nature catalysis

ARTICLES

https://doi.org/10.1038/s41929-020-0471-8

A polyketoacyl-CoA thiolase-dependent pathway for the synthesis of polyketide backbones

Zaigao Tan[1,3], James M. Clomburg[1,2], Seokjung Cheong[1], Shuai Qian[1] and Ramon Gonzalez[1,2] ✉

Polyketides found in nature originate from backbones synthesized through iterative decarboxylative Claisen condensations catalysed by polyketide synthases (PKSs). However, PKSs suffer from complicated architecture, energy inefficiencies, complex regulation, and competition with essential metabolic pathways for extender unit malonyl-CoA, all combining to limit the flux of polyketide biosynthesis. Here we show that certain thiolases, which we term polyketoacyl-CoA thiolases (PKTs), catalyse polyketide backbone formation via iterative non-decarboxylative Claisen condensations, hence offering a synthetic and efficient alternative to PKSs. We show that PKTs can synthesize polyketide backbones for representative lactone, alkylresorcinolic acid, alkylresorcinol, hydroxybenzoic acid and alkylphenol polyketide families, and elucidate the basic catalytic mechanism and structural features enabling this previously unknown activity. PKT-catalysed reactions offer a route to polyketide formation that leverages the simple architecture of thiolases to achieve higher ATP efficiencies and reduced competition with essential metabolic pathways, all of which circumvent intrinsic inefficiencies of PKSs for polyketide product synthesis.

Polyketides represent a large class of secondary metabolites that are produced by bacteria, fungi, plants and animals, and feature important biological and pharmaceutical activities. Over 10,000 structures of polyketides are known so far, of which more than twenty are commercial drugs (for example, erythromycin, tetracycline and lovastatin) with annual sales reaching over US$20 billion dollars[1,2]. At its core, polyketide biosynthesis relies on iterative decarboxylative Claisen condensations that are catalysed by polyketide synthases (PKSs), with possible additional modifications via β-reduction, dehydration and/or enoyl reduction reactions that are catalysed by certain modifying domains of some PKSs. From this relatively small number of chemistries, an incredible amount of structurally diverse polyketides can be formed using various starter and extender compounds and combinatorial arrangement of reaction modules[1]. So far, three distinct types of PKSs have been discovered and classified (type I, II and III PKSs)[3]. Extensive efforts have been devoted to either mining new PKS members from natural hosts or heterologous expression and engineering in industrial organisms such as *Escherichia coli* (*E. coli*), *Saccharomyces cerevisiae* or *Streptomyces coelicolor*[3–8]. However, PKSs suffer from inherent slow kinetics[9] and the high amounts of ATP required to generate the extender unit malonyl-CoA (acetyl-CoA + ATP + HCO_3^- → malonyl-CoA)[10]. The conversion of acetyl-CoA to malonyl-CoA via acetyl-CoA carboxylase (ACC) is also a rate-limiting step that is regulated by many allosteric inhibitors[11]. Although malonyl-CoA (C3) is used as the extender unit, only an acetyl-CoA (C2) unit is incorporated into the growing polyketide backbone, with the remaining C1 unit lost in the form of CO_2 (Fig. 1a). Furthermore, competition with essential metabolic functions (such as the biosynthesis of fatty acids and phospholipids) for the common malonyl-CoA precursor[12] can also limit the flux of polyketide biosynthesis pathways. Finally, many PKSs (especially PKS I) are large multifunctional proteins that have many modules containing multiple domains (with a total size up to the megadalton scale)[13], making the expression and/or engineering of such huge architectures extremely challenging in biomanufacturing.

Here we show that enzymes other than PKSs have the potential to catalyse the iterative Claisen condensation reactions that lead to polyketide synthesis. Specifically, we show that some thiolases (which we term polyketoacyl-CoA thiolases, PKTs) can synthesize polyketide backbones by directly using acetyl-CoA as the extender unit, bypassing the ACC node and avoiding both ATP requirements and carbon losses in the form of CO_2. In our prototype examples, the proposed platform was capable of synthesizing representative polyketides such as triacetic acid lactone, orsellinic acid (ORA) and orcinol, all of which were previously thought to be exclusively produced by PKSs. In addition, the designed platform was found to notably simplify the synthetic architecture for polyketide backbones, as demonstrated in the synthesis of 6-methylsalicylic acid (6-MSA) and *m*-cresol. We further elucidated the basic catalytic mechanism of PKTs and revealed the potential structural features in enabling this previously unknown activity.

Results

Evolutionary analysis of the thiolase superfamily. In PKSs, the core β-ketosynthase (KS) that catalyses iterative decarboxylative Claisen condensations for the synthesis of polyketide backbones (Fig. 1a) belongs to the thiolase superfamily, a group of enzymes that catalyse Claisen condensation reactions and which includes ketoacyl-CoA thiolase (thiolase I), acetoacetyl-CoA thiolase (thiolase II), HMG-CoA synthase (HMGS), β-ketoacyl-ACP synthase I, II and III (KAS I, II and III), and β-ketoacyl-CoA synthase (KCS) (Fig. 1b; see Extended Data Fig. 1a for corresponding reactions)[14]. Phylogenetically, members of this superfamily with the same functionality cluster as a clade, with evolved enzymes diverging from a thiolase-like ancestor similar to archaeal thiolases (Fig. 1b). Furthermore, thiolase I and II cluster closely with archaeal thiolases, suggesting an earlier emergence than other enzymes (Fig. 1b). In contrast to PKS and KAS/KCS, which catalyse irreversible decarboxylative Claisen condensation reactions and use malonyl-ACP/CoA as the extender unit, thiolase

[1]Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, USA. [2]Department of Chemical and Biomedical Engineering, University of South Florida, Tampa, FL, USA. [3]Present address: School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ✉e-mail: ramongonzale@usf.edu

I/II and HMGS catalyse reversible, non-decarboxylative Claisen condensation reactions and use acetyl-CoA as the extender unit (Extended Data Fig. 1a). Moreover, enzymes of the thiolase superfamily exhibit a common fold with a five-layered α–β–α–β–α core structure (Extended Data Fig. 1b)[15,16], and the location of the extensive dimerization interface, placement of the active site and use of the same catalytic cysteine (Cys) residue for covalent attachment of substrates are conserved across members of the superfamily[14].

Considering their divergent evolution and structural similarities, we hypothesized that enzymes other than PKSs in the thiolase superfamily have the potential to catalyse iterative Claisen condensation reactions leading to the synthesis of polyketide backbones (Fig. 1c). Specifically, the deep evolutionary root of thiolase I/II from the archaeal enzyme that led to all thiolase superfamily enzymes raises the possibility that the intrinsic structural characteristics of thiolases could enable polyketide biochemistry. In addition to an improved structure–function understanding of thiolase superfamily enzymes, the use of thiolase I/II (or HMGS) is of interest as they catalyse non-decarboxylative Claisen condensation reactions, which has the potential to avoid the inefficiencies caused by using malonyl-CoA as an extender unit and improve the overall energetics of polyketide biosynthesis (Fig. 1a,c).

Mining thiolases for triacetic acid lactone production. Products naturally synthesized by type III PKSs (PKS IIIs) were initially selected as proxies to confirm whether a pathway based on thiolase I/II (or HMGS) can be employed to synthesize polyketide backbones (Fig. 1c). Specifically, we aimed to establish a proof-of-concept through the synthesis of triacetic acid lactone (TAL), a lactone naturally synthesized by the type III PKS 2-pyrone synthase (2-PS) (Fig. 2a)[17]. Our proposed platform would rely on the condensation of acetoacetyl-CoA with acetyl-CoA followed by cyclization (Fig. 2a), a reaction which is thermodynamically feasible under standard conditions ($\Delta_r G'° = -3.5$ kcal mol^{-1}) (Supplementary Table 1).

We tested the ability of eight representative non-decarboxylative thiolases to condense acetoacetyl-CoA (1 mM) and acetyl-CoA (1 mM) in vitro, including HMGS[18] from *Staphylococcus aureus;* thiolase I/II candidates AtoB[19], EcFadA[20] and PaaJ[21] from *E. coli*; PcaF[22] and FadAx[23] from *Pseudomonas putida*; DcaF[24] from *Acinetobacter* sp. ADP1; and BktB[25] from *Ralstonia eutropha* (6 μM of each, Extended Data Fig. 2). Among them, varying amounts of TAL were observed with purified AtoB (2.9 mg l^{-1}), FadAx (1.9 mg l^{-1}) and BktB (7.7 mg l^{-1}) (Fig. 2b; see mass spectra in Fig. 2c), all of which were higher than *Gerbera hybrida* 2-PS (1 mg l^{-1}, condensing 1 mM acetoacetyl-CoA and 1 mM malonyl-CoA). These results suggest that AtoB, FadAx and BktB might have better catalytic properties than 2-PS/PKS for TAL production from acetoacetyl-CoA. No TAL was observed in reactions with HMGS, EcFadA, PaaJ, DcaF and PcaF (Fig. 2b). Notably, all three TAL-producing thiolases—

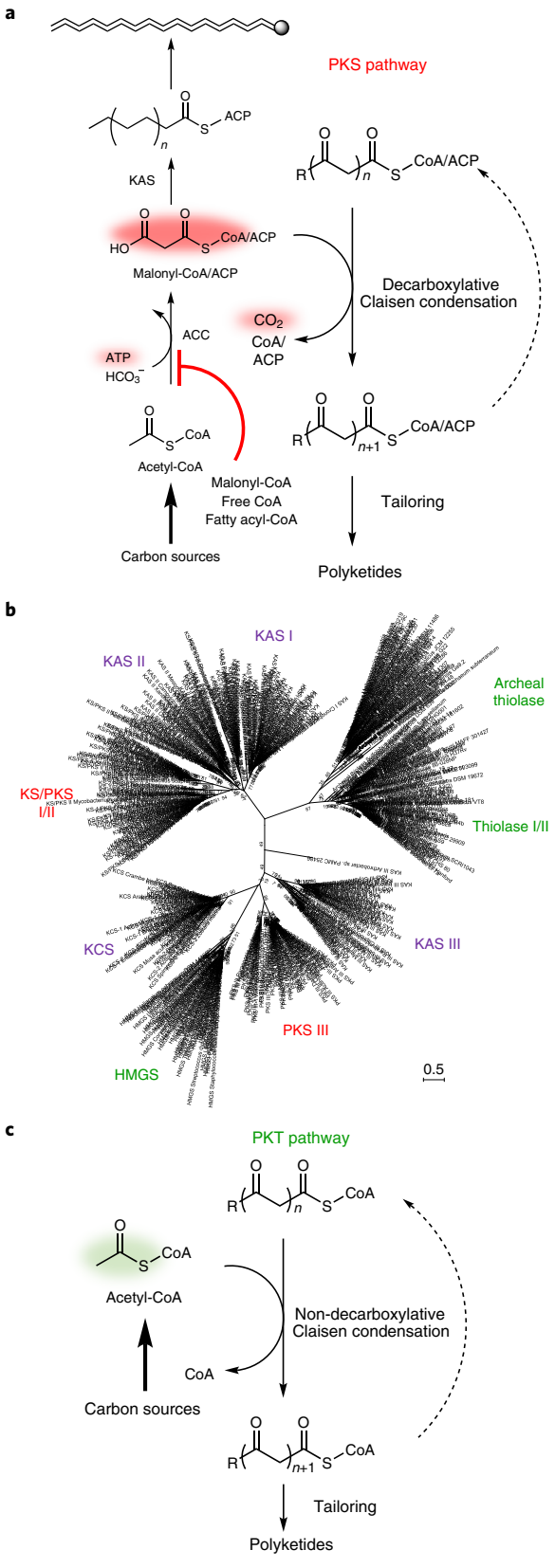


Fig. 1 | Synthesis of polyketide backbones through polyketide synthases and polyketoacyl-CoA thiolases. a, All pathways for the syntheses of the polyketide backbones known so far are mediated by PKSs, which utilize a decarboxylative Claisen condensation mechanism and C3 malonyl-CoA as the extender unit. These pathways consume ATP during generation of the extender unit, release CO_2 during the condensation reaction and compete for malonyl-CoA with other growth-sustaining pathways (for example, biosynthesis of fatty acyl-ACPs and phospholipids). Malonyl-CoA formation is a rate-limiting step that is regulated by many allosteric inhibitors, such as malonyl-CoA, free CoA and long-chain fatty acyl-CoA. **b**, A molecular phylogenetic analysis of the thiolase superfamily, which includes PKS III, β-KS of PKS I/II, thiolases, HMGS, KAS and KCS. The evolutionary history was inferred by using the maximum likelihood method based on the Jones–Taylor–Thornton matrix-based model. The scale bar indicates the average number of amino acid substitutions per site. **c**, The proposed PKT-dependent pathway for the synthesis of polyketide backbones. This pathway utilizes a non-decarboxylative Claisen condensation mechanism and C2 acetyl-CoA as the extender unit. Once formed, the polyketide backbone can be tailored to generate various polyketides. ACP, acyl carrier protein.

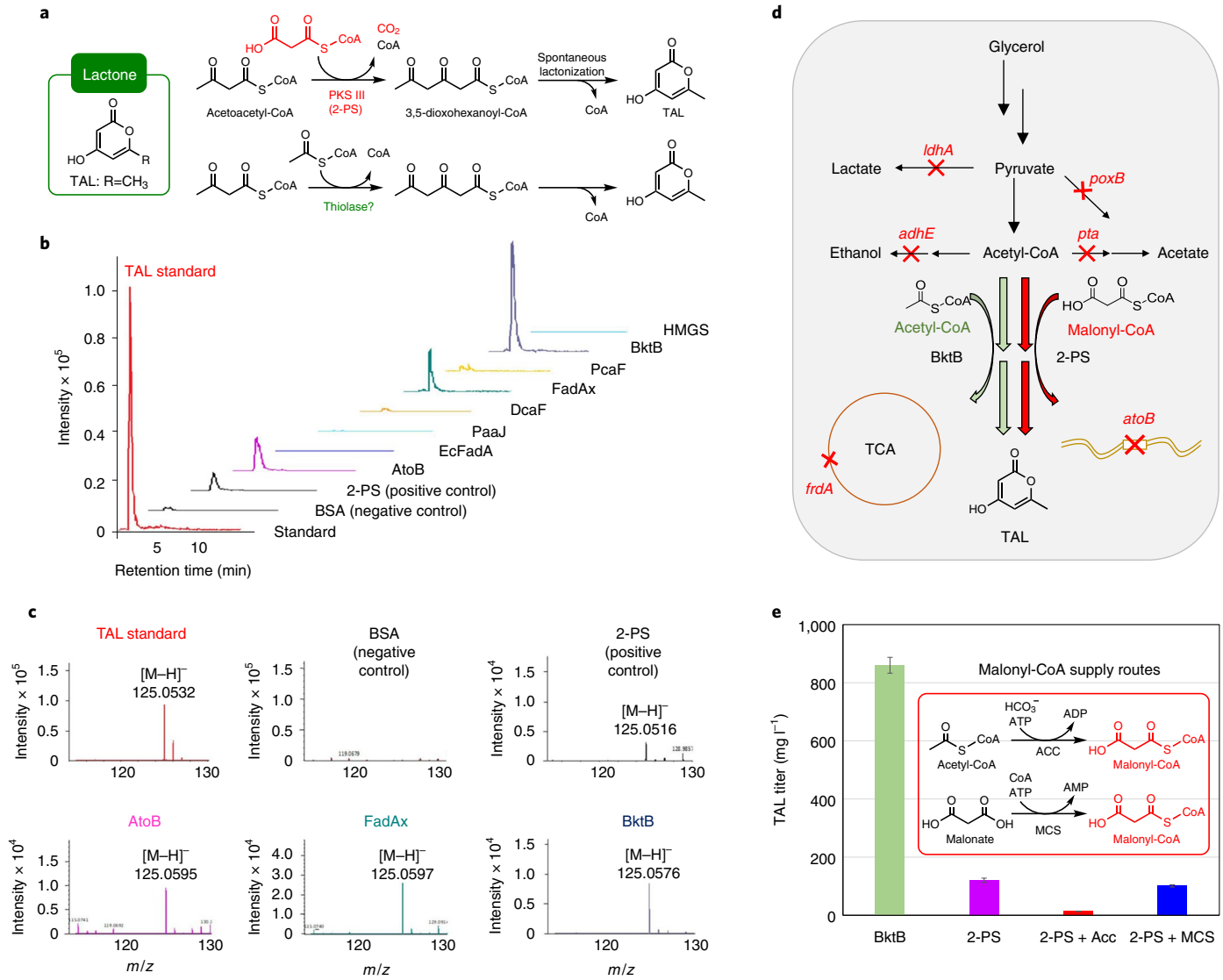


Fig. 2 | Synthesis of polyketide triacetic acid lactone through the proposed PKT pathway. a, Naturally evolved (upper) and proposed (lower) pathways for the synthesis of TAL. **b**, A liquid chromatography profile of in vitro reactions using PKS III 2-PS (positive control) and different thiolases for the synthesis of TAL. The chromatograms are extracted ion chromatograms for TAL. The retention time for TAL is ~2.3 min. Bovine serum albumin (BSA) was used as the negative control. TAL was produced in AtoB, FadAx and BktB samples. **c**, Mass spectrometry identification of TAL produced in AtoB, FadAx and BktB samples under negative mode ($[M^-H]^-$). **d**, Pathway designs for in vivo production of TAL following overexpression of thiolase (*bktB*) and PKS III (*2-ps*) routes in *E. coli* strain JC01 (DE3) Δ*atoB*, which lacks mixed-acid fermentation pathways (Δ*ldhA* Δ*poxB* Δ*adhE* Δ*pta* Δ*frdA*) and native thiolase *atoB* (Δ*atoB*). **e**, TAL titres in various JC01 (DE3) Δ*atoB* strains containing different enzymes expressed from pCDFDuet-1 cassette vector. The inset shows an increasing supply of malonyl-CoA through overexpression of MCS or ACC in PKS III (*2-ps*) overexpression strains. MCS, codon-optimized malonyl-CoA synthetase from *Bradyrhizobium japonicum*; ACC, *E. coli*-native ACC. The error bars represent s.d. calculated from at least three biological replicates.

herein referred to as polyketoacyl-CoA thiolases (PKTs)—shared low sequence similarity with 2-PS, with BktB exhibiting the highest sequence homology at only 13.9% (Extended Data Fig. 3a–c).

As TAL is an important platform molecule from which a wide range of valuable products can be derived[26], we attempted to produce TAL in vivo through a PKT-based pathway using BktB. Although the Gibbs free energy change of TAL formation from acetyl-CoA via non-decarboxylative condensations reflects a thermodynamically unfavourable pathway under standard conditions ($\Delta_rG'^{\circ} = 3.7$ kcal mol^{-1}, Supplementary Table 1), a max–min driving force[27] analysis of the pathway indicates feasibility under high in vivo pools of acetyl-CoA (Extended Data Fig. 4). To this end, five non-essential fermentative pathways that could directly or indirectly lead to the consumption of acetyl-CoA (Δ*ldhA* Δ*poxB* Δ*pta* Δ*adhE* Δ*frdA*) were inactivated in MG1655 derivative strain JC01 (DE3)[23] (Fig. 2d). As AtoB mediated the synthesis of TAL during in vitro experiments (Fig. 2a), the native *atoB* gene was also deleted in strain JC01 (DE3) to ensure TAL production was the result of BktB overexpression (Fig. 2d). We achieved TAL production through the BktB-based pathway, demonstrating a first-of-its-kind, PKS-independent synthetic pathway for TAL biosynthesis (Fig. 2e). The engineered strain JC01 (DE3) Δ*atoB* + BktB produced ~0.86 g l^{-1} of TAL from 14 g l^{-1} of glycerol (yield ≈ 6 × 10^{-2} g/g), which is sevenfold higher than the titre obtained when using PKS III; the engineered strain JC01 (DE3) Δ*atoB* + 2-PS only produced 0.12 g l^{-1} of TAL from the same amount of glycerol (yield ≈ 8.6×10^{-3} g/g)

Table 1 | Comparison of PKSs and PKTs for the synthesis of representative polyketides reported in this study

Polyketides[a]	TAL		ORA		Orcinol		6-MSA		*m*-cresol	
Condensation domain	PKT	PKS III	PKT	PKS III	PKT	PKS III	PKT	KS	PKT	KS
Required modification domain(s)			OAC	OAC	OAC	OAC	TH, KR	TH, KR, AT, ACP	TH, KR	TH, KR, AT, ACP
Starting primer	Acetyl-CoA									
Extender unit	Acetyl-CoA	Malonyl-CoA	Acetyl-CoA	Malonyl-CoA	Acetyl-CoA	Malonyl-CoA	Acetyl-CoA	Malonyl-CoA	Acetyl-CoA	Malonyl-CoA
Number of Steps	3	4	4	5	5	6	5	6	6	7
$\Delta_r G'^{\circ}$ (kcal mol^{-1})[b]	3.7	−15.0	−8.9	−37.0	−7.9	−36.1	−17.8	−45.9	−16.8	−45.0
ATP balance[c]	0	−2	0	−3	0	−3	0	−3	0	−3
Carbon efficacy[d]	1.00	0.75	1.00	0.73	0.88	0.64	1.00	0.73	0.88	0.64
Maximum CO_2 release per product[c]	0	+2	0	+3	+1	+4	0	+3	+1	+4
Type of chemistries involved[e]	NDCC, Lact.	DCC, Lact.	NDCC, Aldol	DCC, Aldol	NDCC, Aldol, Decbxl.	DCC, Aldol, Decbxl.	NDCC, Reduction, Dehydration	DCC, Reduction, Dehydration	NDCC, Reduction, Dehydration, Decbxl.	DCC, Reduction, Dehydration, Decbxl.
Ease of implementation[f]	+++++	+++	++++	++	++++	++	+++	+	+++	+

[a]The number of steps, $\Delta_r G'^{\circ}$, ATP balance and carbon efficiency were determined for generation of polyketide from starting central metabolite acetyl-CoA. [b]$\Delta_r G'^{\circ}$ was estimated under standard conditions (298.15 K, pH 7.3 and an ionic strength of 0.25). Specifically, $\Delta_r G'^{\circ}$ was computed from values of standard Gibbs free energy of formation ($\Delta_f G'^{\circ}$) for compounds that were involved in the reaction. Values for $\Delta_f G'^{\circ}$ were obtained from the MetaCyc database (https://metacyc.org/), which estimates $\Delta_f G'^{\circ}$ using the group contribution method. [c]Plus (+) or minus (−) refers to generation or consumption, respectively. [d]Carbon molar basis; that is, the number of carbons in the product divided by the number of all carbons used. [e]NDCC, non-decarboxylative Claisen condensation; DCC, decarboxylative Claisen condensation, Lact., lactonization; Decbxl., decarboxylation [f]Ease of implementation was estimated by factors such as enzyme availability, the number of required enzymes, ability to express heterologous enzymes and complexity of enzymes with rankings of very high (+++++), high (++++), medium (+++), low (++) and very low (+).

(Fig. 2e), which is consistent with the TAL levels from the wild-type 2-PS-based pathway reported in previous studies[17,28]. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis results showed that 2-PS and BktB were expressed at comparable levels in vivo (Extended Data Fig. 5a), indicating that the lower TAL titres using 2-PS were not a result of low soluble expression.

Considering that 2-PS utilizes malonyl-CoA (opposed to acetyl-CoA) as the extender unit (Fig. 2d), we assessed whether TAL production with this enzyme could be limited by insufficient supply of malonyl-CoA. Two different strategies were used to increase malonyl-CoA levels, namely the heterologous expression of malonyl-CoA synthetase (MCS) and the overexpression of the *E. coli*-native acetyl-CoA carboxylase complex (ACC; see inset of Fig. 2e). Although these strategies led to an increase in the malonyl-CoA supply (Extended Data Fig. 6), TAL titres did not improve (Fig. 2e) and the +ACC actually led to a decrease in TAL production, presumably due to +ACC-mediated cellular toxicity[29]. These findings further suggest that the relatively low TAL titres of the 2-PS/PKS-based pathway probably result from inherent inefficiencies of 2-PS/PKS, such as higher ATP requirements when compared to BktB (a PKT) (Table 1).

Harnessing PKTs for the synthesis of other polyketides. We next tested the potential of a PKT-based platform for the synthesis of alkylresorcinolic acids and alkylresorcinols, which are more structurally complex than TAL and require multiple iterations of non-decarboxylative condensations (Fig. 3a). To this end, ORA and orcinol, which are naturally synthesized by PKS IIIs, were used as proxies in part due to their roles in the synthesis of the important anti-HIV drug daurichromenic acid and related derivatives of interest[30] (Fig. 3a). Orcinol was also the first polyketide to be isolated by Collie and Myers in 1893, when Collie originally proposed the polyketide term to represent natural products derived from CH_2–CO building blocks[31].

Despite this PKT-based route to ORA and orcinol being thermodynamically feasible (Supplementary Table 1), no ORA or orcinol was observed even when using a higher concentration of BktB (12 μM) and a longer reaction time (Fig. 3b). Although this could be due to the inability of BktB to catalyse a second condensation of acetyl-CoA to form 3,5,7-trioxooctanoyl-CoA, the lack of ORA or orcinol formation could also result from the absence of a cyclization activity towards 3,5,7-trioxooctanoyl-CoA. The olivetolic acid cyclase (OAC) from *Cannabis sativa* (*C. Sativa*), which showed cyclase activity towards a 3,5,7-trioxododecanoyl-CoA analogue in previous studies[32,33], was therefore included. Following the addition of OAC (45 μM) and BktB (12 μM), we observed the synthesis of both ORA (~0.02 mg l^{-1}) and orcinol (0.5 mg l^{-1}) from 1 mM acetoacetyl-CoA and 1 mM acetyl-CoA, indicating that PKT BktB can catalyse the iterative addition of acetyl-CoA to a growing polyketoacyl-CoA, with a cyclization enzyme required to observe tailored product formation from the 3,5,7-trioxooctanoyl-CoA backbone (Fig. 3b,c). Interestingly, although BktB and PKS III orcinol synthase (ORS) are both capable of generating the same polyketide backbone, they share less than 10% protein sequence identity (Extended Data Fig. 3d).

We next sought to extend the in vivo PKT-dependent biosynthetic route to ORA and orcinol production in *E. coli* by overexpressing the *bktB* and *oac* genes in the aforementioned JC01 (DE3) Δ*atoB* strain. Both the Gibbs free energy changes and calculated max–min driving force support the feasibility of in vivo synthesis via this pathway (Supplementary Table 1; Extended Data Fig. 4) and the resulting strain produced ~5 mg l^{-1} total ORA and orcinol in 24 h (Fig. 3d; see mass spectra in Extended Data Fig. 7a,b). This engineered *E. coli* strain for ORA/orcinol production further confirms that PKTs such as BktB can iteratively catalyse multiple rounds of non-decarboxylative chain extension with acetyl-CoA as extender unit. The three rounds of iterative addition to a starter molecule observed here are equivalent to the typical condensation cycles catalysed by thirteen of the fifteen known families of plant PKS IIIs for the generation of a variety of polyketides[34]. The titres obtained are comparable with those reported for other polyketides that require three rounds of decarboxylative chain extension using PKS IIIs expressed in *E. coli* strains: for example, olivetolic acid (2.8 mg l^{-1}, tetraketide synthase)[33], phlorisovalerophenone (6.4 mg l^{-1}, valerophenone synthase)[35], resveratrol (3.6 mg l^{-1}, stilbene synthase)[36] and naringenin (5.3 mg l^{-1}, chalcone synthase)[37]. More importantly,

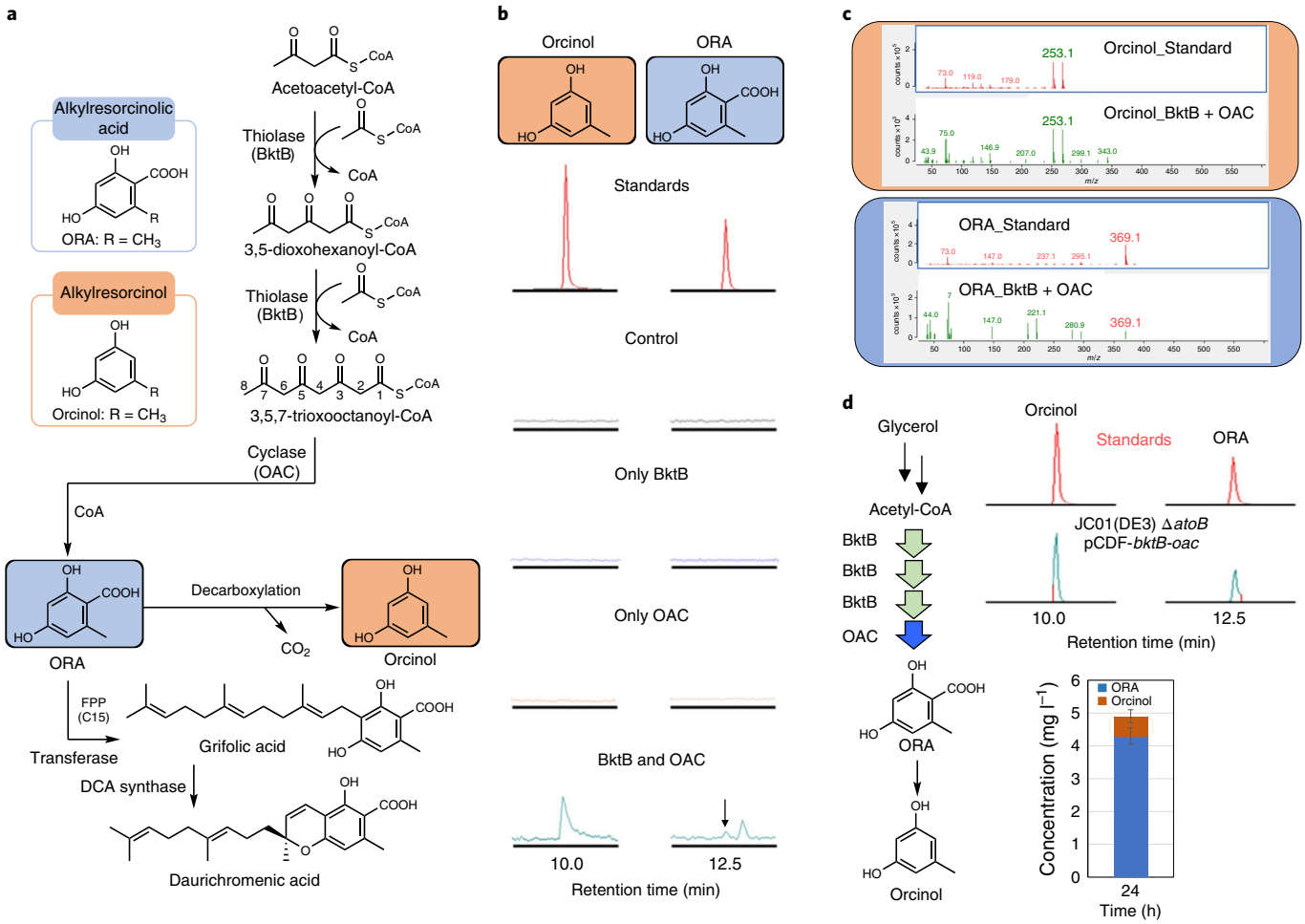


Fig. 3 | Formation of ORA and orcinol through the PKT-based pathway. a, The proposed pathway for the synthesis of ORA and orcinol using BktB (a PKT) and OAC; the resulting ORA can serve as the nucleus for the synthesis of prenylated products grifolic acid and anti-HIV drug daurichromenic acid, which are naturally isolated from *Rhododendron dauricum*. FPP, farnesyl pyrophosphate. **b**, A gas chromatography–mass spectrometry (GC–MS) total ion chromatogram profile of ORA/orcinol standards and in vitro samples. Orcinol and ORA have retention times of ~10 min and ~12.5 min, respectively. **c**, Mass spectrometry identification of orcinol (upper two specta) and ORA (lower two spectra) peaks in the in vitro BktB + OAC sample through comparison to actual standards. **d**, In vivo complete synthesis of ORA and orcinol production from MOPS medium supplemented with 20 g l^{-1} glycerol as the sole carbon source. Genes *bktB* and *oac* were cloned into the pCDFDuet–1 cassette vector and JC01 (DE3) Δ*atoB* was used as the production host strain. ORA and orcinol standards (upper right), a GC–MS profile for ORA/orcinol in vivo samples (middle right) and the calculated concentrations of ORA/orcinol in vivo samples (lower right) are shown. The MS information can be seen in Extended Data Fig. 7a,b. The error bars represent the s.d. calculated from at least three biological replicates.

given that the natural substrate of OAC from *C. sativa* is the longer trioxododecanoyl-CoA (required for olivetolic acid cyclization) instead of the trioxooctanoyl-CoA required for ORA cyclization[32], engineering the binding pocket of OAC to facilitate recognition of shorter trioxooctanoyl-CoA backbone could further improve ORA/orcinol titres.

We also assessed the thermodynamic feasibility of using a PKT-based platform for the synthesis of polyketides that require more than three rounds of condensations. As representative examples, the synthesis of aleosone (naturally produced by PKS III of aloesone synthase through six rounds of condensations[12]) and octaketide SEK4 (naturally produced by PKS III of octaketides synthase through seven rounds of condensations[38]), were both found to be thermodynamically feasible through the PKT-based pathway (Extended Data Fig. 8a,b).

Harnessing PKTs for PK backbones synthesis beyond PKS IIIs. To test whether PKTs can also be employed to generate PK backbones that are naturally synthesized by PKS I/IIs, in addition to the demonstrated PKS III products, we aimed to synthesize the representative polyketide 6-MSA (Fig. 4a). 6-MSA is formed naturally by the PKS I 6-MSA synthase (6-MSAS), which was the first PKS to be purified and the first fungal iterative PKS I whose gene was cloned[39]. 6-MSAS consists of ketoacylsynthase (KS), acyltransferase (AT), thioester hydrolase (TH), ketoreductase (KR) and acyl carrier protein (ACP) domains[40]. All five domains have been found to be indispensable for the functionality of 6-MSAS (~198 kDa in total, Fig. 4b). We designed a de novo PKT-based pathway to synthesize a 6-MSA that is thermodynamically feasible (Fig. 4a; Extended Data Fig. 4) and which consists of only three components, the PKT, as well as TH and KR domains (~136 kDa in total, Fig. 4b). For this, the TH-KR domains from *Aspergillus terreus*[40] were expressed, purified and added to an in vitro reaction mixture which included BktB (12 μM of each). With all three present, we observed the synthesis of 6-MSA (~1 mg l^{-1}) and *m*-cresol (decarboxylated form of 6-MSA, ~0.2 mg l^{-1}) from acetoacetyl-CoA primer (1 mM) and acetyl-CoA extender unit (1 mM) (Fig. 4c, see mass spectra in Extended Data Fig. 7c,d). In addition to this

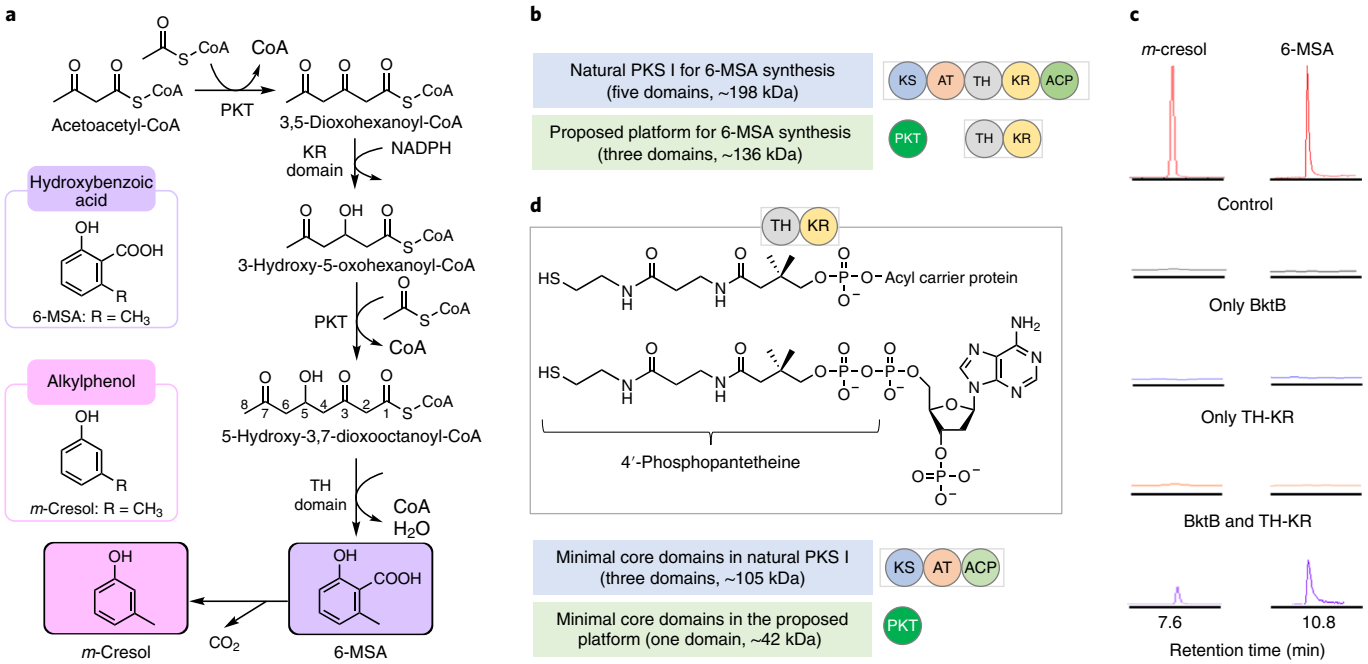


Fig. 4 | Synthesis of polyketides 6-MSA and *m*-cresol. a, The proposed route for the synthesis of 6-MSA and *m*-cresol through the PKT-based pathway. **b**, A comparison of the domain organizations of naturally evolved PKS I (6-MSAS) and an artificially designed PKT platform for 6-MSA synthesis. All five PKS I domains are indispensable to the functionality of 6-MSAS. The proposed PKT platform comprises only three domains: PKT, TH and KR. **c**, A GC–MS total ion chromatogram profile of *m*-cresol and 6-MSA standards and in vitro samples; *m*-cresol and 6-MSA have retenton times of ~7.6 min and ~10.8 min, respectively. **d**, The TH-KR domain can recognize both ACP- and CoA-bound substrates. For PKS I (blue shading), the minimal core contains three functional domains: KS, AT and ACP. By contrast, for the PKT platform (green shading), only one PKT domain is required for functionality. KS, ketoacylsynthase; AT, acyltransferase; TH, thioester hydrolase; KR, ketoreductase; ACP, acyl carrier protein.

demonstration of in vitro production, thermodynamic calculations suggest that 6-MSA/*m*-cresol production starting from acetyl-CoA is a feasible pathway in vivo (Supplementary Table 1; Extended Data Fig. 4). Moreover, the TH-KR domains were found to recognize and catalyse the CoA-bound substrates next to their native ACP-bound substrates, which greatly expands the toolbox for future engineering of CoA-dependent pathways (Fig. 4d).

It is widely accepted for PKS I that each module contains minimally three core functional domains: KS, AT and ACP[41] (~105 kDa in the case of 6-MSAS of *A. terreus*; Fig. 4d). By contrast, for PKT, as it operates on the CoA-bound system, the AT and ACP domains (~60 kDa) become dispensable for polyketide biosynthesis, reducing the core domain size by ~60% (the BktB domain alone is ~42 kDa; Fig. 4d). It is reasonable to assume that such simplicity will considerably simplify the organization and architecture for polyketide synthesis, especially for the large, multimodule PKS Is[13] (Fig. 4d).

In addition to 6-MSA/*m*-cresol, we also assessed the thermodynamic feasibility of using a PKT-based platform for the synthesis of other more structurally complex polyketides that require additional rounds of condensation. Specifically, we analysed PKT-based pathways for the synthesis of norsolorinic acid PK (naturally produced by PKS I of PksA through seven condensations[42]) and tetracenomycin F1 polyketide (naturally produced by PKS II of TCM PKS through nine condensations[43]). We found that both are thermodynamically feasible (Extended Data Fig. 8c,d), further suggesting the broad potential of the proposed platform.

Basic catalytic mechanism of PKT. To elucidate the catalytic mechanism of the newfound PKT activity that supports polyketoacyl-CoA synthesis, we first focused on BktB and its known ketoacyl-CoA thiolase activity, which contains a catalytic triad of a cysteine residue (Cys90) that acts as the nucleophile, a histidine residue (His350) that activates the Cys90 nucleophile and another Cys380 that acts as a base, deprotonating the α-carbon of acetyl-CoA[44] (Fig. 5a). We individually mutated each cysteine residue to serine (C90S, C380S) and histidine to alanine (H350A), finding that these mutations did not affect the correct folding of BktB mutants (Extended Data Fig. 5b). All three mutants showed complete loss of ketoacyl-CoA thiolase activity (Fig. 5b,c). More importantly, each mutation also abolished TAL formation (Fig. 5b,d), demonstrating that the Cys90/His350/Cys380 catalytic triad is also involved in PKT activity. Based on these findings, the PKT-catalysed reaction seems to proceed through a two-step, ping-pong mechanism for the thiolytic condensation, using substrates with multiple keto groups (for example, acetoacetyl-CoA; Fig. 5e). This catalytic mechanism is notably distinct from that of PKS III, where the core catalytic triad is Cys/His/Asn, with Asn (asparagine) playing a critical role in catalysing malonyl-CoA decarboxylation and stabilizing the condensation transition state (Extended Data Fig. 9)[45].

Potential structural features determining PKT activity. The presence of the Cys/His/Cys triad, however, does not seem to solely impart PKT activity, as EcFadA, PaaJ, DcaF and PcaF also share this catalytic triad, despite these enzymes not catalysing TAL formation (Fig. 2b). We hypothesized that the lack of PKT activity in these enzymes may result from an inability to recognize polyketoacyl-CoAs as primers (for example, acetoacetyl-CoA). PaaJ, DcaF and PcaF have been reported as 3-ketoadipyl-CoA thiolases, with succinyl-CoA as the native condensing substrate[21,22,24], suggesting that the binding pocket may accommodate the presence of ω-functionalized succinyl-CoA at the expense of other four-carbon substrates. Consistent with our hypothesis, PaaJ was experimentally validated to only recognize succinyl-CoA as condensation substrate but neither butyryl-CoA nor acetoacetyl-CoA (Fig. 6a). On the other hand, EcFadA, a thiolase with

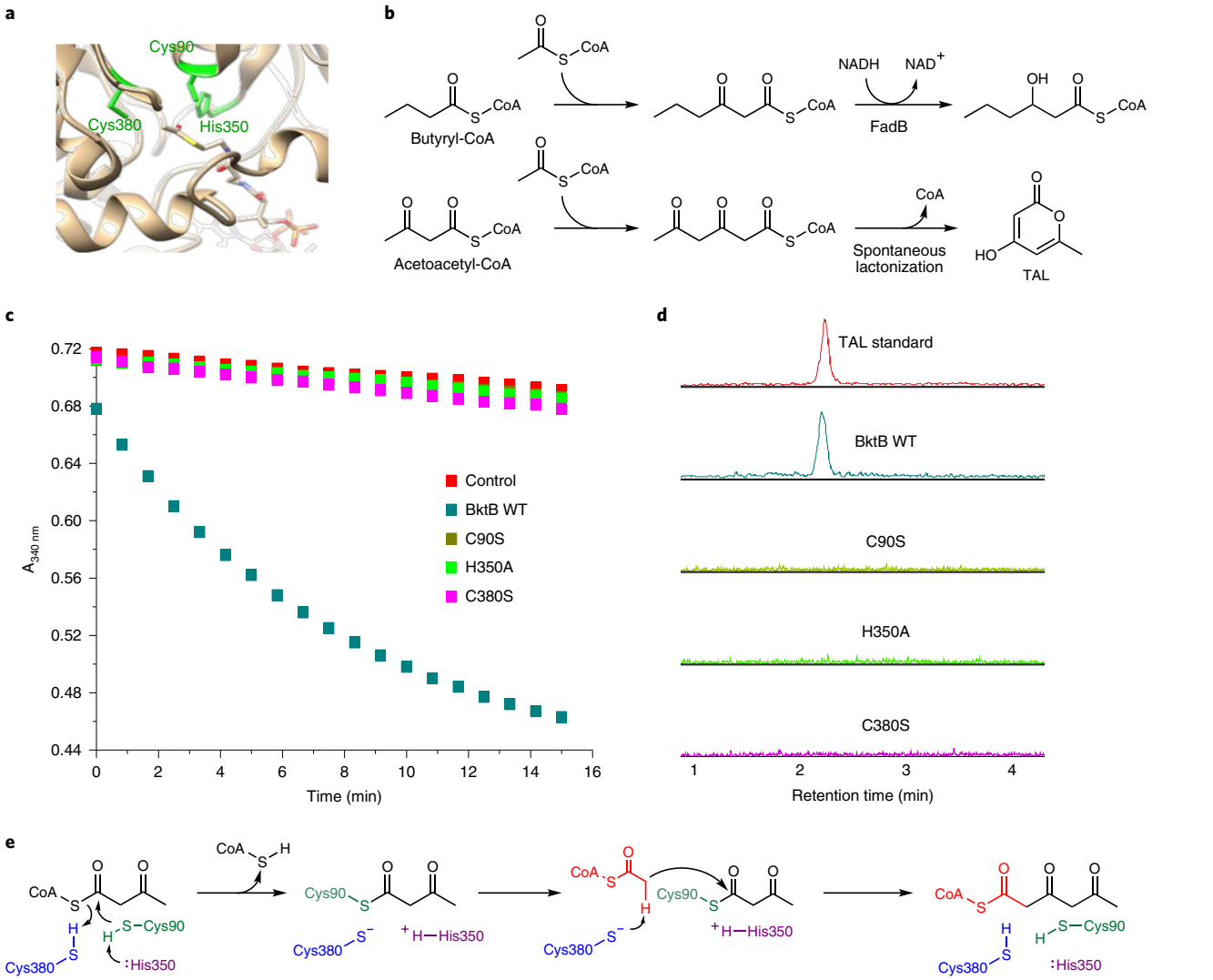


Fig. 5 | Probing the basic catalytic mechanism of PKT activity. a, The catalytic triad (Cys90/His350/Cys380) of BktB. **b**, Top: in vitro analysis of ketoacyl-CoA thiolase activity for BktB mutants, using butyryl-CoA as the condensing primer and acetyl-CoA as the extender unit. Bottom: in vitro analysis of PKT activity for BktB mutants, using acetoacetyl-CoA as the primer and acetyl-CoA as the extender unit for TAL biosynthesis. **c**, All three BktB mutants showed a complete loss of ketoacyl-CoA thiolase activity; NADH consumption was monitored by the decrease in absorbance at 340 nm. WT, wild-type. **d**, All three BktB mutations abolished the PKT activity. **e**, The proposed catalytic mechanism of PKT. The data suggest that a two-step, ping-pong mechanism has been adopted for the condensation reaction. During the first step, Cys90 is activated by His350 through a nucleophilic attack, abstracting the sulfide proton of Cys90, with the activated Cys90 attacking the carbonyl group of the acetoacetyl-CoA substrate to form a covalent acetoacetyl-enzyme intermediate and releasing CoA. In the second step, Cys380 acts as a base, deprotonating the α-carbon of acetyl-CoA and the Cys380-acetyl-CoA nucleophilically attacks the acetoacetyl-enzyme intermediate to yield 3,5-dioxoacyl-CoA.

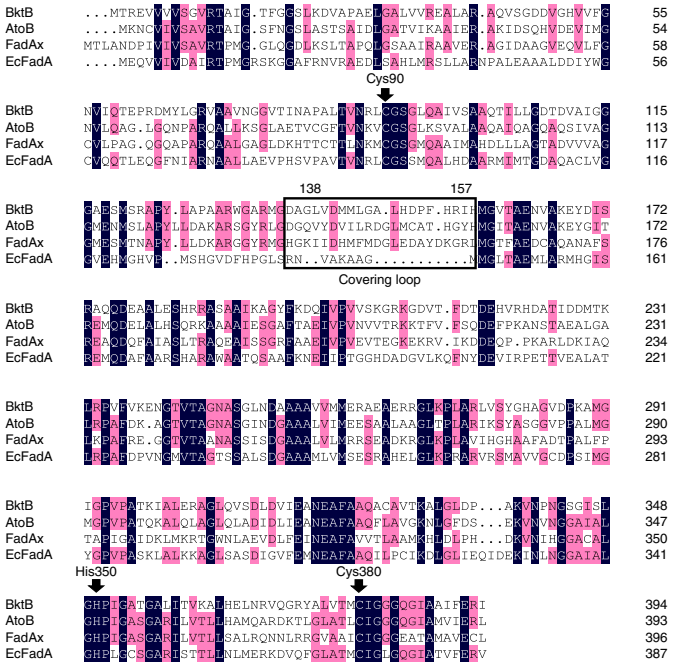
broad chain-length specificity, shows comparable activity towards butyryl-CoA but was incapable of condensing acetoacetyl-CoA with acetyl-CoA (Fig. 6a), which suggests that EcFadA is an ideal non-PKT candidate for comparison with BktB.

Primary sequence alignment of EcFadA and BktB revealed an extra fragment in BktB (amino acids 138–157) that is also present in other thiolases with PKT activity (Fig. 6b), which may comprise the covering loop of thiolases involved in the binding process for branched-chain or linear substrates (Fig. 6b,c)[46]. Inspired by this finding, we performed segment swapping between EcFadA and BktB, in which the EcFadA segment—which consists of nine amino acids (138–146)—was swapped with the corresponding segment in BktB, which consists of 20 amino acids (138–157), resulting in EcFadA$_{_BktB}$. Similar swapping was performed with BktB, yielding BktB$_{_EcFadA}$ (Fig. 6d). Such swapping did not substantially impact the soluble expression levels of these mutants compared with their wild types (Extended Data Fig. 5c). Characterization of EcFadA$_{_BktB}$ revealed a ~30-fold increase in the acetoacetyl-CoA:butyryl-CoA selectivity ratio compared with EcFadA$_{_WT}$ (relative k_{cat}/K_M 0.0367 versus 0.00123), which reflects enhanced substrate affinity (K_M) towards acetoacetyl-CoA and decreased catalytic activity (k_{cat}) towards butyryl-CoA (Fig. 6d). Furthermore, fragment swapping in BktB (BktB$_{_EcFadA}$) nearly abolished PKT activity (Fig. 6d), demonstrating the overall importance of this covering loop. Although EcFadA$_{_BktB}$ increased its efficacy towards acetoacetyl-CoA compared with EcFadA$_{_WT}$ (by ~30-fold), BktB$_{_WT}$ still exhibited higher acetoacetyl-CoA/butyryl-CoA selectivity (relative k_{cat}/K_M, 0.160; Fig. 6d), suggesting that other fragments or secondary structures may also contribute to the PKT activity of BktB, which might deserve further investigation in the future.

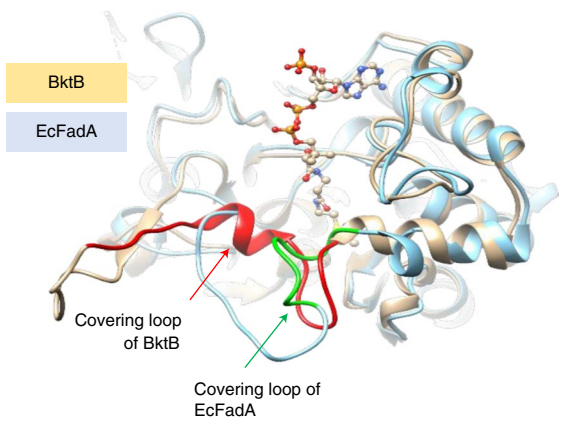
a

Substrate \ Enzyme	BktB (μmol mg^{-1} min^{-1})	EcFadA (μmol mg^{-1} min^{-1})	PaaJ (μmol mg^{-1} min^{-1})
Butyryl-CoA	0.24 ± 0.020	0.16 ± 0.012	(6.7 ± 1.5) × 10^{-3}
Succinyl-CoA	(8.2 ± 1.1) × 10^{-3}	(9.1 ± 1.4) × 10^{-4}	0.11 ± 0.018
Acetoacetyl-CoA	0.11 ± 0.015	(4.1 ± 0.9) × 10^{-3}	(6.0 ± 1.1) × 10^{-3}

b



c



d

Enzyme	Relative k_{cat}/K_M	Substrate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} min^{-1})
BktB$_{WT}$	1	Butyryl-CoA	8.89 ± 0.58	39 ± 3	(2.26 ± 0.04) × 10^5
	(1.60 ± 0.11) × 10^{-1}	Acetoacetyl-CoA	2.46 ± 0.25	68 ± 5	(3.63 ± 0.31) × 10^4
BktB$_{EcFadA}$	1	Butyryl-CoA	0.34 ± 0.05	36 ± 3	(9.45 ± 1.38) × 10^3
	<10^{-5}	Acetoacetyl-CoA	n.d.	n.d.	n.d.
EcFadA$_{WT}$	1	Butyryl-CoA	4.36 ± 0.26	50 ± 4	(8.72 ± 0.59) × 10^4
	(1.23 ± 0.18) × 10^{-3}	Acetoacetyl-CoA	0.39 ± 0.06	3627 ± 35	107.6 ± 16.1
EcFadA$_{BktB}$	1	Butyryl-CoA	0.33 ± 0.03	44 ± 4	(7.46 ± 0.68) × 10^3
	(3.67 ± 0.57) × 10^{-2}	Acetoacetyl-CoA	0.32 ± 0.04	1191 ± 46	271.2 ± 29.9

Fig. 6 | Probing structural determinants of PKT activity. a, An in vitro assay of BktB, EcFadA and PaaJ for the activities of ketoacyl-CoA thiolase (with butyryl-CoA as the condensing primer), 3-ketoadipyl-CoA thiolase (with succinyl-CoA as the condensing primer) and PKT (with acetoacetyl-CoA as the condensing primer). In all cases, acetyl-CoA was used as the extender unit. **b**, Primary sequence alignment of BktB, AtoB, FadAx (PKTs) and EcFadA (non-PKT). The black arrows indicate the locations of the catalytic triad (Cys90/His350/Cys380); the black box indicates the covering loop of the PKTs (note that this fragment is missing in EcFadA). **c**, A secondary structure alignment between PKT BktB and non-PKT EcFadA; EcFadA lacks the covering loop present in BktB. The bound substrate is acetyl-CoA. **d**, In vitro kinetic characterization of ketoacyl-CoA thiolase (with butyryl-CoA as the condensing primer) and PKT (with acetoacetyl-CoA as the condensing primer) activities of BktB and EcFadA wild-types and mutants. The error bars represent s.d. calculated from at least three replicates. n.d., not detected.

Conclusion

Given the divergent evolution and structural similarity within the thiolase superfamily, we hypothesized that polyketide backbones can be formed in a PKS-independent pathway through non-decarboxylative Claisen condensations catalysed by certain thiolases, which we refer to as PKTs. We demonstrated the feasibility of this PKT-based pathway through the synthesis of representative PKs such as lactones (triacetic acid lactone), alkylresorcinolic acids (ORA), alkylresorcinols (orcinol), hydroxybenzoic acid (6-MSA) and alkylphenol (*m*-cresol). The use of PKTs for the synthesis of polyketide backbones differs from all previous studies that exploited thiolases for the production of small organic molecules in several ways[23,47–49]. Compared with PKSs, the PKT-dependent pathway circumvents the intrinsic ATP inefficiencies and minimizes the required domains and catalytic steps from central metabolites to final PKs (Table 1). Furthermore, the PKT-based platform utilizes acetyl-CoA as extender unit, the flux of which is considerably higher than malonyl-CoA extender unit[50], and avoids direct competition with essential metabolic pathways[12]. Although in the illustrative examples of this study all PKTs employ acetyl-CoA as synthetic starting primer (R = CH_3), we anticipate that additional polyketide backbones (R = groups other than CH_3) can also be generated through such PKT-dependent pathways by mining new thiolases with PKT activity from nature or through binding pocket engineering of currently identified thiolases with PKT activity (for example BktB) to accommodate various substrates with modified R groups (Extended Data Fig. 10) and/or used in combination with other modification domains (for example TH, KR domains). By establishing both the basic catalytic mechanism

of PKTs and the importance of secondary structures such as the covering loop in enabling PKT activity, our findings can be expanded to other thiolases to further elucidate their structure-function relationship and harness their full biosynthetic potential to circumvent the aforementioned challenges associated with PKSs. Finally, our results demonstrate that even in the expansive protein sequence space of today, the overriding structural motifs and structure-function relationships observed across proteins from all areas of metabolism (for example thiolase superfamily), can be exploited for novel biochemical functions.

Methods

Strain construction. The *Escherichia coli* K-12 JC01 (DE3) strain was used as the host for all genetic modifications. All constructed strains are listed in Supplementary Table 2. Luria–Bertani medium was used for culturing *E. coli* cells for plasmid construction. When necessary, antibiotics were added at the following final concentrations: ampicillin (100 μg ml^{-1}), spectinomycin (50 μg ml^{-1}), kanamycin (50 μg ml^{-1}) and chloramphenicol (17 μg ml^{-1}).

Plasmid construction. All plasmids used in this study are listed in Supplementary Table 2. Protein sequence information and Genbank accession numbers for all thiolases and other enzymes tested can be found in Supplementary Table 3. The protein structure information of BktB and EcFadA can be found in Supplementary Table 4. For construction of recombinant plasmids, the target genes were inserted into pCDFDuet-1 plasmid (Novagen) digested with appropriate restriction enzymes using Gibson Assembly Cloning Kit (New England Biolabs). *E. coli* genes were polymerase-chain-reaction amplified from *E. coli* MG1655 genomic DNA, whereas heterologous genes were synthesized by GeneArt (Life Technologies) with codon optimization, with the exception of *bktB*, which was directly amplified from genomic DNA of *Ralstonia eutropha*. For polymerase-chain-reaction amplification of *bktB*, a GC buffer or dimethylsulfoxide is required due to high GC content. The resulting Gibson Assembly products were used to transform *E. coli* DH5α competent cell (New England Biolabs) and polymerase-chain-reaction-identified clones were confirmed by DNA sequencing (Eurofins Scientific).

Molecular phylogenetic analysis. The evolutionary history of the thiolase superfamily was inferred by using the maximum likelihood method based on the Jones–Taylor–Thornton matrix-based model. The tree with the highest log-likelihood (−43713.7414) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying the neighbour-joining and BioNJ algorithms to a matrix of pairwise distances that were estimated using a Jones–Taylor–Thornton model and then selecting the topology with the superior log-likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 617 amino acid sequences (see Supplementary Information). All positions containing gaps and missing data were eliminated. There was a total of 86 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

In vitro assays. *E. coli* BL21 (DE3) was used as host strain for expression of His-tagged proteins, from their respective pCDF-P1-X constructs. BL21 (DE3) strains harbouring pCDF-P1-X plasmids were grown at 37 °C in 50 ml fresh Luria–Bertani medium with spectinomycin at an initial $OD_{550} \approx 0.1$. When the OD_{550} of the culture increased to 0.4–0.8, isopropyl-β-D-thiogalactoside was added at final concentration of 0.1 mM for induction. After 4–6 h of induction at 37 °C, cells were harvested by centrifugation. The cell pellet was washed two times using double-distilled H_2O, resuspended in the lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and subjected to sonication using a Sonifier SFX250 (Branson). Following centrifugation (14,000 g, 4 °C, 15 min), the supernatant containing the soluble protein fraction was loaded onto the Ni-NTA Spin Columns (Qiagen). After washing twice with wash buffer (50 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0), the His-tagged enzymes were eluted from the column with elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0). The elution buffer was exchanged with storage buffer (50 mM Tris-HCl, 0.5 mM EDTA, 50 mM NaCl, 10% glycerol, pH 7.9) at 4 °C using Amicon Ultra 0.5 ml centrifugal filters at 30 K (or 10 K for OAC).

The in vitro assays for TAL production were performed in a 200 μl total reaction volume containing 100 mM potassium phosphate buffer (pH 7.0), 3 mM EDTA, 1 mM acetoacetyl-CoA, 1 mM acetyl-CoA (1 mM malonyl-CoA for 2-PS) and 50 μg (~6 μM) of purified enzymes. The reaction mixture was incubated at 22 °C for 30 min and 5 μl H_2SO_4 was added to terminate the reaction. The samples were then analysed through liquid chromatography–mass spectrometry (LC–MS) for qualitative identification of TAL (see below). In vitro quantitative measurements of the specific activity of enzymes for TAL production was performed by monitoring the increase in absorbance at 298 nm (TAL formation)[11,30].

The in vitro assays for ORA/orcinol production were performed in a 200 μl total reaction volume containing 100 mM potassium phosphate buffer (pH 7.0), 3 mM EDTA, 1 mM acetoacetyl-CoA, 1 mM acetyl-CoA, 100 μg of purified BktB (12 μM) and 50 μg of purified OAC (24 μM). The reaction mixture was incubated at 22 °C for 14 h and 5 μl H_2SO_4 was added to terminate the reaction. Samples were then analysed by GC–MS for ORA/orcinol identification.

The in vitro assays for 6-MSA/*m*-cresol production were performed in a 200 μl total reaction volume containing 100 mM potassium phosphate buffer (pH 7.0), 3 mM EDTA, 1 mM acetoacetyl-CoA, 1 mM acetyl-CoA, 100 μg of purified BktB (12 μM) and 200 μg of purified TH-KR (12 μM). The reaction mixture was incubated at 22 °C for 14 h and 5 μl H_2SO_4 was added to terminate the reaction. Samples were then analysed by GC–MS for 6-MSA/*m*-cresol identification.

The in vitro assays for forward ketoacyl-CoA thiolase activity (biosynthetic direction) were performed in a 200 μl total reaction volume containing 100 mM Tris buffer (pH 7.5), 1.5 mM dithiothreitol, 4.5 mM $MgCl_2$, 1 mM NADH, 1 mM butyryl-CoA, 1 mM acetyl-CoA, 10 μg hydroxyacyl-CoA dehydrogenase (FadB from *E. coli*) and 5 μg thiolases. The decrease in absorbance at 340 nm (NADH consumption) was monitored.

The in vitro assays for reverse ketoacyl-CoA thiolase activity (degradative direction) were performed in 200 μl total reaction volume containing 100 mM Tris buffer (pH 7.5), 0.5 mM dithiothreitol, 4.5 mM $MgCl_2$, 0.075 mM acetoacetyl-CoA, 0.2 mM CoA and 5 μg thiolases. The decrease in absorbance at 303 nm (acetoacetyl-CoA consumption) was monitored.

The in vitro assays for forward 3-ketoadipyl-CoA thiolase activity (biosynthetic direction) were performed in a 200 μl total reaction volume containing 100 mM Tris buffer (pH 7.5), 1.5 mM dithiothreitol, 4.5 mM $MgCl_2$, 0.2 mM NADH, 1 mM succinyl-CoA, 1 mM acetyl-CoA, 10 μg 3-hydroxyadipyl-CoA dehydrogenase (PaaH from *E. coli*) and 25 μg thiolases. The decrease in absorbance at 340 nm (NADH consumption) was monitored.

LC–MS identification of TAL. Methanol (100 μl) was added to the in vitro TAL samples, centrifuged at 11,000 g for 5 min to remove the protein pellets, filtered and then subjected to LC–MS analysis using an Agilent 1200 HPLC system and Bruker MicroToF ESI LC–MS System. The column used was Shim-pack XR-ODS II C18, 2.0 mm × 75 mm (Shimadzu). A sample volume of 30 μl was injected into the LC and the LC conditions were as follows: solvent A = 0.1% formic acid in H_2O; solvent B = 100% methanol; flow rate = 0.25 ml min^{-1}; 0–2.5 min, 95% A and 5% B; 2.5–20 min, 95% A and 5% B to 5% A and 95% B; 20–23 min, 5% A and 95% B; 23–24 min, 5% A and 95% B to 95% A and 5% B; 24–30 min, 95% A and 5% B.

GC–MS identification of ORA/orcinol. H_2SO_4 (80 μl) and 340 μl of 30% (wt/v) NaCl solution were added to the in vitro ORA/orcinol samples to adjust the pH and ionic strength; 2 ml of ethyl acetate was added for extraction. Vials were sealed with Teflon-lined septa (Fisher Scientific), secured with caps and rotated at 2,000 r.p.m. for 10 min. The samples were then centrifuged for 2 min at 6,500 g to separate the aqueous and organic layers. After centrifugation, 0.5 ml of the top organic layer was transferred to new 5 ml glass vials and evaporated under a stream of nitrogen; 100 μl pyridine and 100 μl of *N*,*O*-Bis(trimethylsilyl) trifluoroacetamide (BSTFA, Sigma-Alridch) were subsequently added to the dried extract for derivatization at 70 °C for 1 h. After cooling to room temperature, 200 μl of derivatization product was transferred to vials (Fisher Scientific) for GC–MS analysis as below: 1 μl were injected into the GC, which was run in splitless mode using helium gas as a carrier gas with a flow rate of 1 ml min^{-1}. The injector temperature was 280 °C and the oven temperature was initially held at 50 °C for 3 min and then raised to 250 °C at 10 °C min^{-1} and held for 3 min.

Fermentation medium and conditions. MOPS medium supplemented with tryptone (10 g l^{-1}), yeast extract (5 g l^{-1}) and glycerol (20 g l^{-1}) as the carbon source was employed for all fermentations. Fermentations were performed in 25 ml Pyrex Erlenmeyer flasks (narrow mouth/heavy duty rim, Corning) filled with 5–20 ml of the MOPS medium and fitted with foam plugs filling the necks. A single colony of the desired strain was cultivated overnight (14–16 h) in Luria–Bertani medium (with the appropriate antibiotics) and used as the initial inoculum at $OD_{550} \approx 0.05$. Following inoculation, flasks were further incubated at 37 °C and 200 r.p.m. until the OD_{550} reached 0.4–0.8, at which point 0–20 μM isopropyl-β-D-thiogalactoside inducer was added for JC01 (DE3) Δ*atoB* and derivative strains. Sodium malonate (12 mM) was also added when MCS was expressed for malonyl-CoA synthesis. Flasks were then incubated under the same conditions for 48 h post-induction for TAL production or under 22 °C, 72 h for ORA/orcinol production.

HPLC metabolite quantification. The concentration of glycerol, acetic acid and TAL in fermentation broth were determined through ion-exclusion HPLC using a Shimadzu Prominence SIL 20 system (Shimadzu Scientific) equipped with an HPX-87H organic acid column (Bio-Rad) with operating conditions to optimize peak separation (0.3 ml min^{-1} flow rate, 30 mM H_2SO_4 mobile phase, column temperature 42 °C).

Statistical analysis. The two-tailed t-test method was employed to analyse the statistical significance of all data in this study and a *P*-value of <0.05 was deemed statistically significant.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

The authors declare that data supporting the findings of this study are available within the paper and its Supplementary Information. Supplementary Table 3 provides a list of the GenBank accession numbers of the 17 key enzymes used in this study. Data for Figs. 1–6 and Extended Data Fig. 6 are available as Source Data with this paper. All other data are available from the authors on reasonable request.

Received: 17 December 2019; Accepted: 12 May 2020;
Published online: 22 June 2020

References

1. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem.* **48**, 4688–4716 (2009).
2. Ma, S. M. et al. Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589–592 (2009).
3. Yu, D. Y., Xu, F. C., Zeng, J. & Zhan, J. X. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285–295 (2012).
4. Bond, C., Tang, Y. & Li, L. *Saccharomyces cerevisiae* as a tool for mining, studying and engineering fungal polyketide synthases. *Fungal Genet. Biol.* **89**, 52–61 (2016).
5. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
6. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).
7. Huo, L. et al. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412–1436 (2019).
8. Pfeifer, B. A. & Khosla, C. Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **65**, 106–118 (2001).
9. Lowry, B. et al. *In vitro* reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **135**, 16809–16812 (2013).
10. Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* **26**, 90–114 (2009).
11. Brownsey, R., Boone, A., Elliott, J., Kulpa, J. & Lee, W. Regulation of acetyl-CoA carboxylase. *Biochem. Soc. Trans.* **34**, 223–227 (2006).
12. Yan, D. et al. Repurposing type III polyketide synthase as a malonyl-CoA biosensor for metabolic engineering in bacteria. *Proc. Natl Acad. Sci. USA* **115**, 9835–9844 (2018).
13. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
14. Jiang, C., Kim, S. Y. & Suh, D. Y. Divergent evolution of the thiolase superfamily and chalcone synthase family. *Mol. Phylogenetetics Evol.* **49**, 691–701 (2008).
15. Austin, M. B. & Noel, A. J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110 (2003).
16. Haapalainen, A. M., Merilainen, G. & Wierenga, R. K. The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem. Sci.* **31**, 64–71 (2006).
17. Tang, S. Y. et al. Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. *J. Am. Chem. Soc.* **135**, 10099–10103 (2013).
18. Campobasso, N. et al. *Staphylococcus aureus* 3-hydroxy-3-methylglutaryl-CoA synthase: crystal structure and mechanism. *J. Biol. Chem.* **279**, 44883–44888 (2004).
19. Duncombe, G. R. & Frerman, F. E. Molecular and catalytic properties of the acetoacetyl-coenzyme A thiolase of *Escherichia coli*. *Arch. Biochem. Biophys.* **176**, 159–170 (1976).
20. Yang, S. Y., Yang, X. Y. H., Healylouie, G., Schulz, H. & Elzinga, M. Nucleotide-sequence of the *fadA* Gene—primary structure of 3-ketoacyl-coenzyme-A thiolase from *Escherichia coli* and the structural organization of the *fadAB* operon. *J. Biol. Chem.* **265**, 10424–10429 (1990).
21. Teufel, R. et al. Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc. Natl Acad. Sci. USA* **107**, 14390–14395 (2010).
22. Harwood, C. S., Nichols, N. N., Kim, M. K., Ditty, J. L. & Parales, R. E. Identification of the pcaRKF gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. *J. Bacteriol.* **176**, 6479–6488 (1994).
23. Cheong, S., Clomburg, J. M. & Gonzalez, R. Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nat. Biotechnol.* **34**, 556–561 (2016).
24. Parke, D., Garcia, M. A. & Ornston, L. N. Cloning and genetic characterization of *dca* genes required for β-oxidation of straight-chain dicarboxylic acids in *Acinetobacter* sp. strain ADP1. *Appl. Environ. Microbiol.* **67**, 4817–4827 (2001).
25. Slater, S. et al. Multiple β-ketothiolases mediate poly(β-hydroxyalkanoate) copolymer synthesis in *Ralstonia eutropha*. *J. Bacteriol.* **180**, 1979–1987 (1998).
26. Shanks, B. H. & Keeling, P. L. Bioprivileged molecules: creating value from biomass. *Green. Chem.* **19**, 3177–3185 (2017).
27. Noor, E. et al. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* **10**, e1003483 (2014).
28. Xie, D. et al. Microbial synthesis of triacetic acid lactone. *Biotechnol. Bioeng.* **93**, 727–736 (2006).
29. Johnson, A. O. et al. Design and application of genetically-encoded malonyl-CoA biosensors for metabolic engineering of microbial cell factories. *Metab. Eng.* **44**, 253–264 (2017).
30. Qian, S., Clomburg, J. M. & Gonzalez, R. Engineering *Escherichia coli* as a platform for the *in vivo* synthesis of prenylated aromatics. *Biotechnol. Bioeng.* **116**, 1116–1127 (2019).
31. Collie, N. & Myers, W. V. I. I. The formation of orcinol and other condensation products from dehydracetic acid. *J. Chem. Soc. Trans.* **63**, 122–128 (1893).
32. Gagne, S. J. et al. Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *Proc. Natl Acad. Sci. USA* **109**, 12811–12816 (2012).
33. Tan, Z., Clomburg, J. M. & Gonzalez, R. Synthetic pathway for the production of olivetolic acid in *Escherichia coli*. *ACS Synth. Biol.* **7**, 1886–1896 (2018).
34. Lim, Y. P., Go, M. K. & Yew, W. S. Exploiting the biosynthetic potential of type III polyketide synthases. *Molecules* **21**, 806 (2016).
35. Zhou, W. et al. Biosynthesis of phlorisovalerophenone and 4-hydroxy-6-isobutyl-2-pyrone in *Escherichia coli* from glucose. *Microb. Cell Fact.* **15**, 149 (2016).
36. Huang, L., Xue, Z. & Zhu, Q. Method for the production of resveratrol in a recombinant bacterial host cell. US patent WO2006124999A2 (2007).
37. Choi, O. et al. Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **38**, 1657–1665 (2011).
38. Mizuuchi, Y. et al. Novel type III polyketide synthases from *Aloe arborescens*. *FEBS J.* **276**, 2391–2401 (2009).
39. Fujii, I. Functional analysis of fungal polyketide biosynthesis genes. *J. Antibiot.* **63**, 207–218 (2010).
40. Moriguchi, T., Kezuka, Y., Nonaka, T., Ebizuka, Y. & Fujii, I. Hidden function of catalytic domain in 6-methylsalicylic acid synthase for product release. *J. Biol. Chem.* **285**, 15637–15643 (2010).
41. Sabatini, M. et al. Biochemical characterization of the minimal domains of an iterative eukaryotic polyketide synthase. *FEBS J.* **285**, 4494–4511 (2018).
42. Crawford, J. M. et al. Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **461**, 1139–1243 (2009).
43. Shen, B. & Hutchinson, C. R. Deciphering the mechanism for the assembly of aromatic polyketides by a bacterial polyketide synthase. *Proc. Natl Acad. Sci. USA* **93**, 6600–6604 (1996).
44. Kim, E. J., Son, H. F., Kim, S., Ahn, J. W. & Kim, K. J. Crystal structure and biochemical characterization of β-keto thiolase B from polyhydroxyalkanoate-producing bacterium *Ralstonia eutropha* H16. *Biochem. Biophys. Res. Commun.* **444**, 365–369 (2014).
45. Jez, J. M. et al. Structural control of polyketide formation in plant-specific polyketide synthases. *Chem. Biol.* **7**, 919–930 (2000).
46. Blaisse, M. R., Fu, B. & Chang, M. C. Y. Structural and biochemical studies of substrate selectivity in *Ascaris suum* thiolases. *Biochemistry* **57**, 3155–3166 (2018).
47. Clomburg, J. M. et al. Integrated engineering of β-oxidation reversal and ω-oxidation pathways for the synthesis of medium chain omega-functionalized carboxylic acids. *Metab. Eng.* **28**, 202–212 (2015).
48. Dellomonaco, C., Clomburg, J. M., Miller, E. N. & Gonzalez, R. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**, 355–359 (2011).
49. Kim, S., Clomburg, J. M. & Gonzalez, R. Synthesis of medium-chain length (C6-C10) fuels and chemicals via β-oxidation reversal in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **42**, 465–475 (2015).
50. Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y. & Nielsen, J. Microbial acetyl-CoA metabolism and metabolic engineering. *Metab. Eng.* **28**, 28–42 (2015).

Acknowledgements

We thank C. Pennington at Rice University for assistance with LC–MS analysis.

Author contributions

R.G. designed the research. Z.T., S.C. and J.M.C. performed the research. S.Q. performed structure simulation and alignment. Z.T., J.M.C. and R.G. wrote the paper.

Competing interests

The authors have filed a patent application (US patent application no. PCT/US2016/045037).

Additional information

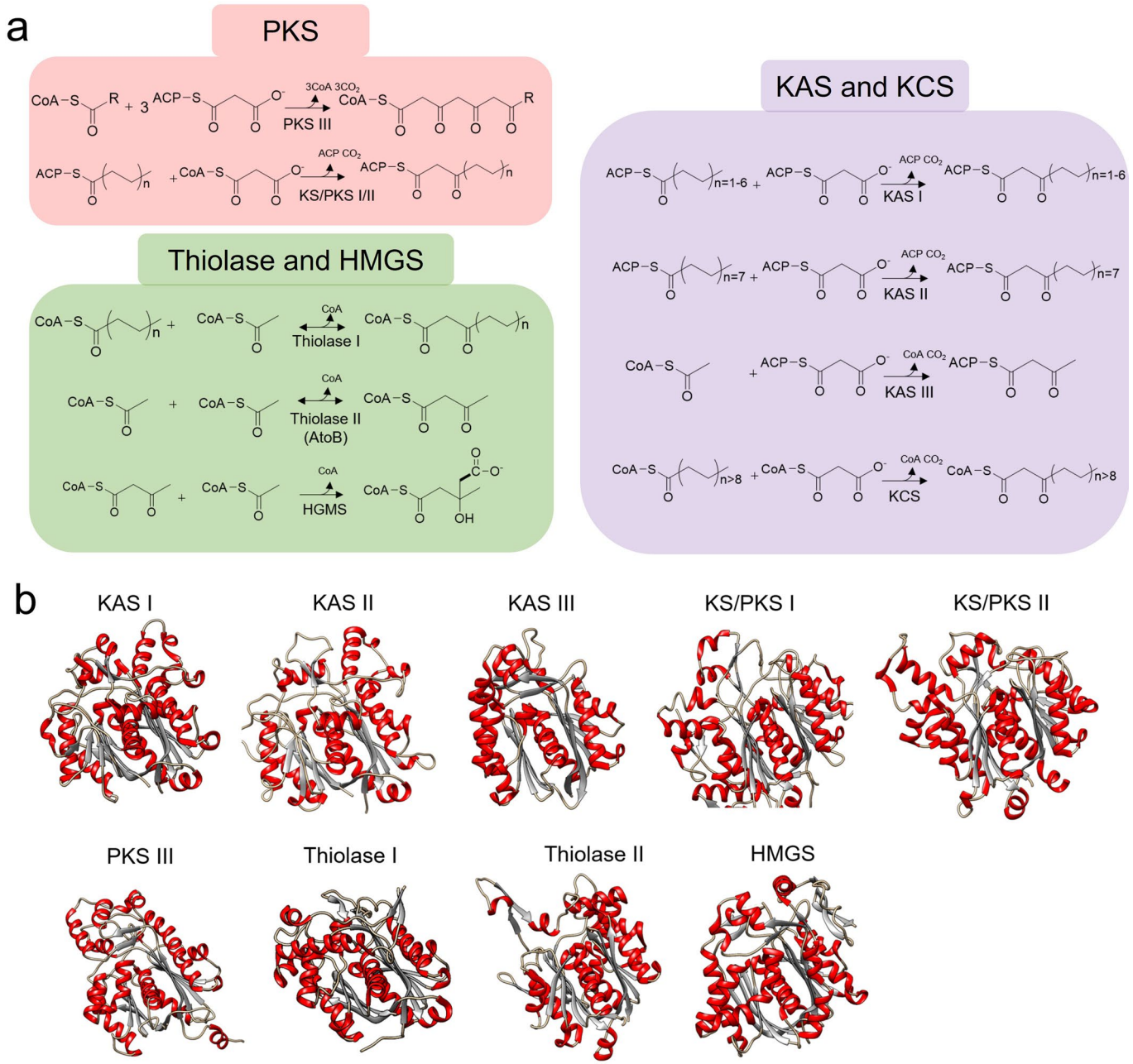
Extended data is available for this paper at https://doi.org/10.1038/s41929-020-0471-8.

Supplementary information is available for this paper at https://doi.org/10.1038/s41929-020-0471-8.

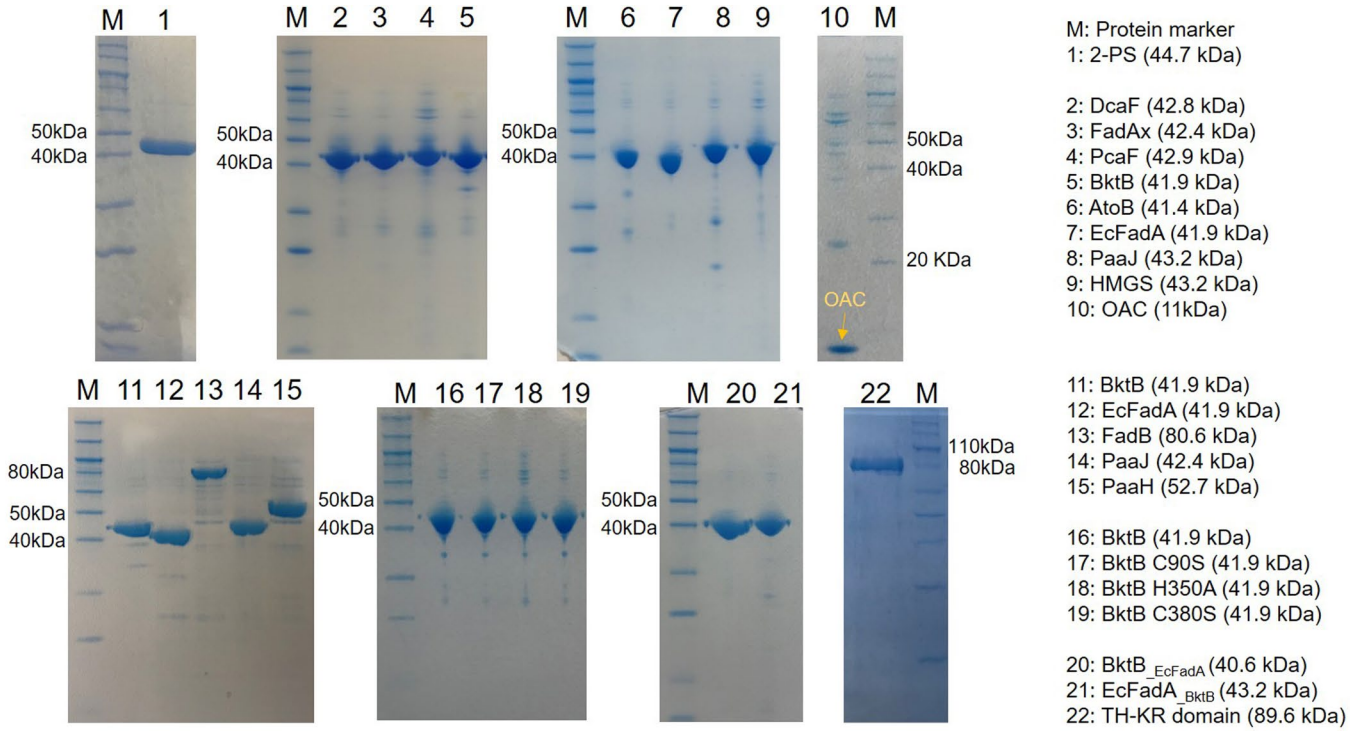
Correspondence and requests for materials should be addressed to R.G.

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

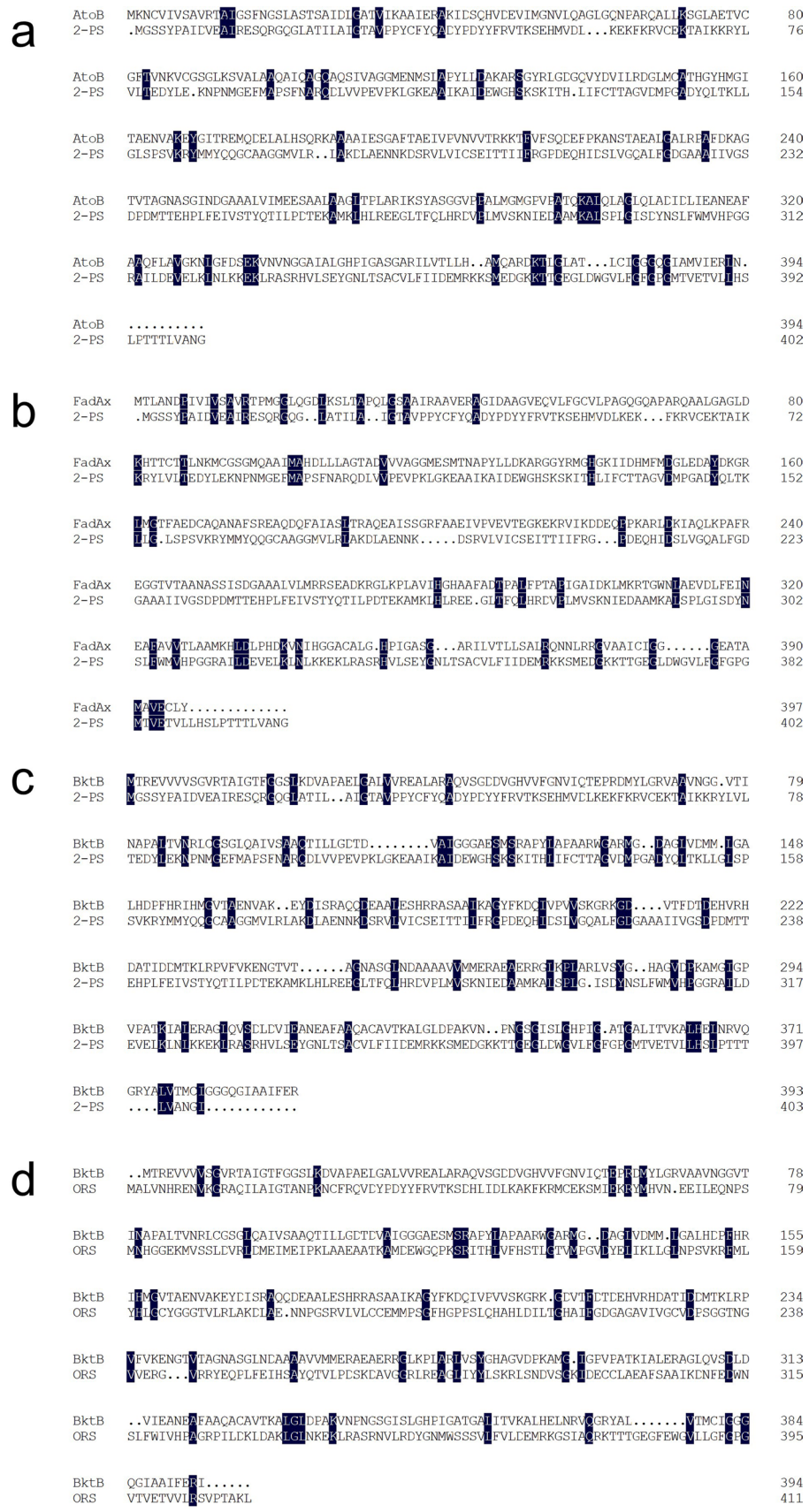
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Extended Data Fig. 1 | Reactions catalysed by members of the thiolase superfamily and structures of corresponding enzymes. a, Reactions catalysed by members of the thiolase superfamily; (**b**) Three-dimensional structures of corresponding enzymes (monomer). The structures shown are those of *Medicago sataiva* chalcone synthase PKS III (1bi5), *Streptomcyes albus* ketosynthase domain of polyketide synthase KS/PKS I (4wky), *Streptomyces coelicolor* KS/PKS II (1tqy), *E. coli* β-ketoacyl-acyl-carrier-protein synthase KAS I (2vb8), *E. coli* KAS II (1kas), *E. coli* KAS III (1hn9), *Arabidopsis thaliana* KCS (modelling structure), human mitochondria thiolase I (4c2j), *E. coli* thiolase II (5f0v), and *Bacillus subtilis* HMGS (4yxq).

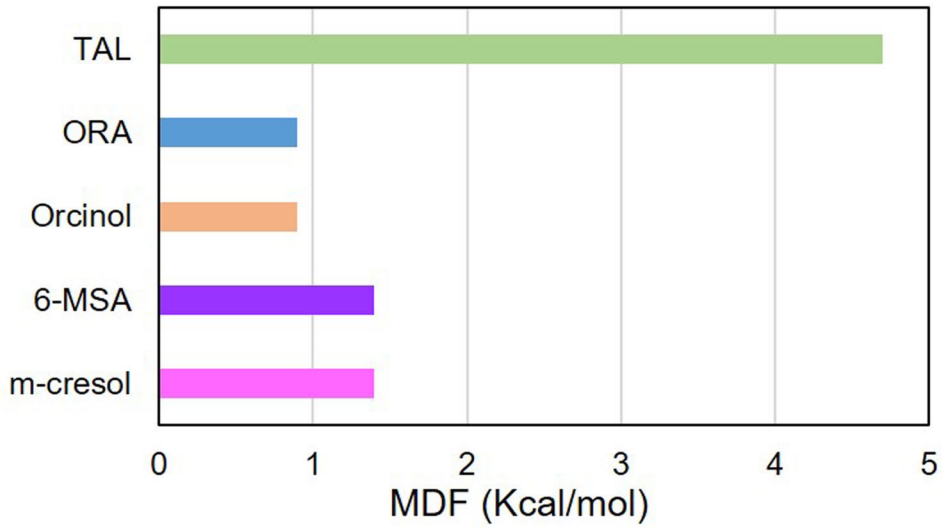


Extended Data Fig. 2 | Purified enzymes used in this study. The sources of enzymes are as follows. PKS III 2-PS was from *Gerbera hybrid*; DcaF was from *Acinetobacter* sp. ADP1; FadAx was from *Pseudomonas putida*; PcaF was from *Pseudomonas putida*; BktB was from *Ralstonia eutropha*; AtoB was from *E. coli*; EcFadA was from *E. coli*; PaaJ was from *E. coli*; HMGS was from *Staphylococcus aureus*; FadB was from *E. coli*; PaaH was from *E. coli*; The expected molecular weights of enzymes with N terminal 6 x His-tag calculated by tool from ExPASy, (https://web.expasy.org/compute_pi/) are indicated in parentheses. BktB (5#, 11#, 16#), PaaJ (8#, 14#) and EcFadA (7#, 12#) appeared more than once as we initially employed them for *in vitro* TAL biosynthesis tests (Fig. 2), and then we further purified and used them for the *in vitro* assay of ketoacyl-CoA thiolase, 3-ketoadipyl-CoA thiolase and PKT activities (see Fig. 6a). For EcFadA WT (7#, 12#) and EcFadA$_{_BktB}$ (21#), the expected MWs from ExPASy calculations are 41.9 kDa and 43.2 kDa, while actual sizes on SDS-PAGE appear ~40 kDa and ~41 kDa.

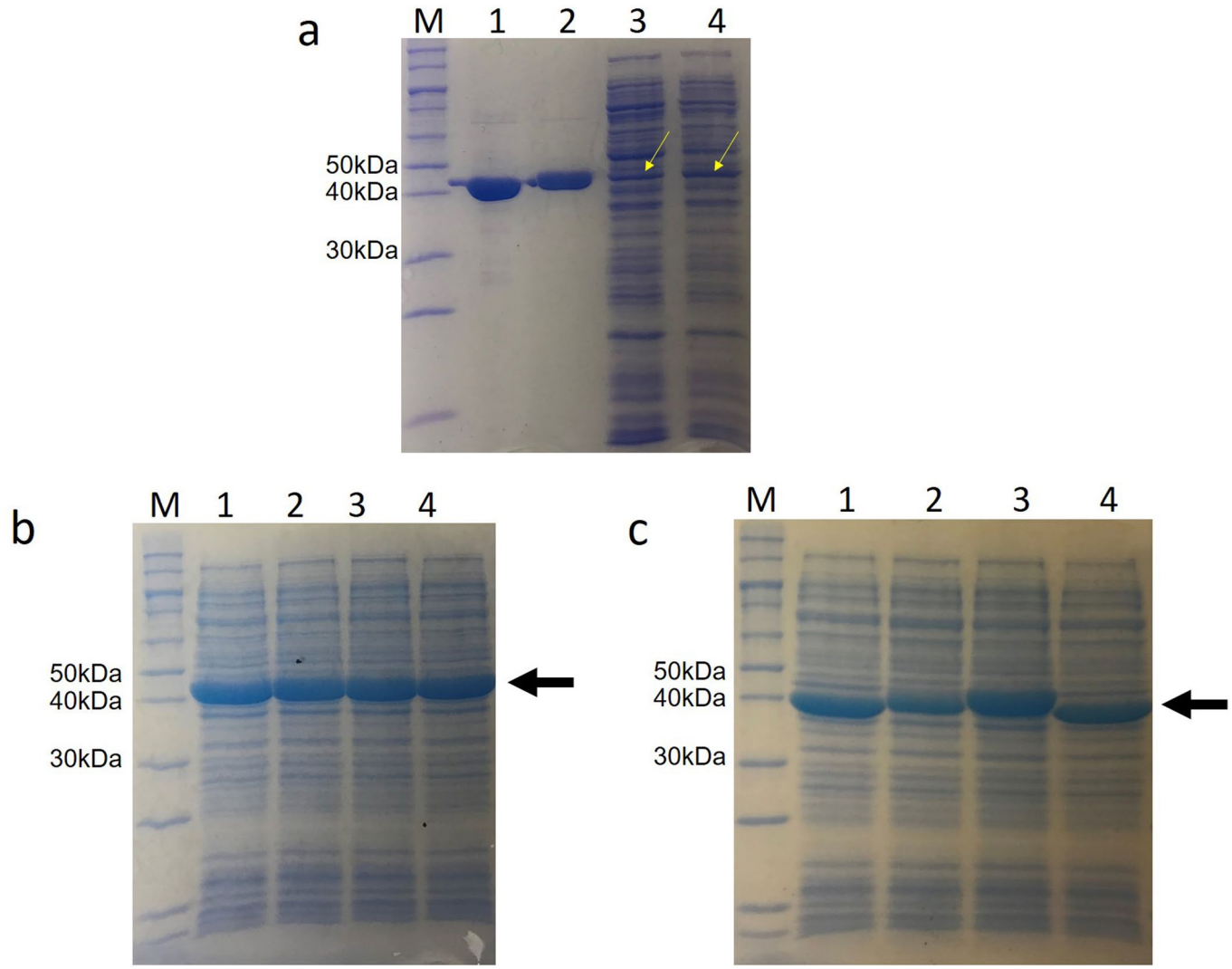


Extended Data Fig. 3 | Primary sequence alignment of thiolases and PKS III. a, Primary sequence alignment of PKS III 2-PS (2-pyrone synthase from *Gerbera hybrid*) with AtoB (from *E. coli*). The sequence identity is only 10.0%; (**b**) Primary sequence alignment of PKS III 2-PS with FadAx (from *Pseudomonas putida*). The sequence identity is only 8.7%; (**c**) Primary sequence alignment of PKS III 2-PS with BktB (from *Ralstonia eutropha*). The sequence identity is only 13.6%. **d**, Primary sequence alignment of PKS III ORS (from *Rhododendron dauricum*) with BktB. The sequence identity is only 9.9%.

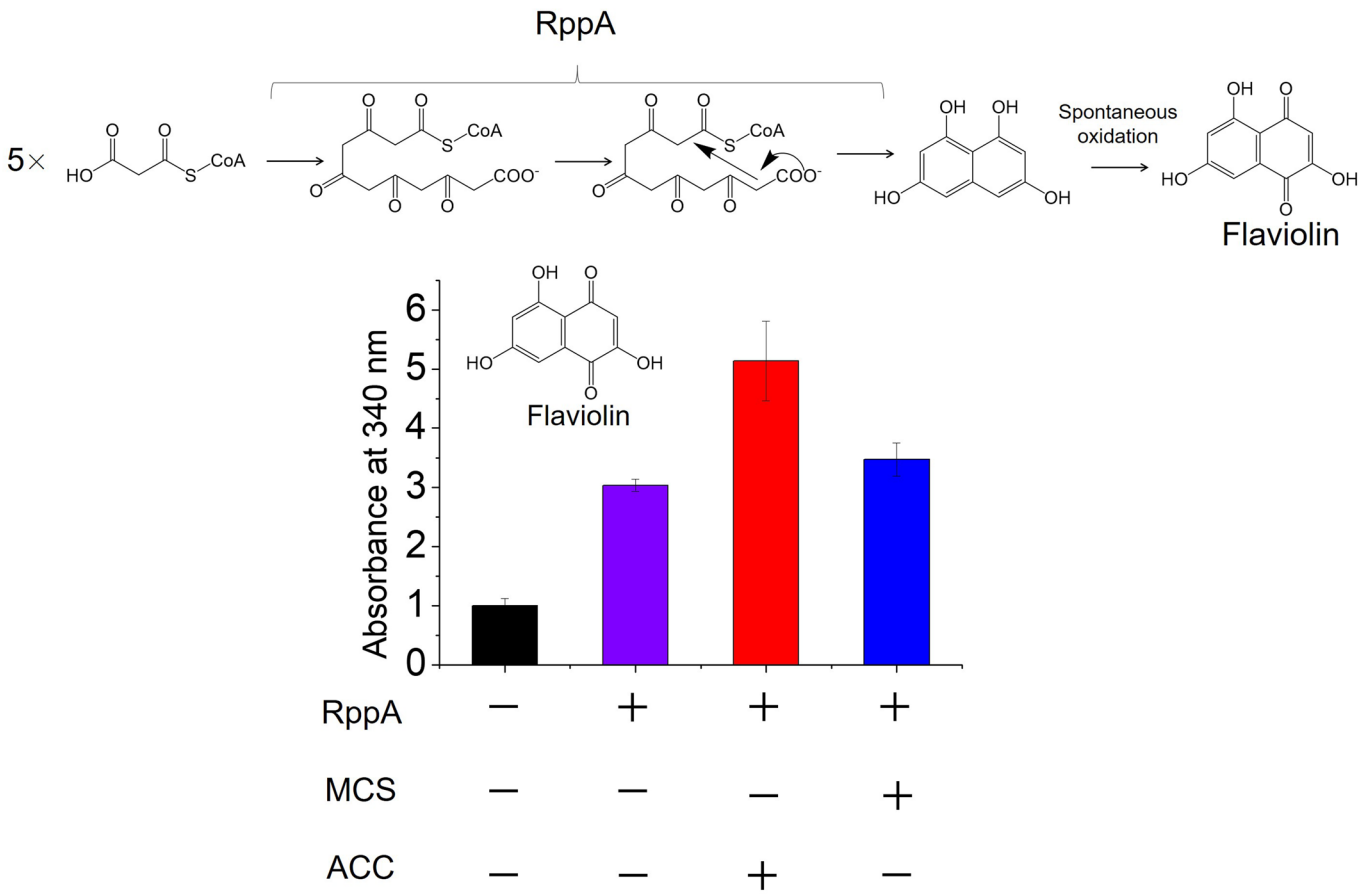
Product	Proposed PKT-based pathway
TAL	3 acetyl-CoA = triacetic acid lactone + 3 CoA
ORA	4 acetyl-CoA = orsellinic acid + 4 CoA
Orcinol	4 acetyl-CoA = orcinol + 4 CoA + CO_2
6-MSA	4 acetyl-CoA + NADPH + H^+ = 6-MSA + 4 CoA + $NADP^+$ + H_2O
m-cresol	4 acetyl-CoA + NADPH + H^+ = m-cresol + 4 CoA + $NADP^+$ + CO_2 + H_2O



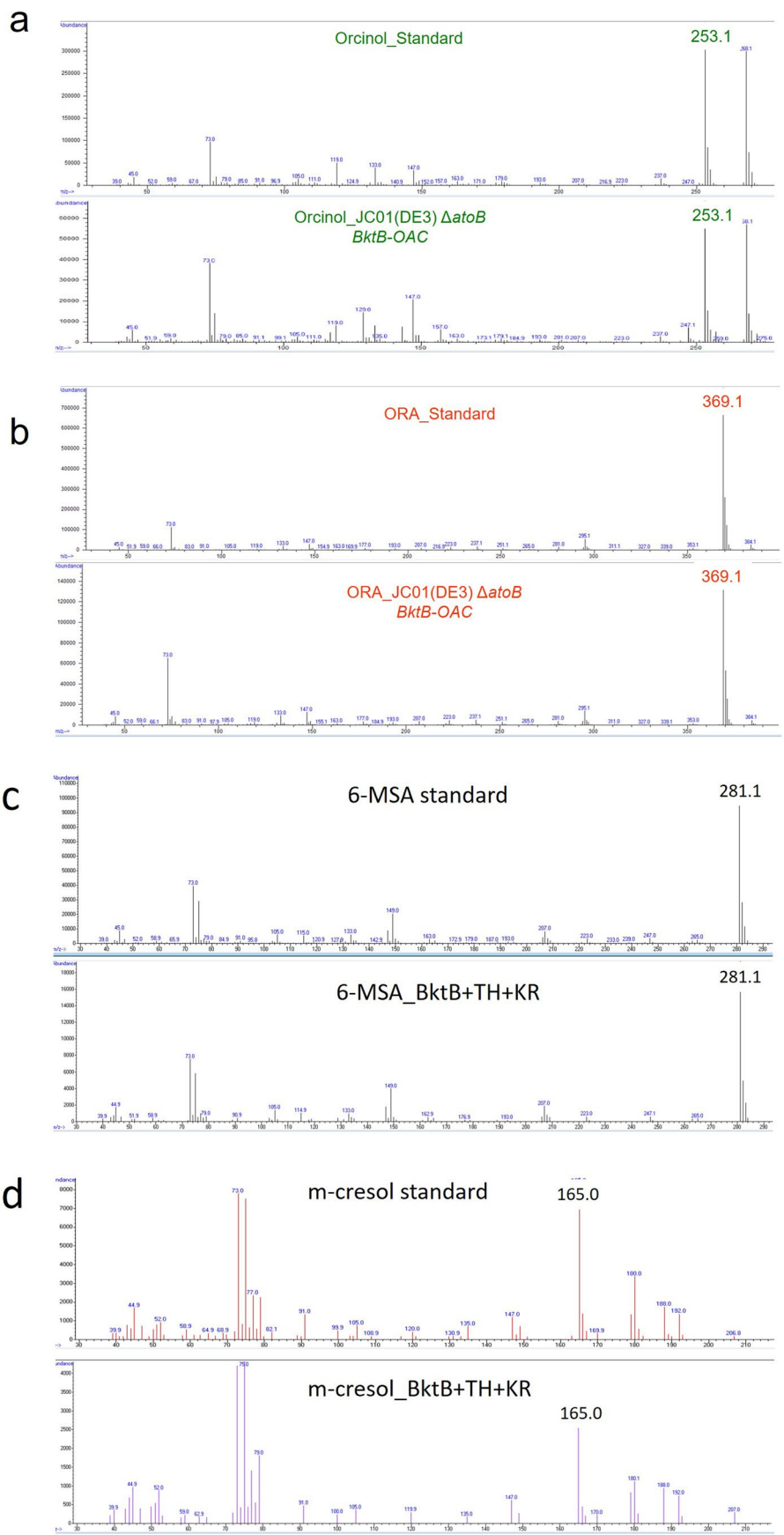
Extended Data Fig. 4 | Max-min Driving Force (MDF) for the synthesis of specified polyketides from acetyl-CoA through the proposed PKT-based pathway. MDF calculations assume minimum and maximum metabolite concentrations of 0.000001 and 0.01 M, respectively, and a fixed NADPH/NADP + ratio of 10.



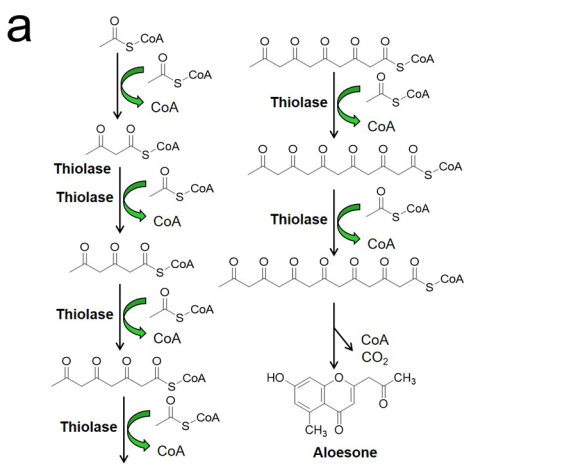
Extended Data Fig. 5 | Summary of SDS-PAGE analysis of samples in this study. a, SDS-PAGE analysis of the soluble expression of BktB and 2-PS in JC01 (DE3) Δ*atoB* host strain. M: marker; 1: purified BktB; 2: purified 2-PS; 3: soluble fraction of JC01 (DE3) Δ*atoB* expressing pCDF-*bktB*, the arrow indicates the soluble BktB protein; 4: soluble fraction of JC01 (DE3) Δ*atoB* harbouring pCDF-*2-ps*, the arrow indicates the soluble 2-PS protein. **b**, SDS-PAGE analysis of the soluble expression of BktB WT and C90S, H350A, C380S mutants. M: marker; 1: soluble fraction of BL21 (DE3) expressing BktB WT; 2: soluble fraction of BL21 (DE3) expressing BktB C90S; 3: soluble fraction of BL21 (DE3) expressing BktB H350A; 4: soluble fraction of BL21 (DE3) expressing BktB C380S. **c**, SDS-PAGE analysis of the soluble expression of EcFadA and BktB mutants after segment swapping. M: marker; 1: soluble fraction of BL21 (DE3) expressing EcFadA WT; 2: soluble fraction of BL21 (DE3) expressing $EcFadA_{\_BktB}$; 3: soluble fraction of BL21 (DE3) expressing BktB WT; 4: soluble fraction of BL21 (DE3) expressing $BktB_{\_EcFadA}$.



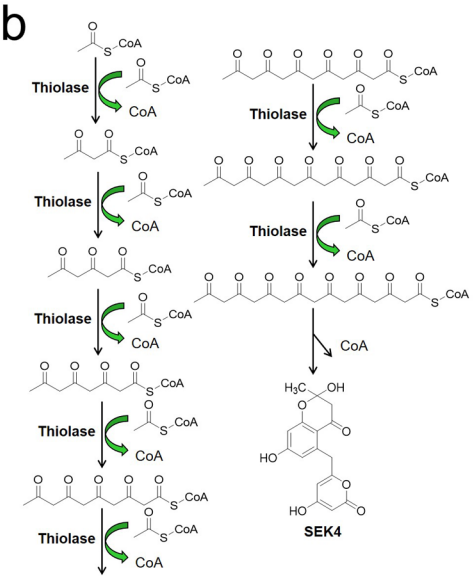
Extended Data Fig. 6 | Assessment of malonyl-CoA availability through the flaviolin biosynthesis pathway. Upper, flaviolin biosynthesis pathway: 5 molecules of malonyl-CoA are condensed by RppA to form flaviolin, which has a specific absorbance at 340 nm. Bottom, RppA was expressed with different MCS/ACC enzymes for characterization of malonyl-CoA availability in strain JC01 (DE3) Δ*atoB*. Engineered *E. coli* strains were cultured in LB-like MOPS medium + 2% (wt/v) glycerol. For strains harbouring MCS, 12 mM malonate sodium was also included. Error bars represent s.d. calculated from at least three biological replicates.

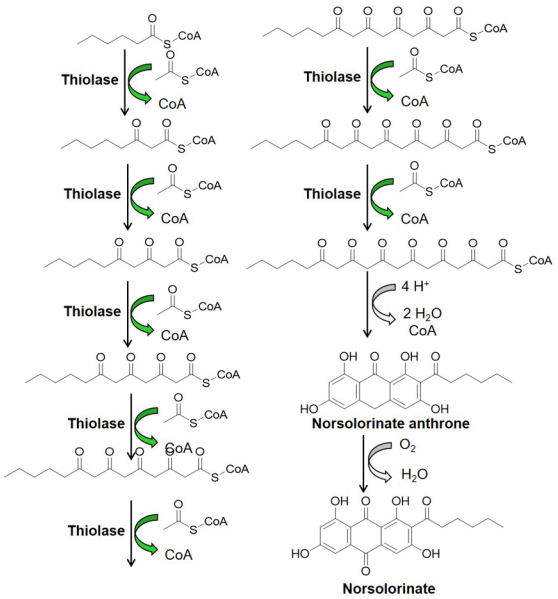


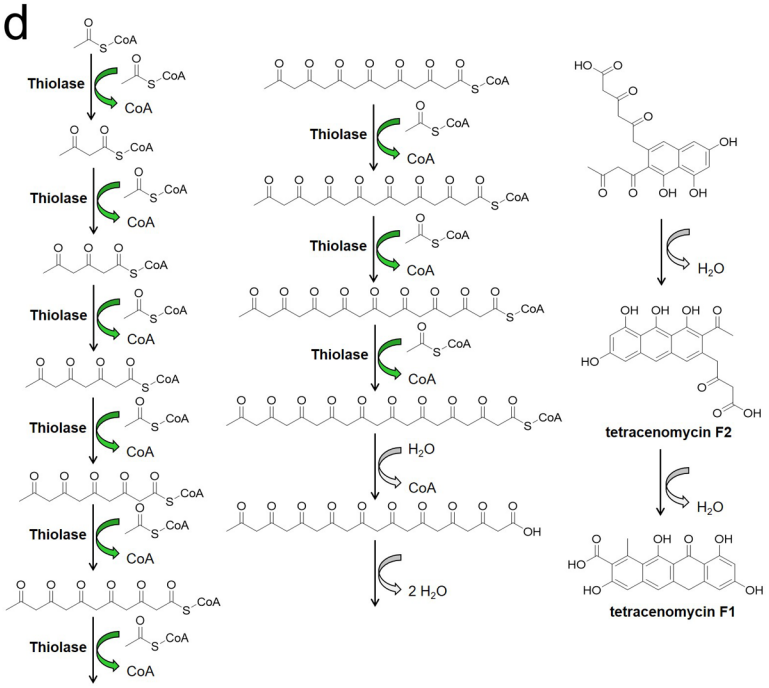
Extended Data Fig. 7 | Mass spectrometry (MS) information of the samples. a, MS information of orcinol standard and orcinol sample extracted from *in vivo* samples; (**b**) MS information of ORA standard and ORA sample extracted from *in vivo* samples; (**c**) MS of 6-MSA standard and 6-MSA sample extracted from *in vitro* samples. **d**, MS of *m*-cresol standard and *m*-cresol sample extracted from *in vitro* samples.



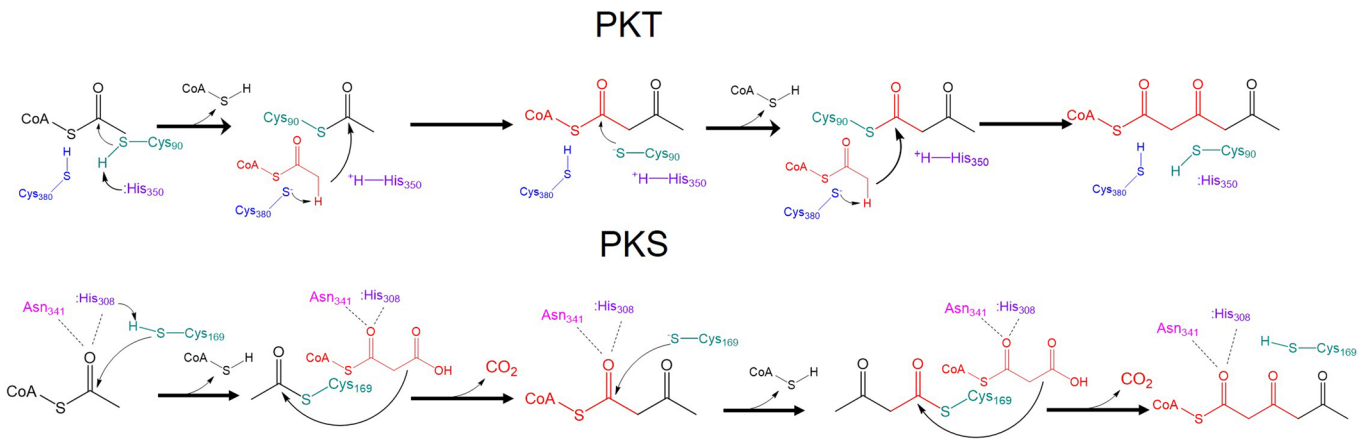
Biosynthesis from acetyl-CoA	Standard $\Delta_r G'^{\circ}$ (kcal/mol)
aloesone	-28.9
SEK4	-9.4
norsolorinate	-116.6
tetracenomycin F1	-96.1



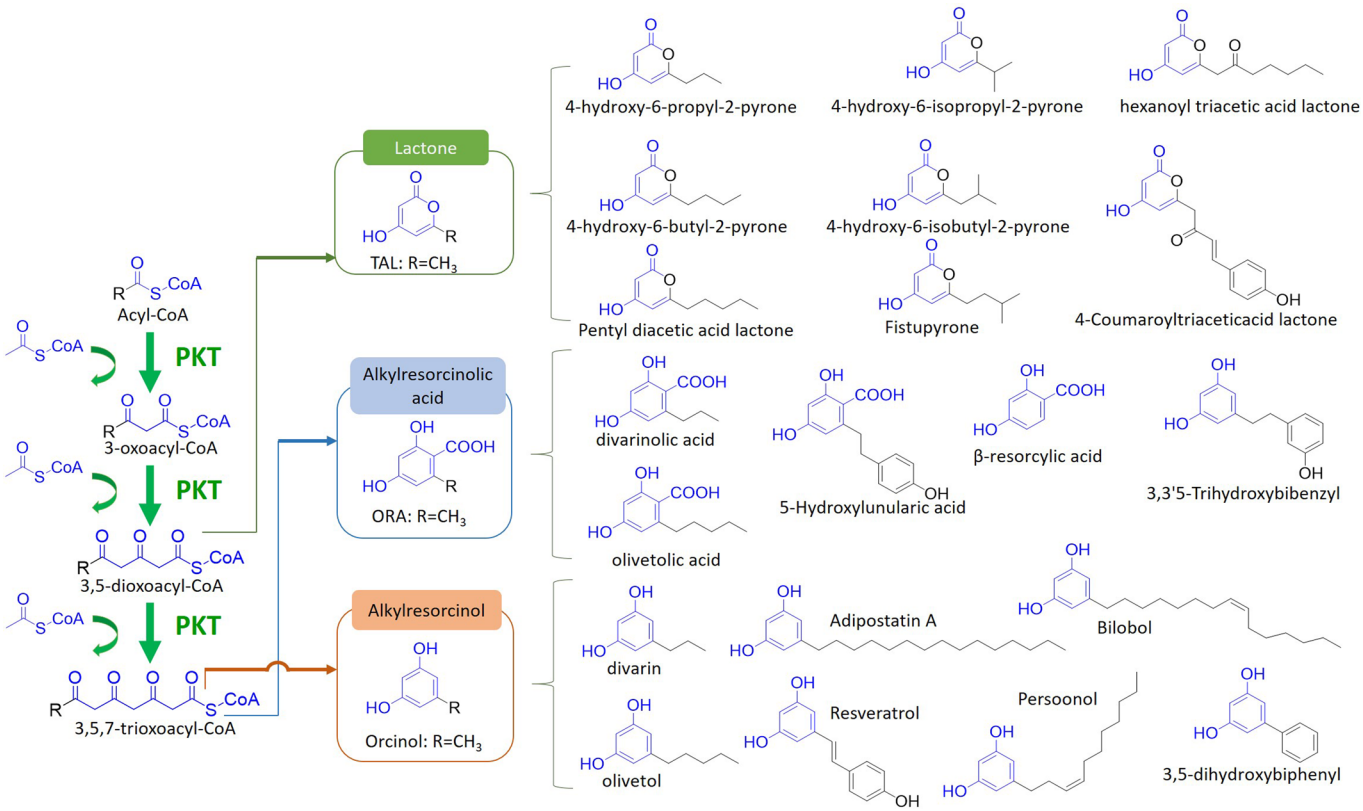




Extended Data Fig. 8 | The proposed PKT-based platforms for aloesone, octaketide SEK4, norsolorinate and tetracenomycin F1 biosynthesis. a, The proposed PKT-based platform for aloesone biosynthesis. **b**, The proposed PKT-based platform for SEK4 biosynthesis. **c**, The proposed PKT-based platform for norsolorinate biosynthesis. **d**, The proposed PKT-based platform for tetracenomycin F1 biosynthesis.



Extended Data Fig. 9 | Catalytic mechanism of polyketoacyl-CoA thiolase (PKT) and representative PKS III 2-PS. Cys/His/Cys is the catalytic triad of PKT activity, enabling a two-step, ping-pong mechanism for the thiolytic condensation reaction using substrates with multiple keto-groups (for example acetoacetyl-CoA). Notably, this catalytic mechanism is distinct from that of PKS III, where the core catalytic triad is Cys/His/Asn, with Asn (asparagine) playing a critical role in catalysing malonyl-CoA decarboxylation and stabilizing the condensation transition state.



Extended Data Fig. 10 | Exploiting PKT to synthesize additional PK backbones and corresponding PKs (R = groups other than CH_3). Synthesis of polyketide backbones for representative lactone, alkylresorcinolic acid, and alkylresorcinol polyketide families through up to 3 rounds of iterative non-decarboxylative Claisen condensations catalyzed by PKTs.

nature research

Corresponding author(s): Ramon Gonzalez

Last updated by author(s): Apr 10, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected using: Agilent MassHunter GC/MS Acquisition B.07.05.2479 (GC-FID/MS), Shimadzu EZStart Version 7.3 Build 18 (HPLC-RID), BioTek Gen5 v2.09.2 (Plate reader), BLASTp suite (NCBI).
Data analysis	Compute pI/Mw tool of ExPASy, DNAMAN 7.0, DynaMut, Chimera 1.13.1 (UCSF), SWISS-MODEL Server SMTL.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data supporting the findings of this study are included in the paper and its supplementary files.

Field-specific reporting

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

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

Life sciences study design

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

Sample size	Sample size of at least three was chosen following all previous publications in similar field.
Data exclusions	No data were excluded from the analyses.
Replication	Replicate experiments were successful.
Randomization	All strains for in vivo culture were randomly picked from the LB with antibiotics agar plate.
Blinding	Investigators were not blinded to strains' genotypes during experiments. All in vivo and in vitro data were analyzed and checked by multiple authors and reviewed by the corresponding author.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging