



# Engineering controllable alteration of malonyl-CoA levels to enhance polyketide production

Received: 25 March 2024

Accepted: 11 April 2025

Published online: 11 June 2025

Sarah H. Klass<sup>1,2,3</sup>, Mia Wesselkamper<sup>4</sup>, Aidan E. Cowan<sup>1,5</sup>, Namil Lee<sup>1,2,6</sup>, Nathan Lanclos<sup>1,4</sup>, Seokjung Cheong<sup>1,3,6</sup>, Zilong Wang<sup>1,2,3</sup>, Yan Chen<sup>1,2</sup>, Jennifer W. Gin<sup>1,2</sup>, Christopher J. Petzold<sup>1,2</sup> & Jay D. Keasling<sup>1,2,3,4,7</sup>✉

Check for updates

Heterologous expression of polyketide synthase (PKS) genes in *Escherichia coli* has enabled the production of various valuable natural and synthetic products. However, the limited availability of malonyl-CoA (M-CoA) in *E. coli* remains a substantial impediment to high-titer polyketide production. Here we address this limitation by disrupting the native M-CoA biosynthetic pathway and introducing an orthogonal pathway comprising a malonate transporter and M-CoA ligase, enabling efficient M-CoA biosynthesis under malonate supplementation. This approach substantially increases M-CoA levels, enhancing fatty acid and polyketide titers while reducing the promiscuous activity of PKSs toward undesired acyl-CoA substrates. Subsequent adaptive laboratory evolution of these strains provides insights into M-CoA regulation and identifies mutations that further boost M-CoA and polyketide production. This strategy improves *E. coli* as a host for polyketide biosynthesis and advances understanding of M-CoA metabolism in microbial systems.

The expression of native or hybrid PKSs in *Escherichia coli* offers a wide range of advantages and opportunities for diverse applications. In their natural hosts, PKSs produce valuable polyketide-derived natural products, including antibiotics, anticancer agents, immunosuppressants and other bioactive compounds<sup>1</sup>. Type III PKSs perform iterative decarboxylative Claisen condensation of acyl-CoA substrates<sup>2</sup>, whereas type I PKSs consist of multiple modular domains responsible for incorporating specific acyl-CoA derivatives into the growing polyketide chain<sup>3</sup>. Hybrid PKSs, combining modules or domains with distinct substrate specificities derived from various natural biosynthetic gene clusters<sup>4</sup>, can be rationally assembled to biosynthesize diverse and complex structures, accessing a broader chemical space than what is achievable by chemical synthesis alone<sup>5–7</sup>. Moreover, hybrid PKSs can be tailored to be compatible with specific microbial chassis, such as *E. coli*, enabling

industrial-scale production in pharmaceuticals, agriculture and biomaterials, exemplified by the production of 6-deoxyerythronolide<sup>8</sup>, resveratrol<sup>9</sup>, naringenin chalcone<sup>10</sup> and various lactones<sup>7,11,12</sup>.

Unlike microbial hosts that naturally produce PKSs, such as those from the *Streptomyces* genus<sup>8,13</sup>, *E. coli* is not a natural PKS host and, despite its versatility, faces a number of engineering challenges, including increased metabolic burden upon protein expression, lack of native phosphopantetheinyl transferases (PPTases) for PKS activation and limited acyl-CoA availability. Strategies to enhance PKS expression and activity include genetic code optimization<sup>14</sup>, improved hybrid PKS junctions<sup>5</sup> and tuning transcription and translation by optimizing promoter strength<sup>15</sup> and gene copy number<sup>16,17</sup>. However, even the most functional PKSs will fail to produce the desired product if the necessary CoA substrates are limited. In *E. coli*, one of the most commonly used

<sup>1</sup>Joint BioEnergy Institute, Emeryville, CA, USA. <sup>2</sup>Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

<sup>3</sup>Department of Chemical & Biomolecular Engineering, University of California, Berkeley, Berkeley, CA, USA. <sup>4</sup>Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA. <sup>5</sup>Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. <sup>6</sup>QB3, University of California, Berkeley, Berkeley, CA, USA. <sup>7</sup>Centre for Biosustainability, Danish Technical University, Lyngby, Denmark. ✉e-mail: [keasling@berkeley.edu](mailto:keasling@berkeley.edu)

PKS substrates, malonyl-CoA (M-CoA), is tightly regulated and primarily used by fatty acid (FA) biosynthesis, leading to a relatively small pool of free M-CoA and low production of M-CoA-based polyketides<sup>18</sup>.

Many strategies for enhancing the substrate pool of M-CoA have been established, as it has been identified as a clear bottleneck limiting the production of M-CoA-derived products such as polyketides and FAs<sup>19</sup>. Some methods have focused on increasing the availability of acetyl-CoA (Ac-CoA), a precursor for M-CoA, and/or overexpressing or supplementing enzymes responsible for converting Ac-CoA to M-CoA, such as the Ac-CoA carboxylase (ACC) complex<sup>20–23</sup>. Other genetic loci that are implicated in the diversion of carbon away from M-CoA have been targeted by CRISPRi to repress expression and successfully shift metabolic flux toward M-CoA synthesis<sup>24</sup>. Alternatively, exogenous supplementation of the potent FA synthase (FAS) inhibitor cerulenin is used to divert M-CoA toward polyketide production; however, cerulenin also covalently binds to the ketosynthase domains of PKSs in a similar manner to FAS, resulting in an inefficient solution to increase polyketide titers<sup>25,26</sup>. Finally, several methods to bypass the endogenous route to M-CoA via the ACC complex have been developed by expressing various malonate assimilation pathways, which have been shown to increase M-CoA levels and their subsequent incorporation into polyketides<sup>27–30</sup>.

Here, we engineer two PKS-compatible *E. coli* strains, K207-3 and BAP1, to enable a simple and controlled strategy of increasing M-CoA levels and various downstream PKS products. Both strains contain the PPTase gene required for PKS activation, *sfp*, from *Bacillus subtilis* and the native *prpE*, whose gene product converts propionate into propionyl-CoA, under an inducible T7 promoter. K207-3 also contains the genes encoding the propionyl-CoA carboxylase from *Streptomyces coelicolor*, enabling the conversion of propionyl-CoA to methylmalonyl-CoA (mM-CoA)<sup>8,31</sup>.

To provide exogenous control over M-CoA levels, we integrate genes for the widely used<sup>30,32–35</sup> malonate assimilation pathway consisting of the malonate importer, *matC*, and the malonate:CoA-ligase, *matB*, into both genomes, resulting in a plasmid-free strain whose M-CoA and mM-CoA levels are directly tunable by exogenous addition of malonate and propionate, respectively. Furthermore, we achieve complete control over the M-CoA pool by inhibiting the endogenous M-CoA biosynthesis by removing *bioH*, whose gene product is required for biotin biosynthesis<sup>36,37</sup> and, consequently, for activating the ACC complex that normally converts Ac-CoA to M-CoA and, thus, is necessary for cell growth<sup>38</sup>. By introducing the orthogonal MatBC pathway into this auxotrophic strain, the exogenous addition of malonate generates M-CoA and rescues cell growth. Similarly, malonate importers have been used in other biotin-independent *E. coli* strains to allow for the overexpression of the biotin-sequestering and, thus, toxic protein streptavidin<sup>39</sup>. Leveraging this malonate-dependent growth phenotype, we perform adaptive laboratory evolution (ALE), leading to evolved strains that accumulate mutations that increase M-CoA and PKS titers relative to the nonmutagenized strain. Among these are mutations previously identified using methods such as CRISPRi<sup>24</sup>, biosensor-guided random transposon libraries<sup>40</sup> or metabolic modeling<sup>23</sup> to improve M-CoA, FA or polyketide titers, in addition to several mutations in related pathways that have not yet been explored in M-CoA or PKS engineering strategies.

This work highlights the power of combining orthogonal control of metabolic pathways with ALE to enhance PKS precursor supply and polyketide titers while also deepening our understanding of M-CoA metabolism in microbial systems.

## Results

### Evaluation of malonate transporters in *E. coli*

Plasmids encoding the malonate transporters MadLM<sup>41</sup>, MdcF<sup>42</sup> and MatC<sup>35</sup> were examined for their ability to import malonate into *E. coli* and convert it to M-CoA by MatB. Production of M-CoA was quantified using the type III PKS1,3,6,8-tetrahydroxynaphthalene (THN) synthase

(RppA), which converts M-CoA to THN, spontaneously forming flaviolin. Flaviolin, a red compound, has distinct absorbance peaks at 340 nm and 520 nm and has served as an indirect indicator of intracellular M-CoA levels available for PKS biosynthesis<sup>43,44</sup>. The arabinose-inducible, highly active RppA variant plasmid (pBADT-RppA-NT)<sup>44</sup> was cotransformed with each of the plasmids pCKmatBC, pBba2c-MatB-MadLM or pBbA2c-MatB-MdcF into BAP1, generating strains BAP1-pMatBC-RppA, BAP1-pMatB-MadLM-RppA and BAP1-pMatB-MdcF-RppA, respectively. Induction with anhydrotetracycline (aTc) caused growth toxicity in MadLM and MdcF, whereas no such toxicity was observed for MatC. These strains were cultured without aTc to mitigate toxicity, relying on leaky expression. Under these conditions, all transporters increased flaviolin production with higher malonate concentrations (Supplementary Fig. 1). MatC was selected for further experiments owing to its lower apparent toxicity.

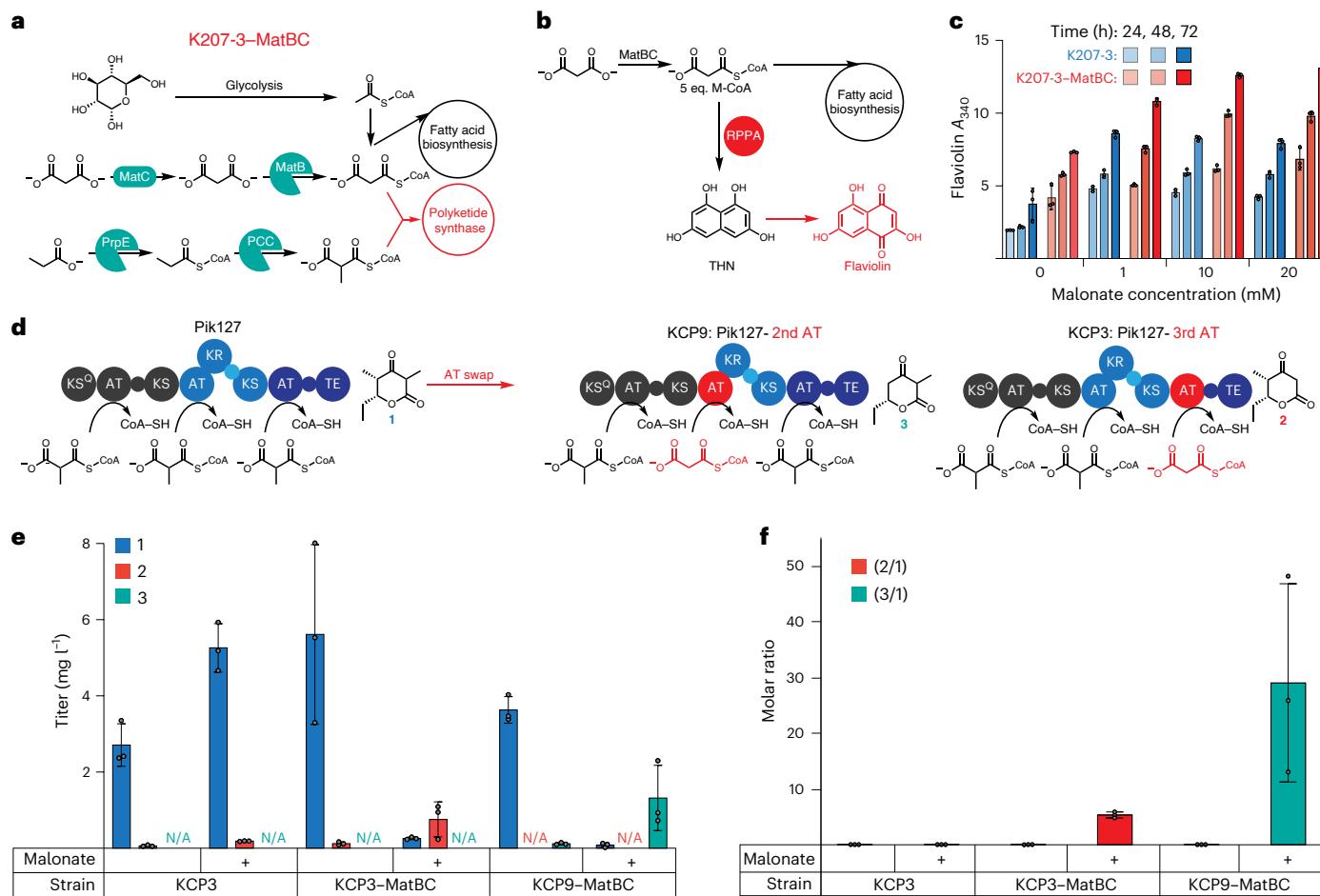
### Engineering controllable levels of M-CoA

The *matC* and *matB* genes from *Rhizobium trifoli* under the control of a *lacUV5* promoter were integrated via homologous recombination into the intergenic region downstream of *ompW*, a safe site for genomic integration and expression of recombinant proteins<sup>45</sup>, creating K207-3-MatBC. This genomic integration enabled tunable increases in M-CoA and mM-CoA levels by adding malonate and propionate to the medium, respectively (Fig. 1a). Similar to the plasmid-based system, K207-3 and K207-3-MatBC were transformed with pBADT-RppA-NT and monitored for their flaviolin production over 3 days in Lysogeny broth (LB) medium (10 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), 0.2% arabinose and 0–20 mM malonate). K207-3-MatBC consistently produced more flaviolin than K207-3 across all malonate concentrations and a 70% increase ( $P = 0.0002$ ) at 20 mM malonate. The increase in flaviolin titers correlated with malonate levels, demonstrating tunable enhancement of polyketide production (Fig. 1b).

### Production of polyketides derived from mM-CoA and M-CoA

PKSs use acyltransferase domains to select CoA substrates to be incorporated into the growing polyketide chain. Replacing native acyltransferase domains with those that use alternative acyl-CoAs can modify the final compound synthesized. However, previous attempts discussed in ref. 46 to replace the mM-CoA-specific acyltransferase domain of the first module (M1) of a shortened pikromycin synthase (Pik127)<sup>47</sup> with an M-CoA-specific acyltransferase in a hybrid version resulted in the formation of both the mM-CoA and M-CoA-derived products, with the mM-CoA product being in excess. This was hypothesized to result from reduced acyltransferase activity or ketosynthase gatekeeping favoring mM-CoA elongation<sup>46,48</sup>. The full pikromycin PKS domain structure details are illustrated in Supplementary Fig. 2.

In addition to these factors, we hypothesized that the discrepancy in expected versus actual product might be due to an overabundance of mM-CoA during polyketide production in K207-3. To test this, *E. coli* strains K207-3 and K207-3-MatBC were transformed with plasmids encoding hybrid Pik127 variants with M-CoA-specific acyltransferase from pikromycin module 3 replacing native domains in module 2 (Pik127-2nd acyltransferase encoded on plasmids pBbS5k\_pik12(pik\_AT3) and pBbA5a\_pik27) or module 3 (Pik127-3rd acyltransferase encoded on plasmids pBbS5k\_pik12 and pBbA5a\_pik27(pik\_AT3)), resulting in the strains KCP9, KCP9-MatBC, KCP3 and KCP3-MatBC, respectively. Notably, the Pik127-3rd acyltransferase construct reinstalls the native ketosynthase–acyltransferase junction and should not suffer ketosynthase gatekeeping. Strains were cultured in EZ-rich medium with propionate (to generate mM-CoA) and ±malonate (to generate M-CoA). Liquid chromatography–mass spectrometry (LC-MS) analysis was performed to quantify the amount of 6-ethyl-3,5-dimethylidihydro-2H-pyran-2,4(3H)-dione (product 1), 6-ethyl-5-methylidihydro-2H-pyran-2,4(3H)-dione (product 2) and 6-ethyl-3-methylidihydro-2H-pyran-2,4(3H)-dione (product 3), and triacetic



**Fig. 1 | Schematic depiction of relevant pathways for MatBC-engineered *E. coli* strains.** **a**, Integration of the MatBC pathway into K207-3 resulted in K207-3-MatBC, enabling modulation of M-CoA and mM-CoA pools by supplementing malonate and propionate, respectively. **b**, Pathway for flaviolin biosynthesis from M-CoA as a proxy for intracellular M-CoA levels. **c**, Absorbance measurements of the culture supernatant from three biological replicates at 340 nm (path length corrected to 0.1 cm<sup>-1</sup>). Maximum absorbance for K207-3-MatBC was observed at 72 h with 20 mM malonate supplementation ( $A_{340} = 13.09 \pm 0.40$ ), representing a 70% increase compared to K207-3 under the same conditions. **d**, Schematic depicting the predicted substrate specificities for Pik127 constructs with the acyltransferase (AT) domain from pikromycin module 3 exchanged at the second and third extension modules. **e**, Titers from three biological replicates of M-CoA-incorporated products 2 and 3 ( $m/z = 155$ ) and the mM-CoA-incorporated product 1 ( $m/z = 169$ ). For the Pik127-3rd acyltransferase construct, product 2

reached a maximum titer of  $0.96 \pm 0.12$  mg l<sup>-1</sup> in KCP3-MatBC supplemented with 20 mM malonate, a 14.8-fold relative to KCP3 without malonate ( $P = 0.0002$ ). For the Pik127-2nd acyltransferase constructs, no product was detected in cultures of KCP9. However, in KCP9-MatBC cultures supplemented with 20 mM malonate, product 3 was observed with titers of  $1.32 \pm 0.85$  mg l<sup>-1</sup>, 33-fold compared to the no-malonate condition, although the significant error prevented statistical significance ( $P = 0.0718$ ). **f**, Molar ratios from three biological replicates of M-CoA-incorporated products (2 and 3) to the mM-CoA-incorporated product (1) were significantly increased in malonate-supplemented KCP3-MatBC and KCP9-MatBC cultures. Ratios of 2/1 (159-fold,  $P < 0.0001$ ) and 3/1 (616-fold,  $P = 0.0475$ ) were observed compared to no-malonate conditions. All data are presented as mean  $\pm$  s.d. Statistical comparisons between groups were performed using GraphPad data analysis software with an unpaired *t*-test, assuming a Gaussian distribution, to calculate *P* values and determine statistical significance.

acid lactone (TAL) was used as an internal standard (Supplementary Figs. 3 and 4).

For the Pik127-3rd acyltransferase strains, supplementing malonate substantially increased the desired M-CoA-derived product (2) in KCP3-MatBC, reaching  $0.96 \pm 0.12$  mg l<sup>-1</sup>, a 14.8-fold change relative to KCP3 without malonate ( $P = 0.0002$ ). The ratio of product (2) to the mM-CoA product (1) was 5.3:1, representing a 150-fold change ( $P < 0.0001$ ) over KCP3 (Fig. 1d,e).

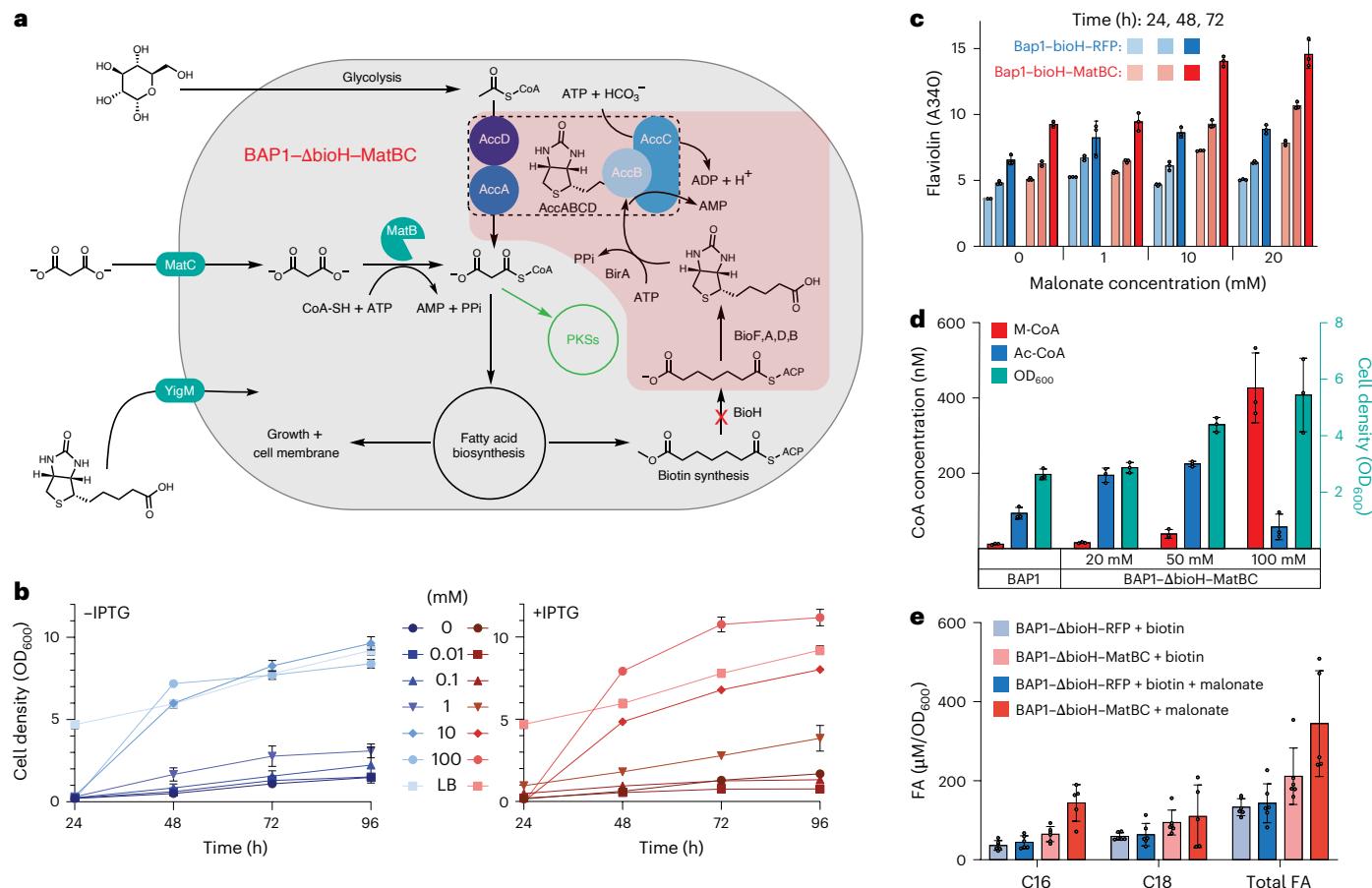
For the Pik127-2nd acyltransferase strains (KCP9 and KCP9-MatBC), products 1 and 3 were only detected in KCP9-MatBC cultures. In this strain, malonate supplementation increased the M-CoA-derived product 3 by 33-fold compared to no supplementation, reaching  $1.32 \pm 0.85$  mg l<sup>-1</sup>. The ratio of M-CoA product 3 to mM-CoA product 1 also improved significantly to 29:1, representing a 600-fold change relative to no supplementation (Fig. 1e). Exact quantification of product 3 was not possible owing to the lack of a specific chemical standard;

instead, quantification relied on the chemically similar standard 6-ethyl-5-methylidihydro-2H-pyran-2,4(3H)-dione (Supplementary Fig. 4). The larger variability in product 3 titers may result from the reduced stability of the hybrid PKS, which contains multiple unnatural junctions, leading to inconsistent protein levels and product yields.

These results suggest that although acyltransferase folding and ketosynthase gatekeeping may influence substrate preference, increasing M-CoA availability via the MatBC pathway and malonate supplementation substantially enhance the incorporation of the intended substrate, M-CoA, into the final polyketide product.

### Engineering and recovery of M-CoA auxotrophy

To precisely control M-CoA levels in *E. coli*, the endogenous M-CoA biosynthesis pathway, mediated by the biotin-activated ACC complex, was inhibited and replaced with the MatBC pathway. The *matB* and *matC* genes were integrated into the BAP1 genome, replacing the essential



**Fig. 2 | Increased polyketide production and growth rescue for malonate-supplemented MatBC strains.** **a**, Integration of *matB* and *matC* at the *bioH* locus generated the biotin auxotrophic strain BAP1-ΔbioH-MatBC. In the absence of exogenous biotin, malonate becomes the sole precursor of M-CoA, as the glucose-derived Ac-CoA pathway is inactivated. **b**, Growth of three biological replicates of BAP1-ΔbioH-MatBC in induced (+10  $\mu$ M IPTG) and uninduced conditions in M9 minimal medium (2% glucose) with increasing malonate concentrations (0–100 mM). Enhanced growth was observed at 1 mM malonate, whereas at  $\geq$ 10 mM malonate, cell density after 48 h at 37 °C was comparable to LB-grown cultures. **c**, Flaviolin production from three biological replicates as a proxy for intracellular M-CoA levels. Absorbance at 340 nm (path length corrected to 0.1 cm<sup>-1</sup>) peaked at 72 h in BAP1-ΔbioH-MatBC supplemented with 20 mM malonate ( $A_{340} = 14.56 \pm 1.06$ ), representing a 60% increase compared to the control strain BAP1-ΔbioH-RFP under the same conditions. **d**, LC-MS quantification of M-CoA and Ac-CoA in BAP1 and BAP1-ΔbioH-MatBC from

three biological replicates cultured in M9 minimal medium with increasing malonate concentrations for 36 h at 37 °C. M-CoA levels in BAP1-ΔbioH-MatBC were significantly higher than in BAP1 at 50 mM (3.3-fold,  $P = 0.016$ ) and 100 mM malonate (36-fold,  $P = 0.0015$ ). When normalized to DCW, M-CoA levels in BAP1-ΔbioH-MatBC were 2.0-fold ( $P = 0.078$ , not significant) and 18.1-fold ( $P = 0.0053$ ) higher than in BAP1 at 50 mM and 100 mM malonate, respectively. **e**, FA titers from three biological replicates measured by GC-MS in BAP1-ΔbioH-RFP and BAP1-ΔbioH-MatBC cultured in M9 minimal medium supplemented with biotin and  $\pm$ 20 mM malonate. In malonate-supplemented cultures, BAP1-ΔbioH-MatBC exhibited significantly higher ( $P < 0.05$ ) C16 and total FA levels than BAP1-ΔbioH-RFP. Total FA content included carbon chain lengths greater than 12, encompassing unsaturated and cyclopropanated isoforms. Statistical comparisons between groups were performed using GraphPad data analysis software with an unpaired *t*-test, assuming a Gaussian distribution, to calculate *P* values and determine statistical significance.

biotin pathway gene *bioH*, to create the strain BAP1-ΔbioH-MatBC. This auxotrophy was rescued either by adding biotin (present in media like LB and transported intracellularly via YigM<sup>49</sup>) or by supplementing malonate, which is imported and converted to M-CoA by the MatBC pathway (Fig. 2a).

Growth rescue by malonate was tested in BAP1-ΔbioH-MatBC using M9 minimal medium with 2% glucose and increasing malonate concentrations (up to 100 mM). The effect of IPTG induction of the *lacUV5* promoter on *matBC* expression was also evaluated. Substantial growth was observed with as little as 1 mM malonate, both with and without IPTG induction. At concentrations of  $\geq$ 10 mM malonate, cell density comparable to LB-grown controls was achieved after 48 h at 37 °C, despite slower growth at high malonate levels during the first 24 h (Fig. 2b).

Flaviolin production, used as a proxy for intracellular M-CoA levels, was measured in BAP1-ΔbioH-MatBC and a control strain,

BAP1-ΔbioH-RFP (where *bioH* was replaced with the red fluorescent protein gene). Both strains, transformed with pBADT-rppA-NT, were cultured in LB (+10  $\mu$ M IPTG and 0.2% arabinose) with increasing malonate concentrations for 3 days. Flaviolin production increased with malonate supplementation only in BAP1-ΔbioH-MatBC. At 20 mM malonate, BAP1-ΔbioH-MatBC produced 64% more flaviolin than BAP1-ΔbioH-RFP ( $P = 0.015$ ; Fig. 2c).

#### Quantification of acyl-CoAs and FAs

To quantify M-CoA levels directly, in the absence of a PKS, BAP1 and BAP1-ΔbioH-MatBC were cultured in M9 minimal medium with 2% glucose and increasing malonate concentrations for 36 h at 37 °C, reaching early stationary phase when M-CoA levels are typically low<sup>50</sup>. In BAP1-ΔbioH-MatBC, both the optical density of a sample measured at a wavelength of 600 nm (OD<sub>600</sub>) and M-CoA levels, measured via LC-MS, increased with malonate supplementation. The

highest M-CoA concentration,  $436.9 \pm 57.5$  nM, was observed with 100 mM malonate. Normalized to cell density, this corresponded to  $0.206 \pm 0.025$  nmol mg<sup>-1</sup> dry cell weight (DCW), an 18.1-fold change compared to  $0.011 \pm 0.09$  nmol mg<sup>-1</sup> DCW in BAP1 (Fig. 2d).

The impact of increased M-CoA levels on FA biosynthesis, FA profiles of BAP1-ΔbioH-MatBC and the control strain BAP1-ΔbioH-RFP were analyzed using gas chromatography (GC)-MS after 24 h of growth in M9 minimal medium at 37 °C with or without malonate supplementation. BAP1-ΔbioH-MatBC nearly depleted all exogenous malonate during the experiment, whereas substantial amounts remained in the medium of BAP1-ΔbioH-RFP (Supplementary Fig. 7). With malonate supplementation, BAP1-ΔbioH-MatBC produced substantially higher levels of C16 FAs and had a greater total FA content, including unsaturated and cyclopropanated FAs, compared to BAP1-ΔbioH-RFP (Fig. 2e).

### Production of uniformly labeled <sup>13</sup>C polyketides

In BAP1-ΔbioH-MatBC, the engineered MatBC pathway is the sole means for generating M-CoA. This design facilitates the production of uniform isotopically labeled polyketides, free from contamination by unlabeled endogenously produced M-CoA without the need to cultivate in [<sup>13</sup>C]glucose (Fig. 3a). Here the flaviolin-producing strain BAP1-ΔbioH-MatBC-RppA was cultivated in M9 minimal medium supplemented with 2% unlabeled glucose with either unlabeled malonate or [<sup>13</sup>C]malonate labeled solely at position 2. A control without supplemental malonate failed to grow and produce flaviolin, while the strains supplemented with malonate turned turbid and red (Supplementary Fig. 5). LC-MS analysis of the cultures revealed a consistent 5-Da mass shift in the peak corresponding to flaviolin when [<sup>13</sup>C]malonate was used compared to the unlabeled counterpart, in accordance with the incorporation of five M-CoA molecules (Fig. 3b).

To demonstrate the necessity of preventing endogenous M-CoA biosynthesis for achieving uniform <sup>13</sup>C labeling, biotin was introduced into the medium along with [<sup>13</sup>C]malonate to enable endogenous M-CoA production. Extracted ion chromatograms (EICs) for masses corresponding to flaviolin labeled with zero to five [<sup>13</sup>C]carbons were examined for cultures fed solely with [<sup>13</sup>C]malonate and those also supplemented with biotin. In the case of [<sup>13</sup>C]malonate alone, a prominent peak at 210 Da was observed, representing five [<sup>13</sup>C]M-CoA molecules with 99.8% purity over other isotopically labeled species. Conversely, when biotin was included in the medium, detectable peaks for masses corresponding to flaviolin with zero to five [<sup>13</sup>C]carbons were evident (Fig. 3c). EICs for masses 205–210 Da were similarly analyzed for controls where only malonate or biotin was added to the culture. In both cases, a predominant peak at 205 Da was observed, accompanied by minor peaks at 206 and 207, which is in line with the natural isotopic abundance of carbon (Supplementary Fig. 6). Over the 3-day period, the final OD<sub>600</sub> for both strains exceeded 5.0, while the flaviolin absorbance at 340 nm exceeded 13.5, similar to what was observed in LB medium (Fig. 2b). Notably, when supplementing with both [<sup>13</sup>C]malonate and biotin over a 72-h period, the ratio of the uniformly labeled 210 Da peak progressively rose from 79.9% on day 1 to 92.6% by day 3 (Supplementary Fig. 6). This suggests that, over time, the M-CoA pool derived from supplemented malonate dominated the M-CoA derived from Ac-CoA.

### ALE of BAP1-ΔbioH-MatBC

ALE was performed on BAP1-ΔbioH-MatBC to enhance M-CoA production and availability by exploiting its growth dependency on exogenous malonate. The BAP1-ΔbioH-MatBC strain was transformed with the mutagenesis plasmid-6 (MP6), which harbors *dnaQ926*, *dam*, *seqA*, *emrR*, *ugi* and *cdaI* and enables a 322,000-fold increase in the mutation rate of *E. coli* chromosomal DNA<sup>51</sup>. The strain was cultured in M9 minimal medium with 2% glucose and 1 mM malonate, the lowest concentration exhibiting growth recovery, in a Chi.Bio turbidostat<sup>52</sup> where the culture's OD<sub>600</sub> was maintained at 0.75. After 7 days, the evolved

culture was plated on M9 minimal agar plates with 3 mM malonate, and six strains (E1\_S1–6) were further cultured and sequenced. Increased growth was observed for four of six strains, indicating positive growth adaptations (Supplementary Fig. 8).

In this first round of ALE, many mutations were observed in all strains relative to the nonmutagenized control. A common mutation in all strains was a single point mutation (C>A) to the lac operator upstream of the *matB* and *matC* genes, which is known to disrupt repressor affinity<sup>53</sup>. PCR amplification and sequencing of the total evolved pool revealed that 99.6% of the sequencing reads contained this lac operator mutation. This result, supported by proteomics data (Supplementary Data 1), indicates that increased expression of MatB and MatC provides a growth advantage in this context.

Building on these findings, we extended ALE to explore additional conditions and a longer time frame. To prevent evolutionary escape and enable prolonged selection, a kanamycin-resistance marker was integrated into the genome, replacing additional biotin biosynthesis genes (*bioA*, *bioB*, *bioC*, *bioD* and *bioF*), creating BAP1-ΔbioH-MatBC-Δbio-kan. This strain was transformed with the MP6 plasmid and evolved for 2 weeks in M9 medium with 2% glucose and 20 mM malonate, a concentration that supports full growth recovery and high polyketide titers (Fig. 2c). After 2 weeks, the culture was split into the following two conditions: low malonate (1 mM) and high malonate (20 mM), for an additional 2 weeks. The experimental workflow is illustrated in Fig. 4a. Serial dilutions in test tubes were performed throughout the 4-week evolution. Individual colonies from each condition (E2\_S1–17) were isolated and sent for whole-genome sequencing.

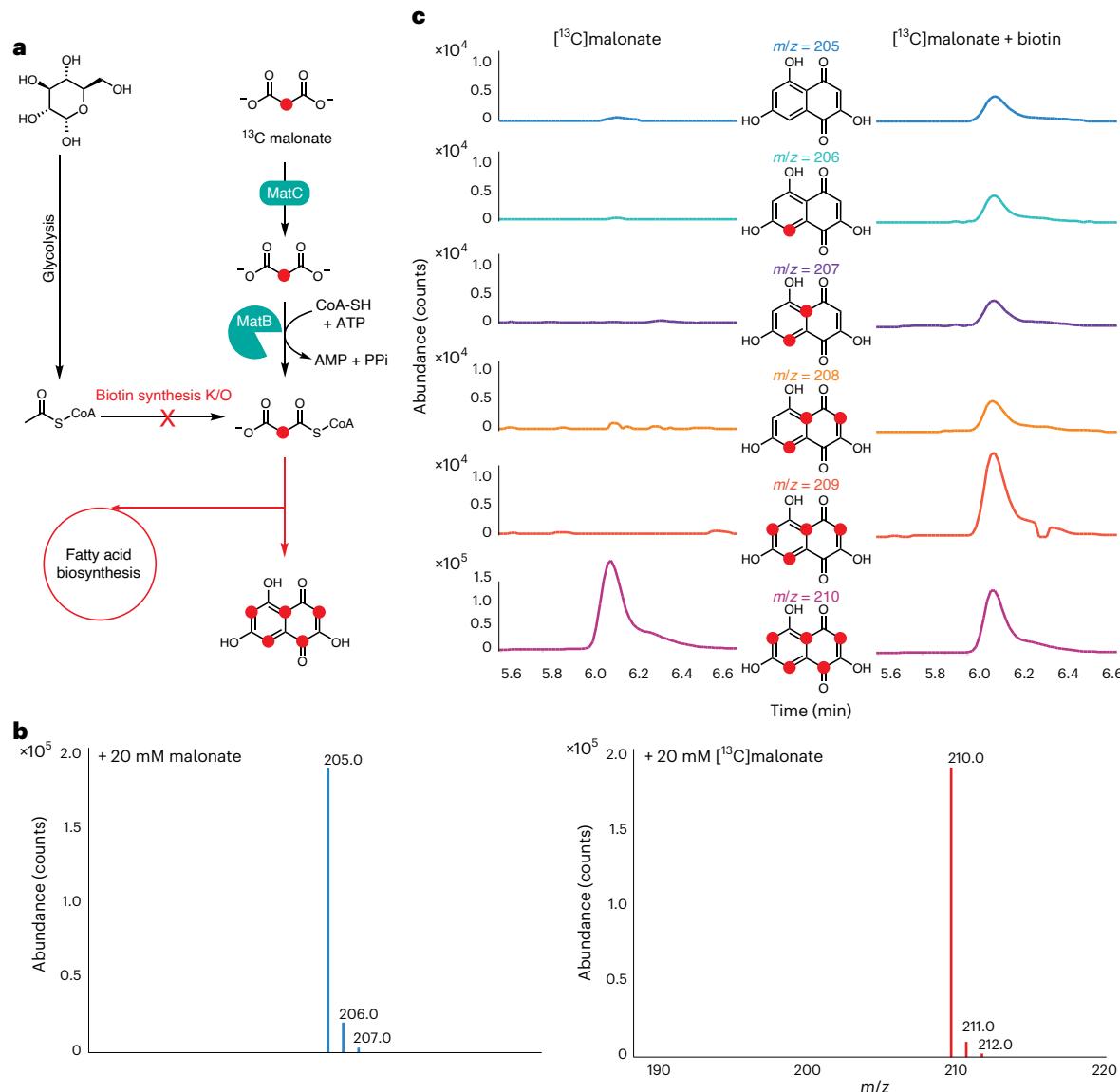
As in the previous ALE experiment, numerous mutations were identified in all evolved strains compared to the nonmutagenized control (Supplementary Data 2). All strains again carried a mutation in the lac operator upstream of *matBC*. However, in the 4-week evolution experiment, a substantially higher number of mutations were observed, with over 1,518 missense, frameshift or stop codon mutations identified across all 23 strains (E1\_S1–6 and E2\_S1–17). These mutations were transformed into an absence–presence matrix and analyzed using hierarchical clustering (Pearson correlation and average linkage method), resulting in a dendrogram with four distinct clusters (Fig. 4b, clades 1–4).

Strains evolved for 1 week in 1 mM malonate (E1\_S1–S7) formed clade 4, which exhibited the fewest average mutations (41.33 per strain). In contrast, clade 2 (E2\_S3, S10, S12, S13 and S14) contained strains that evolved in 20 mM malonate for 2 weeks, followed by 1 mM malonate for an additional 2 weeks, and had the highest average mutations (303 per strain). Clade 1 (E2\_S1, S2, S4, S6, S7 and S11) included strains from both 1 mM and 20 mM malonate conditions during the second 2-week evolution. Clade 3 (E2\_S5, S8, S9, S15, S16 and S17) contained strains that evolved exclusively in 20 mM malonate for 2 or 4 weeks before the split into two conditions.

Given the large number of missense mutations and limited functional characterization, our analysis focused on frameshift and stop codon mutations that resulted in the premature termination of protein translation. Mutations present in at least 50% of strains within a clade were considered representative and used to construct a presence–absence mutation matrix (Fig. 4b), with functional annotations assigned where possible.

Some of the observed mutations have been previously identified to improve free FA or M-CoA levels by their deletion or downregulation in *E. coli*, such as *atoB* (clades 1 and 2), *creC* (clade 2) and *glpG* (clade 3)<sup>54</sup>. The increased free FA titers were proposed to occur through reduction of flux through β-oxidation (*atoB*), downregulation of phosphate starvation response (*creC*) and downregulation of membrane proteins (*glpG* and *creC*), which might be associated with tolerance to free FAs<sup>54–56</sup>.

Additional mutations were observed in pathways previously linked to improved FA or M-CoA titers. Knockouts of iron transport genes,



**Fig. 3 | Production of uniformly labeled  $^{13}\text{C}$ -labeled PKS products using BAP1- $\Delta$ bioH-MatBC.** **a**, Schematic representation of uniform isotopic labeling of M-CoA-derived polyketides. **b**, Extracted mass spectra of flaviolin synthesized in BAP1- $\Delta$ bioH-MatBC supplemented with unlabeled or  $^{13}\text{C}$ -labeled malonate. The incorporation of [ $^{13}\text{C}$ ]malonate results in a 5 Da shift, indicating uniform integration of five  $^{13}\text{C}$ -labeled M-CoA molecules. **c**, Comparative analysis of  $^{13}\text{C}$ -labeled isotopic incorporation in flaviolin produced by BAP1- $\Delta$ bioH-MatBC, with or without restoration of biotin auxotrophy through the exogenous addition of 25 mg l<sup>-1</sup> biotin. After 24 h of cultivation in autoinduction M9

minimal medium at 37 °C, the strains achieved up to 99.8% purity for five-carbon <sup>13</sup>C-labeled flaviolin ( $m/z = 210$ ) in the absence of biotin. By contrast, biotin supplementation reduced the labeling efficiency to 79.9%, with additional peaks corresponding to mono ( $m/z = 206$ ), di ( $m/z = 207$ ), tri ( $m/z = 208$ ) and tetra ( $m/z = 209$ ) incorporation of <sup>13</sup>C-labeled M-CoA. The 340 nm and 600 nm absorbances in a 1 cm light path were recorded at the final 72-h time point. For cultures supplemented with <sup>[13]C</sup>malonate,  $OD_{600} = 5.5$ ,  $A_{340} = 14.78$  and  $A_{340}/OD_{600} = 2.7$ . In cultures supplemented with both <sup>[13]C</sup>malonate and biotin,  $OD_{600} = 5.1$ ,  $A_{340} = 13.65$  and  $A_{340}/OD_{600} = 2.7$ .

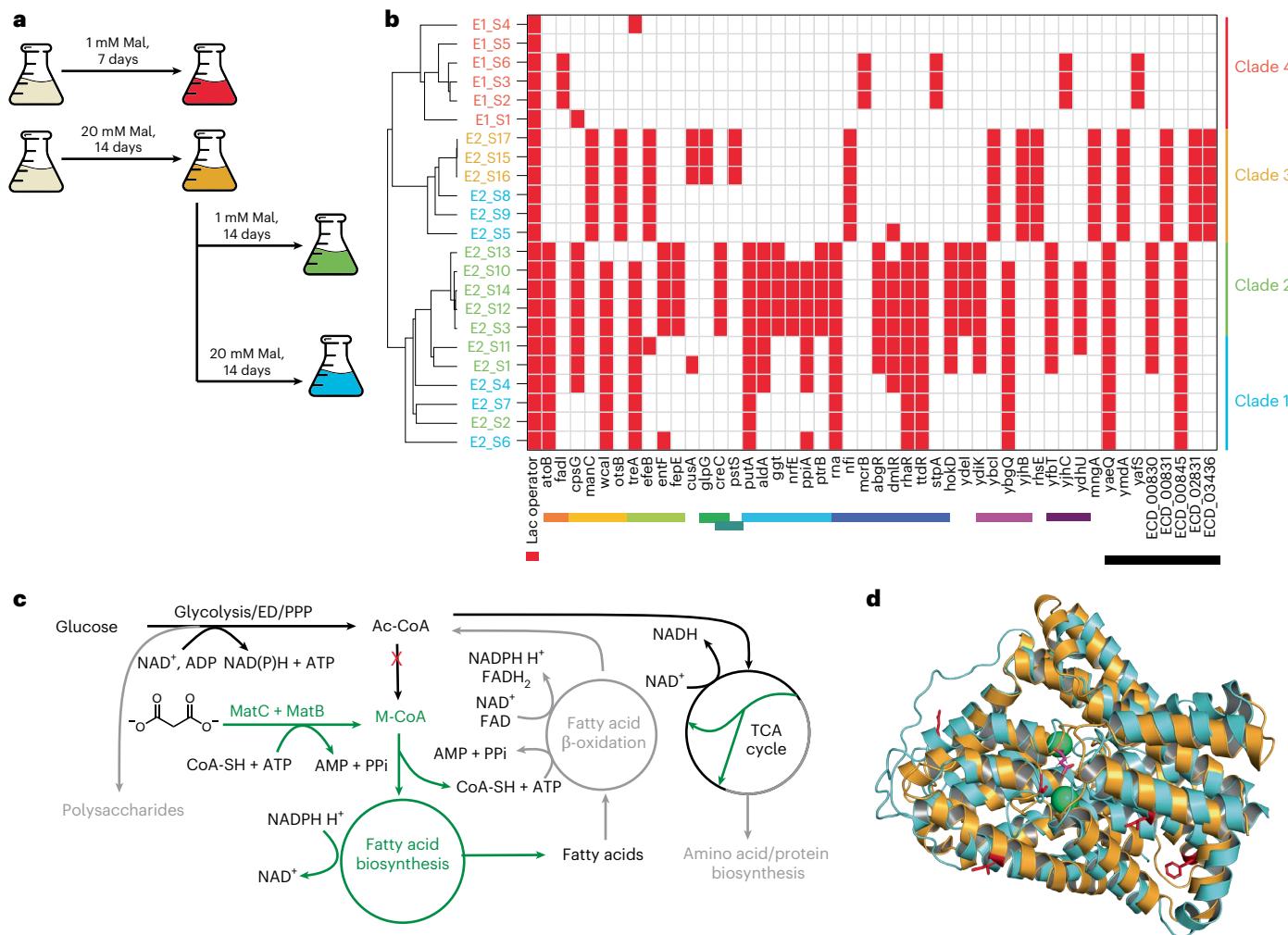
*fhuA* and *tonB*, have been shown to increase intracellular M-CoA levels and polyketide production, whereas iron supplementation reduces these levels<sup>40</sup>. In our study, frameshift or stop codon mutations were observed in genes involved in iron transport and metabolism, such as *efeB* (clades 3 and 1), *entF* (clades 2 and 1) and *fepE* (clade 2).

Similarly, *rcsA* knockouts, which redirect flux toward glycolysis by repressing colanic acid biosynthesis, have been reported to enhance M-CoA and polyketide titers<sup>40</sup>. Correspondingly, our strains harbored mutations associated with colanic acid biosynthesis genes *cpsG* (clades 1 and 2), *manC* (clade 3) and *wcal* (clades 1 and 2).

Clade 4 also possessed a mutation to *fadl*, encoding a thiolase enzyme involved in  $\beta$ -oxidation. The specific role of *fadl* in  $\beta$ -oxidation remains debated, with evidence suggesting it functions on long-chain

FAs under aerobic conditions or short-chain FAs anaerobically<sup>57,58</sup>. However, *fad*/knockouts do not completely disable  $\beta$ -oxidation activity, as both *fad*/and its homolog *fadA* must be inactivated to fully block FA catabolism<sup>59,60</sup>.

Mutations in genes related to amino acid metabolism, protein biosynthesis and nitrogen regulation, which could redirect carbon flux from protein synthesis to lipid production, were observed in all clades, with clades 1 and 2 being particularly enriched. Prior studies have shown that repressing ribosomal protein L23 (*rplW*) or tRNA-associated genes (*tyrU* and *rnpB*) increases FA production<sup>54</sup>. In our strains, frameshift or stop codon mutations in amino acid metabolism genes, such as *aldA* (lactaldehyde dehydrogenase, involved in the methylglyoxal pathway) and *putA* (proline dehydrogenase), were identified. Mutations



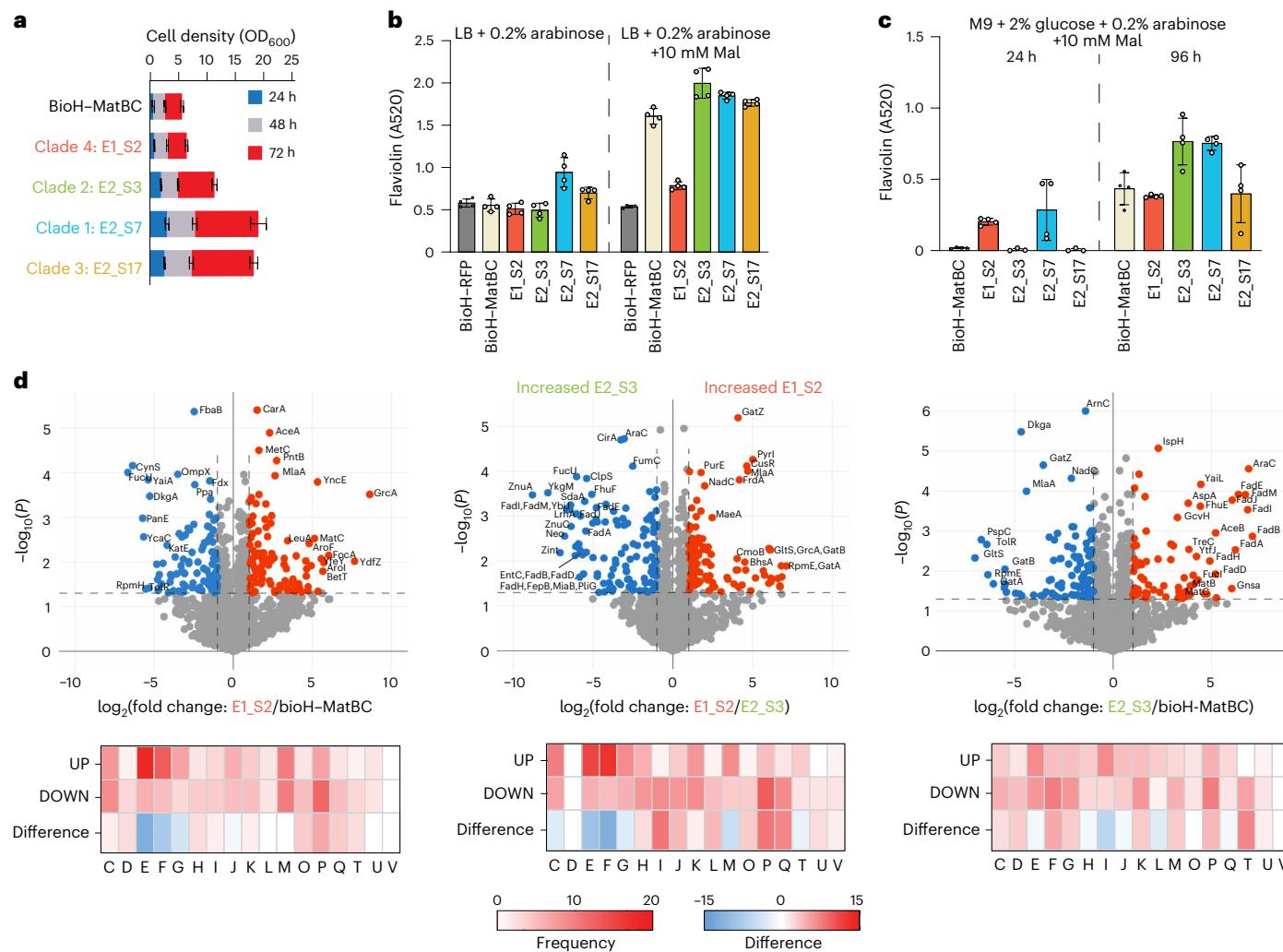
**Fig. 4 | Mutations acquired through ALE. a**, Schematic representation of the adaptive evolution conditions for 23 evolved strains (E1\_S1–S6 and E2\_S1–S17). **b**, Clustering of evolved strains based on 1,518 observed mutations. A dendrogram was constructed using Pearson's correlation and an average linkage method to group strains into four distinct clades. The presence–absence matrix highlights representative frameshift or stop codon mutations occurring in more than 50% of strains within each clade. Red indicates the presence of a mutation, whereas white indicates its absence. Mutations are functionally categorized where possible with boxes of the corresponding colors next to the gene names—MatBC pathway (red), FA  $\beta$ -oxidation (orange), colanic acid/polysaccharide biosynthesis (gold), iron transport and metabolism (lime green), Membrane association (green), phosphate transport (blue–green), amino acid/protein

biosynthesis (light blue), DNase/RNase/transcription regulation (dark blue), putative membrane association (magenta), alternative carbon sources (purple) and unknown function (black). **c**, Metabolic pathway analysis of ALE-derived mutations mapped onto central metabolism. Pathways with increased flux are indicated in green, and those with decreased flux due to frameshift or stop codon mutations are marked in gray. **d**, AlphaFold-predicted structure of MatC (gold), highlighting six missense mutations (A365S, F400S, V87A, I91N, L424P and P372L) observed in six separate strains. These mutations, identified as representative of cluster 1 by PCA analysis (Supplementary Fig. 9), are shown in red. The MatC structure is overlaid on the VcINDY transporter (Protein Data Bank (PDB) ID: **6OL1**; cyan), with bound sodium (green) and succinate (pink) shown for reference. Mal, malonate.

in protein biosynthesis genes, including *ppiA* (peptidyl-prolyl isomerase A), *ptrB* (oligopeptidase B) and *rna* (RNaseI), were also observed. Additionally, a mutation in *nrfE*, a component of the nitrite reductase complex involved in nitrogen regulation, was detected.

A comparison with the mutational profiles of a previously reported ALE of *E. coli* in glucose minimal medium<sup>61,62</sup> revealed no overlap of representative frameshift or stop codon mutations with our study. However, missense mutations in RNA polymerase genes, such as *rpoB* (E546V and E672K), *rpoC*<sup>62</sup>, *rpoD* (S253P) and *rpoA* (G279V)<sup>61</sup>, have been reported to alter promoter binding and broadly impact transcriptional levels, providing fitness advantages. In our dataset, missense mutations were found in *rpoB* (D853N,  $n = 11$ ; K203R,  $n = 7$ , co-occurring with D853N), *rpoC* (L1059P,  $n = 1$ ; S655Y,  $n = 1$ ) and *rpoD* (A375T,  $n = 1$ ). These findings align with previous observations of common mutations in the *rpoBC* operon during ALE<sup>63,64</sup>, suggesting adaptation to minimal glucose media alongside selective pressures such as M-CoA and FA limitations.

The evolved strains also carried numerous other missense mutations with unknown functional impacts. Of particular interest were six unique missense mutations in the MatC malonate importer (A365S, F400S, V87A, I91N, L424P and P372L) found in six strains (E2\_S2, S3, S4, S6, S7 and S16). To visualize the potential locations of these mutations on the MatC protein, AlphaFold 2.3 was used to model the MatC structure, clustering 640 predicted structures into six representative cluster control models (Supplementary Fig. 9). The structure representing cluster 1 closely resembled the crystal structure of the sodium-dependent dicarboxylate transporter (*Vibrio cholerae* INDY (VcINDY)) from *Vibrio cholerae*, a homodimer that undergoes structural rearrangement upon substrate binding<sup>65</sup>. Some mutations, such as P372L, were located on the protein exterior or in flexible regions (V87A and I91N). Others, like F400S (strain E2\_S3), were near the  $\text{Na}^+$  substrate-binding site, whereas A365S and P372L aligned with the VcINDY  $\alpha$ -helical region involved in substrate binding and



**Fig. 5 | Characterization of evolved strains.** **a**, Cell density measurements over 3 days for four biological replicates of representative strains from each clade of the dendrogram in Fig. 4. **b,c**, Flaviolin production from four biological replicates of representative strains from each clade in LB, LB + 10 mM Mal (**b**) and M9 + 10 mM Mal at 24 h and 96 h time points (**c**). Significantly increased ( $P < 0.05$ ) titers compared to the unevolved bioH-MatBC strain were observed for E1\_S2 in M9 + 10 mM Mal at 24 h; E2\_S3 in LB + 10 mM Mal and M9 + 10 mM Mal at 96 h; E2\_S7 in all conditions; and E2\_S17 in LB, LB + 10 mM Mal and M9 + 10 mM Mal at 96 h. All data are presented as mean values  $\pm$  s.d. Statistical comparisons between groups were performed using GraphPad data analysis software with an unpaired *t*-test, assuming a Gaussian distribution, to calculate *P* values and determine statistical significance. **d**, Proteomic analysis from four biological replicates comparing evolved and unevolved strains, showing COG assignments for proteins with significantly increased (fold change  $> 1$ ,  $P < 0.05$ ) or decreased (fold change  $< -1$ ,  $P < 0.05$ ) abundance, determined using

Welch's *t*-test correction. The number of proteins in each COG category is shown for UP and DOWN proteins, along with the difference between these counts (DOWN - UP = difference). COG functional categories—C, energy production and conversion; D, cell cycle control, cell division and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, post-translational modification, protein turnover and chaperones; P, inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport and catabolism; T, signal transduction mechanisms; U, intracellular trafficking, secretion and vesicular transport; V, defense mechanisms. UP, upregulated; DOWN, downregulated.

conformational changes (Fig. 4d). Further structural studies are needed to understand how these mutations affect the function and structure of MatC.

Based on the genome sequencing and proteomics data (Fig. 5d), we propose the following hypothesis for the observed mutations (Fig. 4c): mutations in the lac operator increase *matB* and *matC* expression, enhancing flux toward M-CoA to support FA biosynthesis. Missense mutations to *matC* may further enhance this flux, although biocatalytic studies are needed for confirmation. Reduced glucose diversion to polysaccharide synthesis increases glucose availability for NADPH and ATP production. Decreased protein translation and biosynthesis (amino acids and RNA) align with the proteomic data of our evolved

strains (Fig. 5d), indicating a shift toward the Entner–Doudoroff pathway, which uses fewer enzymatic steps than glycolysis and produces NADPH and NADH.

Mutations in iron transport genes increase flux through the glyoxylate shunt pathway enzymes (AceA and AceB) to bypass iron-dependent enzymes in the tricarboxylic acid (TCA) cycle. Mutations to  $\beta$ -oxidation genes (*atoB* and *fadI*) decrease FA degradation, lowering M-CoA demand. Consistently, CRISPRi knockdowns of *atoB* and other  $\beta$ -oxidation genes (*fadE*, *fadB* and *yqeF*) have been shown to increase FA levels in *E. coli*<sup>54</sup>.

To assess whether *fadI* similarly prevents FA degradation and increases M-CoA, the BAP1– $\Delta$ bioH–MatBC strain was engineered with a

*fad*/knockout, creating BAP1-ΔbioH-MatBC-ΔfadI. M-CoA and Ac-CoA levels were compared among this strain, the unevolved BAP1-MatBC strain and the evolved E1\_S2 strain carrying a *fad* mutation. E1\_S2 exhibited a substantial increase in M-CoA compared to BAP1-MatBC, whereas BAP1-ΔbioH-MatBC-ΔfadI displayed intermediate M-CoA levels, not significantly different from either strain (Supplementary Fig. 10).

### Characterization of evolved strains

To assess phenotypic differences between clades, one strain from each identified clade (E1\_S2, E2\_S3, E2\_S7 and E2\_S17), representing distinct ALE culturing conditions, was cultivated in M9 minimal media with 2% glucose and 10 mM malonate for growth recovery tests (Fig. 5a). In total, 10 mM malonate was chosen as an intermediate concentration between the two cultivation conditions of 1 and 20 mM. Strains from clades 1, 2 and 3 showed substantially improved growth (OD<sub>600</sub>) compared to the unevolved strain, whereas the clade 4 strain exhibited marginally better growth only at early time points. Under its ALE condition (1 mM malonate), clade 4 outperformed the unevolved strain (Supplementary Fig. 8).

To evaluate if evolved strains enhanced M-CoA availability for PKSs and increased polyketide production, the flaviolin production plasmid pBADT-RppA-NT was introduced into one representative strain from each clade (E1\_S2, E2\_S3, E2\_S7 and E2\_S17). These strains, along with controls BAP1-ΔbioH-RFP and unevolved BAP1-ΔbioH-MatBC, were cultured under the following three conditions: M9 media + 10 mM malonate, LB and LB + 10 mM malonate. Cultures were grown in quadruplicate at 37 °C for 4 days, and flaviolin production was measured at 520 nm to reduce media background interference between LB and M9 media.

In LB, E2\_S7 and E2\_S17 produced substantially more flaviolin than other strains, even without malonate. All strains outperformed the RFP control when malonate was added. E1\_S2 (clade 4) produced less flaviolin than the unevolved strain, whereas E2\_S3, E2\_S7 and E2\_S17 showed substantially higher production (Fig. 5b). In M9 media, flaviolin production after 24 h was notable only for E1\_S2 and E2\_S7. However, after 96 h, flaviolin production was observed in all strains, with E2\_S3 and E2\_S7 outperforming the unevolved strain (Fig. 5c). Owing to its superior performance across conditions, E2\_S3 (clade 2) was selected for further study.

To investigate key adaptive mutations in strain E2\_S3, its proteome was compared to the unevolved MatBC strain and E1\_S2 from clade 4, which had the fewest mutations. All strains were cultivated in M9 minimal medium with 2% glucose and 10 mM malonate for 48 h at 37 °C, starting at an OD<sub>600</sub> of 0.001. Both evolved strains shared a mutation in the *lac* operator regulating *matB* and *matC* expression and mutations to β-oxidation genes (E2\_S3, Δ*atoB*; E1\_S2, Δ*fadI*). Proteomic differences were identified using Welch's *t*-test (cutoff of >±1-fold, *P* < 0.05), and differentially expressed proteins were annotated by their Clusters of Orthologous Groups (COG) name with the eggNOG-Mapper V2 tool (<http://eggNOG-mapper.embl.de/>)<sup>66</sup>.

Compared to the unevolved strain, E1\_S2 showed increased expression of genes associated with amino acid transport and metabolism (E) and nucleotide transport and metabolism (F), with decreased expression of inorganic ion transport genes (P). Notably upregulated proteins included MatB (2.43-fold) and MatC (5.1-fold), along with proteins involved in aromatic amino acid biosynthesis (AroL, 5.7-fold; EntB, 5.4-fold; AroF, 4.8-fold; TyrA, 3.2-fold) and the methylerythritol phosphate pathway (Dxr, 4.0-fold; IspH, 3.4-fold; IspU, 3.4-fold).

In the highly mutated strain E2\_S3, compared to the unevolved strain, there was increased expression of genes related to lipid transport and metabolism (I) and decreased expression of genes for signal transduction and nucleotide transport and metabolism (F). Significantly upregulated proteins included MatB (3.9-fold) and MatC (3.6-fold), FA β-oxidation proteins (FadA, FadB, FadD, FadE, FadH, FadI,

FadJ, FadM and TesB; 3.9–7.07-fold), glyoxylate shunt proteins (AceA, 2.6-fold; AceB, 5.2-fold) and iron transport proteins (EntC, 5.2-fold; FhuE, 4.4-fold). Decreased expression of TktB, PoxB, PfkB and PurC (-2.82-fold to -1.09-fold) suggests a shift toward the Entner–Doudoroff pathway, which generates NADPH and ATP with fewer enzymatic steps compared to glycolysis and the pentose phosphate pathway. When comparing E2\_S3 to E1\_S2, the highly mutated E2\_S3 showed reduced expression of amino acid (E) and nucleotide transport genes (F) but higher expression of genes related to lipid metabolism (I), inorganic ion transport and secondary metabolite biosynthesis (Q). Additionally, AceA, AceB and MatB were also increased (Fig. 5d).

Enhanced glyoxylate shunt activity enables Ac-CoA to enter the TCA cycle while avoiding carbon loss as CO<sub>2</sub> and bypassing iron-dependent enzymes. This pathway's activation is further linked to FA presence, regulated by IclR and indirectly by FadR67. Detailed proteomic and COG data are provided in Supplementary Data 1.

Overall, the proteomic data support the proposed pathways most affected by ALE (Fig. 4c). In E2\_S3, mutations in β-oxidation and iron transport likely drive increased expression of related proteins in response to FA accumulation and iron starvation. Increased levels of β-oxidation protein in E2\_S3 compared to E1\_S2 suggest that *atoB* knockouts more effectively block FA degradation than *fad*/mutations, consistent with intermediate M-CoA levels observed in the reengineered strain. Enhanced glyoxylate shunt activity bypasses carbon loss as CO<sub>2</sub> and iron-dependent enzymes in the TCA cycle. Additionally, increased expression of the glyoxylate shunt pathway is known to increase in the presence of FAs via negative regulation by IclR and indirectly by FadR<sup>67</sup>. Detailed proteomic and COG data are available in Supplementary Data 1. Additionally, all plasmids and strains created or used in this study are listed in Supplementary Table 1.

### Discussion

Expressing engineered PKSs in *E. coli* is a sustainable approach for producing valuable small molecules. However, unlike native hosts, *E. coli* often lacks sufficient acyl-CoA precursors, including M-CoA, which is tightly regulated and depleted during the stationary phase<sup>18,19</sup>.

This study introduced *matB* and *matC* into PKS-compatible *E. coli* strains (BAP1 and K207-3), enabling precise control over M-CoA levels. This strategy increased titers of M-CoA-derived compounds, including type III PKS naphthoquinones, FAs and type I PKS products. Hybrid type I PKS constructs (pik127-2nd and pik127-3rd acyltransferases) demonstrated how substrate pool variations substantially affect product composition, emphasizing the importance of optimizing acyl-CoA precursor pools.

To further regulate M-CoA, the endogenous M-CoA pathway in BAP1-ΔbioH-MatBC was disrupted by deleting biotin synthesis genes, allowing malonate supplementation to control growth, M-CoA and polyketide titers. This approach achieved up to an 18.1-fold M-CoA improvement over wild-type strains, comparable to or exceeding prior efforts<sup>20</sup>. Leveraging this control over M-CoA facilitated the production of uniformly labeled <sup>13</sup>C-polyketides without requiring fully isotopically labeled carbon sources. Notably, as most polyketides are derived from a mixture of acyl-CoA starter units, if the PKS encodes for extension by at least one M-CoA, uniform production of the labeled polyketide can be achieved. Compounds enriched with stable isotopes such as <sup>13</sup>C are invaluable as metabolic tracers in medicine<sup>68,69</sup> and research, where uniform labeling ensures accurate results<sup>70,71</sup>. Highly <sup>13</sup>C-enriched compounds are particularly beneficial for *in vivo* <sup>13</sup>C MRI trials, where the enhanced signal improves imaging<sup>72–74</sup>. This method offers a cost-effective, efficient alternative for producing >99% uniformly labeled M-CoA-derived products and FAs, outperforming traditional approaches that rely on fully labeled carbon sources like [<sup>13</sup>C]glucose or [<sup>13</sup>C]acetate<sup>75</sup>.

Using the engineered malonate growth dependency, ALE was performed to identify mutations enhancing M-CoA production while

avoiding or compensating for mutations that often come at a cost in cellular growth<sup>23,24,76</sup>. The evolved strains acquired many mutations and clustered into four distinct clades. Loss-of-function mutations in genes like *atoB*, *creC* and *glpG* aligned with prior research identifying their roles in enhancing FA or M-CoA titers. More interesting was the presence of loss-of-function mutations to genes that have yet to be targeted through any direct approaches but are in relevant pathways such as  $\beta$ -oxidation, colanic acid biosynthesis, iron transport and amino acid/protein biosynthesis, where knockouts or knockdowns of genes have previously been shown to improve FA, M-CoA or polyketide titers. Although some of our observed mutations could be rationalized, many remain unclear. Re-engineering these mutations, particularly to transcription factors (*abgR*, *dmlR*, *ttdR*, *treR* and *rhaR*), could further elucidate their effects.

Finally, the evolved strains were assessed for M-CoA-derived polyketide production. Evolved strains were identified that outperformed the original strains in minimal medium with malonate and in biotin-rich conditions (for example, LB), indicating mutations that enhance M-CoA availability even outside selective conditions. This highlights ALE as a powerful tool for improving M-CoA flux and polyketide biosynthesis while uncovering pathways involved in FA and CoA metabolism.

Regulating M-CoA pools is essential for the biosynthesis of polyketides derived from this precursor. Orthogonal pathways that boost M-CoA levels provide opportunities to investigate, engineer and evolve M-CoA metabolism for improved heterologous PKS expression in *E. coli*. Using ALE to identify mutations that increase flux toward M-CoA and downstream PKS products highlights the utility of biotin/malonate auxotrophic strains. Moreover, ALE facilitates the selection of compensatory mutations that can alleviate growth defects while also enhancing M-CoA production, further supporting the evolution of robust strains for polyketide biosynthesis.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-025-01911-6>.

## References

1. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
2. Katsuyama, Y. & Ohnishi, Y. Type III polyketide synthases in microorganisms. *Methods Enzymol.* **515**, 359–377 (2012).
3. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679 (1991).
4. Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33**, 203–230 (2016).
5. Englund, E. et al. Biosensor guided polyketide synthases engineering for optimization of domain exchange boundaries. *Nat. Commun.* **14**, 4871 (2023).
6. McDaniel, R. et al. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel ‘unnatural’ natural products. *Proc. Natl Acad. Sci. USA* **96**, 1846–1851 (1999).
7. Menzella, H. G. et al. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176 (2005).
8. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
9. Lim, C. G., Fowler, Z. L., Hueller, T., Schaffer, S. & Koffas, M. A. G. High-yield resveratrol production in engineered *Escherichia coli*. *Appl. Environ. Microbiol.* **77**, 3451–3460 (2011).
10. Wu, J., Zhou, T., Du, G., Zhou, J. & Chen, J. Modular optimization of heterologous pathways for de novo synthesis of (2S)-naringenin in *Escherichia coli*. *PLoS ONE* **9**, e101492 (2014).
11. Kao, C. M., Luo, G., Katz, L., Cane, D. E. & Khosla, C. Engineered biosynthesis of a triketide lactone from an incomplete modular polyketide synthase. *J. Am. Chem. Soc.* **116**, 11612–11613 (1994).
12. Ray, K. A. et al. Assessing and harnessing updated polyketide synthase modules through combinatorial engineering. *Nat. Commun.* **15**, 6485 (2024).
13. Gao, X., Wang, P. & Tang, Y. Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **88**, 1233–1242 (2010).
14. Schmidt, M. et al. Maximizing heterologous expression of engineered type I polyketide synthases: investigating codon optimization strategies. *ACS Synth. Biol.* **12**, 3366–3380 (2023).
15. Zhang, J., Bista, R., Miyazawa, T. & Keatinge-Clay, A. T. Boosting titers of engineered triketide and tetraketide synthases to record levels through T7 promoter tuning. *Metab. Eng.* **78**, 93–98 (2023).
16. Choi, G.-S., Choo, H. J., Kim, B.-G. & Ahn, J.-H. Synthesis of acridone derivatives via heterologous expression of a plant type III polyketide synthase in *Escherichia coli*. *Microb. Cell Fact.* **19**, 73 (2020).
17. Choo, H. J. & Ahn, J.-H. Synthesis of three bioactive aromatic compounds by introducing polyketide synthase genes into engineered *Escherichia coli*. *J. Agric. Food Chem.* **67**, 8581–8589 (2019).
18. Cronan, J. E. & Thomas, J. Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods Enzymol.* **459**, 395–433 (2009).
19. Milke, L. & Marienhagen, J. Engineering intracellular malonyl-CoA availability in microbial hosts and its impact on polyketide and fatty acid synthesis. *Appl. Microbiol. Biotechnol.* **104**, 6057–6065 (2020).
20. Zha, W., Rubin-Pitel, S. B., Shao, Z. & Zhao, H. Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. *Metab. Eng.* **11**, 192–198 (2009).
21. Wu, J., Zhang, X., Xia, X. & Dong, M. A systematic optimization of medium chain fatty acid biosynthesis via the reverse  $\beta$ -oxidation cycle in *Escherichia coli*. *Metab. Eng.* **41**, 115–124 (2017).
22. Leonard, E., Lim, K.-H., Saw, P.-N. & Koffas, M. A. G. Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. *Appl. Environ. Microbiol.* **73**, 3877–3886 (2007).
23. Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D. & Koffas, M. A. G. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.* **13**, 578–587 (2011).
24. Wu, J., Du, G., Chen, J. & Zhou, J. Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. *Sci. Rep.* **5**, 13477 (2015).
25. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol.* **6**, 775–784 (1999).
26. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931–943 (2007).
27. Moteallehi-Ardakani, M. H., Asad, S., Marashi, S.-A., Moghaddasi, A. & Zarpavar, P. Engineering a novel metabolic pathway for improving cellular malonyl-CoA levels in *Escherichia coli*. *Mol. Biotechnol.* **65**, 1508–1517 (2023).
28. Sirirunguangs, S. et al. Engineering site-selective incorporation of fluorine into polyketides. *Nat. Chem. Biol.* **18**, 886–893 (2022).
29. Liang, B., Zhang, X., Meng, C., Wang, L. & Yang, J. Directed evolution of tripartite ATP-independent periplasmic transporter for 3-hydroxypropionate biosynthesis. *Appl. Microbiol. Biotechnol.* **107**, 663–676 (2023).

30. Wu, J., Du, G., Zhou, J. & Chen, J. Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab. Eng.* **16**, 48–55 (2013).

31. Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R. & Kealey, J. T. Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production. *J. Ind. Microbiol. Biotechnol.* **30**, 500–509 (2003).

32. Fang, Z., Jones, J. A., Zhou, J. & Koffas, M. A. G. Engineering *Escherichia coli* co-cultures for production of curcuminoids from glucose. *Biotechnol. J.* **13**, e1700576 (2018).

33. Leonard, E. et al. Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Mol. Pharm.* **5**, 257–265 (2008).

34. Boghigian, B. A., Zhang, H. & Pfeifer, B. A. Multi-factorial engineering of heterologous polyketide production in *Escherichia coli* reveals complex pathway interactions. *Biotechnol. Bioeng.* **108**, 1360–1371 (2011).

35. An, J. H. & Kim, Y. S. A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in *Rhizobium trifolii*—cloning, sequencing, and expression of the enzymes in *Escherichia coli*. *Eur. J. Biochem.* **257**, 395–402 (1998).

36. Lin, S., Hanson, R. E. & Cronan, J. E. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nat. Chem. Biol.* **6**, 682–688 (2010).

37. Cao, X., Zhu, L., Hu, Z. & Cronan, J. E. Expression and activity of the BioH esterase of biotin synthesis is independent of genome context. *Sci. Rep.* **7**, 2141 (2017).

38. Cronan, J. E. The classical, yet controversial, first enzyme of lipid synthesis: *Escherichia coli* acetyl-CoA carboxylase. *Microbiol. Mol. Biol. Rev.* **85**, e0003221 (2021).

39. Jeschek, M. et al. Biotin-independent strains of *Escherichia coli* for enhanced streptavidin production. *Metab. Eng.* **40**, 33–40 (2017).

40. Li, H., Chen, W., Jin, R., Jin, J.-M. & Tang, S.-Y. Biosensor-aided high-throughput screening of hyper-producing cells for malonyl-CoA-derived products. *Microb. Cell Fact.* **16**, 187 (2017).

41. Schaffitzel, C., Berg, M., Dimroth, P. & Pos, K. M. Identification of an  $\text{Na}^+$ -dependent malonate transporter of *Malonomonas rubra* and its dependence on two separate genes. *J. Bacteriol.* **180**, 2689–2693 (1998).

42. Hoenke, S., Schmid, M. & Dimroth, P. Sequence of a gene cluster from *Klebsiella pneumoniae* encoding malonate decarboxylase and expression of the enzyme in *Escherichia coli*. *Eur. J. Biochem.* **246**, 530–538 (1997).

43. Yang, 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).

44. Incha, M. R. et al. Leveraging host metabolism for bisdemethoxycurcumin production in *Pseudomonas putida*. *Metab. Eng. Commun.* **10**, e00119 (2020).

45. Bassalo, M. C. et al. Rapid and efficient one-step metabolic pathway integration in *E. coli*. *ACS Synth. Biol.* **5**, 561–568 (2016).

46. Dickinson, M. S., Miyazawa, T., McCool, R. S. & Keatinge-Clay, A. T. Priming enzymes from the pikromycin synthase reveal how assembly-line ketosynthases catalyze carbon–carbon chemistry. *Structure* **30**, 1331–1339 (2022).

47. Miyazawa, T., Fitzgerald, B. J. & Keatinge-Clay, A. T. Preparative production of an enantiomeric pair by engineered polyketide synthases. *Chem. Commun.* **57**, 8762–8765 (2021).

48. Hirsch, M., Fitzgerald, B. J. & Keatinge-Clay, A. T. How *cis*-acyltransferase assembly-line ketosynthases gatekeep for processed polyketide intermediates. *ACS Chem. Biol.* **16**, 2515–2526 (2021).

49. Ringlstetter, S. L. Identification of the biotin transporter in *Escherichia coli*, biotinylation of histones in *Saccharomyces cerevisiae* and analysis of biotin sensing in *Saccharomyces cerevisiae*. <https://doi.org/10.5283/EPUB.15822> (2011).

50. Takamura, Y. & Nomura, G. Changes in the intracellular concentration of acetyl-CoA and malonyl-CoA in relation to the carbon and energy metabolism of *Escherichia coli* K12. *J. Gen. Microbiol.* **134**, 2249–2253 (1988).

51. Badran, A. H. & Liu, D. R. Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).

52. Steel, H., Habgood, R., Kelly, C. L. & Papachristodoulou, A. In situ characterisation and manipulation of biological systems with Chi. *Bio. PLoS Biol.* **18**, e3000794 (2020).

53. Betz, J. L., Sasmor, H. M., Buck, F., Insley, M. Y. & Caruthers, M. H. Base substitution mutants of the lac operator: *in vivo* and *in vitro* affinities for lac repressor. *Gene* **50**, 123–132 (1986).

54. Fang, L. et al. Genome-scale target identification in *Escherichia coli* for high-titer production of free fatty acids. *Nat. Commun.* **12**, 4976 (2021).

55. Baek, J. H., Kang, Y. J. & Lee, S. Y. Transcript and protein level analyses of the interactions among PhoB, PhoR, PhoU and CreC in response to phosphate starvation in *Escherichia coli*. *FEMS Microbiol. Lett.* **277**, 254–259 (2007).

56. Rodríguez-Moyá, M. & Gonzalez, R. Proteomic analysis of the response of *Escherichia coli* to short-chain fatty acids. *J. Proteom.* **122**, 86–99 (2015).

57. Sah-Teli, S. K. et al. Complementary substrate specificity and distinct quaternary assembly of the *Escherichia coli* aerobic and anaerobic  $\beta$ -oxidation trifunctional enzyme complexes. *Biochem. J.* **476**, 1975–1994 (2019).

58. Campbell, J. W., Morgan-Kiss, R. M. & Cronan, J. E. Jr. A new *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic  $\beta$ -oxidation pathway. *Mol. Microbiol.* **47**, 793–805 (2003).

59. Pavoncello, V., Barras, F. & Bouveret, E. Degradation of exogenous fatty acids in *Escherichia coli*. *Biomolecules* **12**, 1019 (2022).

60. Agnew, D. E., Stevermer, A. K., Youngquist, J. T. & Pfleger, B. F. Engineering *Escherichia coli* for production of C12–C14 polyhydroxyalkanoate from glucose. *Metab. Eng.* **14**, 705–713 (2012).

61. Choe, D. et al. Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* **10**, 935 (2019).

62. LaCroix, R. A. et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. *Appl. Environ. Microbiol.* **81**, 17–30 (2015).

63. Tenaillon, O. et al. The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).

64. Sandberg, T. E., Lloyd, C. J., Palsson, B. O. & Feist, A. M. Laboratory evolution to alternating substrate environments yields distinct phenotypic and genetic adaptive strategies. *Appl. Environ. Microbiol.* **83**, e00410–e00417 (2017).

65. Sauer, D. B. et al. Structural basis of ion–substrate coupling in the  $\text{Na}^+$ -dependent dicarboxylate transporter VcINDY. *Nat. Commun.* **13**, 2644 (2022).

66. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. EggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).

67. Gui, L., Sunnarborg, A. & LaPorte, D. C. Regulated expression of a repressor protein: FadR activates *iclR*. *J. Bacteriol.* **178**, 4704–4709 (1996).

68. Brenna, J. T. Use of stable isotopes to study fatty acid and lipoprotein metabolism in man. *Prostaglandins Leukot. Essent. Fatty Acids* **57**, 467–472 (1997).

69. Gázquez, A. et al. In vivo kinetic study of materno–fetal fatty acid transfer in obese and normal weight pregnant women. *J. Physiol.* **597**, 4959–4973 (2019).

70. Wiechert, W.  $^{13}\text{C}$  metabolic flux analysis. *Metab. Eng.* **3**, 195–206 (2001).

71. Long, C. P. & Antoniewicz, M. R. High-resolution  $^{13}\text{C}$  metabolic flux analysis. *Nat. Protoc.* **14**, 2856–2877 (2019).

72. Acién Fernández, F. G. et al. Assessment of the production of  $^{13}\text{C}$  labeled compounds from phototrophic microalgae at laboratory scale. *Biomol. Eng.* **20**, 149–162 (2003).

73. Lippert, A. R., Keshari, K. R., Kurhanewicz, J. & Chang, C. J. A hydrogen peroxide-responsive hyperpolarized  $^{13}\text{C}$  MRI contrast agent. *J. Am. Chem. Soc.* **133**, 3776–3779 (2011).

74. Kurhanewicz, J. et al. Hyperpolarized  $^{13}\text{C}$  MRI: path to clinical translation in oncology. *Neoplasia* **21**, 1–16 (2019).

75. Cronan, J. E. Jr & Batchelor, J. G. An efficient biosynthetic method to prepare fatty acyl chains highly enriched with  $^{13}\text{C}$ . *Chem. Phys. Lipids* **11**, 196–202 (1973).

76. Fowler, Z. L., Gikandi, W. W. & Koffas, M. A. G. Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl. Environ. Microbiol.* **75**, 5831–5839 (2009).

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

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

© The Author(s) 2025

## Methods

### Chemical compounds and standards

All chemicals, reagents and solvents were purchased from commercial suppliers and used without further purification unless otherwise stated. Standard laboratory chemical reagents were obtained from Sigma-Aldrich or Thermo Fisher Scientific. Water used in all experiments was purified using a Milli-Q system. Specific sources for specialized reagents or specific solutions are listed below:

1. The 2.5 M stock solution of [<sup>13</sup>C]malonic acid (Santa Cruz Biotechnology) was prepared by the addition of NaOH to neutral pH.
2. The 2.5 M stock solution of disodium malonate (Spectrum Chemicals) was prepared in Milli-Q H<sub>2</sub>O and adjusted to a neutral pH.
3. M9 minimal medium–1× M9 salts, 2% glucose, 2 mM MgSO<sub>4</sub>, thiamine (1 µg ml<sup>-1</sup>), 0.1 mM CaCl<sub>2</sub> and filter sterilized.
4. 6-Ethyl-3,5-dimethylhydro-2H-pyran-2,4(3H)-dione was prepared as previously reported<sup>77</sup>.
5. 6-Ethyl-5-methylhydro-2H-pyran-2,4(3H)-dione was prepared as previously reported<sup>78</sup>.

### Cloning of expression plasmids

*matB* and *matC* were PCR-amplified from the plasmid pCKmatBC purchased from Addgene (138587)<sup>39</sup>. PCR reactions were performed using Phusion High-Fidelity DNA Polymerase 2X Master Mix (New England Biolabs (NEB)) or PrimeSTAR Max DNA Polymerase 2× Master Mix (Takara). Gibson Assembly was performed using NEBuilder HiFi DNA Assembly to assemble the plasmids—pBbS5k\_pik12(pik\_AT3), pBbA5a\_pik27, pR6K-ompW-MatBC, pR6K-bioH-RFP and pR6K-bioH-matBC. QuikChange Site-Directed Mutagenesis Kit was used to create the pR6K-ompW-RFP\* vector without an NdeI cut site in the backbone, allowing for NdeI/Xhol insertion of genes of interest by restriction digest and T4 ligation. Plasmids pBbA2c-MatB-MadLM and pBbA2c-MatB-MdcF were constructed by digesting the BglBrick vector pBbA2c\_RFP<sup>79</sup> with EcoRI and BamHI. *E. coli*-optimized sequences for the *mdcF* and *madL* and *madM* genes were ordered as gene fragments from Integrated DNA Technologies (IDT) with overhang regions encoding ribosomal binding site (RBS) sequences and EcoRI and BamHI restriction sites. Gene inserts were also restriction-digested and assembled using standard T4 ligation protocols. Subsequently, these intermediate plasmids were sequence-verified and then digested with EcoRI and BglII. The *E. coli* codon-optimized *matB* was also ordered, as well as a gene fragment from IDT with an RBS sequence and corresponding EcoRI and BglII restriction digest sites. The *matB* fragment was also restriction-digested and assembled with the *madL* and *mdcF* backbones using standard T4 ligation protocols. All plasmids and strains used in this study are listed in Supplementary Table 1. Gene-specific primers are listed in Supplementary Table 2. Specific gene sequences are listed in Supplementary Table 3. Restriction enzymes were used as recommended by the manufacturer (NEB). Assembled plasmids containing the R6K backbone were transformed into One Shot Pir2 chemically competent cells (Thermo Fisher Scientific). All other plasmids were transformed into XL1-Blue or NEB 5-α F'/q chemically competent cells. Single colonies of the transformants were cultured and miniprepped using the QIAprep Spin Miniprep Kit. All the plasmid sequences were validated by whole-plasmid sequencing from Primordium Labs.

### Integration of DNA into *E. coli*

Strains BAP1 and K207-3 were transformed with the temperature-sensitive red recombinase plasmid pKD46 (ref. 80) and cultured in LB medium with carbenicillin (100 µg ml<sup>-1</sup>) at 30 °C. Fresh pKD46 transformants were grown to OD<sub>600</sub> ~0.5, induced with 0.2% arabinose for 1 h and made electrocompetent by 100-fold concentration and three washes with ice-cold 10% glycerol.

pR6K plasmids containing homology arms and the gene of interest were PCR-amplified using integration primers (Supplementary Table 2), gel-purified (Zymo Research) and eluted in 10 µl nuclease-free water. Approximately 500 ng of PCR product was electroporated into cold electrocompetent cells in a 0.1 cm gap disposable cuvette using a Bio-Rad Micropulser (settings—EC1, 1.8 kV) and rescued with 1 ml SOC (Super Optimal broth with catabolite repression) medium at 30 °C for 2 h. The culture was concentrated to 100 µl, plated on LB + chloramphenicol (30 µg ml<sup>-1</sup>) and then incubated overnight at 30 °C. Transformants were screened by colony PCR with screen primers (Supplementary Table 2), and positive clones were cultured at 37 °C to remove pKD46.

Antibiotic resistance genes flanked by flippase recognition target sites were excised using the temperature-sensitive flip recombinase plasmid PCP20 (ref. 80). Transformants were plated on LB + carbenicillin (100 µg ml<sup>-1</sup>), grown overnight at 30 °C, and screened via colony PCR. Successful excision was confirmed by sequencing and restored sensitivity to chloramphenicol (30 µg ml<sup>-1</sup>). PCP20 was removed by restreaking on LB and incubating at 42 °C overnight.

For deletion of *fad* and biotin operon genes (*bioA*, *bioB*, *bioC*, *bioD* and *bioF*), pKD13 was amplified with FadI-K/O or Bio-K/O primers to generate dsDNA encoding kanamycin resistance flanked by the corresponding homology arms. PCR products were gel-purified and eluted in 10 µl nuclease-free water. pKD46 transformants of BAP1-ΔbioH-MatBC were grown to OD<sub>600</sub> ~0.5, induced with 0.2% arabinose for 1 h and made electrocompetent by 100-fold concentration and three washes with ice-cold 10% glycerol. Approximately 500 ng of PCR product was electroporated, and cells were rescued in SOC at 30 °C for 2 h. Transformants were screened via colony PCR, and positive clones were cultured at 37 °C to remove pKD46. Antibiotic resistance cassettes were excised using PCP20 as detailed above.

### Screening of malonate importers using plasmid-based systems

Electrocompetent BAP1 cells were transformed with plasmids pCKmatBC, pBbA2c-MatB-MadLM and pBbA2c-MatB-MdcF via electroporation (Micropulser; EC1, 1.8 kV). Cells were rescued in SOC (1 ml, 1 h, 37 °C), plated on LB + chloramphenicol (30 µg ml<sup>-1</sup>) and then incubated overnight at 37 °C. The pBADT-RppA-NT plasmid carrying a truncated *rppA* sequence from *S. coelicolor* for increased enzymatic activity was transformed into BAP1 strains with malonate importer plasmids using the same electroporation protocol<sup>44,81</sup>. Transformants were plated on LB + chloramphenicol (30 µg ml<sup>-1</sup>) + kanamycin (50 µg ml<sup>-1</sup>) and incubated overnight at 37 °C.

Initial attempts to induce the MadLM- and MdcF-containing strains with 10 nM aTc resulted in severe growth defects. As such, only the MatBC plasmid was induced with the addition of IPTG (10 µM) to the media. Cultures were grown in LB + 0.2% arabinose + 10 µM IPTG with 0, 1, 5 or 10 mM malonate. Absorbance at 340 nm was monitored over 3 days. Triplicate cultures were grown in 24-well deep-well plates (30 °C, 200 rpm). Aliquots (200 µl) were collected at 24, 48 and 72 h and centrifuged (5,000g for 10 min), and 100 µl supernatant was transferred to a 96-well black clear-bottom plate for absorbance measurement (340 nm and 520 nm; Biotek Synergy 4 Plate Reader). Supernatants were diluted as needed to remain within the linear detection range (Supplementary Fig. 1).

### Screening of integrated MatBC strains in LB and M9 media

The pBADT-RppA-NT plasmid was transformed into BAP1-ΔbioH-RFP, K207-3, BAP1-ΔbioH-MatBC and K207-3-MatBC strains, with selection on kanamycin (50 µg ml<sup>-1</sup>). Colonies were inoculated into LB + kanamycin (50 µg ml<sup>-1</sup>) and grown overnight at 37 °C. Cultures were diluted 1:100 into fresh LB or M9 minimal media with kanamycin (50 µg ml<sup>-1</sup>), 0.2% L-arabinose, 10 µM IPTG and increasing malonate concentrations (0–20 mM). Triplicate cultures were grown in 24-well

deep-well plates (30 °C at 200 rpm). Aliquots (200  $\mu$ l) were collected at 24, 48 and 72 h and centrifuged (5,000g for 10 min), and 100  $\mu$ l of supernatant was transferred to a 96-well black clear-bottom plate for absorbance measurement (340 nm and 520 nm; Biotek Synergy 4). Samples were diluted as needed to remain within the linear detection range.

To quantify flaviolin concentration, LB blank absorbance was subtracted from samples, and extinction coefficients ( $\epsilon_{340} = 3,068 \text{ M}^{-1} \text{ cm}^{-1}$  and  $\epsilon_{520} = 1,305 \text{ M}^{-1} \text{ cm}^{-1}$ ) were applied for approximation<sup>32</sup>.

### Flaviolin production in M9 minimal medium with malonate or [ $^{13}\text{C}$ ]malonate

A single BAP1- $\Delta$ bioH-MatBC colony carrying pBADT-RppA-NT was inoculated into LB + kanamycin (50  $\mu\text{g ml}^{-1}$ ) and grown overnight at 37 °C. The culture (1 ml) was centrifuged (13,000g), washed three times with M9 buffer, resuspended in 1 ml of M9 buffer and diluted 1:1,000 into 10 ml of fresh M9 medium containing 2% glucose, 0.2% L-arabinose, 10  $\mu\text{M}$  IPTG and kanamycin (50  $\mu\text{g ml}^{-1}$ ).

Aliquots (2 ml) were distributed into glass tubes and supplemented with the following five conditions: (1) control (no malonate/biotin), (2) 20 mM malonate, (3) 20 mM [ $^{13}\text{C}$ ]malonate, (4) 25 mg  $\text{l}^{-1}$  biotin and (5) 20 mM [ $^{13}\text{C}$ ]malonate + 25 mg  $\text{l}^{-1}$  biotin. Cultures were incubated at 37 °C for 3 days, with 50  $\mu\text{l}$  samples taken at 24, 48 and 72 h for LC-MS analysis.

Notably, for condition 1 (no malonate/biotin), the strain does not grow, whereas the strains supplemented with biotin or malonate turn from light pink on day 1 to visibly turbid and dark red by day 3 (Supplementary Fig. 5). OD<sub>600</sub> and flaviolin production were determined for the final 72 h time point by measuring the absorbance at 600 nm and 340 nm using a path length of 0.1 cm<sup>-1</sup>.

For LC-MS analysis, samples were mixed with 50  $\mu\text{l}$  LC-MS grade methanol + 0.1% formic acid and centrifuged (13,000g), and the supernatant was further purified using Amicon Ultra-0.5 centrifugal filters (3 kDa, 14,000g for 15 min), and the flow through was then transferred to LC-MS vials.

Samples were analyzed by a 1260 Infinity High Performance Liquid Chromatography (HPLC) system (Agilent Technologies) coupled to an Agilent LC-MSD and a 6520 Quadrupole Time-of-Flight Mass Spectrometer system (Agilent Technologies). In total, 10  $\mu\text{l}$  of the sample was injected onto a Phenomenex Kinetex XB-C18 (2.6  $\mu\text{m}$ , 100  $\times$  3 mm, 100 Å) LC column and separated with the following HPLC protocol: mobile phase A was composed of 0.1% formic acid in water, and mobile phase B was composed of 0.1% formic acid in methanol. The QTOF-MS acquisition parameters were as follows: drying and nebulizing gases were set to 11 l min<sup>-1</sup> and 30 lb in<sup>-2</sup>, respectively, and a drying-gas temperature of 330 °C. ESI was conducted in the negative ion mode with a capillary voltage of 3,500 V, scanning from 100 to 300  $m/z$ . Data acquisition (Workstation B.08.00) and processing were conducted using the Agilent MassHunter software package.

### Flaviolin production in evolved strains (M9 minimal medium and LB media)

Single colonies of BAP1- $\Delta$ bioH-RFP and BAP1- $\Delta$ bioH-MatBC and the evolved strains E1\_S2, E2\_S3, E2\_S7 and E2\_S17 (kanamycin resistance removed via pcp20-mediated FLP recombination) were transformed with pBADT-RppA-NT and selected on kanamycin. Colonies were inoculated into LB + kanamycin (50  $\mu\text{g ml}^{-1}$ ) and grown overnight at 37 °C.

Cultures (1 ml) were centrifuged (13,000g), washed three times, resuspended in 1 ml M9 minimal medium and then diluted 1:1,000 into four replicates in 24-well deep-well plates. Cultures were incubated at 37 °C, 200 rpm in the following:

1. M9 minimal medium + 10 mM malonate, 2% glucose, 0.2% L-arabinose, 1× trace elements and 50  $\mu\text{g ml}^{-1}$  kanamycin.
2. LB + 0.2% L-arabinose and 50  $\mu\text{g ml}^{-1}$  kanamycin.

### 3. LB + 10 mM malonate, 0.2% L-arabinose and 50 $\mu\text{g ml}^{-1}$ kanamycin.

Cultures were shaken at 37 °C, 200 rpm for 4 days, with 50  $\mu\text{l}$  supernatant samples taken every 24 h, diluted 1:1 with Milli-Q H<sub>2</sub>O and transferred to a 96-well black clear-bottom plate for absorbance measurement (340 nm and 520 nm; Biotek Synergy Plate Reader). Samples exceeding the linear detection range were further diluted. To quantify flaviolin concentration, the media blank absorbance was subtracted. Absorbance at 520 nm was used instead of 340 nm to correct for background differences between LB and M9 minimal media.

### Pikromycin hybrid PKS fermentation

For the pik127 PKSs, the three-module PKS is split across two plasmids. For the hybrid PKS construct that has the M-CoA loading at the first extension module, the combination of the plasmids pBbS5k\_pik12(pik\_AT3) and pBbA5a\_pik27 in the K207-3 and K207-3-MatBC results in the strains KCP9 and KCP9-MatBC. For the hybrid PKS construct that has the M-CoA loading at the second extension module, the combination of the plasmids pBbS5k\_pik12 and pBbA5a\_pik27(pik\_AT3) in the K207-3 and K207-3-MatBC strains is also reported in ref. 78 to create KCP3 and KCP3-MatBC.

Dual-plasmid strains (KCP3, KCP3-MatBC, KCP9 and KCP9-MatBC) were selected on LB + spectinomycin and kanamycin (50  $\mu\text{g ml}^{-1}$ ). Colonies were grown overnight in LB + carbenicillin (100  $\mu\text{g ml}^{-1}$ ) and kanamycin (50  $\mu\text{g ml}^{-1}$ ) at 30 °C and then diluted 1:100 into 2 ml fresh EZ-rich medium with the same antibiotics. Triplicate cultures for each condition were incubated in 24-well deep-well plates at 37 °C, 200 rpm.

At OD<sub>600</sub> ~0.6, cultures were cooled to 18 °C and induced with 0.1 mM IPTG. Upon induction, 20 mM propionate was added to all wells, and 20 mM malonate was added in experimental conditions with corresponding no-malonate controls. Fermentations were run in triplicate. After 6 days (18 °C, 200 rpm), plates were frozen at -80 °C for further analysis.

Frozen cultures were thawed, and 200  $\mu\text{l}$  was transferred to a 96-well plate containing 200  $\mu\text{l}$  of 10  $\mu\text{M}$  internal standard TAL in acetonitrile (ACN) and 2  $\mu\text{l}$  concentrated HCl. Compounds were extracted by incubating at 37 °C, 200 rpm for 30 min. Cell debris was removed by centrifugation (5,000g for 20 min), and supernatants were filtered using Amicon Ultra-0.5 centrifugal filters (3 kDa, 14,000g for 15 min). The flow through was then transferred to LC-MS vials and analyzed using a 1260 Infinity HPLC system (Agilent Technologies) coupled to an Agilent Liquid Chromatography Mass Spectrometry Detector (LC-MSD) and a 6520 Quadrupole Time-of-Flight Mass Spectrometer (QTOF-MS) system (Agilent Technologies). In total, 10  $\mu\text{l}$  of the sample was injected onto a Phenomenex Kinetex XB-C18 (2.6  $\mu\text{m}$ , 100  $\times$  3 mm, 100 Å) LC column and separated with the following HPLC protocol: mobile phase A was composed of 0.1% formic acid in water, and mobile phase B was composed of 0.1% formic acid in methanol. The QTOF-MS acquisition parameters were as follows: drying and nebulizing gases were set to 11 l min<sup>-1</sup> and 30 lb in<sup>-2</sup>, respectively, and a drying-gas temperature of 330 °C. ESI was conducted in the negative ion mode with a capillary voltage of 3,500 V, scanning from 100 to 300  $m/z$ . Data acquisition (Workstation B.08.00) and processing were conducted using Agilent MassHunter software package.

### Free FA quantification

BAP1- $\Delta$ bioH-RFP and BAP1- $\Delta$ bioH-MatBC cultures were grown in M9 minimal medium with 2% glucose, 10  $\mu\text{M}$  IPTG and 20 mM malonate in 24-well plates (2.5 ml, 37 °C, 200 rpm, 20 h). For BAP1- $\Delta$ bioH-RFP, 25 mg  $\text{l}^{-1}$  biotin was added. All strains and conditions were tested in triplicate, with the experiment performed twice ( $n = 6$ ).

For sample preparation, 100  $\mu\text{l}$  of culture was first removed to measure OD<sub>600</sub> for cell density determination. Then, 500  $\mu\text{l}$  of culture

was spiked with 5  $\mu$ l of 25 mM nonanoic acid as an internal standard (final concentration—250  $\mu$ M), followed by acidification with 100  $\mu$ l glacial acetic acid. After vortexing, 2 ml of ethyl acetate was added for FA extraction, and samples were incubated for 30 min at 200 rpm. Following centrifugation at 4,000g for 10 min, the organic layer was transferred to a screw-top tube and left to evaporate overnight in a fume hood.

To derivatize the extracted FAs, 200  $\mu$ l of 1.25 M anhydrous HCl in methanol was added to each dried sample. Tubes were tightly capped and incubated at 80 °C for 1 h. After cooling, 500  $\mu$ l of hexane was added, and samples were vortexed before centrifugation at 15,000g for 1 min. The hexane layer was collected for GC–MS analysis.

For detection, 2  $\mu$ l of the sample was injected using a CTC Analytics PAL System automated sampler onto an Agilent 5973 Network Mass Selective Detector GC–MS. The GC temperature program was as follows: initial  $T = 50$  °C, ramp 10 °C min $^{-1}$  to 160 °C, ramp 30 °C min $^{-1}$  to 230 °C hold 1 min, ramp to 300 °C hold 3 min, return to 50 °C. MS acquisition parameters were as follows: mass detection range,  $m/z = 50$ –350 Da.

EICs of  $m/z$  74 Da, corresponding to the McLafferty rearrangement ion of FA methyl esters, were used to identify peaks. FA peaks were integrated and quantified using calibration curves for C12, C14, C16 and C18 FAs, with nonanoic acid as the internal standard.

### Proteomics analysis

Four biological replicate cultures of BAP1– $\Delta$ bioH–MatBC, BAP1– $\Delta$ bioH–MatBC– $\Delta$ bioH–kan, E1\_S2 and E2\_S3 were grown in 2 ml M9 minimal medium with 2% glucose and 10 mM malonate. All strains were cultivated in M9 minimal medium with 2% glucose and 10 mM malonate for 48 h at 37 °C after inoculation at an  $OD_{600}$  of 0.001. In total, 1 ml of cell culture was collected by centrifugation at 5,000g for 15 min, and the resulting cell pellets were stored at –80 °C until further processing.

Protein was extracted from cell pellets, and tryptic peptides were prepared by following the established proteomic sample preparation protocol<sup>83</sup>. Briefly, cell pellets were resuspended in Qiagen P2 Lysis Buffer (Qiagen) to promote cell lysis. Proteins were precipitated with the addition of 1 mM NaCl and 4 $\times$  vol acetone, followed by two additional washes with 80% acetone in water. The recovered protein pellet was homogenized by pipetting and mixing with 100 mM ammonium bicarbonate in 20% methanol. Protein concentration was determined by the detergent compatible (DC) protein assay (Bio-Rad). Protein reduction was accomplished using 5 mM tris 2-(carboxyethyl)phosphine for 30 min at room temperature, and alkylation was performed with 10 mM iodoacetamide (final concentration) for 30 min at room temperature in the dark. Overnight digestion with trypsin was performed at a 1:50 trypsin-to-total protein ratio. The resulting peptide samples were analyzed on an Agilent 1290 Ultra High Performance Liquid Chromatography (UHPLC) system coupled to a Thermo Scientific Orbitrap Exploris 480 mass spectrometer for discovery proteomics<sup>84</sup>. Briefly, peptide samples were loaded onto an Ascentis ES-C18 Column (Sigma-Aldrich) and were eluted from the column by using a 10-min gradient from 98% solvent A (0.1% FA in H<sub>2</sub>O) and 2% solvent B (0.1% FA in ACN) to 65% solvent A and 35% solvent B. Eluting peptides were introduced to the mass spectrometer operating in positive-ion mode and were measured in data-independent acquisition (DIA) mode with a duty cycle of three survey scans from  $m/z$  380 to  $m/z$  985 and 45 MS2 scans with precursor isolation width of 13.5  $m/z$  to cover the mass range. DIA raw data files were analyzed by an integrated software suite, DIA-NN<sup>85</sup>. The database used in the DIA-NN search (library-free mode) is *E. coli*'s latest UniProt proteome FASTA sequences plus the protein sequences of the heterologous proteins and common proteomic contaminants. DIA-NN determines mass tolerances automatically based on first-pass analysis of the samples with automated determination of optimal mass accuracies. The retention time extraction window was determined individually for all MS runs analyzed via the automated optimization

procedure implemented in DIA-NN. Protein inference was enabled, and the quantification strategy was set to robust LC = high accuracy. The output main DIA-NN reports were filtered with a global false discovery rate (FDR) = 0.01 on both the precursor level and protein group level. The Top3 method, which is the average MS signal response of the three most intense tryptic peptides of each identified protein, was used to plot the quantity of the targeted proteins in the samples<sup>86,87</sup>.

The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE<sup>88</sup> partner repository with the dataset identifier [PXD060741](https://doi.org/10.6019/PXD060741). DIA-NN is freely available for download from <https://github.com/vdemichev/DiaNN>.

Significant differences between the proteomic profiles of the strains were determined by using a Welch's *t*-test and a cutoff of  $>\pm 1$ -fold and a *P* value of  $<0.05$ . Significantly up and down proteins were then annotated when function could be identified using their COG name with the eggNOG-Mapper V2 tool (<http://eggnog-mapper.embl.de/>)<sup>66</sup>. Excel files containing all the significantly up and down proteins for each comparison are provided in Supplementary Data 1.

### Quantification of M-CoA and Ac-CoA

Cultures of BAP1 and BAP1– $\Delta$ bioH–MatBC were grown in 3.5 ml of M9 minimal medium supplemented with 2% glucose and 10  $\mu$ M IPTG, with increasing concentrations of malonate (0–100 mM) prepared from a 2.5 M sodium malonate dibasic stock solution. Each condition was tested in triplicate using 24-well deep-well plates covered with breathable film. Cultures were incubated at 37 °C with shaking at 200 rpm for 36 h. After incubation, 25  $\mu$ l of culture was removed, diluted four-fold and  $OD_{600}$  was measured and converted to DCW using the factor 1  $OD_{600} = 0.396$  g DCW<sup>89</sup>. The remaining culture was centrifuged at 4 °C for 10 min to pellet cells. The supernatant was collected for HPLC analysis of malonate, glucose and acetate concentrations, while the cell pellet was frozen at –80 °C for later analysis. Before LC–MS analysis, frozen cell pellets were thawed on ice. To each pellet, 350  $\mu$ l of a 45:45:10 ACN:MeOH:H<sub>2</sub>O solution containing 0.1% glacial acetic acid was added. Cells were resuspended by repetitive pipetting, followed by centrifugation at 13,000g at 4 °C for 15 min. A 100  $\mu$ l aliquot of the supernatant was transferred to LC–MS vials for subsequent analysis, where 10  $\mu$ l of sample was injected onto a C18 (2.6  $\mu$ m, 100  $\times$  3 mm, 100  $\text{\AA}$ ) LC column, and analyzed with the following HPLC protocol: buffer A—water with 20 mM ammonium acetate; buffer B—MeOH with 20 mM ammonium acetate; flow rate—0.60 ml min $^{-1}$ ; 2% buffer B for 1 min, 2–85% buffer B gradient for 3.0 min, 85–90% buffer B gradient for 2.0 min, 90–95% buffer B gradient for 2.0 min, 95–2% buffer B for 0.2 min, re-equilibration at 2% buffer B for 4.8 min. MS acquisition parameters were as follows: mass-to-charge ratios corresponding to 854 Da (M-CoA) and 810 Da (Ac-CoA) were monitored in SIM positive mode.

### ALE of BAP1– $\Delta$ bioH–MatBC

**One-week mutagenesis turbidostat.** The BAP1– $\Delta$ bioH–MatBC strain was transformed with the mutagenesis-enhancing plasmid MP6, and transformants were selected on LB agar containing chloramphenicol (30  $\mu$ g ml $^{-1}$ ). A single colony was grown overnight in M9 minimal medium with 2% glucose and 10 mM malonate. The culture was then diluted 1:100 into fresh M9 minimal medium supplemented with 2% glucose, 30  $\mu$ g ml $^{-1}$  chloramphenicol, 1% arabinose and 1 mM malonate. The culture was maintained in a Chi.Bio turbidostat at an  $OD_{600}$  of 0.75 through continuous dilution over 7 days<sup>52</sup>.

On day 7, an aliquot of the evolved strain was streaked onto M9 minimal agar plates containing 3 mM malonate and incubated at 37 °C. Individual colonies were subsequently inoculated into LB, M9 minimal medium and M9 minimal medium with 1 mM malonate and grown at 37 °C for 24 h. Mutants were identified by their inability to grow in M9 minimal medium alone but increased growth in malonate-supplemented conditions. The genomes of six mutants with the highest  $OD_{600}$  values (E1\_S1–S6) were selected for sequencing.

**Four-week mutagenesis serial dilution.** The BAP1- $\Delta$ bioH-MatBC- $\Delta$ bio-kan strain was transformed with MP6, and transformants were isolated on LB agar with chloramphenicol (30  $\mu$ g ml $^{-1}$ ). A single colony was grown overnight in M9 minimal medium supplemented with 2% glucose, 50  $\mu$ g ml $^{-1}$  kanamycin and 20 mM malonate. The culture was then diluted 1:100 into 5 ml of fresh M9 medium containing 2% glucose, 30  $\mu$ g ml $^{-1}$  chloramphenicol, 50  $\mu$ g ml $^{-1}$  kanamycin, 1% arabinose and 20 mM malonate in a 30 ml glass culture tube. The culture was maintained through serial dilution over 14 days.

On day 14, an aliquot was saved as a glycerol stock, streaked onto kanamycin ( $50 \mu\text{g ml}^{-1}$ ) agar plates and incubated at  $37^\circ\text{C}$ . The remaining culture was split into two new tubes containing fresh M9 minimal medium with 2% glucose,  $30 \mu\text{g ml}^{-1}$  chloramphenicol,  $50 \mu\text{g ml}^{-1}$  kanamycin and 1% arabinose, with either 1 mM or 20 mM malonate added. These cultures were maintained through serial dilution for an additional 14 days.

At the end of the experiment (day 28), aliquots of evolved strains were saved as glycerol stocks and streaked onto kanamycin agar plates (50  $\mu$ g ml<sup>-1</sup>). Individual colonies from the three plates were inoculated into M9 minimal medium supplemented with 10 mM malonate and grown at 37 °C for 24 h. A total of 17 colonies (E2\_S1–S17) were selected for whole-genome sequencing.

## Whole-genome sequencing

**Samples from 1-week mutagenesis turbidostat.** Using the method adapted from ref. 90, genomic DNA was prepared from cultured strains using a Wizard Genomic DNA Purification Kit. In total, 100 ng of DNA was sheared to 600 bp using the fragmentation enzyme in the xGen DNA Lib Prep EZ Kit (Integrated DNA Technologies), and size was selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing and ligation of Illumina-compatible adapters (Integrated DNA Technologies) using the xGen DNA Lib Prep EZ Kit. Bioanalyzer High Sensitivity DNA Kit (Agilent) and Qubit Fluorometers (Thermo Fisher Scientific) were used to determine the concentration of the libraries. Libraries were sequenced on the Illumina MiSeq.

**Samples from 2- to 4-week mutagenesis serial dilution.** Overnight cultures were diluted to  $4\text{--}6 \times 10^9$  cells per ml and then centrifuged at 13,000g to pellet the cells. The cell pellet was washed once in 1 ml of 1× PBS and resuspended in 500  $\mu\text{l}$  of Zymo 1× DNA/RNA Shield preservative and submitted for bacterial whole-genome sequencing performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

## Hierarchical clustering and dendrogram construction

To analyze the genetic relationships between the 23 *E. coli* strains based on their mutation profiles, a hierarchical clustering analysis was performed using a presence-absence mutation matrix. In this binary matrix, rows represented specific mutations and columns corresponded to individual strains, with a value of 1 indicating the presence of a mutation and 0 indicating its absence.

For data processing, the mutation matrix was transposed so that each strain was represented as a binary vector corresponding to its mutation profile. Hierarchical clustering was conducted on this transposed matrix using the Morpheus online platform (<https://software.broadinstitute.org/morpheus>)<sup>91</sup>. The dendrogram was calculated using one minus the Pearson correlation and an average linkage distance.

## Structure prediction

The sequence for MatC was trimmed at the N and C termini to remove disordered regions. AlphaFold 2.3 was used to predict the protein structure of MatC monomers using the configuration scheme in Table 1. All models were run with refinement disabled and number of recycles set to 1. All five default models from the 128 random seeds predicted were

**Table 1 | AlphaFold 2 configurations for predicting MatC conformations**

Group	max_msa	max_extra_msa	Dropout enabled	Random seeds
1	16	32	Yes	16
2	16	32	No	16
3	32	64	Yes	16
4	32	64	No	16
5	64	128	Yes	16
6	64	128	No	16
7	256	512	Yes	16
8	256	512	No	16

used in data analysis, totaling 640 structures. A pairwise root mean square deviation (RMSD) matrix was constructed using alignments through the PyMOL Python package. A two-dimensional projection of the matrix was constructed using multidimensional scaling and transformed using principal components analysis (PCA) with the Scikit-learn Python package. Points were clustered using the DBSCAN algorithm, and representative cluster control models were selected via a script identifying structures with the most similarity to all others within their respective clusters (Supplementary Fig. 8).

All visualization was done with Python scripts through PyMOL (the PyMOL Molecular Graphics System, v.3.0; Schrödinger).

## Reporting summary

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

## Data availability

The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE<sup>88</sup> partner repository with the dataset identifier [PXD060741](#). DIA-NN is freely available for download from <https://github.com/vdemichev/DiaNN>.

All strains and their corresponding registry IDs used in this study are listed in Supplementary Table 1 and are available through the Joint Bio-Energy Institute's Inventory of Composable Elements, an open-source registry software and platform for managing information about biological parts (<https://public-registry.jbei.org/folders/894>). Additionally, whole-genome sequencing files of the evolved mutants are linked to the corresponding strains.

Data are available from the corresponding authors upon request. Source data are provided with this paper.

## References

77. Englund, E. et al. Expanding extender substrate selection for unnatural polyketide biosynthesis by acyltransferase domain exchange within a modular polyketide synthase. *J. Am. Chem. Soc.* **145**, 8822–8832 (2023).
78. Wang, Z. et al. A microbial platform for recyclable plastics with customizable properties. Preprint at *Research Square* <https://doi.org/10.21203/rs.3.rs-3171588/v1> (2023).
79. Lee, T. S. et al. BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *J. Biol. Eng.* **5**, 12 (2011).
80. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
81. Izumikawa, M. et al. Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from *Streptomyces coelicolor* A3(2). *J. Ind. Microbiol. Biotechnol.* **30**, 510–515 (2003).

82. Krauser, S., Kiefer, P. & Heinzle, E. Multienzyme whole-cell *in situ* biocatalysis for the production of flaviolin in permeabilized cells of *Escherichia coli*. *ChemCatChem* **4**, 786–788 (2012).
83. Chen, Y. et al. Alkaline-SDS cell lysis of microbes with acetone protein precipitation for proteomic sample preparation in 96-well plate format. *PLoS ONE* **18**, e0288102 (2023).
84. Chen, Y., Gin, J. & J Petzold, C. Discovery proteomic (DIA) LC-MS/MS data acquisition and analysis. *protocols.io* <https://doi.org/10.17504/protocols.io.e6nwwk1z7vmk/v2> (2022).
85. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).
86. Ahrné, E., Molzahn, L., Glatter, T. & Schmidt, A. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. *Proteomics* **13**, 2567–2578 (2013).
87. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteom.* **5**, 144–156 (2006).
88. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
89. Myers, J. A., Curtis, B. S. & Curtis, W. R. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys.* **6**, 4 (2013).
90. Zhan, C. et al. Improved polyketide production in *C. glutamicum* by preventing propionate-induced growth inhibition. *Nat. Metab.* **5**, 1127–1140 (2023).
91. Morpheus. [software.broadinstitute.org/morpheus](https://software.broadinstitute.org/morpheus) (accessed 10 September 2024).

## Acknowledgements

We thank R. Kuo, Y. Liang and J. McCauley (Joint BioEnergy Institute) for the development and operation of the whole-genome sequencing pipeline. This material is based upon work supported by the National

Science Foundation (NSF; grant 2036849 to S.H.K. and M.W.), the NSF Graduate Research Fellowship Program (under grant DGE 2146752 to N. Lanclos) and by the Joint BioEnergy Institute (<https://www.jbei.org>), which is supported by the US Department of Energy, Office of Science, Biological and Environmental Research Program under contract DE-AC02-05CH11231, and the Philomathia Foundation (to J.D.K.).

## Author contributions

S.H.K. conceptualized the project, developed the methodology and wrote the original draft of the manuscript. S.H.K., M.W., A.E.C., S.C., N. Lanclos, N. Lee, Z.W., Y.C., J.W.G. and C.J.P. conducted the investigation. S.H.K. and J.D.K. did the writing, reviewing and editing of the manuscript. J.D.K. arranged the resources and provided supervision. All authors revised and approved the manuscript.

## Competing interests

J.D.K. has a financial interest in Demetrix, Lygos, Napigen, Berkeley Yeast, Zero Acre Farms, Ansa Biotechnologies, Apertor Pharmaceuticals, ResVita Bio, BioMia and Cyklos Materials. The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41589-025-01911-6>.

**Correspondence and requests for materials** should be addressed to Jay D. Keasling.

**Peer review information** *Nature Chemical Biology* thanks Vincent de Boer, Nicolai Kallscheuer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

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

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

### 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 Bitek Synergy 4, Agilent MassHunter software package Workstation B.08.00

Data analysis Graph Pad 10.4.1, Agilent MassHunter software package Workstation B.08.00., software suite DIA-NN, Morphesu: <https://software.broadinstitute.org/morpheus>

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

### Data

Policy information about [availability of data](#)

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

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

Data availability statement:

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE14 partner repository with the dataset identifier PXD060741. DIA-NN is freely available for download from <https://github.com/vdemichev/DiaNN>.

All strains and their corresponding registry IDs used in this study are listed in Supplementary Table 1 and are available through the Joint BioEnergy Institute's Inventory of Composable Elements (ICE), an open-source registry software and platform for managing information about biological parts (<https://public-registry.jbei.org/folders/894>). Additionally whole genome sequencing files of the evolved mutants are linked to the corresponding strains.

Data is available from the corresponding authors upon request.

## Research involving human participants, their data, or biological material

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

Reporting on sex and gender

N/A

Reporting on race, ethnicity, or other socially relevant groupings

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

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

## Field-specific reporting

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

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size

for most experiments between 3–6 replicates are used to basic statistical analysis, including the calculation of standard deviation and error bars

Data exclusions

only one data point is excluded in this text, 1 replicate out of 6 was physically dropped on the floor prior to GCMS

Replication

Experiments on production titers were repeated multiple times ( $\geq 2$ ), and while exact titers varied, the trends across strains and supplementation conditions remained consistent. Experiments not replicated included proteomics, whole genome sequencing, adaptive laboratory evolution

Randomization

Samples were allocated based on genetic composition (genomic or plasmid-bearing). Experimental conditions, including media composition and additives, were standardized across all strains to ensure consistency and control for potential confounding variables.

Blinding

Blinding was not implemented in this study as the data collected was primarily empirical and not subject to interpretation by the researchers

## Behavioural & social sciences study design

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

Study description

Research sample

Sampling strategy

Data collection

Timing

Data exclusions

Non-participation

Randomization

# Ecological, evolutionary & environmental sciences study design

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

Study description	<input type="text"/>
Research sample	<input type="text"/>
Sampling strategy	<input type="text"/>
Data collection	<input type="text"/>
Timing and spatial scale	<input type="text"/>
Data exclusions	<input type="text"/>
Reproducibility	<input type="text"/>
Randomization	<input type="text"/>
Blinding	<input type="text"/>

Did the study involve field work?  Yes  No

## Field work, collection and transport

Field conditions	<input type="text"/>
Location	<input type="text"/>
Access & import/export	<input type="text"/>
Disturbance	<input type="text"/>

## 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		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

## Antibodies

Antibodies used	<input type="text"/> N/A
Validation	<input type="text"/> N.A.

## Eukaryotic cell lines

Policy information about [cell lines](#) and [Sex and Gender in Research](#)

Cell line source(s)

Authentication

Mycoplasma contamination

Commonly misidentified lines  
(See [ICLAC](#) register)

## Palaeontology and Archaeology

Specimen provenance

Specimen deposition

Dating methods

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

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

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

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

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

Outcomes

## Dual use research of concern

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

### Hazards

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

No Yes

- Public health
- National security
- Crops and/or livestock
- Ecosystems
- Any other significant area

## Experiments of concern

Does the work involve any of these experiments of concern:

No Yes

- Demonstrate how to render a vaccine ineffective
- Confer resistance to therapeutically useful antibiotics or antiviral agents
- Enhance the virulence of a pathogen or render a nonpathogen virulent
- Increase transmissibility of a pathogen
- Alter the host range of a pathogen
- Enable evasion of diagnostic/detection modalities
- Enable the weaponization of a biological agent or toxin
- Any other potentially harmful combination of experiments and agents

## Plants

Seed stocks

Novel plant genotypes

Authentication

## ChIP-seq

### Data deposition

Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

*May remain private before publication.*


Files in database submission

Genome browser session

(e.g. [UCSC](#))

## Methodology

Replicates

Sequencing depth

Antibodies

Peak calling parameters

Data quality

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

## Magnetic resonance imaging

### Experimental design

Design type

Design specifications

Behavioral performance measures

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI

Used

Not used

### Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

### Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis:  Whole brain  ROI-based  Both

Statistic type for inference

(See [Eklund et al. 2016](#))

Correction

## Models & analysis

n/a Involved in the study

<input type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis

