

- 1 -

**MICROBIAL ENGINEERING FOR THE PRODUCTION OF CHEMICAL AND
PHARMACEUTICAL PRODUCTS FROM THE ISOPRENOID PATHWAY**

Related Applications

5 This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application Serial No. 61/280,877, entitled "Microbial Engineering for the Production of Chemical and Pharmaceutical Products from Isoprenoid Pathway," filed on November 10, 2009 and U.S. Provisional Application Serial No. 61/388,543, entitled "Microbial Engineering for the Production of Chemical and Pharmaceutical Products from Isoprenoid Pathway," filed on September 30, 2010, the entire disclosures of which are incorporated by
10 reference herein in their entireties.

Government Interest

 This work was funded in part by the National Institutes of Health under Grant
15 Number 1-R01-GM085323-01A1. The government has certain rights in this invention.

Field of the Invention

 The invention relates to the production of one or more terpenoids through microbial engineering.
20

Background of the Invention

 Taxol and its structural analogs have been recognized as the most potent and commercially successful anticancer drugs introduced in the last decade.¹ Taxol was first isolated from the bark of the Pacific Yew tree,² and early stage production methods required
25 sacrificing two to four fully grown trees to supply sufficient dosage for one patient.³ Taxol's structural complexity necessitated a complex chemical synthesis route requiring 35-51 steps with highest yield of 0.4%.^{4, 5, 6} However, a semi-synthetic route was devised whereby the biosynthetic intermediate baccatin III was first isolated from plant sources and was subsequently converted to Taxol.⁷ While this approach and subsequent plant cell culture-
30 based production efforts have decreased the need for harvesting the yew tree, production still depends on plant-based processes⁸ with accompanying limitations of productivity and

- 2 -

scalability, and constraints on the number of Taxol derivatives that can be synthesized in search for more efficacious drugs.^{9, 10}

Summary of the Invention

5 Recent developments in metabolic engineering and synthetic biology offer new possibilities for the overproduction of complex natural products through more technically amenable microbial hosts.^{11, 12} Although exciting progress has been made in the elucidation of the biosynthetic mechanism of Taxol in *Taxus*,¹³⁻¹⁶ commercially relevant Taxol-producing strains have eluded prior attempts aiming at the transfer of this complex
10 biosynthetic machinery into a microbial host.^{17, 18} Yet, as with other natural products, microbial production through metabolically engineered strains, offers attractive economics and great potential for synthesizing a diverse array of new compounds with anti-cancer and other pharmaceutical activity.^{19, 20}

 The metabolic pathway for Taxol and its analogs consists of an upstream *isoprenoid*
15 pathway that is native to *E. coli*, and a heterologous downstream terpenoid pathway (Fig. 6). The upstream mevalonic acid (MVA) or methylerythritol phosphate (MEP) pathways can produce the two common building blocks, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), from which Taxol and other isoprenoid compounds are formed.¹² Recent studies have highlighted the engineering of the above upstream pathways to support
20 biosynthesis of heterologous isoprenoids such as lycopene and artemisinic acid.²¹⁻²³ The downstream taxadiene pathway has been reconstructed in *E. coli*, but, to-date, titers have not exceeded 1.3 mg/L.²⁴

 The above rational metabolic engineering approaches focused on either the upstream (MVA or MEP) or the downstream terpenoid pathway, implicitly assuming that
25 modifications are additive, i.e. a linear behavior.²⁵⁻²⁷ While this approach can yield moderate increases in flux, it generally ignores non-specific effects, such as toxicity of intermediate metabolites, cellular effects of the vectors used for expression, and hidden unknown pathways that may compete with the main pathway and divert flux away from the desired target. Combinatorial approaches can avoid such problems as they offer the opportunity to
30 adequately sample the parameter space and elucidate these complex non-linear interactions.^{21, 28, 29, 30} However, they require a high throughput screen, which is often not available for many desirable natural products.³¹ Yet another class of pathway optimization

methods has explored the combinatorial space of different sources of the heterologous genes comprising the pathway of interest.³² Still dependent on a high throughput assay, these methods generally ignore the need for determining an optimal level of expression for the individual pathway genes and, as such, have proven less effective in structuring an optimal pathway.

In the present work, as an example of aspects of the invention, we focus on the optimal balancing between the upstream, IPP-forming pathway with the downstream terpenoid pathway of taxadiene synthesis. This is achieved by grouping the nine-enzyme pathway into two modules - a four-gene, upstream, native (MEP) pathway module and a two-gene, downstream, heterologous pathway to taxadiene (Fig. 1). Using this basic configuration, parameters such as the effect of plasmid copy number on cell physiology, gene order and promoter strength in an expression cassette, and chromosomal integration are evaluated with respect to their effect on taxadiene production. This modular and multivariable combinatorial approach allows us to efficiently sample the main parameters affecting pathway flux without the need for a high throughput screen. The multivariate search across multiple promoters and copy numbers for each pathway module reveals a highly non-linear taxadiene flux landscape with a global maximum exhibiting a 15,000 fold increase in taxadiene production over the control, yielding 300 mg/L production of taxadiene in small-scale fermentations. Further, we have engineered the P450 based oxidation chemistry in Taxol biosynthesis in *E. coli*, with our engineered strains improving the taxadien-5 α -ol production 2400-fold over the state of the art. These improvements unlock the potential for the large scale production of thousands of valuable terpenoids by well-established microbial systems.

In one aspect, the present invention refers to a method for increasing terpenoid production in a cell that produces one or more terpenoids, comprising

providing an *Escherichia coli* (*E. coli*) that produces isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) through an upstream methylerythritol pathway (MEP) and converts the IPP and DMAPP to a terpenoid through a recombinantly expressed downstream terpenoid synthesis pathway; and

controlling the accumulation of indole in the *E. coli* or in a culture of the *E. coli*, thereby increasing terpenoid production in the *E. coli*, wherein the indole accumulates in the *E. coli* culture at less than 100 mg/L, wherein the step of controlling the accumulation of indole in the *E. coli* or in a culture of the *E. coli* comprises balancing the upstream MEP pathway with

the downstream terpenoid synthesis pathway, modifying or regulating an indole pathway, and/or removing the accumulated indole from the cell culture through chemical methods, optionally using absorbents or scavengers.

Also described herein are methods involving recombinantly expressing a taxadiene
5 synthase enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme in a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway. In some examples the cell is a bacterial cell such as an *Escherichia coli* cell. In some examples, the bacterial cell is a Gram-positive cell such as a *Bacillus* cell. In some examples, the cell is a yeast cell such as a *Saccharomyces* cell or a *Yarrowia* cell. In some examples, the cell is an
10 algal cell or a plant cell.

In some examples, the taxadiene synthase enzyme is a *Taxus* enzyme such as a *Taxus brevifolia* enzyme. In some examples, the GGPPS enzyme is a *Taxus* enzyme such

as a *Taxus canadensis* enzyme. In some examples, the gene encoding for the taxadiene synthase enzyme and/or the gene encoding for the GGPPS enzyme and/or the genes encoding for the one or more components of the MEP pathway is expressed from one or more plasmids. In some examples, the gene encoding for the taxadiene synthase enzyme and/or the gene encoding for the GGPPS enzyme and/or the genes encoding for the one or more components of the MEP is incorporated into the genome of the cell.

In some examples, one or more components of the non-mevalonate (MEP) pathway are selected from the group consisting of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. In certain examples, *dxs*, *idi*, *ispD* and *ispF* are overexpressed. For example, *dxs*, *idi*, *ispD* and *ispF* can be overexpressed on the operon *dxs-idi-idpDF*. In some examples, the gene encoding for the taxadiene synthase enzyme and the gene encoding for the GGPPS enzyme are expressed together on an operon.

In some examples, the cell further expresses a taxadiene 5 α -hydroxylase (T5 α OH) or a catalytically active portion thereof. In certain examples, the T5 α OH enzyme or a catalytically active portion thereof is fused to a cytochrome P450 reductase enzyme or a catalytically active portion thereof. For example, the T5 α OH enzyme can be At24T5 α OH-tTCPR.

The expression of the taxadiene synthase enzyme, the GGPPS enzyme and the one or more components of the MEP pathway can be balanced to maximize production of the taxadiene. Methods described herein can further encompass culturing a cell to produce taxadiene or taxadiene-5 α -ol. In some examples, at least 10 mg L⁻¹ of taxadiene is produced. In certain examples, at least 250 mg L⁻¹ of taxadiene is produced. In some examples, at least 10 mg L⁻¹ of taxadiene-5 α -ol is produced. In certain examples, at least 50 mg L⁻¹ of taxadiene-5 α -ol is produced. In some examples, the percentage of taxadiene conversion to taxadiene-5 α -ol and the byproduct 5(12)-Oxa-3(11)-cyclotaxane is at least 50%, at least 75% or at least 95%.

Methods described herein can further comprise recovering the taxadiene or taxadiene-5 α -ol from the cell culture. In some examples, the taxadiene or taxadiene-5 α -ol is recovered from the gas phase while in other examples, an organic layer is added to the cell culture, and the taxadiene or taxadiene-5 α -ol is recovered from the organic layer.

Also described herein are cells that overexpress one or more components of the non-mevalonate (MEP) pathway, and that recombinantly expresses a taxadiene synthase

enzyme and a geranylgeranyl diphosphate synthase (GGPPS) enzyme. In some examples the cell is a bacterial cell such as an *Escherichia coli* cell. In some examples, the bacterial cell is a Gram-positive cell such as a *Bacillus* cell. In some examples, the cell is a yeast cell such as a *Saccharomyces* cell or a *Yarrowia* cell. In some examples, the cell is an algal cell or a plant cell.

In some examples, the taxadiene synthase enzyme is a *Taxus* enzyme such as a *Taxus brevifolia* enzyme. In some examples, the GGPPS enzyme is a *Taxus* enzyme such as a *Taxus canadensis* enzyme. In some examples, the gene encoding for the taxadiene synthase enzyme and/or the gene encoding for the GGPPS enzyme and/or the genes encoding for the one or more components of the MEP pathway is expressed from one or more plasmids. In some examples, the gene encoding for the taxadiene synthase enzyme and/or the gene encoding for the GGPPS enzyme and/or the genes encoding for the one or more components of the MEP is incorporated into the genome of the cell.

In some examples, the one or more components of the non-mevalonate (MEP) pathway is selected from the group consisting of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. In certain examples, *dxs*, *idi*, *ispD* and *ispF* are overexpressed. For example, *dxs*, *idi*, *ispD* and *ispF* can be overexpressed on the operon *dxs-idi-idpDF*. In some examples, the gene encoding for the taxadiene synthase enzyme and the gene encoding for the GGPPS enzyme are expressed together on an operon. In some examples, the expression of the taxadiene synthase enzyme, the GGPPS enzyme and the one or more components of the MEP pathway are balanced to maximize production of the taxadiene.

In some examples, the cell further expresses a taxadiene 5 α -hydroxylase (T5 α OH) or a catalytically active portion thereof. In certain examples, the T5 α OH enzyme or a catalytically active portion thereof is fused to a cytochrome P450 reductase enzyme or a catalytically active portion thereof. For example, the T5 α OH enzyme can be At24T5 α OH-tTCPR. In some examples, the cell produces taxadiene and/or taxadiene-5 α -ol.

Also described herein are methods for selecting a cell that exhibits enhanced production of a terpenoid, including creating or obtaining a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway, producing terpenoid from the cell, comparing the amount of terpenoid produced from the cell to the amount of terpenoid produced in a control cell, and selecting a first improved cell that produces a higher amount

of terpenoid than a control cell, wherein a first improved cell that produces a higher amount of terpenoid than the control cell is a cell that exhibits enhanced production of terpenoid.

In some examples, the cell recombinantly expresses a terpenoid synthase enzyme and/or a geranylgeranyl diphosphate synthase (GGPPS) enzyme. Methods can further
5 comprise altering the level of expression of one or more of the components of the non-mevalonate (MEP) pathway, the terpenoid synthase enzyme and/or the geranylgeranyl diphosphate synthase (GGPPS) enzyme in the first improved cell to produce a second improved cell, and comparing the amount of terpenoid produced from the second improved cell to the amount of terpenoid produced in the first improved cell, wherein a second
10 improved cell that produces a higher amount of terpenoid than the first improved cell is a cell that exhibits enhanced production of terpenoid. In some examples, the terpenoid synthase enzyme is a taxadiene synthase enzyme. The cell can further recombinantly express any of the polypeptides described herein.

Also described herein are isolated polypeptides comprising a taxadiene 5 α -
15 hydroxylase (T5 α OH) enzyme or a catalytically active portion thereof fused to a cytochrome P450 reductase enzyme or a catalytically active portion thereof. In some examples, the cytochrome P450 reductase enzyme is a *Taxus* cytochrome P450 reductase (TCPR). In certain examples, the taxadiene 5 α -hydroxylase and TCPR are joined by a linker such as GSTGS (SEQ ID NO:50). In some examples, the taxadiene 5 α -hydroxylase and/or TCPR are
20 truncated to remove all or part of the transmembrane region. In certain examples, 8, 24, or 42 N-terminal amino acids of taxadiene 5 α -hydroxylase are truncated. In certain examples, 74 amino acids of TCPR are truncated. In some examples, an additional peptide is fused to taxadiene 5 α -hydroxylase. In certain examples, the additional peptide is from bovine 17 α hydroxylase. In certain examples, the peptide is MALLLAVF (SEQ ID NO:51). In certain
25 examples, the isolated polypeptide is At24T5 α OH-tTCPR. Also disclosed herein are nucleic acid molecules that encode for any of the polypeptides described herein and cells that recombinantly express any of the polypeptides described herein.

Also disclosed herein are methods for increasing terpenoid production in a cell that produces one or more terpenoids. The methods include controlling the accumulation of
30 indole in the cell or in a culture of the cells, thereby increasing terpenoid production in a cell. Any of the cells described herein can be used in the methods, including bacterial cells,

such as *Escherichia coli* cells; Gram-positive cells, such as *Bacillus* cells; yeast cells, such as *Saccharomyces* cells or *Yarrowia* cells; algal cells; plant cell; and any of the engineered cells described herein.

In some examples, the step of controlling the accumulation of indole in the cell or in a culture of the cells includes balancing the upstream non-mevalonate isoprenoid pathway with the downstream product synthesis pathways and/or modifying or regulating the indole pathway. In other examples, the step of controlling the accumulation of indole in the cell or in a culture of the cells includes or further includes removing the accumulated indole from the fermentation through chemical methods, such as by using absorbents or scavengers.

The one or more terpenoids produced by the cell(s) or in the culture can be a monoterpene, a sesquiterpene, a diterpene, a triterpene or a tetraterpene. In certain examples, the terpenoids is taxadiene or any taxol precursor.

Also disclosed herein are methods that include measuring the amount or concentration of indole in a cell that produces one or more terpenoids or in a culture of the cells that produce one or more terpenoids. The methods can include measuring the amount or concentration of indole two or more times. In some examples, the measured amount or concentration of indole is used to guide a process of producing one or more terpenoids. In some examples, the measured amount or concentration of indole is used to guide strain construction.

These and other aspects of the invention, as well as various embodiments thereof, will become more apparent in reference to the drawings and detailed description of the invention.

Brief Description of the Drawings

The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

Figure 1. Multivariate-modular isoprenoid pathway engineering reveals strong non-linear response in terpene accumulation. To increase the flux through the upstream MEP pathway, we targeted reported bottleneck enzymatic steps (dxs, idi, ispD and ispF) for overexpression by an operon (dxs-idi-ispDF).²⁸ To channel the overflow flux from the

- 8 -

universal isoprenoid precursors, IPP and DMAPP, towards Taxol biosynthesis, a synthetic operon of downstream genes GGPP synthase (G) and Taxadiene synthase (T)¹⁶ was constructed. The upstream isoprenoid and downstream synthetic taxadiene pathways were placed under the control of inducible promoters to control their relative gene expression. (a)

5 Schematic of the two modules, the native upstream MEP isoprenoid pathway (left) and synthetic taxadiene pathway (right). In *E. coli* biosynthetic network, the MEP isoprenoid pathway is initiated by the condensation of the precursors glyceraldehydes-3 phosphate (G3P) and pyruvate (PYR) from glycolysis. The Taxol pathway bifurcation starts from the universal isoprenoid precursors IPP and DMAPP to form first the “linear” precursor

10 Geranylgeranyl diphosphate, and then the “cyclic” taxadiene, a committed and key intermediate to Taxol. The cyclic olefin taxadiene undergoes multiple rounds of stereospecific oxidations, acylations, benzylation with side chain assembly to, ultimately, form Taxol. (b) Schematic of the multivariate-modular isoprenoid pathway engineering approach for probing the non-linear response in terpenoid accumulation from upstream and

15 downstream pathway engineered cells. Expression of upstream and downstream pathways is modulated by varying the promoter strength (Trc, T5 and T7) or increasing the copy number using different plasmids. Variation of upstream and downstream pathway expression gives different maxima in taxadiene accumulation.

20 **Figure 2. Optimization of taxadiene production by regulating the expression of the up- and down-stream modular pathways.** (a) Response in taxadiene accumulation to the increase in upstream pathway strengths for constant values of the downstream pathway. (b) the dependence on the downstream pathway for constant increases in the upstream pathway strength. Observed multiple local maxima in taxadiene response depends on the increase in

25 the pathway expression strength upstream or downstream. (c) Taxadiene response from strains engineered (17-24) with high upstream pathway overexpressions (20-100) with two different downstream expressions (~30 and ~60) to identify taxadiene response with balanced expressions. Expression of downstream pathway from the low copy plasmid (p5 and p10) under strong promoter T7TG operon was used to modulate these expressions. Note that both

30 upstream and downstream pathway expressed from different plasmids with different promoters can impose plasmid born metabolic burden. (d) Modulating the upstream pathway with increasing promoter strength from chromosome with two different downstream

- 9 -

expressions (~30 and ~60) to identify the missing search space with reduced toxic effects (strains **25-32**). (e) Genetic details of the taxadiene producing strains. The numbers corresponding to different strains and its corresponding genotype, *E. coli* K12mG1655 Δ recA Δ endA, EDE3-*E. coli* K12mG1655 Δ recA Δ endA with T7 RNA polymerase DE3 construct in the chromosome, MEP – dxs-idi-ispDF operon, GT- GPPS-TS operon, TG- TS-GPPS operon, Ch1 – 1 copy in chromosome, Trc – Trc promoter, T5 – T5 promoter, T7 – T7 promoter, p5, p10, p20 – ~5 (SC101), ~10 (p15), and ~20 (pBR322) copy plasmid.

Figure 3. Metabolite inversely correlates with taxadiene production. (a) Mass spectrum of metabolite that was detected to correlate inversely with taxadiene production in the strain constructs of Fig. 2. The observed characteristic peaks of the metabolite are 233, 207, 178, 117, 89 and 62. (b) Correlation between the isoprenoid byproduct of Fig. 3a and taxadiene. Strains **26-29** and **30-32**, all with chromosomally integrated upstream pathway expression, were chosen for consistent comparison. In strains **26-29** and **30-32**, upstream expression increased by changing the promoters from Trc, to T5 and T7 respectively. The two sets of strains differ only in the expression of the downstream pathway with the second set (**30-32**) having twice the level of expression of the first. With the first set, optimal balancing is achieved with strain **26**, which uses the Trc promoter for upstream pathway expression and also shows the lowest metabolite accumulation. With strains **30-32**, strain **31** shows the lowest accumulation of metabolite and highest production of taxadiene. The data demonstrate the inverse correlation observed between the unknown metabolite and taxadiene production.

Figure 4. Upstream and downstream pathway transcriptional gene expression levels and changes in cell physiology of engineered strains. Relative expression of the first genes in the operon of upstream (DXS) and downstream (TS) pathway is quantified by qPCR. Similar expression profiles were observed with the genes in the downstream of the operons. The corresponding strain numbers are shown in the graph. (a) Relative transcript level DXS gene expression quantified from different upstream expressions modulated using promoters and plasmids under two different downstream expressions. (b) Relative transcript level TS gene expression quantified from two different downstream expression modulated using p5T7 and p10T7 plasmids under different upstream expressions. Our gene expression analysis directly supported the hypothesis, with increase in plasmid copy number (5, 10 and 20) and

- 10 -

promoter strength (Trc, T5 and T7) the expression of the upstream and downstream pathways can be modulated. (c) Cell growth of the engineered strains 25-29. The growth phenotype was affected from activation of isoprenoid metabolism (strain 26), recombinant protein expression (strain 25) and plasmid born metabolic burden (control vs engineered strains) and (d) growth phenotypes of strains 17, 22, 25-32. The black color lines are the taxadiene producing engineered strains and the gray color lines are control strains without downstream expression carrying an empty plasmid with promoter and multi cloning sites. The growth was correlated to the activation of the terpenoid metabolism, plasmid born metabolic burden as well the recombinant protein expression.

10

Figure 5. Engineering Taxol p450 oxidation chemistry in *E. coli*. (a) Schematics of the conversion of taxadiene to taxadiene 5 α -ol to Taxol. (b) Transmembrane engineering and construction of one-component chimera protein from taxadiene 5 α -ol hydroxylase (T5 α OH) and Taxus cytochrome p450 reductase (TCPR). **1** and **2** represents the full length proteins of T5 α OH and TCPR identified with 42 and 74 amino acid TM regions respectively, **3** – chimera enzymes generated from the three different TM engineered T5 α OH constructs, (At8T5 α OH, At24T5 α OH and At42T5 α OH constructed by fusing 8 residue synthetic peptide (A) to 8, 24 and 42 AA truncated T5 α OH) through a translational fusion with 74 AA truncated TCPR (tTCPR) using 5 residue GSTGS linker peptide. (c) Functional activity of At8T5 α OH-tTCPR, At24T5 α OH-tTCPR and At42T5 α OH-tTCPR constructs transformed into taxadiene producing strain 18. (d) Time course profile of taxadien-5 α -ol accumulation and growth profile of the strain 18-At24T5 α OH-tTCPR fermented in a 1L bioreactor.

Figure 6. Biosynthetic scheme for taxol production in *E. coli*. Schematics of the two modules, native upstream isoprenoid pathway (left) and synthetic Taxol pathway (right). In *E. coli* biosynthetic network, divergence of MEP isoprenoid pathway initiates from the precursors glyceraldehyde-3 phosphate (G3P) and Pyruvate (PYR) from glycolysis (I-V). The Taxol pathway bifurcation starts from the *E. coli* isoprenoid precursor IPP and DMAPP to “linear” precursor Geranylgeranyl diphosphate (VIII), “cyclic” taxadiene (IX), “oxidized” taxadiene 5 α -ol (X) to multiple rounds of stereospecific oxidations, acylations, benzylation and epoxidation for early precursor Baccatin III (XII) and finally with side chain assembly to Taxol (XIII). DXP - 1-deoxy-D-xylulose-5-phosphate, MEP- 2C-methyl-D-erythritol-4-

30

- 11 -

phosphate, CDP-ME - 4-diphosphocytidyl-2C-methyl-D-erythritol, CDP-MEP - 4-diphosphocytidyl- 2C-methyl-D-erythritol-2-phosphate, ME-cPP - 2C-methyl-D-erythritol-2,4-cyclodiphosphate, IPP- isopentenyl diphosphate, DMAPP - dimethylallyl diphosphate. The genes involved biosynthetic pathways from G3P and PYR to Taxol. DXS-1-deoxy-D-
 5 xylulose-5-phosphate synthase, ispC-1-Deoxy-D-xylulose-5-phosphate reductoisomerase, IspD-4-diphosphocytidyl-2C-methyl-D-erythritol synthase, IspE-4-diphosphocytidyl-2-C-methyl- D-erythritol kinase, IspF-2C-Methyl-D-erythritol-2,4-cyclodiphosphate Synthase, IspG-1-hydroxy-2-methyl-2-(*E*)-butenyl-4- diphosphate synthase, IspH-4-hydroxy-3-methyl-2-(*E*)-butenyl-4-diphosphate reductase IDI-isopentenyl-diphosphate isomerase, GGPPS-
 10 geranyl geranyldiphosphate synthase, Taxadiene synthase, Taxoid 5 α -hydroxylase, Taxoid-5 α -*O*-acetyltransferase, Taxoid 13 α -hydroxylase, Taxoid 10 β -hydroxylase, Taxoid 2 α -hydroxylase, Taxoid 2-*O*-benzoyltransferase, Taxoid 7 β -hydroxylase, Taxoid 10-*O*-acetyltransferase, Taxoid 1 β -hydroxylase*, Taxoid 9 α -hydroxylase, Taxoid 9-keto-oxidase*, Taxoid C4,C20- β - epoxidase*, Phenylalanine aminomutase, Side chain CoA-ligase*, Taxoid
 15 13 *O*-phenylpropanoyltransferase, Taxoid 2'-hydroxylase*, Taxoid 3'-N-benzoyltransferase.^{216,219} * marked genes are yet to be identified or characterized.

Figure 7. Fold improvements in taxadiene production from the modular pathway expression search. (a) Taxadiene response in fold improvements from all the observed
 20 maximas from Figure 2a,b, and c compared to strain 1. The 2.5 fold differences between two highest maximas (strain **17** and **26**) and 23 fold (strain **26** and **10**) with lowest one indicates that missing an optimal response results in significantly lower titers.

Figure 8. Metabolite (a) Correlation between taxadiene to metabolite accumulation. The
 25 metabolite accumulation from the engineered strain is anti-proportionally related to the taxadiene production in an exponential manner. The correlation coefficient for this relation was determined to 0.92 (b) Representative GC-profile from the strains 26-28 to demonstrate the change in taxadiene and metabolite accumulation. Numbers in the chromatogram 1 and 2 corresponding to metabolite and taxadiene peak respectively. (c) GC-MS profile of
 30 metabolite (1) and taxadiene (2) respectively. The observed characteristic peaks of the metabolite are 233, 207, 178, 117, 89 and 62. Taxa-4(20),11,12-diene characteristic ion m/z

- 12 -

272(P⁺), 257 (P⁺-CH₃), 229 (P⁺-C₃H₇); 121, 122, 123 (C-ring fragment cluster).⁶⁰ The peak marked with a star is the internal standard caryophyllene.

Figure 9. GC-MS profiles and taxadiene/taxadien-5 α -ol production from artificial

5 **chimera enzyme engineered in strain 26.** (a) GC profile of the hexane:ether (8:2) extract from three constructs (A-At8T5 α OH-tTCPR, t24T5 α OH-tTCPR and At42T5 α OH-tTCPR) transferred to strain 26 and fermented for 5 days. 1, 2 and 3 labels in the peaks corresponding to the taxadiene, taxadien-5 α -ol and 5(12)-Oxa-3(11)-cyclotaxane (OCT) respectively. (b) The production of taxa-4(20),11,12-dien-5 α -ol and OCT quantified from the three strains. (c) 10 and (d) GC-MS profile of taxa-4(20),11,12-dien-5 α -ol and OCT and the peaks corresponding to the fragmentation was compared with the authentic standards and previous reports^{42,47} GC-MS analysis confirmed the mass spectrum identity to authentic taxa-4(20),11,12-dien-5 α -ol with characteristic ion m/z 288(P⁺), 273 (P⁺-H₂O), 255 (P⁺-H₂O-CH₃).

15 **Figure 10** presents a schematic depicting the terpenoid biosynthetic pathway and natural products produced by this pathway.

Figure 11 presents a schematic depicting modulation of the upstream pathway for amplifying taxadiene production.

20

Figure 12 presents a schematic depicting modulation of the downstream pathway for amplifying taxadiene production.

25 **Figure 13** presents a schematic indicating that the newly identified pathway diversion is not the characteristic of downstream synthetic pathway.

Figure 14. Pathway strength correlates to transcriptional gene expression levels. (c) relative expression of idi, ispD and ispF genes with increasing upstream pathway strength and downstream strength at 31 arbitrary units, and (d) relative expression of idi, ispD and ispF 30 genes with increasing upstream pathway strength and downstream strength at 61 arbitrary units. As expected the gene expression increased as the upstream pathway strength increased. The corresponding strain numbers are indicated in the bar graph. The relative expression was

- 13 -

quantified using the expression of the housekeeping *rrsA* gene. Data are mean +/- SD for four replicates.

Figure 15. Impact of metabolite byproduct Indole accumulation on taxadiene

production and growth. (a) Inverse correlation between taxadiene and Indole. Strains 26 to 28 and 30 to 32, all with chromosomally integrated upstream pathway expression, were chosen for consistent comparison. The two sets of strains differ only in the expression of the downstream pathway with the second set (30 to 32) having twice the level of expression of the first. In strains 26 to 28 and 30 to 32, upstream expression increased by changing the promoters from *Trc*, to *T5* and *T7*, respectively. With the first set, optimal balancing is achieved with strain 26, which uses the *Trc* promoter for upstream pathway expression and also shows the lowest indole accumulation. With strains 30 to 32, strain 31 shows the lowest accumulation of indole and highest production of taxadiene. The fold improvements are relative to strain 25 and 29, respectively, for the two sets. (b) Effect of externally-introduced indole on taxadiene production for the high-producing strain 26. Different concentrations of indole were introduced into cultures of cells cultured in minimal media with 0.5% yeast extract. Taxadiene production was significantly reduced as indole concentration increased from 50 mg/L to 100 mg/L (c) Effect of externally-introduced indole on cell growth for engineered strains of *E. coli*. Data are mean +/- SD for three replicates. Strains devoid of the downstream pathway and with different strengths of the upstream pathway (1, 2, 6, 21, 40 and 100) were selected. Strain 26, the high taxadiene producer, exhibits the strongest inhibition.

Figure 16. Unknown metabolite identified as indole. (A) and (a) Gas chromatogram and mass spectrum of the unknown metabolite extracted using hexane from cell culture. (B) and (b) correspond to the gas chromatogram and mass spectrum of pure indole dissolved in hexane. Further to confirm the chemical identity, the metabolite was extracted from the fermentation broth using hexane extraction and purified by silica column chromatography using hexane:ethylacetate (8:2) as eluent. The purity of the compound was confirmed by TLC and GC-MS. ¹HNMR and ¹³CNMR spectra confirmed the chemical identity of the metabolite as indole. (c) ¹HNMR spectrum of indole extracted from cell culture (CDCl₃, 400 MHz) δ: 6.56 (d, 1H, Ar C-H), 7.16 (m, 3H, Ar C-H), 7.38 (d, 1H, Ar C-H), 7.66 (d, 1H, Ar C-H),

- 14 -

8.05 (b, 1H, Indole NH). (d) ^{13}C NMR δ : 135.7, 127.8, 124.2, 122, 120.7, 119.8, 111, 102.6. (e) is the ^1H NMR spectrum of pure indole.

Figure 17. Fed batch cultivation of engineered strains in 1L-bioreactor. Time courses of taxadiene accumulation (a), cell growth (b), acetic acid accumulation (c) and total substrate (glycerol) addition (d) for strains 22, 17 and 26 during 5 days of fed batch bioreactor cultivation in 1L-bioreactor vessels under controlled pH and oxygen conditions with minimal media and 0.5% yeast extract. After glycerol depletes to ~0.5 to 1 g/L in the fermentor, 3 g/L of glycerol was introduced into the bioreactor during the fermentation. Data are mean of two replicate bioreactors.

Detailed Description of the Invention

Taxol is a potent anticancer drug first isolated as a natural product from the *Taxus brevifolia* Pacific yew tree. However, reliable and cost-efficient production of Taxol or Taxol analogs by traditional production routes from plant extracts is limited. Here, we report a multivariate-modular approach to metabolic pathway engineering to amplify by ~15000 fold the production of taxadiene in an engineered *Escherichia coli*. Taxadiene, the first committed Taxol intermediate, is the biosynthetic product of the non-mevalonate pathway in *E. coli* comprising two modules: the native upstream pathway forming Isopentenyl Pyrophosphate (IPP) and a heterologous downstream terpenoid-forming pathway. Systematic multivariate search identified conditions that optimally balance the two pathway modules to minimize accumulation of inhibitory intermediates and flux diversion to side products. We also engineered the next step, after taxadiene, in Taxol biosynthesis, a P450-based oxidation step, that yielded >98% substrate conversion and present the first example of *in vivo* production of any functionalized Taxol intermediates in *E. coli*. The modular pathway engineering approach not only highlights the complexity of multi-step pathways, but also allowed accumulation of high taxadiene and taxadien-5 α -ol titers (~300mg/L and 60mg/L, respectively) in small-scale fermentations, thus exemplifying the potential of microbial production of Taxol and its derivatives.

This invention is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being

- 15 -

carried out in various ways. Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of “including,” “comprising,” or “having,” “containing,” “involving,” and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

5 Microbial production of terpenoids such as taxadiene is demonstrated herein. When expressed at satisfactory levels, microbial routes reduce dramatically the cost of production of such compounds. Additionally, they utilize cheap, abundant and renewable feedstocks (such as sugars and other carbohydrates) and can be the source for the synthesis of numerous derivatives that may exhibit far superior properties than the original compound. A key
10 element in the cost-competitive production of compounds of the isoprenoid pathway using a microbial route is the amplification of this pathway in order to allow the overproduction of these molecules. Described herein are methods that enhance or amplify the flux towards terpenoid production in *Escherichia coli* (*E. coli*). Specifically, methods are provided to amplify the metabolic flux to the synthesis of isopentenyl pyrophosphate (IPP) (a key
15 intermediate for the production of isoprenoid compounds), dimethylallyl pyrophosphate (DMAPP), geranyl diphosphate (GPP), farnesyl diphosphate (FPP), geranylgeranyl diphosphate (GGPP), and farnesyl geranyl diphosphate (FGPP), paclitaxel (Taxol), ginkgolides, geraniol, farnesol, geranylgeraniol, linalool, isoprene, monoterpenoids such as menthol, carotenoids such as lycopene, polyisoprenoids such as polyisoprene or natural
20 rubber, diterpenoids such as eleutherobin, and sesquiterpenoids such as artemisinin.

Aspects of the invention relate to the production of terpenoids. As used herein, a terpenoid, also referred to as an isoprenoid, is an organic chemical derived from a five-carbon isoprene unit. Several non-limiting examples of terpenoids, classified based on the number of isoprene units that they contain, include: hemiterpenoids (1 isoprene unit), monoterpenoids
25 (2 isoprene units), sesquiterpenoids (3 isoprene units), diterpenoids (4 isoprene units), sesterterpenoids (5 isoprene units), triterpenoids (6 isoprene units), tetraterpenoids (8 isoprene units), and polyterpenoids with a larger number of isoprene units. In some embodiments, the terpenoid that is produced is taxadiene. In some embodiments, the terpenoid that is produced is Citronellol, Cubebol, Nootkatone, Cineol, Limonene,
30 Eleutherobin, Sarcodictyin, Pseudopterოსins, Ginkgolides, Stevioside, Rebaudioside A, sclareol, labdenediol, levopimaradiene, sandracopimaradiene or isopemaradiene.

- 16 -

Described herein are methods and compositions for optimizing production of terpenoids in cells by controlling expression of genes or proteins participating in an upstream pathway and a downstream pathway. The upstream pathway involves production of isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), which can be achieved by two different metabolic pathways: the mevalonic acid (MVA) pathway and the MEP (2-C-methyl-D-erythritol 4-phosphate) pathway, also called the MEP/DOXP (2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate) pathway, the non-mevalonate pathway or the mevalonic acid-independent pathway.

The downstream pathway is a synthetic pathway that leads to production of a terpenoids and involves recombinant gene expression of a terpenoid synthase (also referred to as terpene cyclase) enzyme, and a geranylgeranyl diphosphate synthase (GGPPS) enzyme. In some embodiments, a terpenoid synthase enzyme is a diterpenoid synthase enzyme. Several non-limiting examples of diterpenoid synthase enzymes include casbene synthase, taxadiene synthase, levopimaradiene synthase, abietadiene synthase, isopimaradiene synthase, *ent*-copalyl diphosphate synthase, *syn*-stemar-13-ene synthase, *syn*-stemod-13(17)-ene synthase, *syn*-pimara-7,15-diene synthase, *ent*-sandaracopimaradiene synthase, *ent*-cassa-12,15-diene synthase, *ent*-pimara-8(14), 15-diene synthase, *ent*-kaur-15-ene synthase, *ent*-kaur-16-ene synthase, aphidicolan-16 β -ol synthase, phyllocladan-16 α -ol synthase, fusicocca-2,10(14)-diene synthase, and terpentetriene cyclase.

Surprisingly, as demonstrated in the Examples section, optimization of terpenoid synthesis by manipulation of the upstream and downstream pathways described herein, was not a simple linear or additive process. Rather, through complex combinatorial analysis, optimization was achieved through balancing components of the upstream and downstream pathways. Unexpectedly, as demonstrated in Figures 1 and 2, taxadiene accumulation exhibited a strong non-linear dependence on the relative strengths of the upstream MEP and downstream synthetic taxadiene pathways.

Aspects of the invention relate to controlling the expression of genes and proteins in the MEP pathway for optimized production of a terpenoid such as taxadiene. Optimized production of a terpenoid refers to producing a higher amount of a terpenoid following pursuit of an optimization strategy than would be achieved in the absence of such a strategy. It should be appreciated that any gene and/or protein within the MEP pathway is encompassed by methods and compositions described herein. In some embodiments, a gene

- 17 -

within the MEP pathway is one of the following: *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* or *ispB*. Expression of one or more genes and/or proteins within the MEP pathway can be upregulated and/or downregulated. In certain embodiments, upregulation of one or more genes and/or proteins within the MEP pathway can be combined with downregulation of one or more genes and/or proteins within the MEP pathway.

It should be appreciated that genes and/or proteins can be regulated alone or in combination. For example, the expression of *dxs* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. The expression of *ispC* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. The expression of *ispD* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. The expression of *ispE* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. The expression of *ispF* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*. The expression of *ispG* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispH*, *idi*, *ispA* and *ispB*. The expression of *ispH* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *idi*, *ispA* and *ispB*. The expression of *idi* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *ispA* and *ispB*. The expression of *ispA* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi* and *ispB*. The expression of *ispB* can be upregulated or downregulated alone or in combination with upregulation or downregulation of expression of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi* and *ispA*. In some embodiments, expression of the gene and/or protein of one or more of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, and *idi* is upregulated while expression of the gene and/or protein of *ispA* and/or *ispB* is downregulated.

- 18 -

Expression of genes within the MEP pathway can be regulated in a modular method. As used herein, regulation by a modular method refers to regulation of multiple genes together. For example, in some embodiments, multiple genes within the MEP pathway are recombinantly expressed on a contiguous region of DNA, such as an operon. It should be appreciated that a cell that expresses such a module can also express one or more other genes within the MEP pathway either recombinantly or endogenously.

A non-limiting example of a module of genes within the MEP pathway is a module containing the genes *dxs*, *idi*, *ispD* and *ispF*, as presented in the Examples section, and referred to herein as *dxs-idi-ispDF*. It should be appreciated that modules of genes within the MEP pathway, consistent with aspects of the invention, can contain any of the genes within the MEP pathway, in any order.

Expression of genes and proteins within the downstream synthetic terpenoid synthesis pathway can also be regulated in order to optimize terpenoid production. The synthetic downstream terpenoid synthesis pathway involves recombinant expression of a terpenoid synthase enzyme and a GGPPS enzyme. Any terpenoid synthase enzyme, as discussed above, can be expressed with GGPPS depending on the downstream product to be produced. For example, taxadiene synthase is used for the production of taxadiene. Recombinant expression of the taxadiene synthase enzyme and the GGPPS enzyme can be regulated independently or together. In some embodiments the two enzymes are regulated together in a modular fashion. For example the two enzymes can be expressed in an operon in either order (GGPPS-TS, referred to as "GT," or TS-GGPPS, referred to as "TG").

Manipulation of the expression of genes and/or proteins, including modules such as the *dxs-idi-ispDF* operon, and the TS-GGPPS operon, can be achieved through methods known to one of ordinary skill in the art. For example, expression of the genes or operons can be regulated through selection of promoters, such as inducible promoters, with different strengths. Several non-limiting examples of promoters include Trc, T5 and T7. Additionally, expression of genes or operons can be regulated through manipulation of the copy number of the gene or operon in the cell. For example, in certain embodiments, a strain containing an additional copy of the *dxs-idi-ispDF* operon on its chromosome under Trc promoter control produces an increased amount of taxadiene relative to one overexpressing only the synthetic downstream pathway. In some embodiments, expression of genes or operons can be regulated through manipulating the order of the genes within a module. For example, in

- 19 -

certain embodiments, changing the order of the genes in a downstream synthetic operon from GT to TG results in a 2-3 fold increase in taxadiene production. In some embodiments, expression of genes or operons is regulated through integration of one or more genes or operons into a chromosome. For example, in certain embodiments, integration of the
5 upstream dxs-idi-ispDF operon into the chromosome of a cell results in increased taxadiene production.

It should be appreciated that the genes associated with the invention can be obtained from a variety of sources. In some embodiments, the genes within the MEP pathway are bacterial genes such as *Escherichia coli* genes. In some embodiments, the gene encoding for
10 GGPPS is a plant gene. For example, the gene encoding for GGPPS can be from a species of *Taxus* such as *Taxus canadensis* (*T. canadensis*). In some embodiments, the gene encoding for taxadiene synthase is a plant gene. For example, the gene encoding for taxadiene synthase can be from a species of *Taxus* such as *Taxus brevifolia* (*T. brevifolia*). Representative GenBank Accession numbers for *T. canadensis* GGPPS and *T. brevifolia*
15 taxadiene synthase are provided by AF081514 and U48796, the sequences of which are incorporated by reference herein in their entireties.

As one of ordinary skill in the art would be aware, homologous genes for use in methods associated with the invention can be obtained from other species and can be identified by homology searches, for example through a protein BLAST search, available at
20 the National Center for Biotechnology Information (NCBI) internet site (www.ncbi.nlm.nih.gov). Genes and/or operons associated with the invention can be cloned, for example by PCR amplification and/or restriction digestion, from DNA from any source of DNA which contains the given gene. In some embodiments, a gene and/or operon associated with the invention is synthetic. Any means of obtaining a gene and/or operon associated with
25 the invention is compatible with the instant invention.

In some embodiments, further optimization of terpenoid production is achieved by modifying a gene before it is recombinantly expressed in a cell. In some embodiments, the GGPPS enzyme has one or more of the follow mutations: A162V, G140C, L182M, F218Y, D160G, C184S, K367R, A151T, M185I, D264Y, E368D, C184R, L331I, G262V, R365S,
30 A114D, S239C, G295D, I276V, K343N, P183S, I172T, D267G, I149V, T234L, E153D and T259A. In some embodiments, the GGPPS enzyme has a mutation in residue S239 and/or

- 20 -

residue G295. In certain embodiments, the GGPPS enzyme has the mutation S239C and/or G295D.

In some embodiments, modification of a gene before it is recombinantly expressed in a cell involves codon optimization for expression in a bacterial cell. Codon usages for a variety of organisms can be accessed in the Codon Usage Database (www.kazusa.or.jp/codon/). Codon optimization, including identification of optimal codons for a variety of organisms, and methods for achieving codon optimization, are familiar to one of ordinary skill in the art, and can be achieved using standard methods.

In some embodiments, modifying a gene before it is recombinantly expressed in a cell involves making one or more mutations in the gene before it is recombinantly expressed in a cell. For example, a mutation can involve a substitution or deletion of a single nucleotide or multiple nucleotides. In some embodiments, a mutation of one or more nucleotides in a gene will result in a mutation in the protein produced from the gene, such as a substitution or deletion of one or more amino acids.

In some embodiments, it may be advantageous to use a cell that has been optimized for production of a terpenoid. For example, in some embodiments, a cell that overexpresses one or more components of the non-mevalonate (MEP) pathway is used, at least in part, to amplify isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), substrates of GGPPS. In some embodiments, overexpression of one or more components of the non-mevalonate (MEP) pathway is achieved by increasing the copy number of one or more components of the non-mevalonate (MEP) pathway. For example, copy numbers of components at rate-limiting steps in the MEP pathway such as (*dxs*, *ispD*, *ispF*, *idi*) can be amplified, such as by additional episomal expression.

In some embodiments “rational design” is involved in constructing specific mutations in proteins such as enzymes. As used herein, “rational design” refers to incorporating knowledge of the enzyme, or related enzymes, such as its three dimensional structure, its active site(s), its substrate(s) and/or the interaction between the enzyme and substrate, into the design of the specific mutation. Based on a rational design approach, mutations can be created in an enzyme which can then be screened for increased production of a terpenoid relative to control levels. In some embodiments, mutations can be rationally designed based on homology modeling. As used herein, “homology modeling” refers to the process of

- 21 -

constructing an atomic resolution model of one protein from its amino acid sequence and a three-dimensional structure of a related homologous protein.

In some embodiments, random mutations can be made in a gene, such as a gene encoding for an enzyme, and these mutations can be screened for increased production of a terpenoid relative to control levels. For example, screening for mutations in components of the MEP pathway, or components of other pathways, that lead to enhanced production of a terpenoid may be conducted through a random mutagenesis screen, or through screening of known mutations. In some embodiments, shotgun cloning of genomic fragments could be used to identify genomic regions that lead to an increase in production of a terpenoid, through screening cells or organisms that have these fragments for increased production of a terpenoid. In some cases one or more mutations may be combined in the same cell or organism.

In some embodiments, production of a terpenoid in a cell can be increased through manipulation of enzymes that act in the same pathway as the enzymes associated with the invention. For example, in some embodiments it may be advantageous to increase expression of an enzyme or other factor that acts upstream of a target enzyme such as an enzyme associated with the invention. This could be achieved by over-expressing the upstream factor using any standard method.

Optimization of protein expression can also be achieved through selection of appropriate promoters and ribosome binding sites. In some embodiments, this may include the selection of high-copy number plasmids, or low or medium-copy number plasmids. The step of transcription termination can also be targeted for regulation of gene expression, through the introduction or elimination of structures such as stem-loops.

Aspects of the invention relate to expression of recombinant genes in cells. The invention encompasses any type of cell that recombinantly expresses genes associated with the invention, including prokaryotic and eukaryotic cells. In some embodiments the cell is a bacterial cell, such as *Escherichia* spp., *Streptomyces* spp., *Zymonas* spp., *Acetobacter* spp., *Citrobacter* spp., *Synechocystis* spp., *Rhizobium* spp., *Clostridium* spp., *Corynebacterium* spp., *Streptococcus* spp., *Xanthomonas* spp., *Lactobacillus* spp., *Lactococcus* spp., *Bacillus* spp., *Alcaligenes* spp., *Pseudomonas* spp., *Aeromonas* spp., *Azotobacter* spp., *Comamonas* spp., *Mycobacterium* spp., *Rhodococcus* spp., *Gluconobacter* spp., *Ralstonia* spp., *Acidithiobacillus* spp., *Microlunatus* spp., *Geobacter* spp., *Geobacillus* spp., *Arthrobacter*

- 22 -

spp., *Flavobacterium* spp., *Serratia* spp., *Saccharopolyspora* spp., *Thermus* spp.,
Stenotrophomonas spp., *Chromobacterium* spp., *Sinorhizobium* spp., *Saccharopolyspora*
spp., *Agrobacterium* spp. and *Pantoea* spp. The bacterial cell can be a Gram-negative cell
such as an *Escherichia coli* (*E. coli*) cell, or a Gram-positive cell such as a species of
5 *Bacillus*. In other embodiments, the cell is a fungal cell such as a yeast cell, e.g.,
Saccharomyces spp., *Schizosaccharomyces* spp., *Pichia* spp., *Paffia* spp., *Kluyveromyces*
spp., *Candida* spp., *Talaromyces* spp., *Brettanomyces* spp., *Pachysolen* spp., *Debaryomyces*
spp., *Yarrowia* spp., and industrial polyploid yeast strains. Preferably the yeast strain is a *S.*
cerevisiae strain or a *Yarrowia* spp. strain. Other examples of fungi include *Aspergillus* spp.,
10 *Penicillium* spp., *Fusarium* spp., *Rhizopus* spp., *Acremonium* spp., *Neurospora* spp.,
Sordaria spp., *Magnaporthe* spp., *Allomyces* spp., *Ustilago* spp., *Botrytis* spp., and
Trichoderma spp. In other embodiments, the cell is an algal cell, or a plant cell. It should be
appreciated that some cells compatible with the invention may express an endogenous copy
of one or more of the genes associated with the invention as well as a recombinant copy. In
15 some embodiments, if a cell has an endogenous copy of one or more of the genes associated
with the invention then the methods will not necessarily require adding a recombinant copy
of the gene(s) that are endogenously expressed. In some embodiments the cell may
endogenously express one or more enzymes from the pathways described herein and may
recombinantly express one or more other enzymes from the pathways described herein for
20 efficient production of a terpenoid.

Further aspects of the invention relate to screening for bacterial cells or strains that
exhibit optimized terpenoid production. As described above, methods associated with the
invention involve generating cells that overexpress one or more genes in the MEP pathway.
Terpenoid production from culturing of such cells can be measured and compared to a control
25 cell wherein a cell that exhibits a higher amount of a terpenoid production relative to a
control cell is selected as a first improved cell. The cell can be further modified by
recombinant expression of a terpenoid synthase enzyme and a GGPPS enzyme. The level of
expression of one or more of the components of the non-mevalonate (MEP) pathway, the
terpenoid synthase enzyme and/or the GGPPS enzyme in the cell can then be manipulated
30 and terpenoid production can be measured again, leading to selection of a second improved
cell that produces greater amounts of a terpenoid than the first improved cell. In some
embodiments, the terpenoid synthase enzyme is a taxadiene synthase enzyme.

- 23 -

Further aspects of the invention relate to the identification and characterization (via GC-MS) of a previously unknown metabolite in bacterial *E. coli* cells (Figures 3 and 6). The level of accumulation of the newly identified metabolite, indole, can be controlled by genetically manipulating the microbial pathway by the overexpression, down regulation or mutation of the isoprenoid pathway genes. The metabolite indole anti-correlates as a direct variable to the taxadiene production in engineered strains (Figures 3, 6 and 15). Further controlling the accumulation of indole for improving the flux towards terpenoid biosynthesis in bacterial systems (specifically in cells, such as *E. coli* cells) or other cells, can be achieved by balancing the upstream non-mevalonate isoprenoid pathway with the downstream product synthesis pathways or by modifications to or regulation of the indole pathway. In so doing, the skilled person can reduce or control the accumulation of indole and thereby reduce the inhibitory effect of indole on the production of taxadiene, and other terpenoids derived from the described pathways, such as: monoterpenoids, sesquiterpenoids (including amorphadiene), diterpenoids (including levopimaradiene), triterpenes, and tetraterpenes. Other methods for reducing or controlling the accumulation of indole include removing the accumulated indole from the fermentation through chemical methods such as by using absorbents, scavengers, etc.

In other embodiments, methods are provided that include measuring the amount or concentration of indole in a cell that produces one or more terpenoids or in a culture of the cells that produce one or more terpenoids. The amount or concentration of indole can be measured once, or two or more times, as suitable, using methods known in the art and as described herein. Such methods can be used to guide processes of producing one or more terpenoids, e.g., in process improvement. Such methods can be used to guide strain construction, e.g., for strain improvement.

The identification of the means to achieve this balancing yielded a 15000 fold improvement in the overproduction of terpenoids such as taxadiene, compared to wild type bacterial cells, expressed with a heterologous taxadiene biosynthetic pathway. The production was further increased through modified fermentation methods that yielded concentrations of approximately 2g/L, which is 1500 fold higher compared to any prior reported taxadiene production. As demonstrated herein, by genetically engineering the non-mevalonate isoprenoid pathway in *E. coli* the accumulation of this metabolite can now be controlled which regulates the flux towards the isoprenoid biosynthesis in bacterial *E. coli* cells.

- 24 -

Also demonstrated herein is further channeling of the taxadiene production into the next key precursor to Taxol, taxadien-5 α -ol, achieved through engineering the oxidation chemistry for Taxol biosynthesis. Example 5 presents the first successful extension of the synthetic pathway from taxadiene to taxadien-5 α -ol. Similar to the majority of other terpenoids, the Taxol biosynthesis follows the unified fashion of “two phase” biosynthetic process, (i) the “cyclase phase” of linear coupling of the prenyl precursors (IPP and DMAPP) to GGPP followed by the molecular cyclization and rearrangement for the committed precursor taxadiene (Figure 6, VIII-IX).^{57, 58} After the committed precursor, (ii) the “oxidation phase”, the cyclic olefin taxadiene core structure is then functionalized by seven cytochrome P450 oxygenases together with its redox partners, decorated with two acetate groups and a benzoate group by acyl and aroyl CoA-dependent transferases, keto group by keto-oxidase, and epoxide group by epoxidase lead to the late intermediate baccatin III, to which the C13 side chain is attached for Taxol ((Figure 6, X-XIII).¹⁵ Although a rough sequential order of the early oxidation phase reactions are predicted, the precise timing/order of some of the hydroxylations, acylations and benzylation reactions are uncertain. However it is clear that the early bifurcation starts from the cytochrome p450 mediated hydroxylation of taxadiene core at C5 position followed the downstream hydroxylations using a homologous family of cytochrome p450 enzymes with high deduced similarity to each other (>70%) but with limited resemblance (<30%) to other plant p450's.^{41, 59} In addition, the structural and functional diversity with the possible evolutionary analysis implicit that the taxadiene-5 α -ol gene can be the parental sequence from which the other hydroxylase genes in the Taxol biosynthetic pathway evolved, reflecting the order of hydroxylations.¹⁵

Further aspects of the invention relate to chimeric P450 enzymes. Functional expression of plant cytochrome P450 has been considered challenging due to the inherent limitations of bacterial platforms, such as the absence of electron transfer machinery, cytochrome P450 reductases, and translational incompatibility of the membrane signal modules of P450 enzymes due to the lack of an endoplasmic reticulum.

In some embodiments, the taxadiene-5 α -hydroxylase associated with methods of the invention is optimized through N-terminal transmembrane engineering and/or the generation of chimeric enzymes through translational fusion with a CPR redox partner. In some embodiments, the CPR redox partner is a *Taxus* cytochrome P450 reductase (TCPR; Figure 5b). In certain embodiments, cytochrome P450 taxadiene-5 α -hydroxylase (T5 α OH) is

- 25 -

obtained from *Taxus cuspidate* (GenBank Accession number AY289209, the sequence of which is incorporated by reference herein). In some embodiments, NADPH:cytochrome P450 reductase (TCPR) is obtained from *Taxus cuspidate* (GenBank Accession number AY571340, the sequence of which is incorporated by reference herein).

5 The taxadiene 5 α -hydroxylase and TCPR can be joined by a linker such as GSTGS (SEQ ID NO:50). In some embodiments, taxadiene 5 α -hydroxylase and/or TCPR are truncated to remove all or part of the transmembrane region of one or both proteins. For example, taxadiene 5 α -hydroxylase in some embodiments is truncated to remove 8, 24, or 42 N-terminal amino acids. In some embodiments, the N-terminal 74 amino acids of TCPR are
10 truncated. An additional peptide can also be fused to taxadiene 5 α -hydroxylase. For example, one or more amino acids from bovine 17 α hydroxylase can be added to taxadiene 5 α -hydroxylase. In certain embodiments, the peptide MALLLAVF (SEQ ID NO:51) is added to taxadiene 5 α -hydroxylase. A non-limiting example of polypeptide comprising taxadiene 5 α -hydroxylase fused to TCPR is At24T5 α OH-tTCPR.

15 In some embodiments, the chimeric enzyme is able to carry out the first oxidation step with more than 10% taxadiene conversion to taxadiene-5 α -ol and the byproduct 5(12)-Oxa-3(11)-cyclotaxane. For example, the percent taxadiene conversion to taxadiene-5 α -ol and the byproduct 5(12)-Oxa-3(11)-cyclotaxane can be at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%,
20 approximately 99% or approximately 100%.

 In certain embodiments, the chimeric enzyme is At245 α OH-tTCPR, which was found to be capable of carrying out the first oxidation step with more than 98% taxadiene conversion to taxadiene-5 α -ol and the byproduct 5(12)-Oxa-3(11)-cyclotaxane (OCT; Figure 9a). Engineering of the step of taxadiene-5 α -ol production is critical in the production of
25 Taxol and was found to be limiting in previous efforts to construct this pathway in yeast. The engineered construct developed herein demonstrated greater than 98% conversion of taxadiene *in vivo* with a 2400 fold improvement over previous heterologous expression in yeast. Thus, in addition to synthesizing significantly greater amounts of key Taxol intermediates, this study also provides the basis for the synthesis of subsequent metabolites in
30 the pathway by similar P450 chemistry.

 As used herein, the terms "protein" and "polypeptide" are used interchangeably and thus the term polypeptide may be used to refer to a full-length polypeptide and may also be

- 26 -

used to refer to a fragment of a full-length polypeptide. As used herein with respect to polypeptides, proteins, or fragments thereof, "isolated" means separated from its native environment and present in sufficient quantity to permit its identification or use. Isolated, when referring to a protein or polypeptide, means, for example: (i) selectively produced by expression cloning or (ii) purified as by chromatography or electrophoresis. Isolated proteins or polypeptides may be, but need not be, substantially pure. The term "substantially pure" means that the proteins or polypeptides are essentially free of other substances with which they may be found in production, nature, or *in vivo* systems to an extent practical and appropriate for their intended use. Substantially pure polypeptides may be obtained naturally or produced using methods described herein and may be purified with techniques well known in the art. Because an isolated protein may be admixed with other components in a preparation, the protein may comprise only a small percentage by weight of the preparation. The protein is nonetheless isolated in that it has been separated from the substances with which it may be associated in living systems, i.e. isolated from other proteins.

The invention also encompasses nucleic acids that encode for any of the polypeptides described herein, libraries that contain any of the nucleic acids and/or polypeptides described herein, and compositions that contain any of the nucleic acids and/or polypeptides described herein.

In some embodiments, one or more of the genes associated with the invention is expressed in a recombinant expression vector. As used herein, a "vector" may be any of a number of nucleic acids into which a desired sequence or sequences may be inserted by restriction and ligation for transport between different genetic environments or for expression in a host cell. Vectors are typically composed of DNA, although RNA vectors are also available. Vectors include, but are not limited to: plasmids, fosmids, phagemids, virus genomes and artificial chromosomes.

A cloning vector is one which is able to replicate autonomously or integrated in the genome in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut in a determinable fashion and into which a desired DNA sequence may be ligated such that the new recombinant vector retains its ability to replicate in the host cell. In the case of plasmids, replication of the desired sequence may occur many times as the plasmid increases in copy number within the host cell such as a host bacterium or just a single time per host before the host reproduces by mitosis. In the case of

- 27 -

phage, replication may occur actively during a lytic phase or passively during a lysogenic phase.

An expression vector is one into which a desired DNA sequence may be inserted by restriction and ligation such that it is operably joined to regulatory sequences and may be expressed as an RNA transcript. Vectors may further contain one or more marker sequences suitable for use in the identification of cells which have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., β -galactosidase, luciferase or alkaline phosphatase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies or plaques (e.g., green fluorescent protein). Preferred vectors are those capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably joined.

As used herein, a coding sequence and regulatory sequences are said to be "operably" joined when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences. If it is desired that the coding sequences be translated into a functional protein, two DNA sequences are said to be operably joined if induction of a promoter in the 5' regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably joined to a coding sequence if the promoter region were capable of effecting transcription of that DNA sequence such that the resulting transcript can be translated into the desired protein or polypeptide.

When the nucleic acid molecule that encodes any of the enzymes of the claimed invention is expressed in a cell, a variety of transcription control sequences (e.g., promoter/enhancer sequences) can be used to direct its expression. The promoter can be a native promoter, i.e., the promoter of the gene in its endogenous context, which provides normal regulation of expression of the gene. In some embodiments the promoter can be

- 28 -

constitutive, i.e., the promoter is unregulated allowing for continual transcription of its associated gene. A variety of conditional promoters also can be used, such as promoters controlled by the presence or absence of a molecule.

The precise nature of the regulatory sequences needed for gene expression may vary
5 between species or cell types, but shall in general include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. In particular, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined gene.
10 Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired. The vectors of the invention may optionally include 5' leader or signal sequences. The choice and design of an appropriate vector is within the ability and discretion of one of ordinary skill in the art.

Expression vectors containing all the necessary elements for expression are
15 commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by the introduction into the cells of heterologous DNA (RNA). That heterologous DNA (RNA) is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell.
20 Heterologous expression of genes associated with the invention, for production of a terpenoid, such as taxadiene, is demonstrated in the Examples section using *E. coli*. The novel method for producing terpenoids can also be expressed in other bacterial cells, fungi (including yeast cells), plant cells, etc.

A nucleic acid molecule that encodes an enzyme associated with the invention can be
25 introduced into a cell or cells using methods and techniques that are standard in the art. For example, nucleic acid molecules can be introduced by standard protocols such as transformation including chemical transformation and electroporation, transduction, particle bombardment, etc. Expressing the nucleic acid molecule encoding the enzymes of the claimed invention also may be accomplished by integrating the nucleic acid molecule into the
30 genome.

In some embodiments one or more genes associated with the invention is expressed recombinantly in a bacterial cell. Bacterial cells according to the invention can be cultured in

- 29 -

media of any type (rich or minimal) and any composition. As would be understood by one of ordinary skill in the art, routine optimization would allow for use of a variety of types of media. The selected medium can be supplemented with various additional components. Some non-limiting examples of supplemental components include glucose, antibiotics, IPTG
5 for gene induction, ATCC Trace Mineral Supplement, and glycolate. Similarly, other aspects of the medium, and growth conditions of the cells of the invention may be optimized through routine experimentation. For example, pH and temperature are non-limiting examples of factors which can be optimized. In some embodiments, factors such as choice of media, media supplements, and temperature can influence production levels of terpenoids, such as
10 taxadiene. In some embodiments the concentration and amount of a supplemental component may be optimized. In some embodiments, how often the media is supplemented with one or more supplemental components, and the amount of time that the media is cultured before harvesting a terpenoid, such as taxadiene, is optimized.

According to aspects of the invention, high titers of a terpenoid such as taxadiene, are
15 produced through the recombinant expression of genes associated with the invention, in a cell. As used herein "high titer" refers to a titer in the milligrams per liter (mg L^{-1}) scale. The titer produced for a given product will be influenced by multiple factors including choice of media. In some embodiments, the total taxadiene titer is at least 1 mg L^{-1} . In some
embodiments, the total taxadiene titer is at least 10 mg L^{-1} . In some embodiments, the total
20 taxadiene titer is at least 250 mg L^{-1} . For example, the total taxadiene titer can be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575,
25 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900 or more than 900 mg L^{-1} including any intermediate values. In some embodiments, the total taxadiene titer can be at least 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, or more than 5.0 g L^{-1} including any intermediate values.

30 In some embodiments, the total taxadiene 5α -ol titer is at least 1 mg L^{-1} . In some embodiments, the total taxadiene 5α -ol titer is at least 10 mg L^{-1} . In some embodiments, the total taxadiene 5α -ol titer is at least 50 mg L^{-1} . For example, the total taxadiene 5α -ol titer

- 30 -

can be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, or more than 70 mg L⁻¹ including any intermediate values.

5 The liquid cultures used to grow cells associated with the invention can be housed in any of the culture vessels known and used in the art. In some embodiments large scale production in an aerated reaction vessel such as a stirred tank reactor can be used to produce large quantities of terpenoids, such as taxadiene, that can be recovered from the cell culture. In some embodiments, the terpene is recovered from the gas phase of the cell culture, for
10 example by adding an organic layer such as dodecane to the cell culture and recovering the terpene from the organic layer.

 Terpenoids, such as taxadiene, produced through methods described herein have widespread applications including pharmaceuticals such as paclitaxel (Taxol), artemisinin, ginkgolides, eleutherobin and pseudopterosins, and many other potential pharmaceutical
15 compounds. Further applications include compounds used in flavors and cosmetics such as geraniol, farnesol, geranylgeraniol, linalool, limonene, pinene, cineol and isoprene. Further applications include compounds for use as biofuels such as alcohols of 5, 10, and 15 carbon atom length. It is noted that the above compounds are presently produced as extracts of various plants. Plant extract-based methods are tedious, yield very small amounts and are
20 limited as to the actual molecules that can be so obtained, namely, they do not allow the easy production of derivatives that may possess far superior properties than the original compounds.

Examples

25

METHODS

Strains, plasmids, oligonucleotides and genes

E. coli K12 MG1655 strain was used as the host strain of all the taxadiene strain construction. *E. coli* K12MG1655 Δ(recA,endA) and *E. coli* K12MG1655Δ(recA,endA)ED3
30 strains were provided by Professor Kristala Prather's lab at MIT (Cambridge, MA). Detail of all the plasmids constructed for the study is shown in Table 2. All oligonucleotides used in this study are contained in Table 3.

- 31 -

The sequences of geranylgeranyl pyrophosphate synthase (GGPPS),⁵⁰ Taxadiene synthase (TS),⁵¹ Cytochrome P450 Taxadiene 5 α -hydroxylase (T5 α OH) and *Taxus* NADPH:cytochrome P450 reductase (TCPR)⁴⁶ were obtained from *Taxus canadensis*, *Taxus brevifolia*, *Taxus cuspidate* (Genbank accession codes: AF081514, U48796, AY289209 and AY571340). Genes were custom-synthesized using the plasmids and protocols reported by Kodumal et al.⁵² (Supplementary details Appendix 1) to incorporate *E. coli* translation codon and removal of restriction sites for cloning purposes. Nucleotides corresponding to the 98 and 60 N-terminal amino acids of GGPPS and TS (plastid transit peptide) were removed and the translation insertion sequence Met was inserted.¹⁷

10

Construction of MEP pathway (dxs-idi-idpDF operon).

dxs-idi-ispDF operon was initially constructed by cloning each of the genes from the genome of *E. coli* K12 MG1655 using the primers dxs(s), dxs(a), idi(s), idi(a), ispDF(s) and ispDFI(a) under pET21C+ plasmid with T7 promoter (p20T7MEP).⁵³ Using the primers dxsidiispDFNcoI (s) and dxsidiispDFKpnI(a) dxs-idi-ispDF operon was sub-cloned into and pTrcHis2B (Invitrogen) plasmid after digested with NcoI and KpnI for pTrcMEP plasmid (p20TrcMEP). p20TrcMEP plasmid digested with MluI and PmeI and cloned into MluI and PmeI digested pACYC184-melA(P2A) plasmid to construct p10TrcMEP plasmid. pTrcMEP plasmid digested with BstZ17I and ScaI and cloned into PvuII digested pCL1920 plasmid to construct p5TrcMEP plasmid. For constructing p20T5MEP plasmid initially the dxs-idi-ispDF operon was cloned into pQE plasmid with T5 promoter (pQE-MEP) using the primers dxsidiispDFNcoI (s) and dxsidiispDFXhoI(a). A fraction of the operon DNA with T5 promoter was amplified using the primers T5AgeI(s) and T5NheI(a) from pQEMEP plasmid. The DNA fragment was digested with AgeI/NheI and cloned into the p20T7MEP plasmid digested with SGrAI/NheI enzymes.

25

Construction of Taxadiene pathway (GT and TG operons).

The downstream taxadiene pathways (GT and TG operon) were constructed by cloning PCR fragments of GGPS and TS into the *NcoI* – *EcoRI* and *EcoRI* – *SaII* sites of pTrcHis2B plasmid to create p20TrcGT and p20TrcTG using the primers GGPPSNcoI(s), GGPPSEcoRI(a), TSEcoRI(s), TSSaII(a), TSNcoI(s) TSEcoRI(a) GGPPSEcoRI(s) and GGPPSSaII(a). For constructing p20T5GT, initially the operon was amplified with primers

30

- 32 -

GGPPSNcoI(s) and TSXhoI(a) and cloned into a pQE plasmid under T5 promoter digested with NcoI/XhoI. Further the sequence was digested with XbaI and XhoI and cloned into the pTrc plasmid backbone amplified using the primers pTrcSal(s) and pTrcXba(a). p10T7TG was constructed by subcloning the NcoI/SalI digested TG operon from p20TrcTG into NcoI/SalI digested pACYC-DUET1 plasmid. p5T7TG was constructed by cloning the BspEI/XbaI digested fragment to the XbaI/BspEI digested DNA amplified from pCL1920 plasmid using pCLBspEI(s) and pCLXbaI(a) primers.

Construction of chromosomal integration MEP pathway plasmids

For constructing the plasmids with FRP-Km-FRP cassette for amplifying the sequence for integration, p20T7MEP and p20T5MEP was digested with XhoI/ScaI. FRP-Km-FRP cassette was amplified from the Km cassette with FRP sequence from pkD13 plasmid using the primers KmFRPXhoI(s) and KmFRPScaI(a). The amplified DNA was digested with XhoI/ScaI and cloned into the XhoI/ScaI digested p20T7MEP and p20T5MEP plasmid (p20T7MEPKmFRP and p20T5MEPKmFRP). Similarly the p20TrcMEP plasmid was digested with SacI/ScaI and the amplified DNA using the primers KmFRPSacI(s) and KmFRPScaI(a) was digested, cloned into the p20TrcMEP plasmid (p20TrcMEPKm-FRP).

Chromosomal integration of the MEP pathway cassette (LacIq-MEP-FRP-Km-FRP) cassette

The MEP pathways constructed under the promoters T7, T5 and Trc were localized to the ara operon region in the chromosome with the Kan marker. The PCR fragments were amplified from p20T7MEPKmFRP, p20T5MEPKmFRP and p20TrcMEPKm-FRP using the primers IntT7T5(s), IntTrc(s) and Int(a) and then electroporated into *E. coli* MG1655 recA-end- and *E. coli* MG1655 recA-end- EDE3 cells for chromosomal integration through the λ Red recombination technique.⁵⁴ The site specific localization was confirmed and the Km marker was removed through the action of the FLP recombinase after successful gene integration.

Construction of Taxadiene 5 α -ol pathway

The transmembrane region (TM) of the taxadiene 5 α -ol hydroxylase (T5 α OH) and Taxus Cytochrome P450 reductase (TCPR) was identified using PredictProtein software

- 33 -

(www.predictprotein.org).⁵⁵ For transmembrane engineering selective truncation at 8, 24 and 42 amino acid residues on the N-terminal transmembrane region of taxadiene 5 α -ol hydroxylase (T5 α OH) and 74 amino acid region in the TCPR was performed. The removal of the 8, 24 and 42 residue N-terminal amino acids of taxadiene 5 α -ol hydroxylase (T5 α OH),
 5 incorporation of one amino acid substituted bovine 17 α hydroxylase N-terminal 8 residue peptide MALLLAVF (SEQ ID NO:51) to the N-terminal truncated T5 α OH sequences⁴⁴ and GSTGS peptide linker was carried out using the primer CYP17At8AANdeI(s), CYP17At24AANdeI(s), CYP17At42AANdeI(s) and CYPLinkBamHI(a). Using these primers each modified DNA was amplified, NdeI/BamHI digested and cloned into
 10 NdeI/BamHI digested pACYC DUET1 plasmid to construct p10At8T5 α OH, p10At24T5 α OH and p10At42T5 α OH plasmids. 74 amino acid truncated TCPR (tTCPR) sequence was amplified using primers CPRBamHI(s) and CPRSalI(a). The amplified tTCPR sequence and the plasmids, p10At8T5 α OH, p10At24T5 α OH and p10At42T5 α OH, was digested with BamHI/SalI and cloned to construct the plasmids p10At8T5 α OH-tTCPR, p10At24T5 α OH-tTCPR and p10At42T5 α OH-tTCPR.
 15

Culture growth for screening the taxadiene and taxadiene-5 α -ol analysis

Single transformants of pre-engineered *E. coli* strains harboring the appropriate plasmid with upstream (MEP), downstream taxadiene pathway and taxadiene 5 α -ol were
 20 cultivated for 18h at 30°C in Luria-Bertani (LB) medium (supplemented with appropriate antibiotics, 100 mg/mL carbenecilin, 34 mg/mL chloramphenicol, 25 mg/L kanamycin or 50 mg/L spectinomycin). For small scale cultures to screen the engineered strains, these preinnoculum were used to seed fresh 2-mL rich media (5g/L yeast extract, 10g/L Trypton, 15g/L, glucose, 10g/L NaCl, 100mM HEPS, 3 mL/L Antifoam B, pH 7.6, 100 ug/mL
 25 Carbenicillin and 34 ug/mL chloramphenicol), at a starting A₆₀₀ of 0.1. The culture was maintained with appropriate antibiotics and 100 mM IPTG for gene induction at 22 °C for 5 days.

Bioreactor experiments for the taxadiene 5 α -ol producing strain.

30 The 3-L Bioflo bioreactor (New Brunswick) was assembled as to manufacturer's instructions. One liter of rich media with 1% glycerol (v/v) was inoculated with 50 mL of 8h culture (A₆₀₀ of ~2.2) of the strain 26-At24T5 α OH-tTCPR grown in LB medium containing

- 34 -

the antibiotics (100 mg/mL carbenicillin, 34 mg/mL chloramphenicol) at the same concentrations. 1L-bioreactors with biphasic liquid-liquid fermentation using 20% v/v dodecane. Oxygen was supplied as filtered air at 0.5 v/v/m and agitation was adjusted to maintain dissolved oxygen levels above 50%. pH of the culture was controlled at 7.0 using 10% NaOH. The temperature of the culture in the fermentor was controlled at 30 °C until the cells were grown into an optical density of approximately 0.8, as measured at a wavelength of 600 nm (OD₆₀₀). The temperature of the fermentor was reduced to 22 °C and the cells were induced with 0.1 mM IPTG. Dodecane was added aseptically to 20% (v/v) of the media volume. During the course of the fermentation the concentration of glycerol and acetate accumulation was monitored with constant time intervals. During the fermentation as the glycerol concentration depleted below 0.5 g/L, glycerol (3 g/L) was introduced into the bioreactor.

The fermentation was further optimized using a fed batch cultivation with a defined feed medium containing 0.5% yeast extract and 20% (v/v) dodecane (13.3 g/L KH₂PO₄, 4 g/L (NH₄)₂HPO₄, 1.7 g/L citric acid, 0.0084 g/L EDTA, 0.0025 g/L CoCl₂, 0.015 g/L MnCl₂, 0.0015 g/L CuCl₂, 0.003 g/L H₃BO₃, 0.0025 g/L Na₂MoO₄, 0.008 g/L Zn(CH₃COO)₂, 0.06 g/L Fe(III) citrate, 0.0045 g/L thiamine, 1.3 g/L MgSO₄, 10 g/L glycerol, 5 g/L yeast extract, pH 7.0). The same medium composition was used for the fermentation of strains **17** and **26** with appropriate antibiotics (strain 17: 100 µg/mL carbenicillin and 50 µg/mL spectinomycin; strain 26: 50 µg/mL spectinomycin).

For the taxadien-5 α -ol producing strain, one liter of complex medium with 1% glycerol (v/v) was inoculated with 50 mL of an 8h culture (OD of ~2.2) of strain **26**-At24T5aOH-tTCPR grown in LB medium containing 50 µg/mL spectinomycin and 34 µg/mL chloramphenicol). Oxygen was supplied as filtered air at 0.5 (vvm) and agitation was adjusted to maintain dissolved oxygen levels above 30%. The pH of the culture was controlled at 7.0 using 10% NaOH. The temperature of the culture in the fermentor was controlled at 30°C until the cells were grown into an optical density of approximately 0.8, as measured at a wavelength of 600 nm (OD₆₀₀). The temperature of the fermentor was reduced to 22°C and the pathway was induced with 0.1 mM IPTG. Dodecane was added aseptically to 20% (v/v) of the media volume. During the course of the fermentation, the concentration of glycerol and acetate accumulation was monitored with constant time

- 35 -

intervals. During the fermentation as the glycerol concentration depleted 0.5-1 g/L, 3 g/L of glycerol was introduced into the bioreactor.

GC-MS analysis of taxadiene and taxadiene-5 α -ol

5 For analysis of taxadiene accumulation from small scale culture, 1.5 mL of the culture was vortexed with 1mL hexane for 30 min. The mixture was centrifuged to separate the organic layer. For bioreactor 1 uL of the dodecane layer was diluted to 200 uL using hexane. 1uL of the hexane layer was analyzed by GC-MS (Varian saturn 3800 GC attached to a Varian 2000 MS). The sample was injected into a HP5ms column (30m x 250 uM x 0.25 uM
10 thickness) (Agilent Technologies USA). Helium (ultra purity) at a flow rate 1.0 ml/min was used as a carrier gas. The oven temperature was first kept constant at 50 °C for 1 min, and then increased to 220 °C at the increment of 10 °C/min, and finally held at this temperature for 10 min. The injector and transfer line temperatures were set at 200 °C and 250 °C, respectively.

15 Standard compounds from biological or synthetic sources for taxadiene and taxadiene 5 α -ol was not commercially available. Thus we performed fermentations of taxadiene producing *E. coli* in a 2L bioreactor to extract pure material. Taxadiene was extracted by solvent extraction using hexane, followed by multiple rounds of silica column chromatography to obtain the pure material for constructing a standard curve for GC-MS
20 analysis. We have compared the GC and MS profile of the pure taxadiene with the reported literature to confirm the authenticity of the compound⁶⁰. In order to check the purity we have performed ¹HNMR of taxadiene. Since the accumulation of taxadiene-5 α -ol was very low level we used taxadiene as a measure to quantify the production of this molecule and authentic mass spectral fragmentation characteristics from previous reports⁴².

25

qPCR measurements for transcriptional analysis of engineered strains

Transcriptional gene expression levels of each gene were detected by qPCR on mRNA isolated from the appropriate strains. To prevent degradation, RNA was stabilized before cell lysis using RNeasy Protect Bacteria reagent (Qiagen). Subsequently, total RNA was
30 isolated using RNeasy mini kit (Qiagen) combined with nuclease based removal of genomic DNA contaminants. cDNA was amplified using iScript cDNA synthesis kit (Biorad). qPCR was carried out on a Bio-Rad iCycler using the iQ SYBR Green Supermix (Biorad). The level

- 36 -

of expression of *rrsA* gene, which is not subject to variable expression, was used for normalization of qPCR values.⁵⁶ Table 3 has primers used for qPCR. For each primer pair, a standard curve was constructed with mRNA of *E. coli* as the template.

5 **Example 1: Taxadiene accumulation exhibits strong non-linear dependence on the relative strengths of the upstream MEP and downstream synthetic taxadiene pathways.**

Fig. 1b depicts the various ways by which promoters and gene copy numbers were combined to modulate the relative flux (or strength) through the upstream and downstream pathways of taxadiene synthesis. A total of 16 strains were constructed in order to de-
 10 bottleneck the MEP pathway, as well as optimally balance it with the downstream taxadiene pathway. Fig. 2a,b summarize the results of taxadiene accumulation in each of these strains, with Fig. 2a accentuating the dependence of taxadiene accumulation on the upstream pathway for constant values of the downstream pathway, and Fig. 2b the dependence on the downstream pathway for constant upstream pathway strength (see also Table 1 for the
 15 calculation of the upstream and downstream pathway expression from the reported promoter strengths and plasmid copy numbers³³⁻³⁶). Clearly, there are maxima exhibited with respect to both upstream and downstream pathway expression. For constant downstream pathway expression (Fig. 2a), as the upstream pathway expression increases from very low levels, taxadiene production is increased initially due to increased supply of precursors to the overall
 20 pathway. However, after an intermediate value, further upstream pathway increases cannot be accommodated by the capacity of the downstream pathway. This pathway imbalance leads to the accumulation of an intermediate (see below) that may be either inhibitory to cells or simply indicate flux diversion to a competing pathway, ultimately resulting in taxadiene accumulation reduction.

25 For constant upstream pathway expression (Fig. 2b), a maximum is similarly observed with respect to the level of downstream pathway expression. This is attributed to an initial limitation of taxadiene production by low expression levels of the downstream pathway, which is thus rate limiting with respect to taxadiene production. At high levels of downstream pathway expression we are likely seeing the negative effect of high copy number
 30 on cell physiology, hence, a maximum exists with respect to downstream pathway expression. These results demonstrate that dramatic changes in taxadiene accumulation can be obtained from changes within a narrow window of expression levels for the upstream and

- 37 -

downstream pathways. For example, a strain containing an additional copy of the upstream pathway on its chromosome under Trc promoter control (Strain 8, Fig. 2a) produced 2000 fold more taxadiene than one overexpressing only the synthetic downstream pathway (Strain 1, Fig. 2a). Furthermore, changing the order of the genes in the downstream synthetic operon from GT (GPPS-TS) to TG (TS-GPPS) resulted in 2-3-fold increase (strains 1-4 compared to 5, 8, 11 and 14). The observed results show that the key to taxadiene overproduction is ample downstream pathway capacity and careful balancing between the upstream precursor pathway with the downstream synthetic taxadiene pathway. Altogether, the engineered strains established that the MEP pathway flux can be substantial, if a wide range of expression levels for the endogenous upstream and synthetic downstream pathway are searched simultaneously.

Example 2: Chromosomal integration and fine tuning of the upstream and downstream pathways further enhances taxadiene production.

To provide ample downstream pathway strength while minimizing the plasmid-borne metabolic burden³⁷, two new sets of 4 strains each were engineered (strains 25-28 and 29-32) in which the downstream pathway was placed under the control of a strong promoter (T7) while keeping a relatively low number of 5 and 10 copies, respectively. It can be seen (Fig. 2c) that, while the taxadiene maximum is maintained at high downstream strength (strains 21-24), a monotonic response is obtained at the low downstream pathway strength (strains 17-20, Fig. 2c). This observation prompted the construction of two additional sets of 4 strains each that maintained the same level of downstream pathway strength as before but expressed very low levels of the upstream pathway (strains 25-28 and 29-32, Fig. 2d). Additionally, the operon of the upstream pathway of the latter strain set was chromosomally integrated. It can be seen that not only is the taxadiene maximum recovered, albeit at very low upstream pathway levels, but a much greater taxadiene maximum is attained (300 mg/L). We believe this significant increase can be attributed to a decrease in cell's metabolic burden. This was achieved by 1) eliminating plasmid dependence through integration of the pathway into the chromosome and 2) attaining a fine balance between the upstream and downstream pathway expression.

The 32 recombinant constructs allowed us to adequately probe the modular pathway expression space and amplify ~15000 fold improvement in taxadiene production. This is by

- 38 -

far the highest production of terpenoids from *E. coli* MEP isoprenoid pathway reported (Fig. 3a). Additionally, the observed fold improvements in terpenoid production are significantly higher than those of reported combinatorial metabolic engineering approaches that searched an extensive genetic space comprising up to a billion combinatorial variants of the isoprenoid pathway.³⁰ This suggests that pathway optimization depends far more on fine balancing of the expression of pathway modules than multi-source combinatorial gene optimization. The multiple maxima exhibited in the phenotypic landscape of Fig. 1 underscores the importance of probing the expression space at sufficient resolution to identify the region of optimum overall pathway performance. Fig. 7 depicts the fold improvements in taxadiene production from the modular pathway expression search.

Example 3: Metabolite inversely correlates with taxadiene production and identification of metabolite.

Metabolomic analysis of the previous engineered strains identified an, as yet, unknown, metabolite byproduct that correlated strongly with pathway expression levels and taxadiene production (Fig. 3 and Fig. 8). Although the chemical identity of the metabolite was unknown, we hypothesized that it is an isoprenoid side-product, resulting from pathway diversion and has been anti-correlated as a direct variable to the taxadiene production (Fig. 3 and Fig. 8) from the engineered strains. A critical attribute of our optimal strains is the fine balancing that alleviates the accumulation of this metabolite, resulting in higher taxadiene production. This balancing can be modulated at different levels from chromosome, or different copy number plasmids, using different promoters, with significantly different taxadiene accumulation.

Subsequently the corresponding peak in the gas chromatography-mass spectrometry (GC-MS) chromatogram was identified as indole by GC-MS, ¹H and ¹³C nuclear magnetic resonance (NMR) spectroscopy studies (Fig. 16). We found that taxadiene synthesis by strain 26 is severely inhibited by exogenous indole at indole levels higher than ~100 mg/L (Fig. 15b). Further increasing the indole concentration also inhibited cell growth, with the level of inhibition being very strain dependent (Fig. 15c). Although the biochemical mechanism of indole interaction with the isoprenoid pathway is presently unclear, the results in Fig. 15 suggest a possible synergistic effect between indole and terpenoid compounds of the isoprenoid pathway in inhibiting cell growth. Without knowing the specific mechanism, it

- 39 -

appears strain 26 has mitigated the indole's effect, which we carried forward for further study.

Example 4: Cultivation of engineered strains.

5 In order to explore the taxadiene producing potential under controlled conditions for the engineered strains, fed batch cultivations of the three highest taxadiene accumulating strains (~60 mg/L from strain 22; ~125 mg/L from strain 17; ~300 mg/L from strain 26) were carried out in 1L-bioreactors (Fig. 17). The fed batch cultivation studies were carried out as liquid-liquid two-phase fermentation using a 20%(v/v) dodecane overlay. The organic
10 solvent was introduced to prevent air stripping of secreted taxadiene from the fermentation medium, as indicated by preliminary findings. In defined media with controlled glycerol feeding, taxadiene productivity increased to 174 ± 5 mg/L (SD), 210 ± 7 mg/L (SD), and 1020 ± 80 mg/L (SD), respectively for strains 22, 17 and 26 (Fig. 17a). Additionally, taxadiene production significantly affected the growth phenotype, acetate accumulation and
15 glycerol consumption (Fig. 17b-17d).

Fig. 17c shows that acetate accumulates in all strains initially, however after ~60 hrs acetate decreases in strains 17 and 26 while it continues to increase in strain 22. This phenomenon highlights the differences in central carbon metabolism between high MEP flux strains (26 and 17) and low MEP flux strain (22). Additionally, this observation is another
20 illustration of the good physiology that characterizes a well-balanced, -functioning strain. Acetic acid, as product of overflow metabolism, is initially produced by all strains due to the high initial glycerol concentrations used in these fermentations and corresponding high glycerol pathway flux. This flux is sufficient for supplying also the MEP pathway, as well as the other metabolic pathways in the cell.

25 At ~48 hrs, the initial glycerol is depleted, and the cultivation switches to a fed-batch mode, during which low but constant glycerol levels are maintained. This results in a low overall glycerol flux, which, for strains with high MEP flux (strains 26 and 17), is mostly diverted to the MEP pathway while minimizing overflow metabolism. As a result acetic acid production is reduced or even totally eliminated. Regarding the decline in acetic acid
30 concentration, it is possible that acetic acid assimilation may have happened to some extent, although this was not further investigated from a flux analysis standpoint. Some evaporation and dilution due to glycerol feed are further contributing to the observed acetic acid

- 40 -

concentration decline. In contrast, for strains with low MEP flux (strain 22), flux diversion to the MEP pathway is not very significant, so that glycerol flux still supplies all the necessary carbon and energy requirements. Overflow metabolism continues to occur leading to acetate secretion.

Clearly the high productivity and more robust growth of strain 26 allowed very high taxadiene accumulation. Further improvements should be possible through optimizing conditions in the bioreactor, balancing nutrients in the growth medium, and optimizing carbon delivery.

Example 5: Upstream and downstream pathway expression levels and cell growth reveal underlying complexity.

For a more detailed understanding of the engineered balance in pathway expression, we quantified the transcriptional gene expression levels of *dxs* (upstream pathway) and TS (downstream pathway) for the highest taxadiene producing strains and neighboring strains from Fig 2c and d (strains 17, 22 and 25-32) (Fig. 4a,b). As we hypothesized, expression of the upstream pathway increased monotonically with promoter strength and copy number for the MEP vector from: native promoter, Trc, T5, T7, and 10 copy and 20 copy plasmids, as seen in the DXS expression (Fig. 4a). Thus we found that *dxs* expression level correlates well with the upstream pathway strength. Similar correlations were found for the other genes of the upstream pathway, *idi*, *ispD* and *ispF* (Fig. 14a, b). In the downstream gene expression, a ~2 fold improvement was quantified after transferring the pathway from 5 to 10 copy plasmid (25-28 series and 29-32 series) (Fig. 4b).

While promoter and copy number effects influenced the gene expressions, secondary effects on the expression of the other pathway were also prominent. Fig. 4a shows that for the same *dxs* expression cassettes, by increasing the copy number of the TS plasmid from 5 to 10, *dxs* expression was increased. Interestingly, the 5 copy TS plasmid (strains 25-28 series) contained substantially higher taxadiene yields (Fig. 2d) and less growth (Fig 4c,d) than the 10 copy TS plasmid. Control plasmids that did not contain the taxadiene heterologous pathway, grew two fold higher densities, implying growth inhibition in the strains 25-28 series is directly related to the taxadiene metabolic pathway and the accumulation of taxadiene and its direct intermediates (Fig. 4c). However the strain 29-32 series only showed modest increases in growth yield when comparing the empty control

- 41 -

plasmids to the taxadiene expressing strains (Fig. 4d). This interplay between growth, taxadiene production, and expression level can also be seen with the plasmid-based upstream expression vectors (strain 17 and 22). Growth inhibition was much larger in the 10 copy, high taxadiene producing strain (strain 17) compared to the 20 copy, lower taxadiene
5 producing strain (strain 22) (Fig. 4d). Therefore product toxicity and carbon diversion to the heterologous pathway are likely to impede growth, rather than plasmid-maintenance.

Also unexpected was the profound effect of the upstream expression vector on downstream expression. Fig. 4b would have two straight lines, if there was no cross talk between the pathways. However, ~3 fold changes in TS expression are observed for different
10 MEP expression vectors. This is likely due to significant competition for resources (raw material and energy) that are withdrawn from the host metabolism for overexpression of both the four upstream and two downstream genes.³⁸ Compared to the control strain 25c, a 4 fold growth inhibition was observed with strain 25 indicated that high overexpression of synthetic taxadiene pathway induced toxicity altering the growth phenotype compared to the
15 overexpression of native pathway (Fig. 4c). However, as upstream expression increased, downstream expression was reduced, inadvertently in our case, to desirable levels to balance the upstream and downstream pathways, minimizing growth inhibition (strain 26).

At the extreme of protein overexpression, T7 promoter-driven MEP pathway resulted in severe growth inhibition, due to the synthesis of four proteins at high level (strains 28 and
20 32). Expression of the TS genes by T7 does not appear to have as drastic effect by itself. The high rates of protein synthesis from the T7 induced expression (Fig. 4ab) could lead to the down regulation of the protein synthesis machinery including components of housekeeping genes from early growth phase impairs the cell growth and lower the increase in biomass.^{39,40} We hypothesized that our observed complex growth phenotypes are
25 cumulative effects of (1) toxicity induced by activation of isoprenoid/taxadiene metabolism, and (2) and the effects of high recombinant protein expression. Altogether our multivariate-modular pathway engineering approach generated unexpected diversity in terpenoid metabolism and its correlation to the pathway expression and cell physiology. Rational design of microbes for secondary metabolite production will require an understanding of
30 pathway expression that goes beyond a linear/independent understanding of promoter strengths and copy numbers. However, simple, multivariate approaches, as employed here, can introduce the necessary diversity to both (1) find high producers, and (2) provide a

- 42 -

landscape for the systematic investigation of higher order effects that are dominant, yet underappreciated, in metabolic pathway engineering.

Example 6: Engineering Taxol P450-based oxidation chemistry in *E. coli*.

5 A central feature in the biosynthesis of Taxol is oxygenation at multiple positions of the taxane core structure, reactions that are considered to be mediated by cytochrome P450-dependent monooxygenases.⁴¹ After completion of the committed cyclization step of the pathway, the parent olefin, taxa-4(5),11(12)-diene, is next hydroxylated at the C5 position by a cytochrome P450 enzyme, representing the first of eight oxygenation steps (of the taxane
10 core) on route to Taxol (Fig. 6).⁴² Thus, a key step towards engineering Taxol-producing microbes is the development of P450-based oxidation chemistry *in vivo*. The first oxygenation step is catalyzed by a cytochrome P450, taxadiene 5 α -hydroxylase, an unusual monooxygenase catalyzing the hydroxylation reaction along with double bond migration in the diterpene precursor taxadiene (Fig. 5a). We report the first successful extension of the
15 synthetic pathway from taxadiene to taxadien-5 α -ol and present the first examples of *in vivo* production of any functionalized Taxol intermediates in *E. coli*.

 In general, functional expression of plant cytochrome P450 is challenging⁴³ due to the inherent limitations of bacterial platforms, such as the absence of electron transfer machinery, cytochrome P450 reductases, and translational incompatibility of the membrane signal
20 modules of P450 enzymes due to the lack of an endoplasmic reticulum. Recently, through transmembrane (TM) engineering and the generation of chimera enzymes of P450 and CPR reductases, some plant P450's have been expressed in *E. coli* for the biosynthesis of functional molecules.^{22,44} Still, every plant cytochrome p450 is unique in its transmembrane signal sequence and electron transfer characteristics from its reductase counterpart.⁴⁵ Our
25 initial studies were focused on optimizing the expression of codon-optimized synthetic taxadiene 5 α -hydroxylase by N-terminal transmembrane engineering and generating chimera enzymes through translational fusion with the CPR redox partner from the *Taxus* species, *Taxus* cytochrome P450 reductase (TCPR) (Fig. 5b).^{42,44,46} One of the chimera enzymes generated, At24T5 α OH-tTCPR, was highly efficient in carrying out the first oxidation step
30 with more than 98% taxadiene conversion to taxadien-5 α -ol and the byproduct 5(12)-Oxa-3(11)-cyclotaxane (OCT) (Fig. 9a).

- 43 -

Compared to the other chimeric P450s, At24T5 α OH-tTCPR yielded two-fold higher (21 mg/L) production of taxadien-5 α -ol. As well, the weaker activity of At8T5 α OH-tTCPR and At24T5 α OH-tTCPR resulted in accumulation of a recently characterized byproduct, a complex structural rearrangement of taxadiene into the cyclic ether 5(12)-Oxa-3(11)-cyclotaxane (OCT) (Fig. 9).⁴⁷ The byproduct accumulated at approximately equal amounts as the desired product taxadien-5 α -ol. The OCT formation was mediated by an unprecedented *Taxus* cytochrome P450 reaction sequence involving oxidation and subsequent cyclizations.⁴⁷ Thus, it seems likely that by protein engineering of the taxadiene 5 α -hydroxylases, termination of the reaction before cyclization will prevent the accumulation of such undesirable byproduct and channeling the flux to taxadien-5 α -ol could be achieved.

The productivity of strain 26-At24T5 α OH-tTCPR was significantly reduced relatively to that of taxadiene production by the parent strain 26 (~300 mg/L) with a concomitant increase in the accumulation of the previously described uncharacterized metabolite. No taxadiene accumulation was observed. Apparently, the introduction of an additional medium copy plasmid (10 copy, p10T7) bearing the At24T5 α OH-tTCPR construct disturbed the carefully engineered balance in the upstream and downstream pathway of strain 26. Small scale fermentations were carried out in bioreactors to quantify the alcohol production by strain 26-At24T5 α OH-tTCPR. The time course profile of taxadien-5 α -ol accumulation (Fig. 5d) indicates alcohol production of up to 58 ± 3 mg/L with an equal amount of the OCT byproduct produced. The observed alcohol production was ~2400 fold higher than previous production in *S. cerevisiae*.¹⁷ Further increases of taxadien-5 α -ol production are likely possible through pathway optimization and protein engineering.

The multivariate-modular approach of pathway optimization has yielded very high producing strains of a critical Taxol precursor. Furthermore, the recombinant constructs have been equally effective in redirecting flux towards the synthesis of other complex pharmaceutical compounds, such as mono-, sesqui- and di-terpene (geraniol, linalool, amorphadiene and levopimaradiene) products engineered from the same pathway (unpublished results). Thus, our pathway engineering opens new avenues to bio-synthesize natural products, especially in the context of microbially-derived terpenoids for use as chemicals and fuels from renewable resources. By focusing on the universal terpenoid precursors IPP and DMAPP, it was possible to, first, define the critical pathway modules and then modulate expression such as to optimally balance the pathway modules for seamless

- 44 -

precursor conversion and minimal intermediate accumulation. This approach seems to be more effective than combinatorial searches of large genetic spaces and also does not depend on a high throughput screen.

The MEP-pathway is energetically balanced and thus overall more efficient in converting either glucose or glycerol to isoprenoids. Yet, during the past 10 years, many attempts at engineering the MEP-pathway in *E. coli* to increase the supply of the key precursors IPP and DMAPP for carotenoid^{28, 47}, sesquiterpenoid²³ and diterpenoid⁶¹ overproduction met with limited success. This inefficiency was attributed to unknown regulatory effects associated specifically with the expression of the MEP-pathway in *E. coli*²³. Here we provide evidence that such limitations are correlated with the accumulation of the metabolite indole, owing to the non-optimal expression of the pathway, which inhibits the isoprenoid pathway activity. Taxadiene overproduction (under conditions of indole formation suppression), establishes the MEP-pathway as a very efficient route for biosynthesis of pharmaceutical and chemical products of the isoprenoid family. One simply needs to carefully balance the modular pathways as suggested by our multivariate-modular pathway engineering approach.

For successful microbial production of Taxol, demonstration of the chemical decoration of the taxadiene core by P450 based oxidation chemistry is essential.⁴¹ Cytochrome P450 monooxygenases constitute about one half of the 19 distinct enzymatic steps in the Taxol biosynthetic pathway. Characteristically, these genes show unusual high sequence similarity with each other (>70%) but low similarity (<30%) with other plant P450s.¹⁴ Due to the apparent similarity among Taxol monooxygenases, expressing the proper activity for carrying out the specific P450 oxidation chemistry was a particular challenge. Through TM engineering and construction of an artificial chimera enzyme with redox partner (TCPR), the Taxol cytochrome P450, taxadiene 5 α -hydroxylase, was functionally expressed in *E. coli* and shown to efficiently convert taxadiene to the corresponding alcohol product *in vivo*. Previous *in vitro* studies have described the mechanism of converting taxadiene to taxadien-5 α -ol by native taxadiene 5 α -hydroxylase enzyme, but have not discussed the same conversion *in vivo*.⁴² This oxygenation and rearrangement reaction involves hydrogen abstraction from C20 position of the taxadiene to form an allylic radical intermediate, followed by regio- and stereo-specific oxygen insertion at C5 position to yield the alcohol derivative (Fig. 5a). The observed modest abundance of the enzyme in *Taxus* cells, and the

- 45 -

low k_{cat} values suggested that the 5 α -hydroxylation step of Taxol biosynthesis is slow relative to the downstream oxygenations and acylations in the Taxol pathway.⁴¹ Thus, engineering this step is key to Taxol synthesis, especially in the context of functional engineering of Taxol P450's in prokaryotic host such as *E. coli*. In addition, this step was limiting in previous efforts of constructing the pathway in yeast.¹⁷ The engineered construct in this study demonstrated >98% conversion of taxadiene *in vivo* with product accumulation to ~60 mg/L, a 2400 fold improvement over previous heterologous expression in yeast. This study has therefore succeeded not only in synthesizing significantly greater amounts of key Taxol intermediates but also provides the basis for the synthesis of subsequent metabolites in the pathway by similar P450 chemistry.

Prior studies on structure-activity relationship on Taxol have shown that alterations made either by removal or addition of some of its functional groups did not change materially the activity of the Taxol.^{1, 48} Such studies, however, were limited due to the restricted ability to introduce changes by chemical synthesis. Availability of a microbial path for Taxol synthesis will drastically expand the space of chemical modifications that can be examined, thus increasing the probability of identifying more potent drug candidates. This offers exciting new opportunities for drug development, especially when considering that such drug candidates will also be associated with an efficient production route.

In the past few decades, Taxol has spawned more interest within the scientific communities and general public than any other natural product drug candidate.¹⁰ A major supply crisis is predicted from the projected increase in the use of Taxol or Taxol analogs for cancer chemotherapy, requiring new production routes, such as engineering of Taxol biosynthetic machinery in microbes.⁸ While a few endophytic fungi of *Taxus* species have been isolated capable of producing Taxol naturally, these microbial systems have yet to demonstrate suitability for sustainable production of the drug.⁴⁹ The results reported here represent a disruptive step towards a microbially-derived Taxol or Taxol precursor, by removing the bottlenecks in the committed precursor pathway. Furthermore, the assembly of a synthetic pathway offers new possibilities to tailor Taxol analogs by selectively engineering the pathway, thereby altering the taxane structure. These developments raise optimism for a microbial route for the cost-effective production of Taxol or suitable Taxol precursors.

- 46 -

Table 1. Ranking of upstream and downstream pathway expression in arbitrary units (a.u.). The MEP pathway and GGPP synthase/taxadiene synthase pathway expression levels were estimated using published values of promoter strengths and copy number. Promoter strengths were calculated as $\text{trc}=1$, $\text{T5}=1.96$, $\text{T7}=4.97$, based on Brosius et. al. and Brunner et. al.^{33,34} Gene copy number was assigned by published copy numbers for origin of replication for the different plasmids used, and one copy was used for integrations.³⁵⁻³⁷ Total expression was calculated as the product of promoter strength and gene copy number. Native expression of the MEP pathway was arbitrarily assigned a value of one, and changing the operon order of GGPP synthase and taxadiene was assumed to affect taxadiene synthase expression by 20%.³⁵ These estimates of total expression guided engineering efforts. E - E coli K12 MG1655 with two deletions $\Delta\text{recA}\Delta\text{endA}$; EDE3 - K12 MG1655 $\Delta\text{recA}\Delta\text{endA}$ with a T7 RNA polymerase (DE3) integrated; MEP – *dxs-idi-ispDF* operon; GT - GPPS-TS operon; TG – TS-GPPS operon; Ch1 – 1 copy in chromosome; Trc – *trc* promoter; T5 – T5 promoter; T7 – T7 promoter; p5 – ~5 copy plasmid (pSC101); , p10 - ~10 copy plasmid (p15A); and p20 - ~20 copy plasmid (pBR322).

Table I

Strain #	Pathway engineering	Upstream MEP				Downstream GT or TG				Taxadiene (mg/L)		
		Construct (in addition to native copy) ^{\$}	Copies	Promoter	Expression Strength* (a.u.)	Construct	Copies		Promoter			Expression Strength (a.u.) [#]
							GT/TG					
1	Ep20TrcGT	N/A	0	0	1*	pBR322	20	20	1.00	20	0.02	0.01
2	ECh1TrcMEPp20TrcGT	Chr. ^{\$}	1	1	2	pBR322	20	20	1.00	20	16.00	1.59
3	Ep5TrcMEPp20TrcGT	pSC101	5	1	6	pBR322	20	20	1.00	20	2.55	0.21
4	Ep10TrcMEPp20TrcGT	p15A	10	1	11	pBR322	20	20	1.00	20	1.93	0.323
5	Ep20TrcTG	N/A	0	0	1	pBR322	20	20	1.20	24	0.19	0.01
6	Ep20T5GT	N/A	0	0	1	pBR322	20	20	1.96	39	4.36	0.533
7	Ep20T5GTTTrcT	N/A	0	0	1	pBR322	20	20	2.96	59	1.74	0.265
8	ECh1TrcMEPp20TrcTG	Chr.	1	1	2	pBR322	20	20	1.20	24	45.44	2.28
9	ECh1TrcMEPp20T5GT	Chr.	1	1	2	pBR322	20	20	1.96	39	16.52	0.84
10	ECh1TrcMEPp20T5GT-TrcT	Chr.	1	1	2	pBR322	20	20	2.96	59 [#]	2.52	0.30
11	Ep5TrcMEPp20TrcTG	pSC101	5	1	6	pBR322	20	20	1.20	24	7.41	0.63
12	Ep5TrcMEPp20T5GT	pSC101	5	1	6	pBR322	20	20	1.96	39	21.23	5.86
13	Ep5TrcMEPp20T5TG-TrcT	pSC101	5	1	6	pBR322	20	20	2.96	59	1.40	0.10
14	Ep10TrcMEPp20TrcTG	p15A	10	1	11	pBR322	20	20	1.20	24	2.36	0.29
15	Ep10TrcMEPp20T5GT	p15A	10	1	11	pBR322	20	20	1.96	39	8.91	2.94
16	Ep10TrcMEPp20T5GT-TrcT	p15A	10	1	11	pBR322	20	20	2.96	59	3.40	0.39

17	EDE3p10TrcMEPp5T7TG	p15A	10	1	11	pSC101	5	5.96	31	125.00	8.37
18	EDE3p20TrcMEPp5T7TG	pBR322	20	1	21	pSC101	5	5.96	31	58.00	3.07
19	EDE3p20T5MEPp5T7TG	pBR322	20	1.96	40	pSC101	5	5.96	31	44.00	2.88
20	EDE3p20T7MEPp5T7TG	pBR322	20	4.97	100	pSC101	5	5.96	31	32.00	6.63
21	EDE3p5TrcMEPp10T7TG	pSC101	5	1	6	p15A	10	5.96	61	7.00	1.40
22	EDE3p20TrcMEPp10T7TG	pBR322	20	1	21	p15A	10	5.96	61	59.00	5.57
23	EDE3p20T5MEPp10T7TG	pBR322	20	1.96	40	p15A	10	5.96	61	58.00	5.68
24	EDE3p20T7MEPp10T7TG	pBR322	20	4.97	100	p15A	10	5.96	61	20.00	0.73
25	EDE3p5T7TG	N/A	0	0	1	pSC101	5	5.96	31	19.00	8.23
26	EDE3Ch1TrcMEPp5T7TG	Chr.	1	1	2	pSC101	5	5.96	31	297.00	10.21
27	EDE3Ch1T5MEPp5T7TG	Chr	1	1.96	3	pSC101	5	5.96	31	163.00	10.84
28	EDE3Ch1T7MEPp5T7TG	Chr	1	4.97	6	pSC101	5	5.96	31	26.00	0.32
29	EDE3p10T7TG	N/A	0	0	1	p15A	10	5.96	61	8.00	0.39
30	EDE3Ch1TrcMEPp10T7TG	Chr	1	1	2	p15A	10	5.96	61	30.00	1.59
31	EDE3Ch1T5MEPp10T7TG	Chr	1	1.96	3	p15A	10	5.96	61	40.00	0.56
32	EDE3Ch1T7MEPp10T7TG	Chr	1	4.97	6	p15A	10	5.96	61	17.00	0.41

*A value of 1 was given to account for the native copies of the MEP pathway.

§ MEP construct is localized in the chromosome.

p20T5GT-TrcT - An additional copy of gene T under separate promoter control (Trc) together operon GT (under T5 promoter) on the same plasmid. For the calculation strength, we have added the value as equivalent two separate operons ($\text{TrcT} + \text{T5GT} = (20 \times 1.96 + 20 \times 1 = 59)$) since our studies shows that expression of T was limited compared to G.

No	Plasmid	Origin of replication	Antibiotic marker
1	p20T7MEP	pBR322	Amp
2	p20TrcMEP	pBR322	Amp
3	p20T5MEP	pBR322	Amp
4	p20T7MEPKmFRP	pBR322	Km
5	p20T5MEPKmFRP	pBR322	Km
6	p20TrcMEPKm-FRP	pBR322	Km
7	p10TrcMEP	p15A	Cm
8	p5TrcMEP	SC101	Spect
9	p20TrcGT	pBR322	Amp
10	p20TrcTG	pBR322	Amp
11	p20T5GT	pBR322	Amp
12	p10T7TG	p15A	Cm
13	p5T7TG	SC101	Spect
14	p10At8T5 α OH-tTCPR	p15A	Cm
15	p10At24T5 α OH-tTCPR	p15A	Cm
16	p10At42T5 α OH-tTCPR	p15A	Cm

Table 2. Detail of all the plasmids constructed for the study

SEQ ID NO	Primer Name	Sequences
1	dxsNdeI(s)	CGGCATATGAGTTTTGATATTGCCAAATACCCG
2	dxsNheI(a)	CGGCTAGCTTATGCCAGCCAGGCCTTGATTTTG
3	idiNheI (s)	CGCGGCTAGCGAAGGAGATATACATATGCAAACGGAAC ACGTCATTTTATTG
4	idiEcoRI(a)	CGGAATTCGCTCACAACCCCGGCAAATGTCGG
5	ispDFEcoRI(s)	GCGAATTCGAAGGAGATATACATATGGCAACCACTCATT TGGATGTTTG

- 50 -

6	ispDFXhoI(a)	GCGCTCGAGTCATTTTGTTCCTTAATGAGTAGCGCC
7	dxxsidiispDFNcoI(s)	TAAACCATGGGTTTTGATATTGCCAAATACCCG
8	dxxsidiispDFKpnI(a)	CGGGGTACCTCATTTTGTTCCTTAATGAGTAGCGC
9	dxxsidiispDFXhoI(a)	CGGCTCGAGTCATTTTGTTCCTTAATGAGTAGCGC
10	T5AgeI(s)	CGTAACCGGTGCCTCTGCTAACCATGTTTCATGCCTTC
11	T5NheI(a)	CTCCTTCGCTAGCTTATGCCAGCC
52	GGPPSNcoI(s)	CGTACCATGGTTGATTTCAATGAATATATGAAAAGTAAG GC
12	GGPPSEcoRI(a)	CGTAGAATTCACCTCACAACCTGACGAAACGCAATGTAATC
13	TXSEcoRI(s)	CGTAGAATTCAGAAGGAGATATACATATGGCTAGCTCTA CGGGTACG
14	TXSsalI(a)	GATGGTCGACTTAGACCTGGATTGGATCGATGTAAAC
15	TXSNcoI(s)	CGTACCATGGCTAGCTCTACGGGTACG
16	TXSEcoRI(a)	CGTAGAATTCTTAGACCTGGATTGGATCGATGTAAAC
17	GGPPSEcoRI(s)	CGTAGAATTCAGAAGGAGATATACATATGTTTGATTCA ATGAATATATGAAAAGTAAGGC
18	GGPPSSalI(a)	GATGGTCGACTCACAACCTGACGAAACGCAATGTAATC
19	TSXhoI(a)	GATGCTCGAGTTAGACCTGGATTGGATCGATGTAAAC
20	pTrcSal(s)	GCCGTCGACCATCATCATCATC
21	pTrcXba(a)	GCAGTCTAGAGCCAGAACCGTTATGATGTCGGCGC
22	pCLBspEI(s)	CGTGTCCGGAGCATCTAACGCTTGAGTTAAGCCGC
23	pCLXbaI(a)	GCAGTCTAGAGGAAACCTGTCGTGCCAGCTGC
24	KmFRPXhoI(s)	GACGCTCGAGGAGCAATAACTAGCATAACCCCTTGGGGC CTCTAAACGGGTCTTGAGGGGTTTTTGTCTGTGTAGGCT GGAGCTGCTTCG
25	KmFRPSalI(a)	GACGAGTACTGAACGTCGGAATTGATCCGTCGAC
26	KmFRPSacI(s)	GACGGAGCTCGAGCAATAACTAGCATAACCCCTTGGGGC CTCTAAACGGGTCTTGAGGGGTTTTTGTCTGTGTAGGCT GGAGCTGCTTCG
27	IntT7T5(s)	ATGACGATTTTTGATAATTATGAAGTGTGTTTGTTCATTG

- 51 -

		CATTAATTGCGTTGCGCTCACTG
28	IntTrc(s)	ATGACGATTTTGTATAATTATGAAGTGTGGTTTGTTCATTG GCATCCGCTTACAGACAAGCTGTG
29	Int(a)	TTAGCGACGAAACCCGTAATACACTTCGTTCCAGCGCAG CCGACGTCGGAATTGATCCGTCGAC
30	CYP17At8AANdeI(s)	CGTACATATGGCTCTGTTATTAGCAGTTTTTGTGGCGAAA TTTAACGAAGTAACCCAGC
31	CYP17At24AANdeI(s)	CGTACATATGGCTCTGTTATTAGCAGTTTTTTTTTAGCATC GCTTTGAGTGCAATTG
32	CYP17At42AANdeI(s)	CGTACATATGGCTCTGTTATTAGCAGTTTTTTTTTCGCTCG AAACGTCATAGTAGCCTG
33	CYPLinkBamHI(a)	CGCGGGATCCGGTGCTGCCCGGACGAGGGAACAGTTTGA TTGAAAACCC
34	CPRBamHI(s)	CGCGGGATCCCGCCGTGGTGGAAGTGATACACAG
35	CPRSall(a)	CGCGGTCGACTTACCAAATATCCCGTAAGTAGCGTCCAT C
36	DXS qPCR (s)	A T T C A A A A G C T T C C G G T C C T
37	DXS qPCR (a)	A T C T G G C G A C A T T C G T T T T C
38	TS qPCR (s)	G A C G A A C T G T C A C C C G A T T T
39	TS qPCR (a)	G C T T C G C G G G T A G T A G A C A G
40	rrsA qPCR (s)	A G G C C T T C G G G T T G T A A A G T
41	rrsa qPCR (a)	A T T C C G A T T A A C G C T T G C A C

Table 3. Details of the primer used for the cloning of plasmids, chromosomal delivery of the MEP pathway and qPCR measurements.

Table 4. Protein and Codon optimized nucleotide sequences.**GGPP synthase**

MFDFNEYMKSKAVAVDAALDKAIPLEYPEKIHESMRYSLLAGGKRVRPALCIAACE
 LVGGSQDLAMPTACAMEMIHTMSLIHDDLPCMDNDDFRRGKPTNHKVFGEDEVAVL
 AGDALLSFAFEHIAVATSKTVPSDRTLRLVISELGKTIGSQGLVGGQVVDITSEGDANV
 DLKTLEWIIHHKTAVLLECSVVSGILGGATEDEIARIRRYARCVGLLFQVVDDILDV
 TKSSEELGKTAGKDLLTDKATYPKLMGLEKAKEFAAELATRAKEELSSFDQIKAAPL
 LGLADYIAFRQN (SEQ ID NO:42)

ATGTTTGATTTCAATGAATATATGAAAAGTAAGGCTGTTGCGGTAGACGCGGCTC
 TGGATAAAGCGATTCCGCTGGAATATCCCGAGAAGATTCACGAATCGATGCGCT
 ACTCCCTGTTAGCAGGAGGGAAACGCGTTCGTCGCGCATTATGCATCGCGGCCTG
 TGAACGCTCGGCGGTTACAGGACTTAGCAATGCCAACTGCTTGCGCAATGGA
 AATGATTCACACAATGAGCCTGATTCATGATGATTTGCCTTGATGGACAACGAT
 GACTTTCGGCGCGGTAAACCTACTAATCATAAGGTTTTTGGCGAAGATACTGCAG
 TGCTGGCGGGCGATGCGCTGCTGTCGTTTGCCTTCGAACATATCGCCGTCGCGAC
 CTCGAAAACCGTCCCGTCGGACCGTACGCTTCGCGTGATTTCCGAGCTGGGAAAG
 ACCATCGGCTCTCAAGGACTCGTGGGTGGTCAGGTAGTTGATATCACGTCTGAGG
 GTGACGCGAACGTGGACCTGAAAACCTGGAGTGGATCCATATTCACAAAACGG
 CCGTGCTGCTGGAATGTAGCGTGGTGTGAGGGGGGATCTTGGGGGGCGCCACGG
 AGGATGAAATCGCGCGTATTCGTCGTTATGCCCCTGTGTTGGACTGTTATTTCA
 GGTGGTGGATGACATCCTGGATGTCACAAAATCCAGCGAAGAGCTTGGCAAGAC
 CGCGGGCAAAGACCTTCTGACGGATAAGGCTACATACCCGAAATTGATGGGCTT
 GGAGAAAGCCAAGGAGTTCGCAGCTGAACTTGCCACGCGGGCGAAGGAAGAAC
 TCTCTTCTTTCGATCAAATCAAAGCCGCGCCACTGCTGGGCCTCGCCGATTACAT
 TCGGTTTCGTCAGAAC (SEQ ID NO:43)

Taxadiene synthase

MSSSTGTSKVVSETSSSTIVDDIPRLSANYHGDWLWHHNVQITLETTFRESSTYQERADE
 LVVKIKDMFNALGDGDISPASYDTAWVARLATISSDGSEKPRFPQALNWVFNNQLQ
 DGSWGIESHFSLCDRLNNTNSVIALSVWKTGHSQVQQGAEFIAENLRLLNEEDELSP
 DFQIIFPALLQKAKALGINLPYDLPIFYKYSTTREARLTDVSAAADNIPANMLNALEGL
 EEVIDWNKIMRFQSKDGSFLSSPASTACVLMNTGDEKCFTFLNNLLDKFGGCVPCM
 YSIDLLERLSLVDNIEHLGIGRHFKEIKGALDYVYRHWSERGIGWGRDSLVPDLNTT
 ALGLRTLRLMHGYNVSSDVLNNFKDENGRFPSSAGQTHVELRSVVNLFRAFDLAFPD
 ERAMDDARKFAEPLYREALATKISTNTKLFKEIEYVVEYPWHMSIPRLEARSYIDSYD
 DNYVWQRKTLRYMPSLSNSKCLELAKLDFNIVQSLHQEELKLLTRWWKESGMADI
 NFTRHRVAEYVFSSATFEPEYSATRIAFTKIGCLQVLFDDMADIFATLDELKSFTGEGV
 KRWDTSLLHEIPECMQTCFKVWFKLMEEVNNDVVKVQGRDMLAHIRKPWELYFNC
 YVQEREWLEAGYIPTFEEYLKTYAISVGLGPCTLQPILLMGELVKDDVVEKVHYPSN
 MFELVSLSWRLTNDTKTYQAEKARGQQASGIACYMKDNPGATEEDAIIKHICRVVDR
 ALKEASFYFKPSNDIPMGCKSFIFNLRLCVQIFYKFIDGYGIANEEIKDYIRKVYIDPI
 QV (SEQ ID NO:44)

ATGTCTAGCTCTACGGGTACGTCTAAAGTCGTGAGTGAAACCTCATCGACGATCG
 TGGACGATATTCCACGCTTGTCGGCGAACTATCATGGAGATCTGTGGCATCATAA

- 53 -

CGTCATTCAGACATTGGAAACCCCGTTTCGCGAAAGTAGCACCTACCAGGAACG
GGCAGATGAATTAGTCGTGAAAATCAAAGATATGTTTAATGCATTAGGAGATGG
AGACATCTCGCCAGCGCATATGATACGGCGTGGGTGGCTCGGTTGGCCACGATT
AGCTCCGATGGCAGTGAAAAGCCGCGTTTCCCGCAGGCGCTGAACTGGGTGTTT
AATAATCAATTGCAGGATGGCAGCTGGGGCATTGAATCTCACTTTAGCCTCTGTG
ACCGGTTACTCAACACGACAAACTCCGTAATTGCGTTGTCAGTTTGGAAAACGGG
CCATAGCCAGGTTCAACAGGGCGCGGAATTTATCGCTGAAAATCTGCGCCTGCTG
AACGAGGAGGACGAACTGTCACCCGATTTTCAGATTATTTTTCCGGCTTTACTCC
AGAAAGCCAAAGCCTTAGGCATCAACCTGCCATATGATCTGCCGTTTCATCAAGTA
TCTGTCTACTACCCGCGAAGCCCGTCTCACTGACGTCTCTGCGGCGGCGGACAAT
ATTCCAGCGAACATGCTGAACGCACTGGAAGGGCTGGAAGAGGTTATCGACTGG
AATAAAATCATGCGCTTCCAAAGCAAGGACGGTAGCTTCTTAAGCAGCCCAGCA
TCTACTGCTTGTGTTCTGATGAATACCGGAGACGAAAAGTGCTTTACGTTTCTGA
ACAATCTGCTGGACAAATTTGGGGGTTGTGTTCCCTTGATGTATTCCATTGATCTG
TTGGAACGCTGTCGCTGGTCGATAACATTGAACACTTAGGTATCGGCCGCCACT
TCAAACAAGAAATCAAGGGGGCGTGGATTATGTATACCGTCATTGGAGCGAGC
GTGGTATTGGTTGGGGGCGCGATAGCTTGGTACCTGATCTGAACACCACTGCTTT
GGGACTGCGCACTCTTCGTATGCACGGATAACAACGTTAGTTCCGATGTCTCTCAAT
AATTTCAAGGACGAGAACGGCCGTTTTTTCAGCTCGGCCGGTCAGACGCATGTTG
AACTGCGGTCCGTAGTCAATCTCTTTCGCGCTAGTGATCTGGCCTTCCCCGACGA
GCGCGCTATGGACGATGCACGGAAGTTTGCCGAGCCGTATCTCCGCGAAGCCCT
GGCCACCAAAATTTCAACCAACACCAAGCTTTTCAAAGAAATTGAGTATGTAGT
AGAGTATCCGTGGCATATGTCTATTCCGCGCCTGGAAGCCCGCTCGTATATCGAT
TCTTACGATGACAATTATGTGTGGCAACGCAAAACACTGTACCGTATGCCAGCC
TGTCAAATAGTAAGTGTCTGGAGCTGGCGAAACTGGATTTCAACATTGTGCAATC
CCTGCACCAAGAAGAGCTGAAATTACTGACTCGCTGGTGGAAAGGAATCCGGCAT
GGCAGACATCAATTTTACGCGTCACCGTGTGTCAGAGGTGTACTTCTCCTCGGCG
ACCTTTGAGCCGGAGTATTCGGCCACACGTATTGCATTTACCAAGATTGGCTGCC
TTCAGGTGCTTTTTTGACGATATGGCGGATATTTTTGCGACACTTGATGAGCTTAA
ATCATTTACCGAAGGCGTGAAAGCGTTGGGATACCTCTCTGTTGCATGAAATCCCC
GAATGTATGCAGACCTGCTTCAAAGTTTGGTTCAAACACTGATGGAAGAAGTGAAC
AACGACGTCGTGAAAGTTCAGGGTCGTGATATGTTAGCACACATCCGCAAGCCG
TGGAAGTCTATTTCAATTGCTATGTGCAGGAGCGTGAATGGTTAGAAGCGGGCT
ACATTCCTACCTTCGAAGAGTACTTAAAAACCTATGCCATTTCCGTCCGTTTAGG
CCCGTGCACTCTGCAGCCTATCTTGCTGATGGGTGAGCTGGTAAAGGATGATGTG
GTGGAAGAAAGTTCACTACCCGTCGAATATGTTTGAACCTGGTAAGTCTGAGTTGGC
GTCTGACAAACGACACCAAAACGTACCAGGCAGAAAAGGCACGTGGGCAACAG
GCAAGCGGTATCGCGTGTATATGAAGGATAATCCGGGCGCTACTGAGGAAGAT
GCCATTAAGCATATCTGCCGTGTTGTGGATCGCGCTCTTAAAGAAGCGTCATTTCG
AATATTTTAAACCTAGTAATGATATTCCGATGGGTTGTAAGTCATTCAATTTCAAT
CTTCGCCTGTGCGTGCAAATTTTTTACAAATTTATTGACGGCTACGGAATCGCCA
ACGAAGAAATCAAAGACTATATTCGTAAAGTTTACATCGATCCAATCCAGGTC
(SEQ ID NO:45)

Cytochrome P450 Taxadiene 5 α -hydroxylase (T5 α OH)

MDALYKSTVAKFNEVTQLDCSTESFSIALSAIAGILLLLLFRSKRHSSSLKLPPGKLGIP
FIGESFIFLRALRSNSLEQFFDERVKKFGLVFKTSLIGHPTVVLCPAGNRLILSNEEKL
VQMSWPAQFMKLMGENSVATRRGEDHIVMRSALAGFFGPGALQSYIGKMNTEIQS

- 54 -

HINEKWKGKDEVNVLPLVREL VFNISAILFFNIYDKQE QDRLHKLLETILVGSFALPID
LPGFGFHRALQGRAKLNKIMLSLIKRRKEDLQSGSATATQDLLSVLLTFRDDKGTP
LNDEILDNFSSLLHAS YDTTSPMALIFKLLSSNPECYQKV VQEQLILSNKEEGEEIT
WKDLKAMKYTWQVAQETLRMFPPVFGTFRKAITDIQYDGYTIPKGWKLWTTYST
HPKDLYFNEPEKFMPSRFDQEGKHVAPYTFLPFGGGQRSCVGVWEFSKMEILLFVHHF
VKTFSSYTPVDPDEKISGDPLPLPSKGFSIKLFPRP (SEQ ID NO:46)

ATGGATGCCCTCTATAAGTCTACCGTGGCGAAATTTAACGAAGTAACCCAGCTGG
ATTGCAGCACTGAGTCATTTAGCATCGCTTTGAGTGCAATTGCCGGGATCTTGCT
GTTGCTCCTGCTGTTTCGCTCGAAACGTCATAGTAGCCTGAAATTACCTCCGGGC
AAACTGGGCATTCCGTTTATCGGTGAGTCTTTATTTTTTTGCGCGCGCTGCGCAG
CAATTCTCTGGAACAGTTCTTTGATGAACGTGTGAAGAAGTTCGGCCTGGTATTT
AAAACGTCCCTTATCGGTACCCCGACGGTTGTCTGTGCGGGCCCCGCAGGTAATC
GCCTCATCTGAGCAACGAAGAAAAGCTGGTACAGATGTCTGGCCGGCGCAGT
TTATGAAGCTGATGGGAGAGAACTCAGTTGCGACCCGCCGTGGTGAAGATCACA
TTGTTATGCGCTCCGCGTTGGCAGGCTTTTTCGGCCCGGGAGCTCTGCAATCCTAT
ATCGGCAAGATGAACACGGAAATCCAAAGCCATATTAATGAAAAGTGGAAGG
GAAGGACGAGGTTAATGTCTTACCCCTGGTGCGGGAACTGGTTTTTAACATCAGC
GCTATTCTGTTCTTTAACATTTACGATAAGCAGGAACAAGACCGTCTGCACAAGT
TGTTAGAAACCATTTCTGGTAGGCTCGTTTGCCTTACCAATTGATTTACCGGGTTTC
GGGTTTCACCGCGCTTTACAAGGTCGTGCAAACTCAATAAAATCATGTTGTGCGC
TTATTAACAAAACGTAAAGAGGACTTACAGTCGGGATCGGCCACCGCGACGCAGG
ACCTGTTGTCTGTGCTTCTGACTTTCCGTGATGATAAGGGCACCCCGTTAACCAA
TGACGAAATCCTGGACAACCTTTAGCTCACTGCTTCACGCCTCTTACGACACCACG
ACTAGTCCAATGGCTCTGATTTTCAAATTACTGTCAAGTAACCCCTGAATGCTATC
AGAAAGTCGTGCAAGAGCAACTCGAGATTCTGAGCAATAAGGAAGAAGGTGAA
GAAATTACCTGGAAAGATCTTAAGGCCATGAAATACACGTGGCAGGTTGCGCAG
GAGACACTTCGCATGTTTCCACCGGTGTTCTGGGACCTTCCGCAAAGCGATCACGG
ATATTCAGTATGACGGATACACAATCCCGAAAGGTTGGAAACTGTTGTGGACTA
CCTATAGCACTCATCCTAAGGACCTTTACTTCAACGAACCGGAGAAATTTATGCC
TAGTCGTTTTCGATCAGGAAGGCAAACATGTTGCGCCCTATACCTTCCTGCCCTTT
GGAGGCGGTCAGCGGAGTTGTGTGGGTTGGGAGTTCTCTAAGATGGAGATTCTC
CTCTTCGTGCATCATTTTCGTGAAAACATTTTCGAGCTATACCCCGGTTCGATCCCG
ATGAAAAAATTTCCGGCGATCCACTGCCGCCGTTACCGAGCAAAGGGTTTTCAAT
CAAACGTTCCTCGTCCG (SEQ ID NO:47)

Taxus NADPH:cytochrome P450 reductase (TCPR)

MQANSNTVEGASQGKSLDISRLDHIFALLNKGKGGDLGAMTGSALILTENSQNLMI
LTTALAVLVACVFFVWRRGSDTQKPAVRPTPLVKEEDEEEEDDSAKKKVTIFFGT
QTGTAEGFAKALAEAKARYEKAVFKVVDLDNYAADDEQYEEKLKEKLAFFMLA
TYGDGEPTDNAARFYKWFLEKEREPLWSDLTYGVFGLGNRQYEHFNKVAKAVDE
VLIEQGA KRLVPVGLGDDDDQCIEDDFTAWREQVWPELDQLLRDEDEPTSATPYTA
AIPEYRVEIYDSVVSVEETHALKQNGQAVYDIHHPCRSNVAVRRELHTPLSDRSCH
LEFDISDTGLIYETGDHVGVTHTENSIETVEEAAKLLGYQLDTIFSVHGDKEGTPLGG
SSLPPFP GPCTLR TALARYADLLNPPRKA AFLALAAHASDPAEAERLKFLSSPAGKD
EYSQWVTASQRSLL EIMAEFSAKPPLGVFFAAIAPRLQPRYYSISSSPRFAPSRIHVT
ALVYGPSPTGRIHKGVCSNWMKNSLPSEETHDCSWAPVFVRQSNFKLPADSTTPIVM
VGPGTGFAFPFRGFLQERAKLQEAGEKLGPVLFVFGCRNRQMDYTYEDELKGYVEKG

- 55 -

ILTNLIVAFSREGATKEYVQHKMLEKASDTWSLIAQGGYLYVCGDAKGMARDVHR
TLHTIVQEQUESVDSSKAFLVKKLQMDGRYLRIW (SEQ ID NO:48)

ATGCAGGCGAATTCTAATACGGTTGAAGGCGCGAGCCAAGGCAAGTCTCTTCTG
GACATTAGTCGCCTCGACCATATCTTCGCCCTGCTGTTGAACGGGAAAGGCGGAG
ACCTTGGTGCGATGACCGGGTCGGCCTTAATTCTGACGGAAAATAGCCAGAACTT
GATGATTCTGACCACTGCGCTGGCCGTTCTGGTCGCTTGCGTTTTTTTTTCGTTT
GGCGCCGTGGTGGAAGTGATACACAGAAGCCCCGCCGTACGTCCCACACCTCTTG
TTAAAGAAGAGGACGAAGAAGAAGAAGATGATAGCGCCAAGAAAAAGGTCACA
ATATTTTTTGGCACCCAGACCGGCACCGCCGAAGGTTTCGCAAAGGCCTTAGCTG
AGGAAGCAAAGGCACGTTATGAAAAGGCGGTATTTAAAGTCGTGGATTTGGATA
ACTATGCAGCGGATGACGAACAGTACGAAGAGAAAGTTGAAAAAGGAAAAGCTA
GCGTTCTTCATGCTCGCCACCTACGGTGACGGCGAACCGACTGATAATGCCGCTC
GCTTTTATAAATGGTTTCTCGAGGGTAAAGAGCGCGAGCCATGGTTGTCAGATCT
GACTTATGGCGTGTTTGGCTTAGGTAACCGTCAGTATGAACACTTTAACAAGGTC
GCGAAAGCGGTGGACGAAGTGCTCATTGAACAAGGCGCCAAACGTCTGGTACCG
GTAGGGCTTGGTGATGATGATCAGTGCATTGAGGACGACTTCACTGCCTGGAGA
GAACAAGTGTGGCCTGAGCTGGATCAGCTCTTACGTGATGAAGATGACGAGCCG
ACGTCTGCGACCCCGTACACGGCGGCTATTCCAGAATACCGGGTGGAATCTAC
GACTCAGTAGTGTCGGTCTATGAGGAAACCCATGCGCTGAAACAAAATGGACAA
GCCGTATACGATATCCACCACCCGTGTCGCAGCAACGTGGCAGTACGTCGTGAG
CTGCATAACCCCGCTGTCGGATCGTAGTTGTATTTCATCTGGAATTCGATATTAGTG
ATACTGGGTAAATCTATGAGACGGGCGACCACGTTGGAGTTCATACCGAGAATTC
AATTGAAACCGTGGAAGAAGCAGCTAAACTGTTAGGTTACCAACTGGATACAAT
CTTCAGCGTGATGGGGACAAGGAAGATGGAACACCATTGGGCGGGAGTAGCCT
GCCACCGCCGTTTCCGGGGCCCTGCACGCTGCGGACGGCGCTGGCACGTTACGC
GGACCTGCTGAACCCTCCGCGCAAAGCCGCCTTCCTGGCACTGGCCGCACACGC
GTCAGATCCGGCTGAAGCTGAACGCCTTAAATTTCTCAGTTCTCCAGCCGGAAAA
GACGAATACTCACAGTGGGTCACTGCGTCCCAACGCAGCCTCCTCGAGATTATGG
CCGAATTCCCCAGCGCGAAACCGCCGCTGGGAGTGTTTTTCGCCGCAATAGCGCC
GCGCTTGCAACCTAGGTATTATAGCATCTCCTCCTCCCCGCGTTTCGCGCCGTCTC
GTATCCATGTAACGTGCGCGCTGGTCTATGGTCCTAGCCCTACGGGGCGTATTCA
TAAAGGTGTGTGCAGCAACTGGATGAAGAATTCTTTGCCCTCCGAAGAAACCCA
CGATTGCAGCTGGGCACCGGTCTTTGTGCGCCAGTCAAACCTTTAAACTGCCCGCC
GATTCGACGACGCCAATCGTGATGGTTGGACCTGGAACCGGCTTCGCTCCATTT
GCGGCTTCCTTCAGGAACGCGCAAACTGCAGGAAGCGGGCGAAAAATTGGGCC
CGGCAGTGCTGTTTTTTGGGTGCCGCAACCGCCAGATGGATTACATCTATGAAGA
TGAGCTTAAGGGTTACGTTGAAAAAGGTATTCTGACGAATCTGATCGTTGCATTT
TCACGAGAAGGCGCCACCAAAGAGTATGTTACGCACAAGATGTTAGAGAAAGCC
TCCGACACGTGGTCTTTAATCGCCCAGGGTGGTTATCTGTATGTTTGCGGTGATG
CGAAGGGTATGGCCAGAGACGTACATCGCACCCCTGCATACAATCGTTCAGGAAC
AAGAATCCGTAGACTCGTCAAAAGCGGAGTTTTTAGTCAAAAAGCTGCAAATGG
ATGGACGCTACTTACGGGATATTTGG (SEQ ID NO:49)

References

1. Kingston, D. G. The shape of things to come: structural and synthetic studies of taxol and related compounds. *Phytochemistry* 68, 1844-54 (2007).
- 5 2. 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, 2325-7 (1971).
3. Suffness M, W. M. Discovery and development of taxol. (ed. (ed), S. M.) (CRC Press, Boca Raton, 1995).
- 10 4. Nicolaou, K. C. et al. Total synthesis of taxol. *Nature* 367, 630-4 (1994).
5. Holton, R. A. et al. First total synthesis of taxol. 2. Completion of the C and D rings. *Journal of the American Chemical Society* 116, 1599-1600 (1994).
6. Walji, A. M. & MacMillan, D. W. C. Strategies to Bypass the Taxol Problem. Enantioselective Cascade Catalysis, a New Approach for the Efficient Construction of Molecular Complexity. *Synlett* 18, 1477-1489 (2007).
- 15 7. Holton RA, B. R., Boatman PD. Semisynthesis of taxol and taxotere (ed. M, S.) (CRC Press, Boca Raton, 1995).
8. Frense, D. Taxanes: perspectives for biotechnological production. *Applied microbiology and biotechnology* 73, 1233-1240 (2007).
- 20 9. Roberts, S. C. *Production and engineering of terpenoids in plant cell culture*. *Nature Chemical Biology* 3, 387-395 (2007).
10. Goodman, J. & Walsh, V. The story of taxol : nature and politics in the pursuit of an anti-cancer drug (Cambridge University Press, Cambridge ; New York, 2001).
11. Tyo, K. E., Alper, H. S. & Stephanopoulos, G. N. Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol* 25, 132-7 (2007).
- 25 12. Ajikumar, P. K. et al. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol Pharm* 5, 167-90 (2008).
13. Jennewein, S. & Croteau, R. Taxol: biosynthesis, molecular genetics, and biotechnological applications. *Appl Microbiol Biotechnol* 57, 13-9 (2001).
- 30 14. 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 U S A* 101, 9149-54 (2004).
15. Croteau, R., Ketchum, R. E. B., Long, R. M., Kaspera, R. & Wildung, M. R. Taxol biosynthesis and molecular genetics. *Phytochemistry Reviews* 5, 75-97 (2006).
- 35 16. Walker, K. & Croteau, R. Taxol biosynthetic genes. *Phytochemistry* 58, 1-7 (2001).
17. Dejong, J. M. et al. Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 93, 212-24 (2006).
18. Engels, B., Dahm, P. & Jennewein, S. Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metabolic engineering* 10, 201-206 (2008).
- 40 19. Chang, M. C. Y. & Keasling, J. D. Production of isoprenoid pharmaceuticals by engineered microbes. *Nature chemical biology* 2, 674-681 (2006).
20. Khosla, C. & Keasling, J. D. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* 2, 1019-25 (2003).
- 45 21. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 23, 612-6 (2005).

22. Chang, M. C., Eachus, R. A., Trieu, W., Ro, D. K. & Keasling, J. D. Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nat Chem Biol* 3, 274-7 (2007).
23. Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* 21, 796-802 (2003).
24. Huang, Q., Roessner, C. A., Croteau, R. & Scott, A. I. Engineering *Escherichia coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorg Med Chem* 9, 2237-42 (2001).
25. Kim, S. W. & Keasling, J. D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol Bioeng* 72, 408-15 (2001).
26. Farmer, W. R. & Liao, J. C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nature Biotechnology* 18, 533-537 (2000).
27. Farmer, W. R. & Liao, J. C. Precursor balancing for metabolic engineering of lycopene production in *Escherichia coli*. *Biotechnology progress* 17 (2001).
28. Yuan, L. Z., Rouviere, P. E., Larossa, R. A. & Suh, W. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab Eng* 8, 79-90 (2006).
29. Jin, Y. S. & Stephanopoulos, G. Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. *Metabolic Engineering* 9, 337-347 (2007).
30. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* (2009).
31. Klein-Marcuschamer, D., Ajikumar, P. K. & Stephanopoulos, G. Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene. *Trends in Biotechnology* 25, 417-424 (2007).
32. Sandmann, G. Combinatorial biosynthesis of carotenoids in a heterologous host: a powerful approach for the biosynthesis of novel structures. *ChemBioChem* 3 (2002).
33. Brosius, J., Erfle, M. & Storella, J. Spacing of the -10 and -35 regions in the *tac* promoter. Effect on its *in vivo* activity. *J Biol Chem* 260, 3539-41 (1985).
34. Brunner, M. & Bujard, H. Promoter recognition and promoter strength in the *Escherichia coli* system. *Embo J* 6, 3139-44 (1987).
35. Nishizaki, T., Tsuge, K., Itaya, M., Doi, N. & Yanagawa, H. Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis*. *Appl Environ Microbiol* 73, 1355-61 (2007).
36. Sorensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 115, 113-28 (2005).
37. Jones, K. L., Kim, S. W. & Keasling, J. D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* 2, 328-38 (2000).
38. Hoffmann, F. & Rinas, U. Stress induced by recombinant protein production in *Escherichia coli*. *Advances in Biochemical Engineering Biotechnology* 89, 73-92 (2005).
39. Hoffmann, F., Weber, J. & Rinas, U. Metabolic adaptation of *Escherichia coli* during temperature-induced recombinant protein production: 1. Readjustment of metabolic enzyme synthesis. *Biotechnology and bioengineering* 80 (2002).
40. Chang, D. E., Smalley, D. J. & Conway, T. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Molecular microbiology* 45, 289-306 (2002).

41. Kaspera, R. & Croteau, R. Cytochrome P450 oxygenases of Taxol biosynthesis. *Phytochemistry Reviews* 5, 433-444 (2006).
42. Jennewein, S., Long, R. M., Williams, R. M. & Croteau, R. Cytochrome p450 taxadiene 5 α -hydroxylase, a mechanistically unusual monooxygenase catalyzing the first
5 oxygenation step of taxol biosynthesis. *Chem Biol* 11, 379-87 (2004).
43. Schuler, M. A. & Werck-Reichhart, D. FUNCTIONAL GENOMICS OF P450 S. *Annual Review of Plant Biology* 54, 629-667 (2003).
44. Leonard, E. & Koffas, M. A. G. Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Applied and Environmental Microbiology* 73, 7246 (2007).
45. Nelson, D. R. Cytochrome P450 and the individuality of species. *Archives of Biochemistry and Biophysics* 369, 1-10 (1999).
46. Jennewein, S. et al. Coexpression in yeast of *Taxus* cytochrome P450 reductase with cytochrome P450 oxygenases involved in Taxol biosynthesis. *Biotechnol Bioeng* 89, 588-98 (2005).
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, 6067-75 (2008).
48. Shigemori, H. & Kobayashi, J. Biological activity and chemistry of taxoids from the
20 Japanese yew, *Taxus cuspidata*. *J Nat Prod* 67, 245-56 (2004).
49. Xu, F., Tao, W., Cheng, L. & Guo, L. Strain improvement and optimization of the media of taxol-producing fungus *Fusarium maire*. *Biochemical Engineering Journal* 31, 67-73 (2006).
50. Hefner, J., Ketchum, R. E. & Croteau, R. Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from *Taxus canadensis* and assessment of the role of this prenyltransferase in cells induced for taxol production. *Arch Biochem Biophys* 360, 62-74 (1998).
51. 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, 9201-4
30 (1996).
52. Kodumal, S. J. et al. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proceedings of the National Academy of Sciences* 101, 15573-15578 (2004).
53. Tyo, K. E. J., Ajikumar, P. K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnology* (2009).
54. Datsenko, K. A. & Wanner, B. L. 6640-6645 (*National Acad Sciences*, 2000).
55. Rost, B., Yachdav, G. & Liu, J. The predictprotein server. *Nucleic acids research* 32, W321 (2004).
56. Shalel-Levanon, S., San, K. Y. & Bennett, G. N. Effect of ArcA and FNR on the
40 expression of genes related to the oxygen regulation and the glycolysis pathway in *Escherichia coli* under microaerobic growth conditions. *Biotechnology and bioengineering* 92 (2005).
57. Heinig, U. & Jennewein, S. Taxol: A complex diterpenoid natural product with an evolutionarily obscure origin. *African Journal of Biotechnology* 8, 1370-1385 (2009).
- 45 58. Walji, A. M. & MacMillan, D. W. C. Strategies to Bypass the Taxol Problem. *Enantioselective Cascade Catalysis, a New Approach for the Efficient Construction of Molecular Complexity*. *Synlett* 18, 1477-1489 (2007).

- 59 -

59. 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, 663-72 (2004).
- 5 60. Williams, D. C. et al. Heterologous expression and characterization of a "Pseudomature" form of taxadiene synthase involved in paclitaxel (Taxol) biosynthesis and evaluation of a potential intermediate and inhibitors of the multistep diterpene cyclization reaction. *Arch Biochem Biophys* 379, 137-46 (2000).
61. Morrone D. et al. Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl Microbiol Biotechnol.* 85(6):1893-906 (2010)
- 10

Having thus described several aspects of at least one embodiment of this invention, it is to be appreciated various alterations, modifications, and improvements will readily occur to those skilled in the art. Such alterations, modifications, and improvements are intended to be part of this disclosure, and are intended to be within the spirit and scope of the invention. Accordingly, the foregoing description and drawings are by way of example only. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

All references disclosed herein are incorporated by reference in their entirety for the specific purpose mentioned herein.

What is claimed is:

25

CLAIMS

1. A method for increasing terpenoid production in a cell that produces one or more terpenoids, comprising

providing an *Escherichia coli* (*E. coli*) that produces isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) through an upstream methylerythritol pathway (MEP) and converts the IPP and DMAPP to a terpenoid through a recombinantly expressed downstream terpenoid synthesis pathway; and

controlling the accumulation of indole in the *E. coli* or in a culture of the *E. coli*, thereby increasing terpenoid production in the *E. coli*, wherein the indole accumulates in the *E. coli* culture at less than 100 mg/L, wherein the step of controlling the accumulation of indole in the *E. coli* or in a culture of the *E. coli* comprises balancing the upstream MEP pathway with the downstream terpenoid synthesis pathway, modifying or regulating an indole pathway, and/or removing the accumulated indole from the cell culture through chemical methods, optionally using absorbents or scavengers.

2. The method of claim 1, wherein the cell overexpresses one or more components of the MEP pathway.

3. The method of claim 2, wherein the cell recombinantly expresses a geranylgeranyl diphosphate synthase (GGPPS) enzyme.

4. The method of any one of claims 2-3, wherein the one or more components of the MEP pathway is selected from the group consisting of *dxs*, *ispC*, *ispD*, *ispE*, *ispF*, *ispG*, *ispH*, *idi*, *ispA* and *ispB*.

5. The method of claim 4, wherein the *E. coli* comprises a heterologous *dxs*-*idi*-*ispDF* operon.

6. The method of claim 1, wherein the upstream MEP pathway is balanced with respect to the downstream terpenoid synthesis pathways by one or more of:

increasing gene copy number for one or more upstream or downstream pathway enzymes,

increasing or decreasing the expression level of the upstream and/or downstream pathway genes, as individual genes or as operons, using promoters with different strengths,

increasing or decreasing the expression level of the upstream and/or downstream pathway genes, as individual genes or as operons, using modifications to ribosomal binding sites,

replacing native genes in the upstream and/or downstream pathways with heterologous genes coding for homologous enzymes,

codon-optimization of one or more heterologous enzymes in the upstream and/or downstream pathways,

amino acid mutations in one or more genes of the upstream and/or downstream pathways, or

modifying the order of the upstream and/or downstream pathway genes in a heterologous operon.

7. The method of any one of claims 1-6, wherein the one or more terpenoids is a monoterpene, a sesquiterpene, a diterpene, a triterpene or a tetraterpene.

8. The method of claim 7, wherein the one or more terpenoids is a taxadiene or any taxol precursor.

9. The method of any one of claims 1-8, further comprising measuring the amount or concentration of indole in the *E. coli* or in a culture of the *E. coli*.

10. The method of claim 9, wherein the method comprises measuring the amount or concentration of indole two or more times.

11. The method of claim 9 or claim 10, wherein the measured amount or concentration of indole is used to guide a process of producing one or more terpenoids and/or to guide strain construction.