Kauniolide synthase is a P450 with unusual hydroxylation and cyclization-elimination activity

Qing Liu1,2, Arman Beyraghdar Kashkooli1, David Manzano3,4, Irini Pateraki2, Lea Richard1, Pim Kolkman1, Maria Fátima Lucas5, Victor Guallar5,6, Ric C.H. de Vos7, Maurice C.R. Franssen8, Alexander van der Krol1 & Harro Bouwmeester1,9

Guaiainolides are an important class of sesquiterpene lactones with unique biological and pharmaceutical properties. They have been postulated to be derived from germacranoieties, but for years no progress has been made in the elucidation of their biosynthesis that requires an unknown cyclization mechanism. Here we demonstrate the isolation and characterization of a cytochrome P450 from feverfew (Tanacetum parthenium), kauniolide synthase. Kauniolide synthase catalyses the formation of the guaiainolide kauniolide from the germacranooid substrate costunolide. Unlike most cytochrome P450s, kauniolide synthase combines stereoselective hydroxylation of costunolide at the C3 position, with water elimination, cyclization and regioselective deprotonation. This unique mechanism of action is supported by silico modelling and docking experiments. The full kauniolide biosynthesis pathway is reconstructed in the heterologous hosts Nicotiana benthamiana and yeast, paving the way for biotechnological production of guaiainolide-type sesquiterpene lactones.
Sesquiterpene lactones are C15 terpenoids and constitute a major class of plant secondary metabolites with diverse chemical structures. They are present in plant species in the Acanthaceae, Apiaceae and Asteraceae with over 4000 different structures reported so far. Sesquiterpene lactones are classified in six bicyclic or tricyclic classes named guaianolides, pseudogaia-

Results

CYP71 activity on costunolide and parthenolide. In feverfew extracts, multiple costunolide- and parthenolide-derived products are detected. Previously, it was shown that feverfew P450s acting on germacrene A, germacrene A acid, costunolide and parthe-

Functional characterization of selected CYP71s in yeast. Candidate P450 cDNAs were expressed in yeast WAT11 strain and microsomes of the transgenic yeast were isolated for in vitro testing of the enzymatic activity towards costunolide or parthe-

that kauniolide biosynthesis does not require a cyclase as sug-
ggested before. The full kauniolide biosynthesis pathway is suc-
cessfully reconstructed in both the heterologous hosts Nicotiana benthamiana and yeast. Modelling is used to support elucidation of the putative enzymatic steps catalysed by KLS. The isolation of the KLS gene opens up opportunities for metabolic engineering of guaianolide sesquiterpene lactones.

that kauniolide biosynthesis does not require a cyclase as sug-
gested before. The full kauniolide biosynthesis pathway is suc-
cessfully reconstructed in both the heterologous hosts Nicotiana benthamiana and yeast. Modelling is used to support elucidation of the putative enzymatic steps catalysed by KLS. The isolation of the KLS gene opens up opportunities for metabolic engineering of guaianolide sesquiterpene lactones.

that kauniolide biosynthesis does not require a cyclase as sug-
gested before. The full kauniolide biosynthesis pathway is suc-
cessfully reconstructed in both the heterologous hosts Nicotiana benthamiana and yeast. Modelling is used to support elucidation of the putative enzymatic steps catalysed by KLS. The isolation of the KLS gene opens up opportunities for metabolic engineering of guaianolide sesquiterpene lactones.

that kauniolide biosynthesis does not require a cyclase as sug-
gested before. The full kauniolide biosynthesis pathway is suc-
cessfully reconstructed in both the heterologous hosts Nicotiana benthamiana and yeast. Modelling is used to support elucidation of the putative enzymatic steps catalysed by KLS. The isolation of the KLS gene opens up opportunities for metabolic engineering of guaianolide sesquiterpene lactones.
can sometimes be improved by conjugation to glutathione or cysteine\(^\text{13}\). This conjugation occurs non-enzymatically, is irreversible and in case of cysteine adds an exact mass of \([M] = 121.01464\) and results in a shorter retention time (RT)\(^\text{19}\). The reaction product of the enzyme encoded by \(\text{Tp8879}\) using costunolide as substrate was therefore incubated with cysteine. Analysis of the cysteine conjugates revealed two new peaks at RT \(= 10.07\) min \((\text{[M+H]}^+ = 370.16827)\) and RT \(= 21.41\) min \((\text{[M+H]}^+ = 352.1577)\). The product eluting at RT \(= 10.07\) was identified as 3\(\beta\)-hydroxycostunolide-cysteine (Fig. 2d, e) according to Liu et al.\(^\text{14}\). More interestingly, the mass, fragmentation spectrum and RT of the product eluting at RT \(= 21.41\) min matched that of a kauniolide-cysteine standard (Fig. 2a, b). Indeed, co-injection of kauniolide-cysteine and the enzyme products incubated with cysteine showed co-elution, showing that the enzyme encoded by \(\text{Tp8879}\) catalyses ring closure, resulting in kauniolide formation from costunolide (Supplementary Fig. 2). Moreover, the peak intensity of kauniolide-cysteine was 6.7 times higher than that of 3\(\beta\)-hydroxyxystunolide-cysteine, indicating that 3\(\beta\)-hydroxyxystunolide is a minor product (Fig. 2d, e). We therefore called \(\text{Tp8879}\) kauniolide synthase (\(\text{TpKLS}\)). When microsomes expressing \(\text{TpKLS}\) were incubated with parthenolide, no peaks were detected, not even after incubation of the enzyme products with cysteine. This suggests that costunolide is the natural substrate for \(\text{TpKLS}\) and that \(\text{TpKLS}\) is the first committed enzyme in the biosynthesis towards guaianolides, revealing that guaianolide-type sesquiterpene lactones (like kauniolide) are derived from germacrane-type sesquiterpene lactones (i.e., costunolide).

The production of 3\(\beta\)-hydroxyxystunolide, though as a minor product, from feeding costunolide to \(\text{TpKLS}\) suggests that this enzyme can perform hydroxylation of costunolide at the C3

---

**Fig. 1** Biosynthetic pathway of costunolide and proposed (bio)synthetic pathways for kauniolide. **a** Biosynthetic pathway of costunolide\(^\text{13}\). **b** Conversion of costunolide to parthenolide by PTS\(^\text{14}\) and chemical biomimetic synthesis of kauniolide from parthenolide\(^\text{15}\). **c** Proposed biosynthesis of kauniolide\(^\text{17}\) and proposed biosynthetic pathway of leucodin in chicory\(^\text{2}\). **d** Biosynthetic pathway of kauniolide, through 3\(\alpha\)-hydroxyxystunolide as intermediate, as identified in the present study. GAS germacrene A synthase, GAO germacrene A oxidase, COS costunolide synthase, PTS parthenolide synthase, p-TSA p-toluene sulphonic acid, CH\(_2\)Cl\(_2\) dichloromethane, RT room temperature, POCl\(_3\) phosphoryl chloride.
Moreover, 3-hydroxyxystunolide has been suggested to be an intermediate in kauniolide biosynthesis\(^\text{17}\). We have previously identified a P450 in feverfew which can hydroxylate costunolide at C3 (Tp\(\beta\)-hydroxylase\(^\text{19}\)), but does not produce kauniolide. To assess if 3\(\beta\)-hydroxyxystunolide is an intermediate in the production of kauniolide we incubated TpKLS with 3\(\beta\)-hydroxycostunolide and found no conversion into kauniolide. To assess if 3\(\beta\)-hydroxycostunolide cannot serve as a substrate for TpKLS (Supplementary Fig. 3). The product of the enzyme encoded by Tp8886 enzyme could not be identified due to its low abundance.

**In silico docking of costunolide in TpKLS.** To get further insight into the mechanism by which kauniolide is formed by TpKLS, we used in silico modelling of TpKLS, followed by costunolide substrate docking experiments. As validation for the modelling approach we used two other P450 enzymes that use costunolide as substrate and for which the site of costunolide oxidation is known, i.e., epoxidation of C4–C5 by TpPTS and hydroxylation of C3 by costunolide 3\(\beta\)-hydroxylase\(^\text{13,14}\). The protein structure models of TpPTS, costunolide 3\(\beta\)-hydroxylase and TpKLS were generated using the crystal structure of *Homo sapiens* P450 2C9\(^\text{20}\) (Protein Data Bank (PDB) code 4GQS), *Oryctolagus cuniculus* P450 2C5\(^\text{21}\) (PDB code 1NR6) and *Homo sapiens* dual specificity protein phosphatase 13 (PDB code 2PQ5_A), respectively (Supplementary Table 2). Substrate docking of costunolide was simulated in PELE (Protein Energy Landscape Exploration) for each of the three modelled enzyme structures\(^\text{22,23}\). Two models were used: (i) a free simulation where the substrate has no constraints, and (ii) simulations with a defined constraint, where the substrate was not allowed to move beyond 15 Å from the haem\(^{\text{Fe(IV)}-\text{O}^2}\) (noted as haem oxyanion). Performing a Monte Carlo (MC) move where the substrate and the protein are first perturbed and then relaxed, the substrate binding in PELE is scored by the enzyme–substrate interaction energy, as described by the OPLS (optimized potential for liquid simulations) force field. PELE was run on each system for maximally 5000 MC steps, after which the distance from the haem oxyanion to the sesquiterpene lactone and docking orientation was scored, thereby performing a population analysis (Fig. 3b, c).

The in silico docking studies of costunolide with TpKLS indicate that the C3 position of costunolide is highly favoured to interact with the haem oxyanion in the TpKLS model structure (Fig. 3b), for both the constrained and unconstrained models (1379 and 1417 occasions, respectively). The calculated distance of costunolide C3 to the haem oxyanion in the model is 3.6 Å which translates into a hydrogen atom abstraction coordinate \(\sim 2.2 \text{ Å}\), sufficient to trigger hydroxylation at costunolide C3. Also, in the in silico docking experiments of costunolide with costunolide-3\(\beta\)-hydroxylase, the PELE simulations indicated a preferred regioselective C3 orientation of costunolide towards the haem oxyanion which is in agreement with its enzymatic activity (Fig. 3b, d; Supplementary Fig. 4). Moreover, the in silico docking of costunolide in TpKLS indicates that a-hydroxylation of C3 is favoured (Fig. 3d; Supplementary Fig. 4), while for costunolide-
3β-hydroxylase, in agreement with its enzymatic function, β-hydroxylation of C3 is preferred (Fig. 3c, e). Presumably, α-hydroxylation of costunolide at C3 is directly followed by protonation, dehydration, cyclization and deprotonation in the enzymatic cavity, resulting in the formation of kauniolide. In contrast, the less preferred β-hydroxylation—which seems to occur, though at a much lower rate—at costunolide C3 apparently cannot be protonated and is removed from the enzymatic cavity as a minor side product.

Although for these two enzymes the predictions from the in silico modelling match with the actual enzyme activity, for TpPTS the modelling suggests epoxidation of C9 while the enzyme epoxidizes C4–C5 of costunolide (Supplementary Fig. 5). The number of predictions for this preferred costunolide C9
orientation towards the haem of \textit{TpTPS} was much lower (~252) than the number of preferred orientations of costunolide C3 towards the haem of TpKLS (1417) and costunolide 3\(\beta\)-hydroxylase (365), suggesting that the model of costunolide docking to TpPTS is less reliable than for the other two enzymes.

3\(\alpha\)-Hydroxycostunolide is intermediate substrate of TpKLS. We therefore, according to in silico docking studies, hypothesized that 3\(\alpha\)-hydroxycostunolide serves as the intermediate substrate in kauniolide biosynthesis. Hence, we chemically synthesized 3\(\alpha\)-hydroxycostunolide (analysed by gas chromatography–mass spectrometry (GC-MS); Supplementary Fig. 6) and fed it to TpKLS. GC-MS analysis of extracts from feeding experiments showed conversion of 3\(\alpha\)-hydroxycostunolide to kauniolide by TpKLS (Fig. 4a, b). No kauniolide was detected in samples where 3\(\alpha\)-hydroxycostunolide was incubated with EV (microsomes of yeast with empty vector) (Fig. 4c). We therefore suggest that conversion of costunolide into kauniolide starts with hydroxylation at the C3 position, predominantly in \(\alpha\)-orientation and with \(\beta\)-orientation as a minor side reaction. This is immediately followed by a ring closure reaction for 3\(\alpha\)-hydroxycostunolide, which results in kauniolide formation, while the other intermediate, 3\(\beta\)-hydroxycostunolide, leaves the enzymatic cavity as a side product. The occasional release of the intermediate 3\(\beta\)-hydroxycostunolide from the enzymatic cavity would explain the presence of 3\(\beta\)-hydroxycostunolide as a side product of TpKLS.

**Fig. 4** The 3\(\alpha\)-hydroxycostunolide serves as the intermediate compound in kauniolide formation. a GC-MS chromatogram of kauniolide standard. b GC-MS chromatogram of product of assay with 3\(\alpha\)-hydroxycostunolide as substrate fed to yeast microsomes expressing Tp8879: production of kauniolide (nominal mass 230). c Negative control; feeding 3\(\alpha\)-hydroxycostunolide to microsomes containing EV (microsomes of yeast with empty vector) does not lead to kauniolide formation.
In planta reconstruction of kauniolide biosynthetic pathway. Previously, we reconstituted the full biosynthesis pathway towards costunolide and parthenolide by transient gene expression in *N. benthamiana*. To obtain production of kauniolide in *N. benthamiana*, the *TpKLS* was cloned into a binary plant expression vector under control of the Rubisco small subunit promoter (RBC) which was used for agrobacterium transformation. All genes of the biosynthetic pathway towards kauniolide (*TpGAS*, *CiGAO*, *CiCOS* and *TpKLS*) were subsequently transiently expressed in *N. benthamiana* leaves by co-agro-infiltration. In addition, overexpression of the *Arabidopsis thaliana* 3-hydroxy-3-methylglutaryl-CoA reductase (*AtHMGR*) was used to boost farnesyl-diphosphate production, which is the precursor of all sesquiterpenoids. *N. benthamiana* leaves were harvested 4 days post agro-infiltration (4 dpi), extracted with methanol/formic acid (0.1%) and extracts were measured by LC-Orbitrap-FTMS using targeted analysis for free and conjugated products (costunolide, kauniolide and costunolide/kauniolide-cysteine, and costunolide/kauniolide-glutathione conjugates). No free kauniolide was detected, but both costunolide-cysteine and a mass presumed to be kauniolide-cys ([M+H]+ = 352.1577) were detected in the leaf extracts (Fig. 5b), indicating that kauniolide produced in *N. benthamiana* leaves is conjugated to cysteine.

To confirm the identity of the compound with mass 352.1577, kauniolide-cysteine was (non-enzymatically) conjugated in vitro to cysteine to produce kauniolide-cys. Indeed, retention time and molecular mass ([M+H]+ = 352.1577) of kauniolide-cys were identical to the product obtained in planta upon expression of the kauniolide biosynthetic pathway (Fig. 5a–c). Neither kauniolide-cysteine nor kauniolide-glutathione were detected in control samples (costunolide pathway co-expressed with empty vector).

Reconstruction of kauniolide biosynthetic pathway in yeast. Recently, reconstruction of the biosynthetic pathways of the sesquiterpene lactones, costunolide and artemisinin, in yeast was demonstrated. To achieve the same for kauniolide, we cloned the four biosynthetic genes (*TpGAS*, *TpGAO*, *TpCOS* and *TpKLS*) into two dual yeast expression vectors and transformed yeast cells by plasmid transformation under the control of a galactose-
inducible promoter. Yeast cultures were induced by galactose and after 72 h cells plus medium were extracted with ethyl acetate. The extracts were concentrated, dried over a Na₂SO₄ column and the dehydrated sample was injected into GC-MS for analysis. Yeast cells expressing TpGAS, TpGAO and TpCOS produced costunolide, confirming previous results²⁰. The GC-MS chromatogram from the yeast cells expressing TpGAS, TpGAO, TpCOS and TpKLS showed a new peak at 19.10 min, with the same retention time and mass spectrum as the kauniolide standard (Fig. 6a, b). This indicates a successful production of free kauniolide in yeast cells. No conjugated kauniolide/costunolide was detected by GC-MS.

**Discussion**

Several sesquiterpene lactones have been identified in feverfew⁹ and several biosynthetic pathway enzymes and their corresponding genes have been characterized such as germacrene A synthase, costunolide synthase, parthenolide synthase and costunolide 3β-hydroxylase¹³,¹⁹. Nevertheless, the biosynthesis of several feverfew sesquiterpene lactones (e.g., santamarine, reynosin and artecanin) is still a mystery. Costunolide may play a central role in sesquiterpene lactone biosynthesis since several P450s have been shown to use this compound as substrate for oxidation¹⁸,¹⁹. Furthermore, other researchers have proposed that costunolide might be the branching point for biosynthesis of other classes of sesquiterpene lactones such as guaianolides and eudesmanolides²,¹⁸,²⁶.

Here, we demonstrate the catalytic activity of a CYP71 (TpKLS) responsible for the production of a guaianolide-type sesquiterpene lactone, kauniolide, from a germacranolide-type sesquiterpene lactone, costunolide. Guaianolide biosynthesis has been suggested to start with C3 hydroxylation of costunolide, followed by cyclase activity by a separate enzyme¹⁷. The cyclization would involve the following steps: (i) protonation of the newly introduced hydroxyl group; (ii) loss of water, producing an allylic carbocation C3⁺−C4=C5/C3=C4−C5⁺; (iii) nucleophilic attack of the C1–C10 double bond to the C5 carbocation, forming the annellated 5–7 bicyclic system and a tertiary carbocation at C10; and (iv) regioselective deprotonation at C1 giving a C1–C10 double bond. It is likely that at least some of these steps occur in a concerted fashion¹⁷. Here we show that all these reaction steps take place within a single P450. The production of 3β-hydroxycostunolide as a minor side product of TpKLS enzyme activity confirms that the enzyme hydroxylates costunolide at the
have different functions for different substrates in literature. A conserved active site threonine in Costunolide was indeed docked in a rather peculiar way in the enzyme cavity which allows such a reaction to occur. C3-hydroxyl group results in a suitable orientation of the substrate in the active site of TpKLS for further postulated protonation. In contrast, the correct orientation in the active site of TpKLS for further conversion to kauniolide. Indeed, protein modelling suggests that C3-hydroxylase activity with a preference for C3 position. It has been shown that a different CYP71 from feverfew (costunolide 3β-hydroxylase), stereoselectively adds a hydroxyl group in β-orientation to the C3 position of costunolide but does not convert it to kauniolide. Using yeast microsome feeding assays, we show that 3β-hydroxycostunolide cannot be converted to kauniolide by TpKLS. Indeed, protein modelling suggests that C3-β-hydroxylation of costunolide does not allow the correct orientation in the active site of TpKLS for further postulated protonation (Fig. 3d). In contrast, α-orientation of the C3-hydroxyl group results in a suitable orientation of the substrate in the enzyme cavity which allows such a reaction to occur. Costunolide was indeed docked in a rather peculiar way in TpKLS. A conserved active site threonine in α-helix I appears to have different functions for different substrates in literature and may have an additional role in this system. It may well be that the conserved threonine of α-helix I protonates costunolide after the hydroxylation step. Protonation by this threonine does not occur in the other discussed enzymes. After ring closure the conformation of the product is very different from the substrate, enabling it to move out of the active site, after which the threonine can be protonated again. Perhaps the threonine oxyanion is involved in the regioselective deprotonation of the intermediate carbocation, leading to kauniolide. We therefore suggest that TpKLS has costunolide C3-hydroxylase activity with a preference for α- over the β-orientation, while the 3α-hydroxycostunolide is subsequently converted to kauniolide by the same enzyme. We note that modelling results are based on a heterologous P450 template model. Future acquisition of the TpKLS crystal structures may substantially improve the in silico simulation.

Other enzymes in the feverfew sesquiterpene lactone biosynthesis pathway also display a number of sequential activities, all initiated by oxidation. For instance, the P450 TpGAO converts germacrene A into germacrene A alcohol, germacrene A aldehyde and finally germacrene A acid. A similar type of single-enzyme multiple reactions also occurs in the closely related Artemisia annua, in which amorphadiene is converted by CYP71AV1 into amorphadiene alcohol, amorphadiene aldehyde and finally artemisinic acid. Also, the hydroxylation of germacrene A acid by COS is followed by spontaneous lactone formation. However, in all these examples the sequential enzyme activities target the same position in the substrate. What makes TpKLS unique is that this enzyme not only carries out the conventional P450 catalysed hydroxylation, but also the protonation and dehydration, cyclization and subsequent deprotonation (Fig. 7).

Phylogenetic analysis of the cytochrome P450s from feverfew and some other Asteraceae plants shows that TpKLS clusters more closely with COS proteins than with Pts or costunolide 3β-hydroxylase and the GAOs (Supplementary Fig. 8). Analysis of more than 2.3 million expressed sequenced tags (ESTs) from different Asteraceae species (http://compgenomics.ucdavis.edu) identified contigs from these species having significant homology with TpKLS (but still with less than 65% homology) (Supplementary Fig. 7, Supplementary Table 1). Indeed, some of these Asteraceae species are known to produce guaianolide-type sesquiterpene lactones such as Cichorium spp. (lactucin, lactuopicrin and 11β,13-dihydrolactucin), Cynara carduncula (aguerin B, cynaropicrin). This suggests that a specific and quite unique protein with KLS activity has evolved in feverfew out of COS, but that putative KLSs in other Asteraceae may have evolved independently. This is in stark contrast with the strong sequence similarity of COS genes in Asteraceae species, resulting in a strong clustering in phylogenetic analysis (Supplementary Fig. 8).

Feverfew indeed contains guaianolide sesquiterpene lactones, artemisinin and tanaparthin-β-peroxide. Artescanin has anti-proliferative properties on leukaemia cancer cells. Artescanin biosynthesis requires multiple oxidation steps of kauniolide, likely catalysed by P450s, and these may be identified by feverfew P450
expression profiling and comparison with the accumulation pattern of artecanin (Supplementary Fig. 9). Artecanin shows a stable accumulation level during development of feverfew flowers, but suddenly increases late during development of the flowers. The increased production of artecanin in this late stage of development may be due to reduced expression of PTS14 and therefore reduced competition with PTS for the costunolide substrate. Our elucidation of the biosynthesis of kauniolide, the most basic guaianolide, will provide opportunities for bioengineering of valuable guaianolides like artecanin.

Methods
RNA-sequencing and expression analysis. The T. parthenium EST library used in the present study was made from flower messenger RNA as previously reported. Later Illumina reads of ovary developmental stages were mapped against this library and ExPress (https://pachterlab.github.io/ExPress/) was accommodated for expression level estimation in different developmental stages. Deseq2 was used to check differential expression levels for three biological replicates. A negative binomial distribution test with the false discovery rate threshold of 0.05 was used for Deseq2. Mean expression values of CYP71D1 and CYP71D5 genes which were previously cloned into the pESC-Trp yeast expression vector (Agilent Technologies). CICOS was subcloned from pYED601 using digestion by NotI and Piel and inserted into pESC-Ura. Subsequently, TpKLS was cloned by addition of BamHI and Kpnl sites into CicOS-pESC-Ura. WAT11 yeast was transformed with CIGA0+TqGAS-pESC-Trp and CicOS-pESC-Ura and colonies were selected on SD minimal medium (containing 0.67% Digo yeast nitrogen base without amino acids, 2% glucose and 2% purified agar) which was supplemented by amino acids but omitting t-tryptophan and uracil for auxotrophic selection of transformants. The combination of CIGA0+TqGAS-pESC-Trp and CicOS-pESC-Ura with empty multiple cloning site 2 (MSC2) was used as control represented by GAS+GAS+COS-EV (empty vector). Then, 2000 ml of each medium (medium and cells) was extracted with twice of 500 ml of ethyl acetate. Extracts were concentrated to 100 µl using a rotary evaporator and a N2 flow and used for GC-MS analysis.

Yeast microsome isolation and in vitro microsome assay. Microsome isolation of transformed yeasts was done according to ref.39 with some modifications. Transformed yeast single colonies were pre-cultured in 50 ml SGI medium for 3 days at 30 °C and shaking at 300 rpm. Then, a 250 ml YPL medium containing 25 g sugar was added for induction. Expression was kept for 24 h under the same conditions. Subsequently, the cells were harvested and chilled on ice for 20 min, followed by centrifugation at 4900 × g for 10 min. Cell pellets were resuspended in 100 ml extraction buffer (50 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.6 M sorbitol and 10 mM β-mercaptoethanol) and kept for 10 min at room temperature. Again, the cells were centrifuged (4900 × g for 10 min) and cells were washed three times (3 ml) with extraction buffer, but omitting β-mercaptoethanol. The centrifuge tube was rinsed with 2 ml of the same buffer and the mixture was transferred to a 50 ml Falcon tube. Approximately 25 ml of glass beads (450 pm) were added to the tubes and cells were lysed by vigorous shaking in a cold room for 10 min. The cell lysate was transferred to a 25 ml centrifuge tube and centrifuged at 10,500 × g for 10 min. Later, the supernatant was again centrifuged at 195,000 × g for 2 h. The pellets (microsomal fractions) were resuspended in a 4 ml solution of ice-chilled 50 mM Tris-HCl (pH 7.5) containing 1 mM EDTA and 20% (v/v) glycerol using a glass Tenbroek homogenizer. Aliquots of these microsomal fractions were kept at −80 °C. Yeast microsome in vitro assays were done in a mixture of 28.8 µl isolated microsomal fractions, 4 µl substrate (10 mM in DMSO), 40 µl NADPH (10 mM in 100 mM potassium phosphate buffer), 8 µl potassium phosphate buffer (1 M, pH 7.3), and 90.4 µl of water which was incubated for 2.5 h at 25 °C with shaking (200 rpm). The obtained mixture was centrifuged again at 195,000 × g for 10 min and the supernatant was passed through a 0.22 µm filter before injection into LC-MS. Yeast microsome in vitro assays with 3a-hydroxycostusolone as substrate were performed as described above, but after 2.5 h of incubation at 25 °C and shaking at 200 rpm, samples were extracted 3 times with 2 ml ethyl acetate. The ethyl acetate phase was passed through a Pasteur pipet, plugged with glass wool and filled with Na2SO4 to dehydrate the extract. The extract was then concentrated and injected into GC-MS.

Plasmid construction for expression in N. benthamiana. Heterologous transient expression of CIGA0 was done according to ref.4. Briefly, candidate P450 CDs were cloned into Impactvector 1.1 under the control of Rubisco (rbc) promoter. Other pathway genes (GAO, COS, PTS) were also cloned into the same expression vector. TpGAS was also cloned into Impactvector 1.5 to fuse it with the RBC promoter and the Cos/x mitochondrial targeting sequence as it was shown before that mitochondrial targeting of sesquiterpene synthases results in improved sesquiterpene production. Later, the CDS was transformed into the pBinPlus binary vector by an LR reaction (Gateway-Leaf Clonase TM II) to put the CDS between the right and left borders of transfer DNA (T-DNA) for plant transformation. All these constructs were finally transformed into an AGL-0 Agrobacterium strain. An LR reaction (Gateway-Leaf Clonase TM II) was carried out to clone each gene into the pBinPlus binary vector between the right and left borders of the T-DNA for plant transformation.

Transient expression in N. benthamiana. Transient heterologous expression of biosynthetic pathway genes was done by Agrobacterium-mediated transformation (agro-infiltration) of N. benthamiana plants. Transformed Agrobacteria were grown at 28 °C at 250 rpm for 48 h in LB media with proper antibiotics. Then, Agrobacterium cells were harvested at 4000 × g for 20 min at room temperature. Cells were then resuspended in agro-infiltration buffer (10 mM MES, 10 mM CaCl2 and 100 µM acetosyringone). The final cell density (OD0) in the 600 nm wavelength absorbance was set to 0.5. Gene dosage for control treatments were adjusted by addition of representative number of empty vector(s). Cultures were mixed on a roller-mixer for 2.5 h (50 rpm). The 4-week-old N.benthamiana plants were agroinfiltrated with a needleless 1 ml syringe injection to the abaxial side of the leaf. Transformed plants were kept for another four and half days and then harvested for metabolites analysis.

LC-MS analysis of yeast microsome assay and leaf extracts. In order to analyse the products formed in the microsome assays (except when 3a-hydroxycostusolone was used as substrate; see above) we used a LC-LTQ-Orbitrap FTMS system.
et al.13. In brief, cysteine (150 mM) in 7 µL potassium phosphate buffer (100 mM; Cysteine conjugation as described above. Sartorius, Germany) and were kept in injection vials for LC-Orbitrap-MS analysis of supernatant was passed through a 0.22 µm inorganic membrane and was concentrated and injected into GC-MS. For this quantum mechanical model of the haem, Compound I was created (Fe$^{2+}$ was replaced by Fe$^{3+}$ and O$_2$ was bonded to Fe$^{3+}$) in Maestro. Costunolide was found in this modified Pwothenolide synthase before QSite45,46 kept the substrate and residues inside the cysteine frozen during the calculations. After inserting the haem from Parthenolide synthase in kaunolide synthase and costunolide 3β-Hydroxylase, the input for a molecular dynamics simulation using Desmond 49-50. Explicit water molecules with 150 mM of sodium chloride were simulated in a box of 10 Å surrounding the protein. A NPT system was maintained, providing a constant number of particles, pressure, and temperature. Standard pressure (−1 bar) and a temperature of 300 K were applied. A 100 N m$^{-1}$ constraint was set to Compound I and a 50 N m$^{-1}$ constraint was set on the protein backbone. The simulation duration was 5 ns maximally. System volume and energy were investigated afterwards using the Quality Analysis tool of Desmond48,50. At a potential energy minimum as close to the end of the simulation as possible, one simulation frame was selected. Compound I from the first simulation frame was inserted in the selected final simulation frame. For kaunolide synthase and parthenolide synthase, the final docking frame was used as input for a subsequent simulation with PELE22,23, aiming at refining the binding mode of costunolide in the CYP450s. The costunolide 3β-hydroxylase with docked costunolide was used as input directly. A force constant of 400 N m$^{-1}$ was set on the bond of 2.45 Å between the iron-chelating cysteine and iron cation.

**Enzyme-substrate interaction modelling.** In preparation of binding simulations, costunolide was docked in KLS, costunolide 3β-hydroxylase and PTS. Hereto, Glide was used to produce a receptor grid of the active site, using as centre the position of the ligand in the homology template. Polar side chains were allowed to rotate in the grid. Glide was used to dock costunolide in extra precision mode, using the partial charges from the quantum mechanical description of costunolide. A total of 5000 Initial docking poses were produced. The 100 best poses were then minimized. One final docking pose was returned by Glide. This system was used as input for a molecular dynamics simulation using Desmond 49-50. Explicit water molecules with 150 mM of sodium chloride were simulated in a box of 10 Å surrounding the protein. A NPT system was maintained, providing a constant number of particles, pressure, and temperature. Standard pressure (−1 bar) and a temperature of 300 K were applied. A 100 N m$^{-1}$ constraint was set to Compound I and a 50 N m$^{-1}$ constraint was set on the protein backbone. The simulation duration was 5 ns maximally. System volume and energy were investigated afterwards using the Quality Analysis tool of Desmond48,50. At a potential energy minimum as close to the end of the simulation as possible, one simulation frame was selected. Compound I from the first simulation frame was inserted in the selected final simulation frame. For kaunolide synthase and parthenolide synthase, the final docking frame was used as input for a subsequent simulation with PELE22,23, aiming at refining the binding mode of costunolide in the CYP450s. The costunolide 3β-hydroxylase with docked costunolide was used as input directly. A force constant of 400 N m$^{-1}$ was set on the bond of 2.45 Å between the iron-chelating cysteine sulphur and iron cation.

PELE generates a MC chain where each step uses a ligand and backbone perturbation followed by a relaxation step involving a side chain prediction and a full system minimization. Each ligand perturbation involved 50 different independent trials with: (i) a rotational change of 0.02 radian (70% probability) or 0.25 radian (30% probability); (ii) a translational displacement of 1.0 Å (70% probability) or 0.5 Å (30% probability). The backbone perturbation followed an anisotropic network model approach using the lowest 6 eigenvectors. All side chains within 6 Å from the substrate were included in the relaxation step. PELE ran on the system for maximally 5000 steps. Two different set of simulations were run: one without a constraint on the substrate and one with a restricted search region where the substrate was not allowed to move beyond 15 Å from Compound I. To characterize the enzyme-substrate interactions, PELE used the 2005 OPLS-AA force field and a surface-SGB implicit model solvent.

**Data availability** The CDNA sequence of the candidate genes has been submitted in GenBank under the primary accession numbers MF197558 (TpP8879 (kaunolide synthase)) and MF197559 (β-hydroxylase). The CDNA sequences were also deposited in David Nelson’s cytchrome P450 database (http://drnelson.uthsc.edu/ctyochromP450.html), TpP8879 and TpP8868 were assigned the names CYP17B6Z and CYP17D5, respectively. The RNA-seq dataset (www.terpmed.eu) and computer codes used for this study are available from the corresponding author on reasonable request.
Opportunity, Challenge and Scope of Natural Products in

References

1. Chatuverdi, D. In Opportunity, Challenge and Scope of Natural Products in Medicinal Chemistry (eds Tiwari V. & Mishra B.) 313–334 (Research Signpost, Trichur, 2013).

2. de Kraker, J.-W., Franssen, M. C. R., Joerink, M., de Groot, A. & Bouwmeester, H. J. Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chickory. Plant Physiol. 129, 257–268 (2002).

3. Lepoittevin, J.-P., Berl, V. & Giménez-Arnau, E. α-methylene-γ -butyrolactones: versatile side bioactive natural products. Chem. Rec. 9, 258–270 (2009).

4. Simonsen, H. T., Weitzel, C. & Christensen, S. B. in Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes (eds Kishan Gopal Ramawat & Jean-Michel Mérillon) 3069–3098 (Springer, Berlin Heidelberg, 2013).

5. Piet, D. P., Schrijvers, R., Franssen, M. C. R. & de Groot, A. Biotransformation of allylically bound sesquiterpene lactones from Cichorium intybus. Tetrahedron 51, 7325–7334 (1995).

6. Lepoittevin, J.-P., Berl, V. & Giménez-Arnau, E. α-methylene-γ -butyrolactones: versatile side bioactive natural products. Chem. Rec. 9, 258–270 (2009).

7. Cantrell, C. L. et al. Antimycobacterial activities of dehydrocostus lactone and 11β-hydroxydehydrocostus lactone, inhibit prostate cancer cell migration. J. Med. Food 15, 24–32 (2012).

8. Boeswester, H. J. et al. Isolation and characterization of two germacrene A synthase cDNA clones from chicory. Plant Physiol. 129, 134–144 (2002).

9. Majdi, M. et al. Biosynthesis and localization of parthenolide in glandular trichomes of feverfew (Tanacetum parthenium L. Schultz Bip. and its combinatorial effect with benzindazole. Phytomedicine 20, 59–66 (2012).

10. Kim, J. et al. The hexane extract of Saussurea lappa and its active principle, dehydrocostus lactone, inhibit prostate cancer cell migration. J. Med. Food 15, 24–32 (2012).

11. Boeswester, H. J. et al. Isolation and characterization of two germacrene A synthase cDNA clones from chicory. Plant Physiol. 129, 134–144 (2002).

12. Majdi, M. et al. Biosynthesis and localization of parthenolide in glandular trichomes of feverfew (Tanacetum parthenium L. Schultz Bip.). Phytochemistry 72, 1739–1750 (2011).

13. Nguyen, D. T. et al. Biochemical conservation and evolution of germacrene A oxidase in Asteraceae. J. Biol. Chem. 285, 16588–16598 (2010).

14. Liu, Q. et al. Reconstitution of the costunolide biosynthetic pathway in yeast and Nicotiana benthamiana. PLOS ONE 6, e23255 (2011).

15. Liu, Q. et al. Elucidation and in planta reconstitution of the parthenolide biosynthetic pathway. Metab. Eng. 23, 145–153 (2014).

16. Stai, Y.-D. J. et al. Biomimetic synthesis of asaglin from parthenolide. J. Org. Chem. 77, 7103–7107 (2012).

17. Piet, D. P., Schrijvers, R., Franssen, M. C. R. & de Groot, A. Biotransformation of germacrene epoxides by Cichorium intybus. Tetrahedron 65, 6303–6314 (1999).

18. Piet, D. P., Franssen, M. C. R. & de Groot, A. Biotransformation of allylically activated (E)-E-cycloeca-1, 6-dienols by Cichorium intybus. Tetrahedron 52, 11273–11280 (1996).

19. Ikezawa, N. et al. Lettuce costunolide synthase (CYP71β1L2) and its homolog (CYP71β1L1) from sunflower catalyze distinct regio- and stereoselective hydroxylation in sesquiterpene lactone metabolism. J. Biol. Chem. 286, 21601–21611 (2011).

20. Hullaert, J., Laplace, D. R. & Wisse, J. M. A three-step synthesis of the guaianaloid ring system. Eur. J. Org. Chem. 2014, 3997–3100 (2014).

21. Reynald, R. L., Sensen, S., Stout, C. D. & Johnson, E. F. Structural characterization of human cytochrome P450 2C19 active site differences between P450s 2C8, 2C9, and 2C19. J. Biol. Chem. 287, 44581–44591 (2012).

22. Wester, M. R. et al. Structure of mammalian cytochrome P450 2C5 complexed with dicyclofenac at 2.1 Å resolution: evidence for an induced biomolecular systems at your

Acknowledgements

We would like to thank Dr. Justin T. Fischekidd and Professor Yue Chen for providing sesquiterpene lactone standards. Q.L was funded as a part of Termed (Plant Terpenoids for Human Health: a chemical and genonomic approach to identify and produce bioactive compounds) (Project ID 227448) and A.B.K. was funded by Iranian Ministry of Science Research and Technology.
Author contributions
A.B.K., Q.L., A.v.d.K. and H.B. designed the research experiments. Q.L., A.B.K., L.R., D. M. and I.P. performed gene cloning. D.M. did the gene expression analyses by qPCR. Q. L. and L.R. did transient expression in *N. benthamiana*, yeast microsome assays and LC-Oribitrap-FTMS. A.B.K. performed yeast transformation, chemical synthesis, GC-MS analysis, analysed assembled RNA-seq expression data, phylogenetic and transcript analysis. A.B.K., A.v.d.K., V.G. and M.F.L. designed the modelling experiments and P.K. conducted protein docking. Q.L., A.B.K., D.M., I.P., L.R., P.K., M.F.L., R.d.V., M.C.R.F., A.v.d.K. and H.B. interpreted the data and reviewed the manuscript. A.B.K. and Q.L. wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06565-8.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

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

© The Author(s) 2018