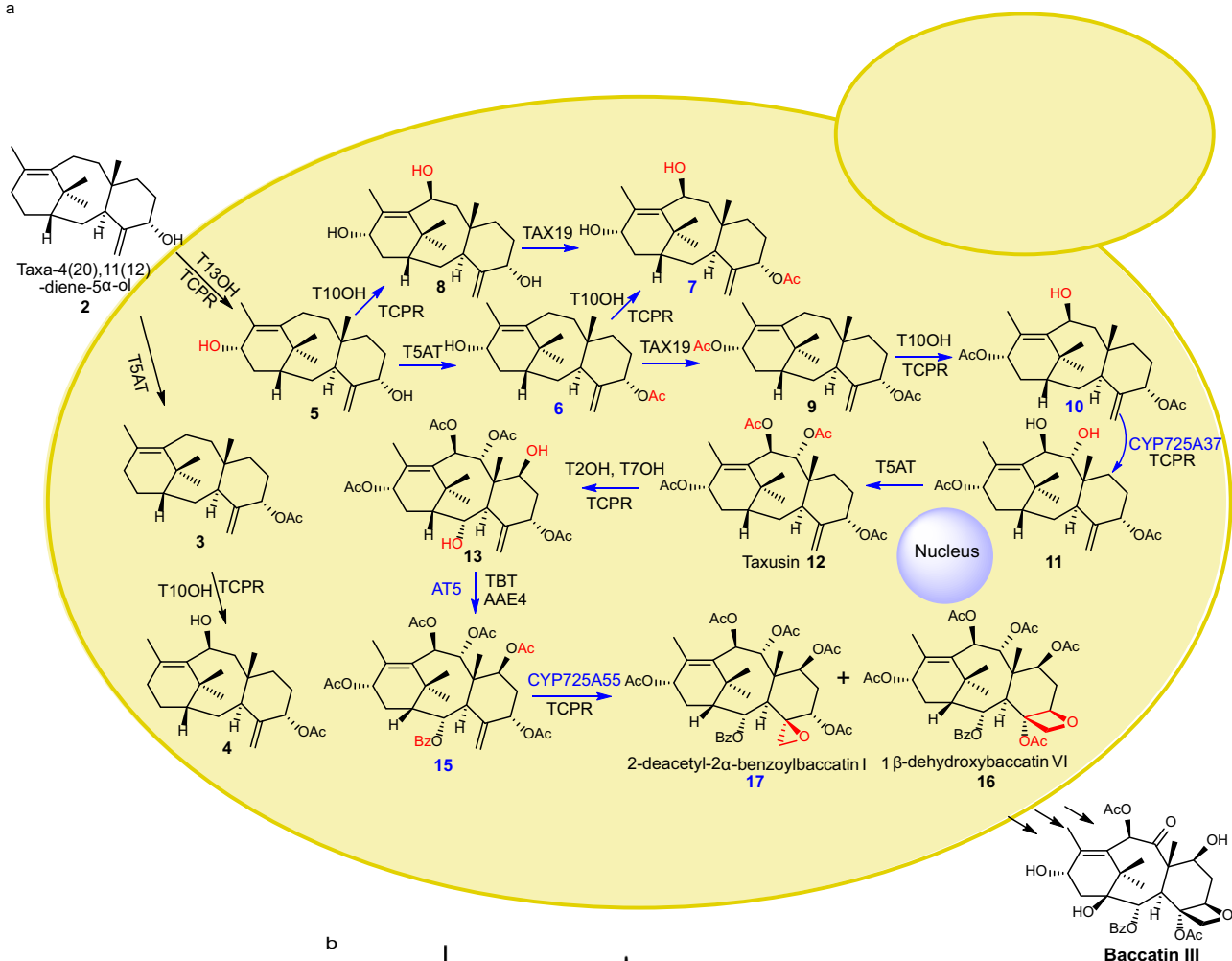
Over the past two decades, numerous efforts employing chemistry, phytochemistry, biochemistry and molecular biology have been undertaken to reveal the biosynthetic pathway for Taxol. In particular, an overwhelming majority of the genes (12 in 14) involved in the biosynthetic pathway of Taxol were identified by Prof. Rodney Croteau group¹⁵. However, except the biosynthesis of the key diterpenoid core skeleton, taxadiene, and the following two steps of oxygenation, little progress leading to the tri-oxygenated taxoid intermediates and beyond has been made until this study. In this study, we focused on the downstream steps of **2**, which was the proposed first hydroxylation product of taxa-4(5),11(12)-diene. Followed by step-wise re-characterizing key enzymes, identifying uncharacterized enzymes via integrated assays of in vivo substrate feeding of cells and in vitro substrate incubation with crude protein lysates based on their corresponding gene-expressed *S. cerevisiae* cells, pathways for biosynthesis of the highly oxygenated tetracyclic core skeleton of Taxol, 1 β -dehydroxybaccatin VI (**16**) from **2** was elucidated (Fig. 4). Along with the reviewing process of this manuscript, similar or alternative catalysis processes for the oxetane ring formation were published by three independent studies^{38–40}, of which, the study of Jiang et al.⁴⁰, not only supported our findings of enzyme characterization but also proved that this biosynthesis pathway network elucidated via *S. cerevisiae* chassis cells is largely similar to that via the *tobacco* chassis.

While realizing that the commonly focused di-oxygenated taxoid **4** might not be the key intermediate for the biosynthesis of Taxol, we retreated back to hypothesizing the di-oxygenated taxoid **5** transformed from mono-oxygenated taxoid **2** as the key intermediate. We re-characterized the function of several previously reported P450s and acetyltransferases^{16,18,19,22,33}, revealed their reaction orders via both in vitro and in vivo catalysis (Fig. 4a and Supplementary Fig. 1a, b), and finally the tri-oxygenated taxoid derivatives **7**, **8**, and **10** biosynthesized from **5** were identified (Fig. 2a).

We further identified taxoid C9 α -hydroxylase *CYP725A37* catalyzing the formation of tetra-oxygenated taxoid **11** from **10** (Fig. 2d). Then, via re-characterizing the function of *T5AT*, we revealed a biosynthetic pathway for the formation of tetra-oxygenated intermediate **12** (Fig. 4a), an abundant metabolite in the heartwood of yew but long been ignored in Taxol biosynthesis^{20,21}. Thus, the previously reported formation of the hexa-oxygenated taxoid intermediate **13** from **12** was confirmed in the Taxol biosynthesis pathway²⁰ (Fig. 4a). Furthermore, we proved that the hexa-oxygenated taxoid **13** can be acetylated by a missing acetyltransferase *AT5* and the previously reported *TBT*²⁴ to form **15** (Fig. 4a).

The oxetane motif in Taxol is essential for the binding with α , β -tubulin dimer in the process of inhibiting microtubule dynamics in actively replicating cells⁴¹. So far, in this proposed and rebuilt synthetic pathway from **2** to **15**, all of the taxoid intermediates share the same diterpenoid taxadiene core skeleton with increasing levels of oxygenation and various acylation modifications. We identified *CYP725A55* which transforms **15** to 1 β -dehydroxybaccatin VI (**16**) via the critical enzymatic oxetane ester formation reaction (Fig. 3). This highly oxygenated Taxol intermediate **16** shares the same Taxol tetracyclic core skeleton as that of the well-characterized baccatin III and all Taxol/Taxol derivatives (Fig. 4a).

a



b

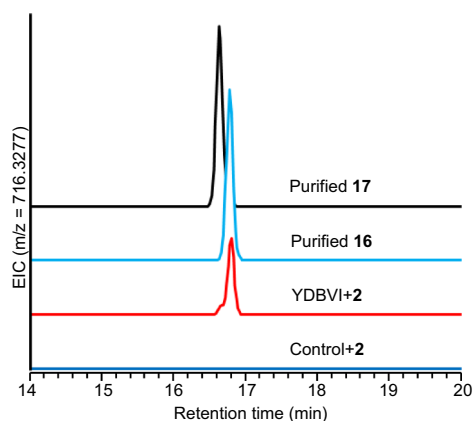


Fig. 4 | The elucidated pathway of 1 β -dehydroxybaccatin VI (16) and its biosynthesis in engineered yeast. a The biosynthetic schematic of 16 from 2 in engineered yeasts. Reactions, enzymes and new compounds characterized in this study were marked in blue. New formed groups in compounds were marked with red. **b** LC-MS chromatograms show the formation of 16 in engineered yeast YDBVI

co-expressing T13OH, TSAT, TAX19, T10OH, CYP725A37, T2OH, T7OH, AAE4, TBT, AT5, CYP725A55, and TCPR. AAE4, benzoyl-CoA synthetase. *S. cerevisiae* chassis strain YBD80 was used as the control. EIC extracted ion chromatogram. Source data are provided as a Source Data file.

Although either epoxidation-acyl rearrangement or stepwise oxetane formation-acylation process was proposed over the years for the biosynthesis of oxetane ester^{26,36}, neither has ever been experimentally proved. The isotope labeling (²H or ¹⁸O) and feeding experiments suggest a different catalyzing process in our enzymatic reaction system (Fig. 3f). The C4-C20 double bond was oxidized by CYP725A55 to form a C4 cation intermediate, which was transferred to oxetane ester 1 β -dehydroxybaccatin VI (16) via acyl rearrangement or captured by C20 oxygen to generate the epoxide derivative 2-deacetyl-2 α -benzoylbaccatin I (17) (Fig. 3e, f). For the second acyl rearrangement step

to form oxetane ring, there are two possible mechanisms, either concerted or stepwise processes. The oxygen isotope experiment was carried out to characterize this process. Thematically, the ¹⁸O-labeled acetyl rearranges to the C4 hydroxyl group to form only one oxetane product 26a via a concerted process, while both of the labeled and regular oxygen atom can equally attack the C4 cation to form two different oxetane derivatives (26a and 26b) with the ¹⁸O at the different positions via a stepwise process. According to the MS analysis of the de-acetyl product 28a, one of the oxygens was labeled and almost no regular oxygen atom attacked product was formed, which is consist

with the concerted rearrangement mechanism (Fig. 3f and Supplementary Fig. 11). These results indicated that the oxetane ester was overwhelmingly formed via a cascade oxidation-concerted acyl rearrangement mechanism in our enzymatic reaction system (Fig. 3f). It is interesting to notice that Jiang et al. also proposed an oxetane ring formation mechanism of oxidation-acyl rearrangement without using epoxide as an intermediate based on density functional theory (DFT) calculations⁴⁰, supporting our cascade oxidation-concerted acyl rearrangement mechanism.

The rebuilding of biosynthetic pathway of Taxol has already achieved the de novo biosynthesis of **2** in *E. coli* strains¹⁴. In this synthetic pathway from **2** to **16**, intermediates, **6**, **10** and **15** have not been reported in plants until now, while **11**, **12**, and **16** were isolated from *Taxus* plants previously (Supplementary Table 3)³⁴. The targeted compound **16** and almost all of the supposed intermediates including **6**, **9**, **10**, **12**, **13**, and **15** could be detected by sequentially co-expressing these 12 involved enzymes, testifying the elucidated biosynthetic pathway of **16** from **2** in living cells (Supplementary Fig. 12d). The intermediate **11** was exceptionally undetected in the metabolites of these recombinant yeast strains, which may result from its efficient acetylation catalyzed by TSAT as observed in the in vivo reaction (Supplementary Fig. 8c).

It is particularly interesting to notice that, in contrast to baccatin III, which bears a hydroxy, a ketone and a hydroxy at C7, C9 and C13 positions, respectively, compound **16** bears an acetyl group at each of these three positions. It seems that the formation of some intermediates need be protected by acetylation at C7, C9 and C13 positions and then the deacetylation at C7, C9, and C13 of **16** would be involved in the downstream biosynthesis of baccatin III from **16**. The adding and removing-protective-group strategy has been well applied in chemical synthesis and also found in the biosynthetic pathways of some complex plant natural products to protect unstable intermediates and inhibit the formation of byproducts via acetylation and deacetylation. For example, in the biosynthetic pathway of noscapine, an acetyl group was introduced at C8 by an O-acetyltransferase AT1 to function as a protective group to avoid hemiacetal formation in the next oxidation step by CYP82X1 and subsequently hydrolyzed by a carboxylesterases CXE1⁴². And an acetyl group of the intermediate vinorine serves as a stabilizing group to prevent ring-opening and was then hydrolyzed by an acetylmalan esterase at the last stage of ajmaline biosynthetic pathway⁴³. Similarly, an acetyl group of C21 introduced by acetyltransferase CsL21AT possibly acted as protecting group to stabilize the hemiacetal ring, which led to an increased yield of target kihadalactone A⁴⁴. Moreover, acetylation might also play important roles in the transport or storage of intermediates in the biosynthetic pathway⁴². So, these introduced acetyl groups at C7, C9, and C13 may also stabilize these highly oxygenated intermediates to avoid unexpected rearrangement reactions in the biosynthesis of Taxol and related taxoids or benefit the transport of these intermediates.

Several studies have found that much higher yield of 5(12)-oxa-3(11)-cyclotaxane (OCT) or iso-oxa-cyclotaxane (iso-OCT) (85% to 95% in total) rather than **2** (5% to 15%) were produced by heterogeneously expressing TSOH in different chassis, such as *S. cerevisiae*, *E. coli* or *Nicotiana benthamiana*^{45,46}. Structurally, different from the eight carbon-B ring system in taxoids, OCT and iso-OCT have two fused rings with C3(11) bond. It seems highly improbable to disrupt the C3(11) bond and transfer 3(11)-cyclotaxane core to the taxa-11(12)-ene core⁴⁷. Although it seems these OCTs may only be the bio-precursors of C3(11)-cyclotaxanes^{34,47}, there is still a possibility that the downstream products of OCTs can rearrange to taxoids in the complex biological environments in *Taxus* plants. It is worthy to study the downstream products of OCTs and the relationship of 3(11)-cyclotaxane core and taxa-11(12)-ene core. Recently, TSOH was proposed to catalyze oxetane ring formation via the rearrangement of unstable epoxide

intermediate from endotaxadiene and exotaxadiene based on the detection of a taxologenic oxetane among the found four products catalyzed by TSOH³⁸. In another study³⁹, taxologenic oxetane formation was also proposed to be catalyzed by TSOH together with a 2-oxoglutarate/Fe(II)-dependent dioxygenase. Although these proposed mechanisms, different from that of the cascade oxidation-concerted acyl rearrangement catalyzed by CYP725A55 shown in this study and Jiang et al.⁴⁰, need further experimental confirmations, the presence of both multi-products of TSOH and different mechanisms potential for oxetane ring formation implicated that the biosynthetic pathway of Taxol is more likely a complex, matrix rather than a linear pathway. Future comprehensive verifications are needed.

In summary, our systematic studies clearly elucidated the complete biosynthetic route to the highly oxygenated 1 β -dehydroxybaccatin VI with the Taxol tetracyclic core skeleton from the mono-oxygenated taxoid taxa-4(20),11(12)-diene-5 α -ol, laying down a solid foundation for the breakthrough of complete elucidation of the biosynthetic pathway of Taxol. Of course, further efforts are required to elucidate the complete biosynthetic pathway of baccatin III from 1 β -dehydroxybaccatin VI and other parallel biosynthetic pathway of Taxol. It is also interesting to further confirm whether Taxoid 1 β -hydroxylase and 9 α -dioxygenase are necessary³⁹ or unwanted⁴⁰ to biosynthesize baccatin III in *N. benthamiana* and *Taxus* in the future. While an important experience learnt from the recent emergence of breakthroughs³⁸⁻⁴⁰, including this study, cross-disciplinary investigations and collaborations, particularly between the chemists and biochemists with the phytochemists and plant molecular physiologists are essential in order to completely fulfill this long-awaited endeavor.

Methods

Media, strains, and chemicals

Saccharomyces cerevisiae strain BY4742 was purchased from EURO-SCARF. These engineered strains were cultured in SC minimal medium containing 6.7 g/L yeast nitrogen base without amino acids (291940, BD), 20 g/L glucose (10010518, Sinopharm Chemical Reagent) with corresponding deficient amino acid or YPD containing 10 g/L yeast extract (LP0021B, Thermo Fisher Scientific), 20 g/L bacteriological peptone (LP0137, Thermo Fisher Scientific), and 20 g/L glucose at 30°C. *Escherichia coli* TOP10 and BL21 (DE3) were cultured in Luria-Bertani (LB) medium containing 5 g/L yeast extract, 10 g/L peptone (LP0042B, Thermo Fisher) and 10 g/L sodium chloride (10019318, Sinopharm Chemical Reagent) at 37°C. Agar (10000582, Sinopharm Chemical Reagent) was added for SC minimal medium, YPD and LB plates with 20 g/L. Primers used in this research were ordered from Sangon Biotech (Shanghai) and listed in Supplementary Data 2. Ethyl acetate (G23272C), petroleum ether (G84208B), tetrahydrofuran (O11342848), triethylchlorosilane (92440B), 4-dimethylaminopyridine (14766 A), chromatography silica gel (G72651A), triethyl amine (17471H), triethylchlorosilane (92440B), pyridine (G14230A), tetrabutylammonium fluoride (14363 G), CDCl₃(87070 S), CD₃OD (84556 S), D₆-DMSO (38114 L), Pd(OAc)₂ (BD01121063), 1,4-Benzoquinone (12960B), and anisole (O11072489) were purchased from Shanghai Titan Scientific. Dichloromethane (80047318), MeOH(10014118), EtOH (10009218), acetic anhydride (10000318), ethylene diamine tetraacetic acid (10009717), acetic acid (10000218), *n*-hexane (80068618), sodium bicarbonate (O1-S0181), sodium sulfate (10020518), hydrazine hydrate (C196715000), HCl (10011018), and KOH (10017018) were purchased from Sinopharm Chemical Reagent. Dimethyl sulfoxide (A600163), ampicillin (A100339), dithiothreitol (A620058), and benzylsulfonyl fluoride (A610425) were purchased from Sangon Biotech (Shanghai). D₆-acetic anhydride (175641) and acetyl-CoA (A2056) were purchased from Sigma-Aldrich. 1,3-Dicyclohexylcarbodiimide (N806920) was purchased from Macklin. C₆D₆ (B100912) was purchased from Aladdin. Benzoyl CoA (ZL-23055) was purchased from Shanghai ZZBIO. Nicotinamide adenine dinucleotide

phosphate (60302ES02) was purchased from Yeasen Biotechnology (Shanghai). The high performance liquid chromatography (HPLC) grade acetonitrile (CAEQ-4-003306-4000) was purchased from Anpel Laboratory Technologies (Shanghai).

Taxus pan-transcriptome analysis

A total of 43 publicly available transcriptome/RNA-seq datasets from *Taxus* species (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=25628>) were collected from National Center for Biotechnology Information SRA (Supplementary Data 3). Nine initial assemblies were performed for the same sequencing platform of the datasets within same species. For Illumina sequencing platform data assembly, these raw sequencing datasets were processed using Trimmomatic⁴⁸ (version 0.38) to remove adapters and trim low-quality sequences with the following parameters (ILLUMINACLIP: NexteraPE_TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:4:15 LEADING:3 TRAILING:3 MINLEN:36). Subsequently, assembly was performed using Trinity⁴⁹ (version 2.9.0) with the following parameters (--max_memory 300 G --min_contig_length 200 --CPU 40 --bflyCPU 40 --inchworm_cpu 40) to obtain initial assembly results. Finally, seven Trinity assembly datasets were obtained. For 454 life science sequencing platform data assembly, the NCBI SRA data files were converted to SFF (Standard Flowgram Format) using sratoolkit⁵⁰ (version 2.10.5). Then the SFF files were assembled using Newbler (version 3.0, Roche) with “-m -cpu 64 -cdna” parameter to generate assembled isotigs transcripts. To merge these assemblies, the seven Trinity assembly datasets and two Newbler assembly datasets were combined. The contigs was shredded into fragments of 2000 bp length with 200 bp overlap. The resulted fragments assembly was performed using Newbler (version 3.0, Roche) with the following parameters: -o *Taxus* -m -cpu 64 -cdna. For deduplication, the initial nine assembly datasets and the merged assembly datasets were combined. CD-HIT-EST⁵¹ (version 4.8.1, parameters: -c 1 -aS 1 -M 0 -G 0 -d 0 -r 1 -g 1 -T 40 -i Transcripts.fasta -o cdhit) was utilized to obtain the final non-redundant *Taxus* transcriptome. For transcript completeness assessment, sequence similarity searches were conducted using the Diamond⁵² (version 0.9.36) alignment against the Uniprot Swiss-Prot database⁵³ (Database version: 202001). Trinity⁴⁹ (version 2.9.0) analyze_blastPlus_topHit_coverage tool was employed to assess the completeness of non-redundant transcripts. Finally, coding sequences of the transcripts were predicted using TransDecoder (version 5.0.5) with default parameters.

RNA-seq and de novo transcriptome assembly

Taxus yunnanensis Cheng et L. K. Fu twigs (including wood and needles) were homogenized using a chilled mortar and pestle, and total RNA was isolated using RNeasy Pure Plant Kit (DP441, TIANGEN BIOTECH) according to the manufacturer's instructions. For PacBio ISO-SEQ data analysis, SMRTlink (version 8.0.0, Pacific Biosciences) was used to filter and process raw subreads, using the Iso-seq3 pipeline to obtain highly accurate long reads. These raw subreads were used to generate circular consensus sequence (CCS) reads using the following parameters: --min-rq 0.9. The resulting CCS reads were then stripped of adapter sequences using LIMA. The poly-A tails were then removed using the Refine tool. Full-length non-chimeric (FLNC) reads were classified based on the presence or absence of 5' primer, 3' primer and the poly(A) tail. FLNC reads were clustered using the Iso-seq3 cluster module in the SMRTlink to generate consensus sequences, which were subsequently polished using the raw subreads. For Illumina RNA-seq data analysis, these raw reads were extracted from the sequencing results, and FastQC (Version 0.11.9) with default parameters was used to evaluate quality of the raw reads. Adaptor sequences in the raw reads were trimmed using Trimmomatic⁴⁸ (Version 0.38). Moreover, the bases at the begin and end of each reads with quality below 3 were cut. The

Trimmomatic slides from the 5' end in windows and the window size was 4 bases. When the average quality in the window is lower than the setting threshold of 15, the read is cut. Only the reads with length >36 bp was kept. The obtained data was subjected for de novo transcriptome assembly using Trinity⁴⁹ (Version 2.9.0).

Cloning of candidate genes

RNA isolated from *Taxus yunnanensis* Cheng et L. K. Fu twigs was converted into cDNA using the PrimeScript RT reagent Kit with gDNA eraser (RR047A, Takara). In general, the open reading frames (ORFs) of candidate genes were cloned from *Taxus* twig cDNA with PrimeSTAR HS DNA Polymerase kit (R010A, Takara). Then PCR products were linked into the pMD18T vector (DI01A, Takara) and transformed into *E. coli* TOP10 cells. Transformants were plated for selection on LB plates with 100 µg/mL ampicillin and positive colonies were verified by colony PCR and Sanger sequencing.

Construction of yeast strains

S. cerevisiae codon-optimized genes for GGPP synthase (GGPPs), TS, T13OH, T10OH, T7OH, T5AT, TAX19, CYP725A37, TCPR, AAE4, and TBT were synthesized by Tsingke Biotechnology or Shanghai Genaray Biotech and were listed in Supplementary Data 4. *S. cerevisiae* and *E. coli* strains used in this study were listed in Supplementary Table 4. PrimeSTAR HS DNA Polymerase kit was used for all PCR amplification steps according to the manufacturer's instruction. HiPure Gel DNA Pure Mini Kit (D2111-03, Guangzhou Magen Biotechnology) was used to purify DNA products according to the manufacturer's instruction. The details for construction of yeast strains can be found in Supplementary Note 1. While the overall strategy is as following:

(1) The strong promoters (GAL1p and GAL10p) and terminators commonly used in yeast metabolic engineering were chosen to express heterologous gene. In order to change the galactose-induced promoters (GAL1p and GAL10p) into constitutive expression, the GAL-promoter negative regulator, *GAL80* of BY4742 was deleted with an URA3 marker and the resulted strain YBD80 was chosen as the chassis cell for heterologous expression of candidate genes.

(2) Three selective markers, *i.e.*, HIS3, LEU and KanMX, were used for the construction of genetic engineered yeast strains based on the chassis strain YBD80 and the details were described in the Supplementary Table 4. Because multiple genes could be simultaneously integrated in the chromosome of yeast, we achieved the integration of 12 genes by three-round integrations in chassis strain YBD80 using 3 different makers (HIS3, LEU and KanMX), without recycling usage of the markers.

(3) For clone verification, in case of integration of one or two genes, five colonies were picked for verification, while in the case of integration of multiple genes (>3), 20 colonies were picked for verification.

In vivo feeding assays

A single colony of yeast strain was incubated in 1 mL YPD medium and then grown at 30 °C, 250 rpm. Substrate (25 or 500 µM in dimethyl sulfoxide, exact concentrations for each individual assays were listed in the Supplementary Table 5) was added to the medium at 24 h. Metabolite was extracted using 1 mL ethyl acetate (EA) at 48 h and the pooled extracts were evaporated. The residue was dissolved in EA and analyzed by Gas chromatography flame ionization detector (GC-FID) or HPLC. The control strain (YBD80 or YTCPR) was treated in parallel. The concentration of product was estimated on the conversion rate and the assumption that the substrate and corresponding product shared the same chromatographic response (the estimated product concentrations for each of the reactions were shown in the Supplementary Table 5).

In vitro enzymatic assays

In order to eliminate the effect of time-consuming ultracentrifugation and redissolution in the process of microsomes preparation on the activity of enzyme, crude protein lysate of yeast was used for these in vitro enzymatic assays. A single colony of yeast strain was incubated in 50 mL YPD medium and grown 48 h at 30 °C, 250 rpm. Then yeast cell was incubated in 200 mL YPD medium after centrifugation at room temperature and grown 24 h at 30 °C, 250 rpm. Next, yeast cell was incubated in 200 mL YPD medium after centrifugation at 4 °C and grown 24 h at 16 °C, 110 rpm. Finally, collected yeast cell was pulverized with liquid nitrogen and resuspended with 100 mM potassium phosphate buffer (pH 7.4) containing 1 mM ethylene diamine tetraacetic acid, 0.5 mM dithiothreitol, 1 mM benzylsulfonyl fluoride. After removing cell fragments using centrifugation, the supernatant was used as crude protein lysate. These in vitro enzymatic assays were conducted in 300 μ L reactions of containing 100 mM potassium phosphate buffer (pH 7.4), 1 mM NADPH (for P450s) or acetyl-CoA (for acyltransferases), 25 or 500 μ M substrate, and 12 mg crude protein lysate. The crude protein lysate from control strain (YBD80 or YTCPR) was treated in parallel. For the preparation of crude protein of acyltransferases, a single colony of *E. coli* strain was incubated in 4 mL LB medium and grown 12 h at 37 °C, 200 rpm. Then *E. coli* strain was incubated in 50 mL LB medium and grown at 37 °C, 200 rpm until OD₆₀₀ was 0.6. Next, protein expression was induced by 0.2 mM isopropyl β -D-1-thiogalactopyranoside at 16 °C, 110 rpm for 18 h. The collected cells were suspended in 100 mM phosphate buffer (pH 7.4) and then disrupted with an Ultrasonic Cell Disruption System (JY92-IIDN, Shanghai Jingxin Industrial Development) (300 W, ultrasonic time: 3 s; rest time: 3 s, 10 min). After removing the cell debris by centrifugation and the supernatant was used for enzymatic assays. These in vitro enzymatic assays of acyltransferases were conducted in 600 μ L reactions of containing 100 mM potassium phosphate buffer (pH 7.4), 1 mM acetyl-CoA, 1 mM benzoyl CoA, 500 μ M substrate, 12 mg crude protein lysate from YTBT and 1 mg *E. coli* crude protein lysate. The crude protein lysate from control *E. coli* strain with empty vector was treated in parallel. The reaction system was mixed at 30 °C for 12 h. The reactants were extracted with equal volume EA and analyzed by GC-FID or HPLC.

GC-FID analysis

GC-FID analysis was performed on a GC2010 pro (Shimadzu, Suzhou, China) equipped with a flame ionization detector using a TG-5MS capillary column (30 m \times 0.25 mm \times 0.25 μ m, Thermo Fisher Scientific). The oven temperature was kept at 150 °C for 2 min, increased to 250 °C at a rate of 10 °C/min, held 250 °C for 20 min, then increased to 280 °C at a rate of 10 °C/min, and finally held 280 °C for 5 min. The temperature of inlet and detection were 250 °C and 260 °C, respectively. The sample volume of per injection was 2 μ L in split ratio of 20. The flow rate of carrier gas (nitrogen) was kept at 1.25 mL/min.

HPLC analysis

HPLC analysis was performed on a Shimadzu LC 20A system (Shimadzu, Kyoto, Japan) equipped with a diode array detector, an auto-sampler, a binary pump and a thermostatically controlled column compartment using a Welch Boltimate™ C18 column (2.7 μ m, 2.1 mm \times 100 mm). Water (A) and acetonitrile (B) were used in the gradient elution system. The gradient elution program was carried out as follows: 0–2 min (15% B), 15 min (98% B), and 15.01–17 min (15% B). The flow rate was kept at 0.45 mL/min and the temperature of column compartment was set as 35 °C. Products were detected at 203 nm.

HPLC/ESIMS analysis

The sample preparation for HPLC/ESIMS analysis is same to GC-FID or HPLC analysis. The chromatographic analysis was carried out on a Dionex Ultimate 3000 RSLC (HPG) ultra-performance liquid

chromatography system (Thermo Fisher Scientific) using a Welch Boltimate™ C18 column (2.7 μ m, 2.1 mm \times 100 mm). Water (A) and acetonitrile (B) were used in the gradient elution system. The injection volume was 1 μ L. The gradient elution program was carried out as follows: 0–1 min (5% B), 26 min (95% B), and 26.01–28 min (95% B), 28.01–30 min (5% B). The flow rate was kept at 0.35 mL/min and the temperature of column compartment was set as 35 °C. Products were also detected at 203 nm. The HR-ESIMS data were acquired using a Q Exactive quadrupole orbitrap high-resolution mass spectrometry with a HESI ionization source (spray voltage 3.5 kV, Thermo Fisher Scientific) in the positive ionization mode. The sheath and aux gas were maintained at 40 and 10 (arbitrary units), respectively. The capillary temperature and heater temperature were set at 300 °C and 350 °C, respectively. The S-Lens RF level was set at 50. The Orbitrap mass analyzer was operated at a resolving power of 70,000 in full-scan mode (scan range: 200–1000 m/z ; automatic gain control (AGC) target: 1e6) with a dynamic exclusion setting of 4.0 s. Xcalibur 4.2 software was used for compound analysis. The number of technical and/or biological replicates was indicated in the Figure legend. For the MS analysis of biosynthetic intermediates in yeast cell, the MS analysis of purified standards were performed at the same time. Characteristic ion fragment and retention time were used to identify compounds.

Chemical synthesis of 2

Taxa-4(5), 11(12)-diene (**1**) was extracted from the medium of strain YTA with petroleum ether (PE) and purified with chromatography silica gel using PE as eluent. The solution of **1** (2.0 g, 7.4 mmol), Pd(OAc)₂ (412 mg, 25 mol%), 1,4-Benzoquinone (1.76 g, 14.4 mmol), anisole (3.2 mL, 29.6 mmol) in acetic acid (200 mL) was degassing and charged with N₂, then the system was stirred at 50 °C overnight⁵⁴. The reaction system was cooled to room temperature to quench the reaction. Water (300 mL) and *n*-hexane (500 mL) were added and the organic phase was separated, and the aqueous phase was extracted with *n*-hexane (3 \times 500 mL). The combined organic phase was washed with saturated sodium bicarbonate, brine, dried with anhydrous sodium sulfate and filtered. The solvent was removed in vacuum and the resulted residue was purified with chromatography silica gel (PE/EA = 50/1) to obtain 5 α -acetoxy-taxa-4(20),11(12)-diene (**3**) as yellow oil (800 mg).

To the solution of 5 α -acetoxy-taxa-4(20),11(12)-diene (800 mg, 2.42 mmol) in THF /MeOH (1/1, 10 mL), KOH (1.0 g, 1.8 M in final reaction system) was added in one portion, the resulted reaction system was stirred at room temperature for 24 h. The solvent was removed in vacuum and the resulted residue was purified with chromatography silica gel (PE/EA = 30/1) to obtain taxa-4(20),11(12)-diene-5 α -ol (**2**) as colorless solid (588 mg, 33% yield for 2 steps).

Chemical synthesis of 5 α -trideuterated acetoxy taxa-4(20),11(12)-diene-13 β -ol (**18**)

Triethyl amine (0.69 mL, 10.2 mmol), triethylchlorosilane (0.43 mL, 2.3 mmol) was dropped to the solution of taxa-4(20),11(12)-diene-5 α ,13 β -diol (**5**) (650 mg, 2.1 mmol) in dichloromethane (15 mL) at 0 °C, subsequently. Then 4-dimethylaminopyridine (131 mg, 1.0 mmol) was added in one portion and the resulted reaction system was stirred at room temperature with 2 h. Then water was added to quench the reaction and separated. The aqueous phase was extracted with dichloromethane (3 \times 30 mL), the combined organic phase was washed with brine, dried with sodium sulfate, filtered and concentrated in vacuum. The resulted residue was purified on chromatography silica gel (PE/EA = 20/1) to get the compound 13 α -triethylsiloxytaxa-4(20), 11(12)-diene-5 α -ol (**29**) (300 mg, 34% yield).

The mixture of compound **29** (300 mg, 0.72 mmol), pyridine (2 mL) and D₆-acetic anhydride (2 mL) was stirred at room temperature overnight and the concentrated in vacuum, the residue was purified on chromatography silica gel (PE/EA = 20/1) to provide the compound 5 α -trideuterated acetoxy-13 α -triethylsiloxytaxa-4(20),11(12)-diene (**30**) (240 mg).

Tetrabutylammonium fluoride (1.0 mL, 1 M solution in THF, 1 mM) was dropped to the solution of compound **30** (240 mg, 0.5 mmol) in THF (5 mL) at room temperature and the reaction system was stirred with 1 h and quenched with water (10 mL). The organic phase was separated and the aqueous phase was extracted with EA (3 × 20 mL), the combined organic phase was washed with brine, dried with sodium sulfate, filtered and concentrated in vacuum. The resulting residue was purified on chromatography silica gel (PE/EA = 7/1) to obtain the desired deuterated substrate as colorless solid (150 mg, 67% yield for two steps).

Chemical synthesis of ¹⁸O-labeled compound **23**

To the solution of **29** (1.2 g, 2.87 mmol) and acetic acid-¹⁸O₂ (732 mg, 4.6 mmol) in anhydrous dichloromethane (16 mL), the triethyl amine (2.0 mL, 14 mmol), 1,3-dicyclohexylcarbodiimide (1.18 g, 5.74 mmol), 4-dimethylaminopyridine (351 mg, 2.87 mmol) were added sequentially at room temperature. The resulted reaction mixture was stirred at room temperature with 2 days and then quenched by water (30 mL). The organic phase was separated and the aqueous phase was extracted with dichloromethane (3 × 40 mL). The combined organic phase was washed with brine, dried with anhydrous sodium sulfate, filtered and concentrated in vacuum. The resulted residue was purified on silica chromatography gel (PE/EA = 15/1) to obtain the inseparable mixture of **29** and ¹⁸O-labeled 5 α -acetoxy-13 α -triethylsilyloxytaxa-4(20),11(12)-diene-13 α -ol (**31**) (1.2 g).

The mixture of **29** and **31** was dissolved in anhydrous THF (10 mL) and Tetrabutylammonium fluoride (10.0 mL, 1 M solution in THF, 10.0 mmol) was dropped at room temperature. The reaction system was stirred at room temperature with 1 h and quenched with water (20 mL). The organic phase was separated and the aqueous phase was extracted with ethyl acetate (3 × 20 mL), the combined organic phase was washed with brine, dried with sodium sulfate, filtered and concentrated in vacuum. The resulted residue was purified on chromatography silica gel (PE/EA = 7/1) to obtain the desired ¹⁸O-labeled substrate **23** as colorless solid (422 mg, 42% yield for two steps).

Chemical deacetylation of ¹⁸O-labeled 1 β -dehydroxybaccatin VI (**26a**) to prepare ¹⁸O-labeled 4,7,9,10,13-pentadeacetyl-1 β -dehydroxybaccatin VI (**28a**)

To the solution of ¹⁸O-labeled 1 β -dehydroxybaccatin VI (**26a**) (2.4 mg, 3.4 μ mol) in ethanol (200 μ L), hydrazine hydrate (200 μ L) was added. The resulted solution was stirred at room temperature over 32 h and then diluted with water (2 mL), neutralized with 1 M HCl aqueous solution⁵⁵. The resulted aqueous solution was extracted with ethyl acetate (3 × 10 mL). The combined organic phase was washed with brine and dried with sodium sulfate, filtered and purified by prep-HPLC to obtain the desired ¹⁸O-labeled 4,7,9,10,13-pentadeacetyl-1 β -dehydroxybaccatin VI (**28a**).

Chemical introducing acetyl group to synthesize compounds **15**, **20**, and **25**

The mixture of compound **14** (11.8 mg, 17.3 μ mol) in acetic anhydride (0.5 mL) and pyridine (0.5 mL) was stirred at room temperature overnight and then concentrated in vacuum, the residue was purified by prep-HPLC to give the 2 α -benzoyloxy-7 β -acetoxytaxusin (**15**) (9.4 mg) in 85% yield. Similarly, 2 α -benzoyloxy-5 α -trideuterated acetyl-taxusin-7 β -ol (**19**) (6.0 mg, 8.8 mmol) was stirred in acetic anhydride (0.5 mL) and pyridine (0.5 mL) to give 2 α -benzoyloxy-7 β -acetoxy-5 α -trideuterated acetyl-taxusin (**20**) (4.8 mg) with 82% yield. ¹⁸O-labeled 2 α -benzoyloxy-7 β -acetoxytaxusin (**25**) (4.2 mg) was synthesized from ¹⁸O-labeled 2 α -benzoyloxy-acetoxytaxusin-7 β -ol (**24**) (4.7 mg, 7.3 μ mol) in 83% yield as the above procedure.

Isolation and purification of standards

For the identification of products, the in vivo feeding systems were amplified correspondingly. Specifically, a single colony of yeast strain was incubated in 50 mL YPD medium in 250 mL shake flask and then grown at 30 °C, 250 rpm about 16 h. Then, 2 mL culture containing yeast strain was incubated in 200 mL medium in 1 L shake flask and yeast strain was grown at 30 °C, 250 rpm. In order to achieve the conversion efficiency in tube, low concentration substrate (1%, v/v, 2.5–50 μ M in dimethyl sulfoxide) was added to the medium at 24 h. Product was extracted with EA at 60 h and purified with the silica gel column using different ratios of PE and EA as eluent. The purity of the product was judged by NMR analysis. If necessary, further purification was performed using Agilent 1260 Infinity II preparative liquid chromatograph system using a Welch Ultimate XB-C18 (5 μ m, 21.2 mm × 20 mm) preparative column. Water (A) and acetonitrile (B) were used in the gradient elution system and products were detected at 203 nm. For the detail of isolation and purification of product, please see Supplementary Note 2.

NMR analysis

All the ¹H NMR, ¹³C NMR, 2D-COSY NMR, 2D-NOE NMR, 2D-HSQC NMR and 2D-HMBC NMR were recorded on Bruker AM-400, 500-MHz or Bruker AV Neo 500-, 600-MHz spectrometers in the CDCl₃, CD₃OD, C₆D₆ or D₆-DMSO and the data was analyzed and processed by using MestRenova (version 14.1.0). The chemical shift (δ) of ¹H NMR was given in ppm relative to TMS (δ = 0.00 ppm), CHCl₃ (δ = 7.26 ppm), C₆H₆ (δ = 7.16 ppm) or DMSO (δ = 2.50 ppm). The chemical shift (δ) of ¹³C NMR was given in ppm relative to CDCl₃ (δ = 77.16 ppm), CD₃OD (δ = 49.00 ppm), C₆D₆ (δ = 128.06 ppm) or D₆-DMSO (δ = 39.52 ppm). All the ²H NMR were recorded on Bruker Bruker AM-400, 500 MHz spectrometers in CHCl₃, the chemical shift (δ) of ²H NMR was given in ppm relative to added standard CDCl₃ (δ = 7.26 ppm). The details of the NMR data including the methodologies for the determination the stereochemistry of key hydroxyl group and selectively acylation position were shown in Supplementary Note 2. All the NMR spectra and the key COSY, NOE and HSBC were supplied in Supplementary Figs. 13–115. NMR data of 1 β -dehydroxybaccatin VI (**16**) is also available in Supplementary Table 6.

Data availability

RNA sequencing data generated in this study have been deposited at NCBI under BioProject no. [PRJNA1062083](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1062083). CYP725A37, codon-optimized CYP725A37, CYP725A55 and AT5 have been deposited Genbank under accession numbers [PP197199](https://www.ncbi.nlm.nih.gov/nuccore/PP197199), [PP197200](https://www.ncbi.nlm.nih.gov/nuccore/PP197200), [PP197201](https://www.ncbi.nlm.nih.gov/nuccore/PP197201) and [PP197202](https://www.ncbi.nlm.nih.gov/nuccore/PP197202), respectively. Sequences of genes used in this study were obtained from Genbank under the following accession numbers: GGPPs (D28748 [<https://www.ncbi.nlm.nih.gov/nuccore/D28748.1>]), TS (U48796), T13OH (AY056019), T10OH (AF318211), T10OH2 (AY563635), T7OH (AY307951), T2OH (AY518383), T5AT (AF190130), TAX19 (AY628434), TCPR (AY571340), AAE4 (MN961507), and TBT (AF297618). Source data are provided with this paper.

References

1. 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).
2. Sabzehzari, M., Zeinali, M. & Naghavi, M. R. Alternative sources and metabolic engineering of Taxol: Advances and future perspectives. *Biotechnol. Adv.* **43**, 107569 (2020).
3. Min, L. et al. Strategies and lessons learned from total synthesis of taxol. *Chem. Rev.* **123**, 4934–4971 (2023).
4. Nicolaou, K. C. et al. Total synthesis of taxol. *Nature* **367**, 630–634 (1994).

5. Holton, R. A. et al. First total synthesis of taxol .2. Completion of the C-Ring and D-Ring. *J. Am. Chem. Soc.* **116**, 1599–1600 (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. Kanda, Y. et al. Two-phase synthesis of taxol. *J. Am. Chem. Soc.* **142**, 10526–10533 (2020).
8. Yukimune, Y., Tabata, H., Higashi, Y. & Hara, Y. Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in *Taxus* cell suspension cultures. *Nat. Biotechnol.* **14**, 1129–1132 (1996).
9. Baloglu, E. & Kingston, D. G. A new semisynthesis of paclitaxel from baccatin III. *J. Nat. Prod.* **62**, 1068–1071 (1999).
10. Horwitz, S. B. How to make taxol from scratch. *Nature* **367**, 593–594 (1994).
11. Koeppe, A. E. et al. Cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene is the committed step of taxol biosynthesis in *Pacific Yew*. *J. Biol. Chem.* **270**, 8686–8690 (1995).
12. Hefner, J. et al. Cytochrome P450-catalyzed hydroxylation of taxa-4(5),11(12)-diene to taxa-4(20),11(12)-dien-5 alpha-ol: the first oxygenation step in taxol biosynthesis. *Chem. Biol.* **3**, 479–489 (1996).
13. Biggs, B. W. et al. Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **113**, 3209–3214 (2016).
14. Ajikumar, P. K. et al. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70–74 (2010).
15. Guerra-Bubb, J., Croteau, R. & Williams, R. M. The early stages of taxol biosynthesis: an interim report on the synthesis and identification of early pathway metabolites. *Nat. Prod. Rep.* **29**, 683–696 (2012).
16. Jennewein, S., Wildung, M. R., Chau, M., Walker, K. & Croteau, R. Random sequencing of an induced *Taxus* cell cDNA library for identification of clones involved in Taxol biosynthesis. *Proc. Natl Acad. Sci. USA* **101**, 9149–9154 (2004).
17. Jennewein, S., Long, R. M., Williams, R. M. & Croteau, R. Cytochrome p450 taxadiene 5alpha-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis. *Chem. Biol.* **11**, 379–387 (2004).
18. Schoendorf, A., Rithner, C. D., Williams, R. M. & Croteau, R. B. Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from *Taxus* and functional expression in yeast. *Proc. Natl Acad. Sci. USA* **98**, 1501–1506 (2001).
19. Jennewein, S., Rithner, C. D., Williams, R. M. & Croteau, R. B. Taxol biosynthesis: taxane 13 alpha-hydroxylase is a cytochrome P450-dependent monooxygenase. *Proc. Natl Acad. Sci. USA* **98**, 13595–13600 (2001).
20. Chau, M. & Croteau, R. Molecular cloning and characterization of a cytochrome P450 taxoid 2alpha-hydroxylase involved in Taxol biosynthesis. *Arch. Biochem. Biophys.* **427**, 48–57 (2004).
21. Chau, M., Jennewein, S., Walker, K. & Croteau, R. Taxol biosynthesis: molecular cloning and characterization of a cytochrome P450 taxoid 7 beta-hydroxylase. *Chem. Biol.* **11**, 663–672 (2004).
22. Walker, K., Schoendorf, A. & Croteau, R. Molecular cloning of a taxa-4(20),11(12)-dien-5alpha-ol-O-acetyl transferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Arch. Biochem. Biophys.* **374**, 371–380 (2000).
23. 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).
24. Walker, K. & Croteau, R. Taxol biosynthesis: molecular cloning of a benzoyl-CoA:taxane 2alpha-O-benzoyltransferase cDNA from *Taxus* and functional expression in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97**, 13591–13596 (2000).
25. 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).
26. Ondari, M. E. & Walker, K. D. The taxol pathway 10-O-acetyltransferase shows regioselective promiscuity with the oxetane hydroxyl of 4-deacetyl taxanes. *J. Am. Chem. Soc.* **130**, 17187–17194 (2008).
27. Walker, K., Fujisaki, S., Long, R. & Croteau, R. Molecular cloning and heterologous expression of the C-13 phenylpropanoid side chain-CoA acyltransferase that functions in Taxol biosynthesis. *Proc. Natl Acad. Sci. USA* **99**, 12715–12720 (2002).
28. Walker, K. D., Klettke, K., Akiyama, T. & Croteau, R. Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in Taxol biosynthesis. *J. Biol. Chem.* **279**, 53947–53954 (2004).
29. Ramirez-Estrada, K. et al. Transcript profiling of jasmonate-elicited *Taxus* cells reveals a beta-phenylalanine-CoA ligase. *Plant Biotechnol. J.* **14**, 85–96 (2016).
30. Sanchez-Munoz, 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).
31. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* **33**, 377–383 (2015).
32. Kaspera, R., Cape, J. L., Faraldos, J. A., Ketchum, R. E. & Croteau, R. B. Synthesis and in vitro evaluation of taxol oxetane ring D precursors. *Tetrahedron Lett.* **51**, 2017–2019 (2010).
33. Chau, M., Walker, K., Long, R. & Croteau, R. Regioselectivity of taxoid-O-acetyltransferases: heterologous expression and characterization of a new taxadien-5alpha-ol-O-acetyltransferase. *Arch. Biochem. Biophys.* **430**, 237–246 (2004).
34. Wang, Y. F. et al. Natural taxanes: developments since 1828. *Chem. Rev.* **111**, 7652–7709 (2011).
35. Min, Z. D. J., Liang, H. & Studies, J. Y. On the Taxane diterpenes of the heart-wood from *Taxus mirei*. *Acta Pharm. Sin.* **24**, 673–677 (1989).
36. Guerittevoegelein, F., Guenard, D. & Potier, P. Taxol and derivatives - a biogenetic hypothesis. *J. Nat. Prod.* **50**, 9–18 (1987).
37. Srividya, N. et al. Biochemical characterization of acyl activating enzymes for side chain moieties of Taxol and its analogs. *J. Biol. Chem.* **295**, 4963–4973 (2020).
38. 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 (2023).
39. Zhang, Y. J. et al. Synthetic biology identifies the minimal gene set required for paclitaxel biosynthesis in a plant chassis. *Mol. Plant* **16**, 1951–1961 (2023).
40. Jiang, B. et al. Characterization and heterologous reconstitution of biosynthetic enzymes leading to baccatin III. *Science* **383**, 622–629 (2024).
41. Löwe, J., Li, H., Downing, K. H. & Nogales, E. Refined structure of $\alpha\beta$ -tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).
42. Dang, T. T., Chen, X. & Facchini, P. J. Acetylation serves as a protective group in noscapine biosynthesis in opium poppy. *Nat. Chem. Biol.* **11**, 104–106 (2015).
43. Ruppert, M. et al. Functional expression of an ajmaline pathway-specific esterase from *Rauvolfia* in a novel plant-virus expression system. *Planta* **222**, 888–898 (2005).
44. Pena, R. D. et al. Complex scaffold remodeling in plant triterpene biosynthesis. *Science* **379**, 361–368 (2023).
45. Edgar, S. et al. Mechanistic insights into taxadiene epoxidation by taxadiene-5 α -hydroxylase. *ACS Chem. Biol.* **11**, 460–469 (2016).
46. Li, J. H. et al. Chloroplastic metabolic engineering coupled with isoprenoid pool enhancement for committed taxanes biosynthesis in *Nicotiana benthamiana*. *Nat. Commun.* **10**, 4850 (2019).

47. Rontein, D. et al. CYP725A4 from yew catalyzes complex structural rearrangement of taxa-4(5),11(12)-diene into the cyclic ether 5(12)-oxa-3(11)-cyclotaxane. *J. Biol. Chem.* **283**, 6067–6075 (2008).
 48. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
 49. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
 50. Leinonen, R., Sugawara, H. & Shumway, M. International Nucleotide Sequence Database Collaboration The sequence read archive. *Nucleic Acids Res.* **39**, D19–D21 (2010).
 51. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
 52. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
 53. Bateman, A. et al. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
 54. Wilde, N. C., Isomura, M., Mendoza, A. & Baran, P. S. Two-phase synthesis of (-)-taxuyunnanin D. *J. Am. Chem. Soc.* **136**, 4909–4912 (2014).
 55. Lin, H. X., Jiang, Y., Chen, J. M., Chen, J. K. & Chen, M. Q. Synthesis and crystal structure of 2-debenzoyl and 4-deacetyl 1-deoxybaccatin VI derivatives. *J. Mol. Struct.* **738**, 59–65 (2005).
- performed the bioinformatics analyses; X.Y., D.M., G.Z., and Z.Z. supervised and coordinated the experiments; C.Y., Y.W., D.M., G.Z., and Z.Z. wrote the manuscript. C.Y., Y.W., and Z.S. contributed equally to this work.

Competing interests

These authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-46583-3>.

Correspondence and requests for materials should be addressed to Xing Yan, Dawei Ma, Guoping Zhao or Zhihua Zhou.

Peer review information *Nature Communications* thanks Hui Ming Ge, Reuben Peters and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Acknowledgements

This work was financially supported by the National Key Research and Development Program of China (Grant No. 2018YFA0900700, Z.Z.), the Strategic Biological Resources Service Network Plan (Grant No. KFJ-BRP-009, Z.Z.), the Strategic Priority Research Program (Grant No. XDB27020206, Z.Z.) and International Partnership Program (153D31KYSB20170121, G.Z.) of the Chinese Academy of Sciences, and the National Natural Science Foundation of China (Grant No. 31921006, Z.Z.) as well as GsynBioT (Shanghai) Co., Ltd. We thank Prof. David Nelson from the committee of standardized cytochrome P450 Nomenclature for naming P450s.

Author contributions

C.Y., Z.S., and P.W. performed the biological experiments; C.Y., Y.W., Z.S., L.X., and W.L. carried out the chemical experiments; C.Y. and X.Y.