Construction and modular implementation of the THETA cycle for synthetic CO_2 fixation

Check for updates

Shanshan Luo [1] ✉, Christoph Diehl [1], Hai He [1], YoungJun Bae[1,2],
Melanie Klose[1], Peter Claus[3], Niña Socorro Cortina [1,4],
Celia Alvarez Fernandez [1,5], Helena Schulz-Mirbach [1], Richard McLean[1],
Adán Andrés Ramírez Rojas [6], Daniel Schindler [6,7], Nicole Paczia [3] &
Tobias J. Erb [1,7] ✉

Synthetic biology offers the opportunity to build solutions for improved capture and conversion of carbon dioxide (CO_2) that outcompete those evolved by nature. Here we demonstrate the design and construction of a new-to-nature CO_2-fixation pathway, the reductive tricarboxylic acid branch/4-hydroxybutyryl-CoA/ethylmalonyl-CoA/acetyl-CoA (THETA) cycle. The THETA cycle encompasses 17 enzymes from 9 organisms and revolves around two of the most efficient CO_2-fixing enzymes described in nature, crotonyl-CoA carboxylase/reductase and phosphoenolpyruvate carboxylase. Here using rational and machine learning-guided optimization approaches, we improved the yield of the cycle by two orders of magnitude and demonstrated the formation of different biochemical building blocks directly from CO_2. Furthermore, we separated the THETA cycle into three modules that we successfully implemented in vivo by exploiting the natural plasticity of *Escherichia coli* metabolism. Growth-based selection and/or ^{13}C-labelling confirmed the activity of three different modules, demonstrating the first step towards realizing highly orthogonal and complex CO_2-fixation pathways in the background of living cells.

Biological carbon fixation assimilates more than 380 gigatons of carbon dioxide (CO_2) annually and plays an essential role in the global carbon cycle as a major carbon sink[1]. This makes biological carbon fixation key for any efforts of re-balancing the global carbon cycle and mitigating the effects of climate change[2]. In nature, plants, algae and other autotrophs convert CO_2 into organic molecules through CO_2-fixation pathways. So far, seven natural CO_2-fixation pathways have been discovered[3,4]. All of them have different physiological properties and are adapted to certain environmental conditions[5]. Despite this biological diversity, nature has only occupied a very limited solution space, indicating that many theoretically possible CO_2-fixation pathways have not been explored or were selected for by evolution[6].

With the emergence of synthetic biology, it has become possible to realize these new-to-nature solutions from the first principles by combining freely the entire repertoire of enzymes, reactions and mechanisms[7]. So far, more than 30 new-to-nature CO_2-fixation pathways have been designed theoretically[6,8,9]. Two oxygen-insensitive designs, the crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle and the reductive glyoxylate and pyruvate synthesis-malyl-CoA-glycerate (rGPS-MCG) cycle, have been realized

in vitro[10,11], offering different possibilities for sustainable biosynthesis and cell-free biology[12,13].

Yet, an open challenge is the feasibility of new-to-nature pathways in vivo. Importantly, this challenge scales with the complexity and orthogonality of the pathway design, as the potential to create more unwanted interactions with the native genetic and metabolic network of a cell increases with increasing size of the metabolic network[14]. While naturally existing CO_2-fixation pathways have been successfully transferred to *Escherichia coli* and other model organisms through the expression of a handful genes[15–18], the proof of principle is still outstanding for new-to-nature CO_2-fixation pathways that feature several orthogonal metabolites and reactions that do not share any overlap with the central carbon metabolism of the host. The implementation of new-to-nature CO_2-fixation pathways in vivo, however, is a necessary step, not only to broaden their applications and impact, but also to examine their real-life performance and compare them with their natural counterparts to gain more insights into the evolution and limitation of carbon fixation in the context of living cells.

In this Article, we show the successful design, realization and optimization of a synthetic CO_2-fixation pathway, the reductive tricarboxylic acid branch/4-hydroxybutyryl-CoA/ethylmalonyl-CoA/acetyl-CoA (THETA) cycle, by using rational design and machine learning. The THETA cycle converts CO_2 directly into the central building block acetyl-CoA, which makes it potentially a versatile platform for the synthesis of value-added compounds from CO_2 (ref. [12]). Notably, we also demonstrate how the THETA cycle can be modularized and transferred to *E. coli*, which paves the way for the complete implementation and adaptive evolution of this new-to-nature CO_2-fixation cycle in the future.

Results

Design of the THETA cycle

For the design of a new-to-nature CO_2-fixation pathway, we decided to rely on carboxylases with favourable catalytic properties. Among all naturally existing carboxylases, phosphoenolpyruvate (PEP) carboxylase (Ppc; reaction number 2 in THETA) and crotonyl-CoA carboxylase/reductase (Ccr; reaction number 11) stand out in respect to kinetic, thermodynamic and mechanistic considerations[8,19]. Ppc and Ccr are the fastest CO_2-fixing enzymes known so far, outcompeting ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) by more than one order of magnitude[8,10]; both enzymes catalyse reactions favouring the carboxylation direction under physiological conditions ($\Delta_r G'^m$ (Ppc carboxylation) of −44.4 kJ mol⁻¹, $\Delta_r G'^m$ (Ccr carboxylation) of −30.2 kJ mol⁻¹; Supplementary Fig. 1), adding a strong thermodynamic driving force, and show no side-reactivity or inhibition by oxygen, which made them prime candidates for our design.

Ppc catalyses the carboxylation of PEP into the C_4 compound oxaloacetate (OAA). It serves in anaplerosis of the tricarboxylic acid (TCA) cycle, facilitates CO_2 shuttling in C_4 and Crassulacean acid metabolism plants, and operates in the autotrophic dicarboxylate/4-hydroxybutyrate (DC/4HB) cycle in archaea. Ccr on the other hand catalyses the reductive carboxylation of the C_4-compound crotonyl-CoA into the C_5-compound (2S)-ethylmalonyl-CoA. The enzyme operates in the anaplerotic ethylmalonyl-CoA pathway for acetyl-CoA assimilation[20], as well as in polyketide biosynthesis[21,22], and has been used in the successful construction of a synthetic CO_2-fixation pathway recently[10].

To design a CO_2-fixation cycle based on these two carboxylation reactions, we first expanded each carboxylation reaction into a pathway module, and then linked the two modules together to form a continuous cycle with acetyl-CoA as the output molecule (Fig. 1). Based on the Ppc reaction, we drafted a pathway module (module 1) to convert the C_3-compound pyruvate into the C_4-compound succinate through PEP synthase (reaction number 1), Ppc and three reactions from the reductive branch of the TCA cycle. Expanding the Ccr reaction with three

reactions of the ethylmalonyl-CoA pathway, and three reactions of the 3-hydroxypropionate (3HP) bicycle gave rise to the other carboxylation module (module 3) that converts the C_4-compound crotonyl-CoA into the C_2-compound acetyl-CoA and the C_3-compound pyruvate. To link these two carboxylation modules, we designed another module (module 2) that transforms succinate to crotonyl-CoA via succinyl-CoA ligase (SucCD; reaction number 6) and four reactions from both of the DC/4HB and 3HP/4HB cycles.

Together, these three modules (comprising 17 reactions) form a CO_2-fixation cycle with the overall reaction: $CO_2 + HCO_3^- + 4ATP + 3NADPH + 2NADH + CoA \rightarrow$ acetyl-CoA + $FADH_2$ (Fig. 1), which we call the THETA cycle (Fig. 1) (for its comparison with natural and realized synthetic CO_2-fixation pathways, see Supplementary Table 1). Max–min driving force (MDF) analysis[23] within physiological metabolite concentration ranges confirmed its thermodynamic feasibility for in vivo implementation, for example, in *E. coli* (Supplementary Fig. 1). This cycle was recently also identified through a computational approach, where it was named the reductive citramalyl-CoA cycle[9], but was not demonstrated experimentally, thus far.

Construction of the THETA cycle in vitro

To demonstrate the feasibility of the THETA cycle in vitro, we identified suitable enzyme candidates on the basis of their availability, activities, specificities and stabilities (for parameters of all enzymes, see Supplementary Table 2). We then cloned, expressed and purified all proteins as His-tag fusions with the exception of Ppc, malate dehydrogenase (reaction number 3) and fumarase (reaction number 4) that we purchased commercially. For the conversion of methylsuccinyl-CoA into mesaconyl-C1-CoA in vitro, we decided to use an engineered methylsuccinyl-CoA dehydrogenase (Mcd) homologue, methylsuccinyl-CoA oxidase (Mco, reaction number 14) that can directly use O_2 as an electron acceptor[10,24] in combination with catalase (reaction number 18) (Fig. 2a). For the reduction of fumarate into succinate, we employed fumarate reductase from *Trypanosoma brucei* (Frd, reaction number 5), which is the only single-subunit, membrane-independent and NADH-dependent Frd described so far[25].

To validate the cycle, we combined all 17 enzymes (specific activities summarized in Supplementary Table 2) and cofactors in 0.5 ml 100 mM 3-(N-morpholino)propanesulfonic acid (MOPS) buffer (pH 7.0) with 20 mM NaH¹³CO₃ as the CO_2 source. We added 200 µM pyruvate to start the cycle and monitored all CoA esters with liquid chromatography–mass spectrometry (LC–MS). We detected single-labelled acetyl-CoA (8.5 µM in total) after 60 min (for ¹³C-labelling patterns, see Supplementary Fig. 2), indicating that this first version of the THETA cycle (THETA 1.0) was indeed functional (Fig. 2b and Supplementary Fig. 3b). However, we also observed accumulation of an unwanted side product that is not part of the reaction sequence and that we further identified as malyl-CoA (Supplementary Fig. 3c). Formation of malyl-CoA resulted from promiscuous activity of SucCD (reaction number 6) with malate[26] and was further increased when ATP regeneration was used (THETA 1.1, Supplementary Fig. 3c). To recycle this dead-end metabolite, we added malyl-CoA thioesterase (reaction number 19) as the proofreading enzyme for version 2.0, and also introduced a NADH/NADPH regeneration system alongside the existing ATP regeneration system (Fig. 2a). THETA 2.0 produced 38 µM acetyl-CoA within 60 min as the predominant product without substantial accumulation of malyl-CoA or any other CoA ester (Fig. 2b and Supplementary Fig. 4b).

To identify potential bottlenecks in the reaction sequence, we started the cycle from different intermediates: pyruvate, fumarate or succinate (200 µM each). Acetyl-CoA production was 3.7-fold higher when the cycle was started from succinate, indicating that Frd was one of the limiting steps (Supplementary Fig. 4c). This observation is in line with the very low specific activity of Frd (22 mU mg⁻¹)[25], which is almost one order of magnitude lower than other enzymes in the THETA cycle. We noticed that fumarate serves as a redox
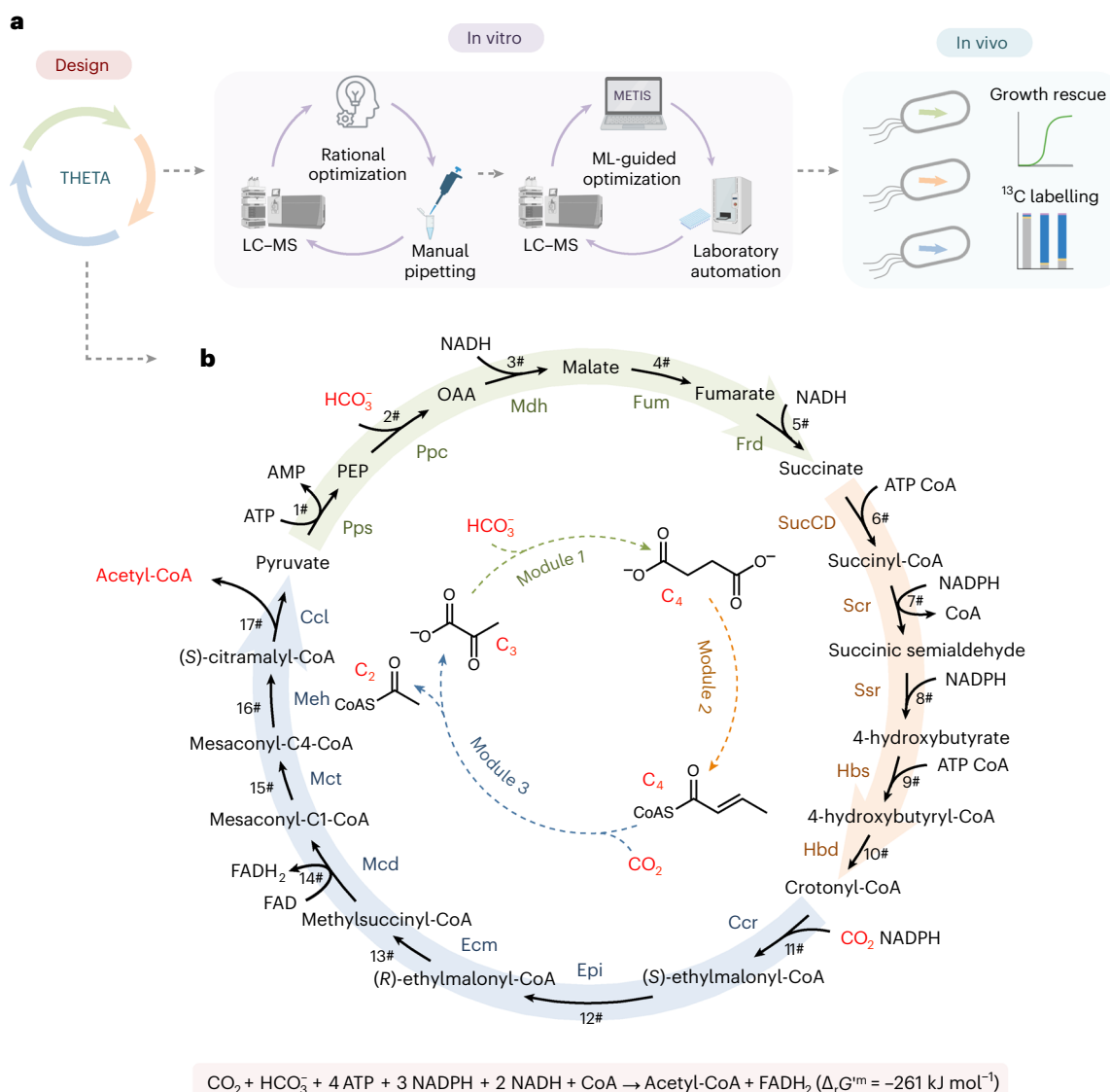
Fig. 1 | The THETA cycle and its realization. a, The overall strategy used in this work to establish the THETA cycle. The THETA cycle was designed in three modules, constructed in vitro, optimized through rational approaches and machine learning, before its three modules were implemented in *E. coli*. **b**, The scheme of the THETA cycle. Inner circle: simplified scheme of the THETA cycle, consisting of three pathway modules. Outer circle: detailed scheme of the THETA cycle. The conditions for $\Delta_r G'^m$ are at 25 °C, 1 bar pressure, pH 7.0, pMg 3, ionic strength 0.25 M and 1 mM for all reactants. The calculation details of $\Delta_r G'^m$ can be found at ref. 47. For enzyme abbreviations, see Supplementary Table 2. ML, machine learning.

substrate in different enzymes, including dihydroorotate reductase type 1a (DHOD1a)[27], which opened the possibility to realize a dihydroorotate-based Frd bypass. In this bypass, fumarate reduction is driven through oxidation of dihydroorotate by DHOH1a, while dihydroorotate itself is regenerated by DHOD1b with NADH[28] (Fig. 2a). We tested four different DHOD1a and two DHOD1b homologues, and chose DHOD1a from *Trypanosoma cruzi* and DHOD1b from *Bacillus subtilis* due to their favourable kinetics (Supplementary Fig. 5). Using the Frd bypass in THETA 3.0 substantially improved acetyl-CoA production by ~5-fold (~200 μM acetyl-CoA in 60 min) (Fig. 2b), independent of the starting substrate (pyruvate, fumarate or succinate), suggesting that the Frd step was no longer a bottleneck (Supplementary Fig. 6). The CO_2-fixation rate of THETA 3.0 reached 2.7 nmol min^{-1} mg^{-1} of core cycle proteins. This rate is comparable to the efficiency of CBB cycle in cell extracts (up to 3 nmol min^{-1} mg^{-1} CBB cycle protein) and other synthetic CO_2-fixation pathways[10,29], although the productivity and stability of the system were still very limited compared with the natural counterpart.

Optimization and extension of the THETA cycle in vitro

When testing THETA 3.0 with different concentrations of pyruvate as starting substrate (100–500 μM), acetyl-CoA production was capped at 300 μM (Supplementary Fig. 7), suggesting that the system was still limited, but without clear indication of the potential bottleneck(s). To exploit the full potential of the THETA cycle, we thus decided to improve the 34-component system through an active learning-guided optimization workflow (METIS)[30]. This workflow is able to search the combinatorial space of a complex in vitro system over several rounds for (local) optima. METIS can be used with lab automation for prototyping and analysis of different combinations of components, followed by machine learning-guided prediction of an improved set of combinations.

To explore the combinatorial space for improved acetyl-CoA production using METIS in combination with an Echo liquid handler, we limited each THETA cycle component to a maximum of eight different concentrations and downsized the assay volume to 10 μl. Testing 30 different combinations in triplicates over nine iterative
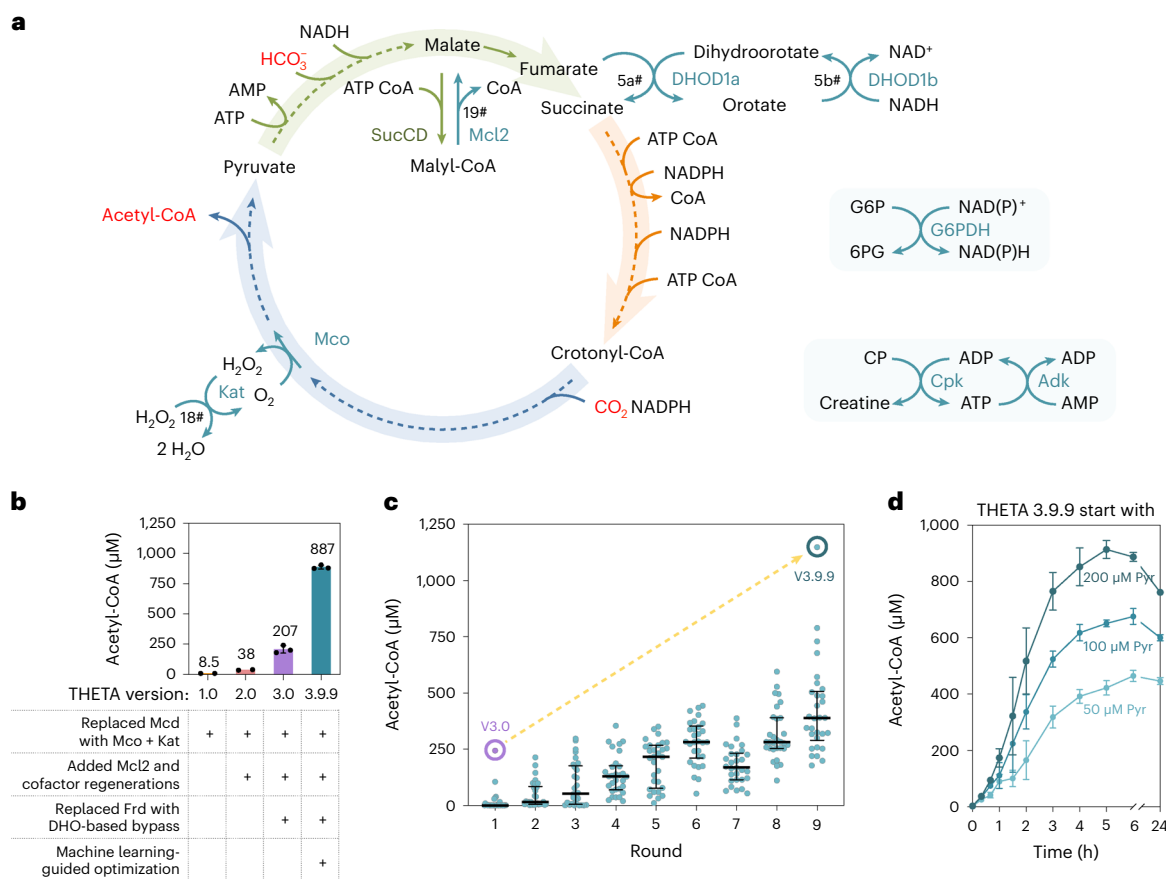
Fig. 2 | In vitro demonstration and optimization of THETA. a, A simplified scheme of THETA 3.9.9. Modifications from the original design (shown in Fig. 1b) are marked in teal. **b**, The increase of acetyl-CoA production from THETA version 1.0 to 3.9.9. Optimization was achieved through rational design (THETA 1.0 to 3.0) and machine learning-guided approaches (THETA 3.9.9). Manually confirmed test results are shown. **c**, Optimization of the THETA cycle in vitro by active learning with 30 different conditions in triplicates tested for each of the nine rounds. Assays (10 μl) were pipetted by an Echo liquid handler, started with 200 μM pyruvate and stopped after 3 h. Formed acetyl-CoA was quantified by LC–MS. As a control, V3.0 (purple circle, same composition as in the manual experiment shown in **b**) was prepared through the same procedure and is shown in comparison with the best THETA version (3.9.9, teal circle, same

composition as in the manual experiment shown in **b**). **d**, Manually repeated experiments (120 μl scale) for V3.9.9 started with the original (200 μM pyruvate) and two more substrate concentrations. Samples were withdrawn at different timepoints and quantified by LC–MS. Data in **b** and **d** represent the mean ± s.d. obtained in $n = 3$ independent experiments except for THETA 1.0 and 2.0 in **b** (where data represent the mean of $n = 2$ independent experiments). Data in **c** represent the median with interquartile range derived from $n = 30$ different testing conditions, with each individual data point representing the mean obtained in $n = 3$ independent experiments. G6P, glucose-6-phosphate; 6PG, 6-phosphogluconolactone; G6PDH, glucose-6-phosphate dehydrogenase; CP, creatine phosphate; Cpk, creatine phosphokinase; Adk, adenylate kinase; DHO, dihydroorotate; Pyr, pyruvate.

rounds yielded 1,150 μM acetyl-CoA from 200 μM substrate (pyruvate) in the best condition (THETA 3.9.9). This represents a ~5-fold increase from the initial condition (THETA 3.0, Fig. 2c) and a 135-fold improvement over THETA 1.0. Increasing the concentration of Mco in combination with lowering the concentration of 4-hyroxybutyryl-CoA synthetase (Hbs, reaction number 9) and coenzyme B12 was beneficial for increased productivity of the system (Supplementary Fig. 8). A similar pattern was recently observed during optimization of the CETCH cycle[30], which overlaps partially with the THETA cycle, indicating that these beneficial traits are common features rather than pathway-specific observations.

We repeated THETA 3.9.9 (and two other variants; Supplementary Fig. 9) at 120 μl volume to verify scalability within experimental error (80–90%) and also tested different starting concentrations of pyruvate. While starting with 100 μM and 50 μM pyruvate resulted in lower absolute product yields (680 μM and 470 μM acetyl-CoA, respectively) (Fig. 2d), total turnover (product-to-substrate ratio is the number of turns of the cycle) was increased to 7 and 9, respectively, reaching values comparable to other in vitro CO2-fixation systems that were reported recently[10,11].

To expand the product scope of the THETA cycle, we tested different outputs besides acetyl-CoA. On the one hand, malonyl-CoA was chosen due to its role as a central building block for different value-added compounds[31], such as polyketides[12,32], fatty acids[33,34], and fatty-acid-derived fuels and chemicals[33,34]. On the other hand, glyoxylate/glycolate was selected since it can be further used to build up more complex C4 molecules and anaplerotic reaction sequences[13]. For malonyl-CoA production, we expanded the reaction sequence of THETA 3.9.9 by an engineered propionyl-CoA carboxylase (Pcc*) that accepts acetyl-CoA[35], which increased the cycle output by an additional 9% (THETA 3.9.9.mc; Fig. 3a,b), probably due to the constant removal of acetyl-CoA from the reaction equilibrium of citramalyl-CoA lyase (Ccl, reaction number 17).

For the production of glycolate we replaced the reaction sequence of THETA 3.9.9 from OAA to succinate by a reaction sequence of the glyoxylate cycle that we had optimized in a previous study[13] (Fig. 3c). With this modification, THETA 3.9.9.gc produced more than 530 μM glycolate from 50 μM succinate, increasing total turnover of the system to greater than ten (Fig. 3d). Overall, these experiments demonstrate that the THETA cycle is a robust CO2-fixation network that can be flexibly
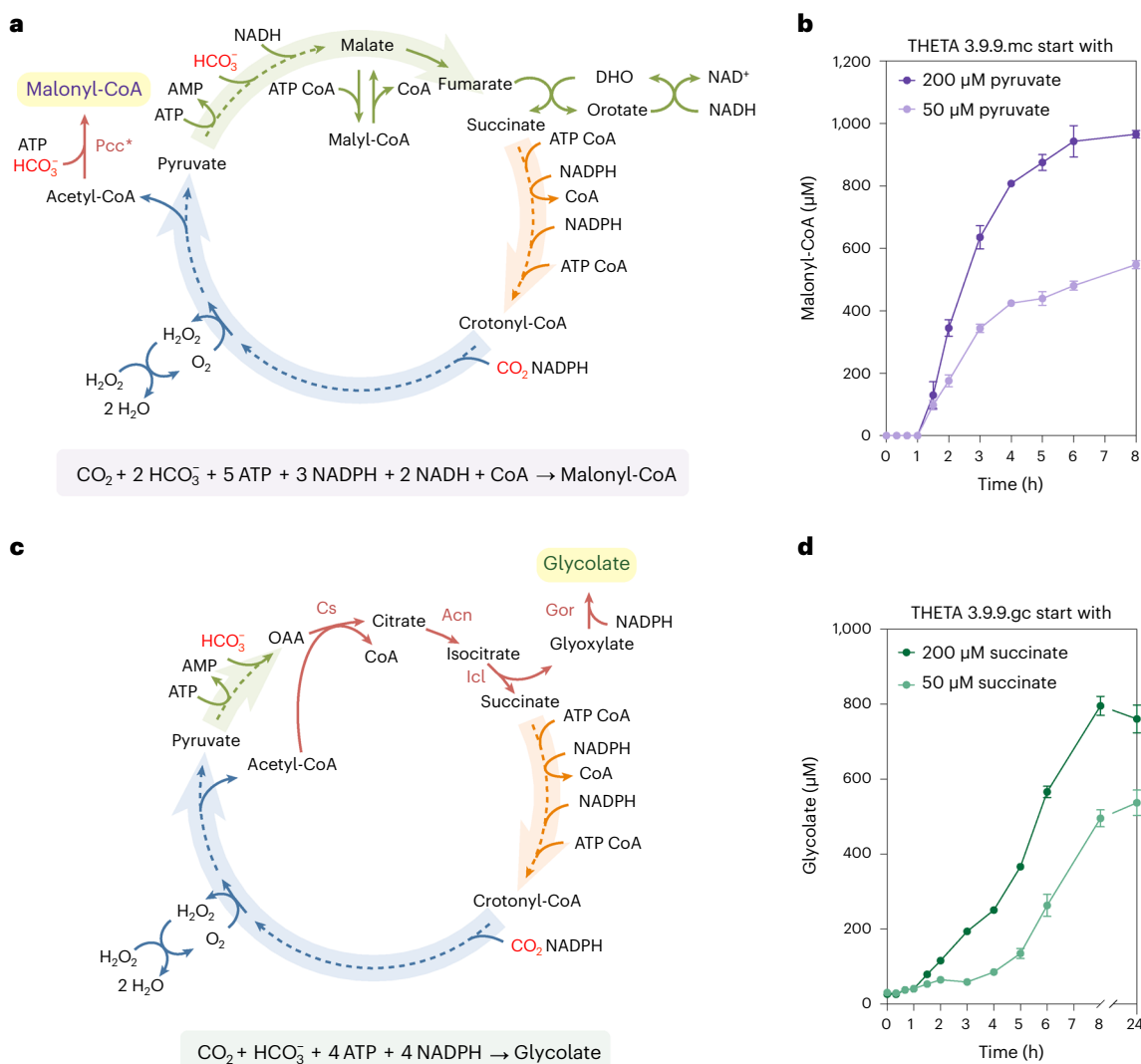
Fig. 3 | Extension of the THETA cycle in vitro. a, A simplified scheme of THETA 3.9.9.mc. Modifications from THETA 3.9.9 in **a** and **c** are marked in red. ATP and NAD(P)H regeneration systems in **a** and **c** are the same as in THETA 3.9.9 and are not shown. **b**, For the production of malonyl-CoA from CO_2, we followed the same protocol used in Fig. 2d and added Pcc* (PccD407I from *Methylobacterium extorquens*) to carboxylate acetyl-CoA. The cycle was tested with starting concentrations of 200 µM and 50 µM pyruvate, which had provided the highest yield and the highest turnover rates, respectively, for THETA version 3.9.9 (Fig. 2d). **c**, A simplified scheme of THETA 3.9.9.gc. **d**, For the production of glycolate, we replaced part of the reaction sequence as shown in **c** and used 50 µM and 200 µM succinate to start the reaction sequence. Because the release of the primary CO_2-fixation product (glyoxylate) occurs at the level of the Icl reaction, we used succinate as starting substrate (instead of pyruvate). Data in **b** and **d** represent the mean ± s.d. obtained in *n* = 3 independent experiments. Pcc, propionyl-CoA carboxylase; Cs, citrate synthase; Acn, aconitase; Icl, isocitrate lyase; Gor, glyoxylate reductase.

adopted in respect to its metabolic output, setting the stage for the integration of the THETA cycle (and its subparts) into other synthetic and native metabolic networks.

Implementation of the THETA cycle in vivo: module 1

Next, we focused on integrating (sub)parts of the THETA cycle into the native metabolic network of *E. coli*. To that end, we aimed at realizing the pathway's three different modules (module 1, from pyruvate to succinate; module 2, from succinate to crotonyl-CoA; and module 3, from crotonly-CoA to pyruvate and acetyl-CoA) individually in *E. coli* to demonstrate in vivo feasibility of the THETA cycle (Fig. 4a).

For in vivo integration of module 1, we sought to use a succinyl-CoA auxotroph, in which module 1 could support succinyl-CoA formation from pyruvate. We created this succinyl-CoA auxotroph on the basis of an acetyl-CoA auxotrophic strain JCL301 (Δ*aceEF* Δ*poxB* Δ*pflB*)[36], in which the TCA cycle is disconnected from glycolysis and requires

external feeding of acetate. In this strain, we additionally blocked succinyl-CoA formation through the (oxidative) TCA cycle, as well as the glyoxylate shunt by deleting *sucAB* and *aceA* in JCL301 (Fig. 4b) (for accession numbers of genes for in vivo implementation, see Supplementary Table 3). The resulting strain SL1 was not able to synthesize succinyl-CoA, an essential metabolite for the synthesis of methionine, lysine and diaminopimelic acid without supplementation of succinate in minimal medium containing glucose and acetate. Note that all enzymes of module 1, except Frd, are already part of *E. coli*'s native network, which required only the introduction of this reaction to realize module 1. Heterologous expression of NADH-dependent Frd from *T. brucei* or overexpression of L-aspartate oxidase (NadB) from *E. coli* (oxidizing aspartate under reduction of fumarate) restored growth of SL1 on minimal medium with 1% (w/v) glucose and 10 mM acetate (Fig. 4c and Supplementary Fig. 10), successfully demonstrating the integration of module 1 in vivo. ^{13}C-labelling experiments with
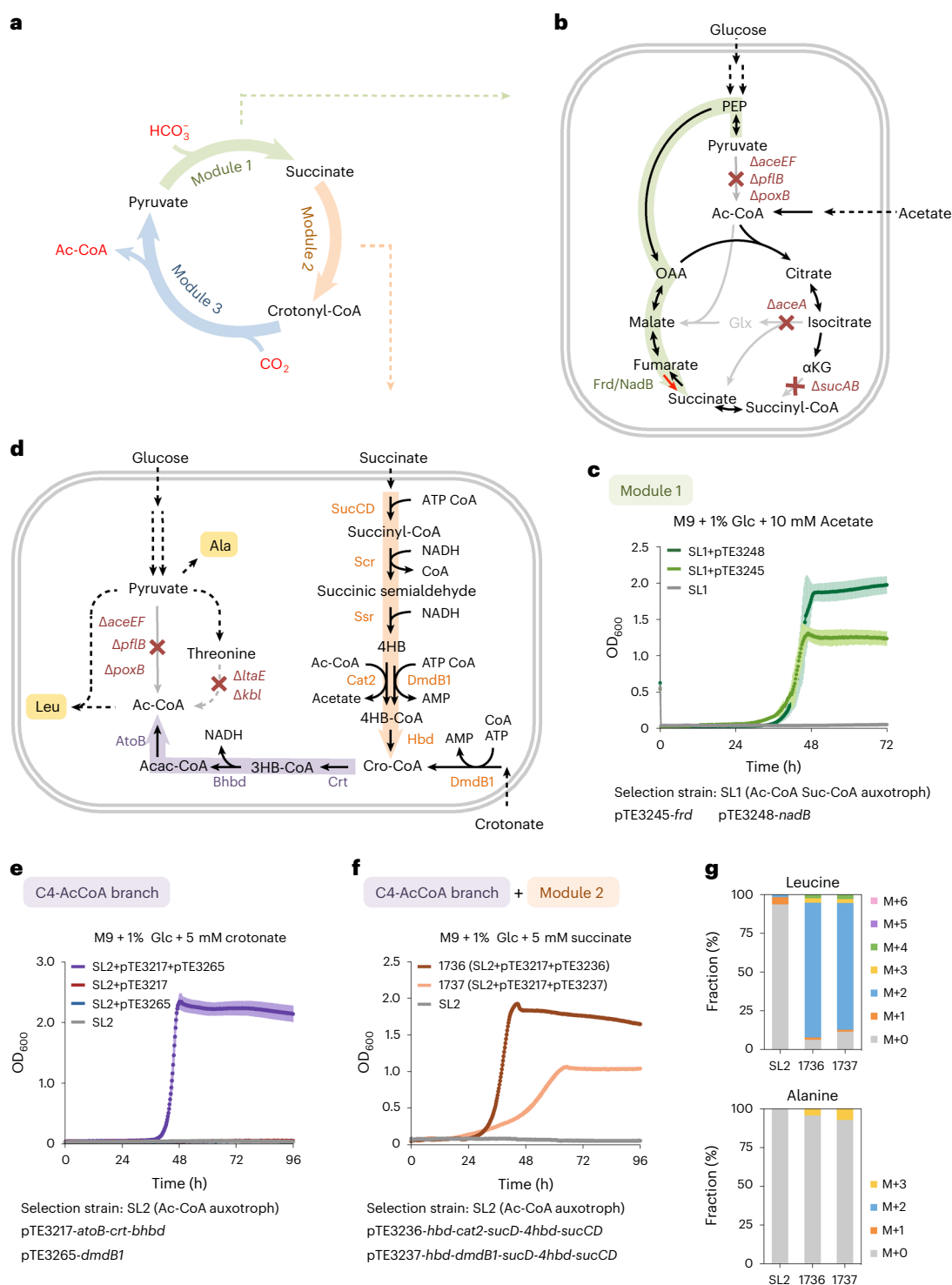
Fig. 4 | Implementation of module 1 and module 2 of THETA in vivo.
a, A simplified scheme of the THETA cycle, consisting of three modules. **b**, The scheme for implementing module 1 in vivo. **c**, Heterologous expression of Frd (*T. brucei*) or overexpression of NadB (*E. coli*) rescued the growth of the acetyl-CoA and succinyl-CoA auxotroph strain SL1 on M9 with 1% glucose and 10 mM acetate without supplementing succinate. **d**, The scheme for implementing module 2 in vivo. A short reaction sequence (C_4-acetyl-CoA shunt) is added to channel crotonyl-CoA, the output molecule of module 2, into acetyl-CoA. **e**, Expression of crotonyl-CoA ligase (DmdB1 from *P. aeruginosa*) and the C_4-acetyl-CoA shunt (pTE3217) enabled the acetyl-CoA auxotroph strain SL2 convert crotonate to acetyl-CoA and grow without supplementing acetate. **f**, Expression of module 2 (pTE3236 or pTE3237) and the C_4-acetyl-CoA shunt (pTE3217) enabled the acetyl-CoA auxotroph strain SL2 to convert succinate to

acetyl-CoA and grow without supplementing acetate. The growth was achieved through short-term evolution on plates. *sucD* encodes Scr and *4hbd* encodes Ssr. **g**, ^{13}C labelling of proteinogenic leucine (Leu) and alanine (Ala) after cultivating the strains with [U-^{13}C]succinate (SL2 on M9 with 1% glucose, 10 mM acetate and 10 mM [U-^{13}C]succinate; 1736 and 1737 on M9 with 1% glucose and 10 mM [U-^{13}C]succinate). The labelling pattern confirmed operation of module 2 in vivo, since leucine (derived from one acetyl-CoA and two pyruvate) was mainly labelled twice while alanine (derived from pyruvate only) was predominantly non-labelled. The growth data represent the mean ± s.d. obtained in *n* = 3 independent experiments. For enzyme abbreviations, see Supplementary Table 2. Glx, glyoxylate; αKG, α-ketoglutarate; Ac-CoA, acetyl-CoA; Suc-CoA, succinyl-CoA; Acac-CoA, acetoacetyl-CoA; 3HB-CoA, 3-hydroxybutyryl-CoA; Cro-CoA, crotonyl-CoA; 4HB-CoA, 4-hydroxybutyryl-CoA.

[U-^{13}C]glucose and non-labelled acetate and CO_2 showed expected succinate labelling patterns (Supplementary Fig. 11), further validating our results.

Implementation of the THETA cycle in vivo: module 2

To establish module 2 (succinate to crotonyl-CoA) in vivo, we sought to use an acetyl-CoA auxotroph and channel crotonyl-CoA, the output molecule of module 2, into acetyl-CoA through a short reaction sequence (C_4-acetyl-CoA shunt; Fig. 4d). We used strain JCL301 (see above) as a starting point and increased tightness of selection by additionally deleting *kbl* and *ltaE* to block conversion of threonine into acetyl-CoA, resulting in strain SL2 (Fig. 4d). We introduced crotonyl-CoA ligase (DmdB1 from *Pseudomonas aeruginosa*) and the three enzymes of the C_4-acetyl-CoA shunt, acetyl-CoA acetyltransferase (AtoB), 3-hydroxybutyryl-CoA dehydrogenase (Bhbd), and crotonase (Crt) via expression plasmids (pTE3265-*dmdB1* and pTE3217-*atoB-crt-bhbd*). The corresponding strain was able to grow on minimal medium with 1% glucose and 5 mM crotonate, indicating that the C_4-acetyl-CoA shunt was functional in vivo (Fig. 4e and Supplementary Fig. 12).

We next aimed at integrating the reactions of module 2 together with the C_4-acetyl-CoA shunt into strain SL2 to restore growth of the strain on succinate. Module 2, as prototyped in vitro, consists of five enzymes (SucCD, Scr, Ssr, Hbs and Hbd; reaction numbers 6–10) that convert succinate into crotonyl-CoA. We assembled the corresponding genes into two operons, expressed the five enzymes under non-selecting conditions in JCL301 and tested for activities in crude cell extracts. Analysis showed that Ssr, Hbs and Hbd activities were relatively low (below 9 mU mg^{-1} lysate) (Supplementary Figs. 13 and 14). To improve enzyme activities of module 2, we switched to Scr and Ssr from *Porphyromonas gingivalis*, replaced Hbs by a 4-hydroxybutyryl-CoA (4HB-CoA) transferase (Cat2 from *P. gingivalis*) and moved Hbd to the start of the operon. These changes resulted in one to two orders of magnitude increased enzyme activity in crude cell extracts (Supplementary Figs. 13 and 14). We further combined the two operons onto one expression plasmid (pTE3236) and created another version of the plasmid, in which we replaced Cat2 by DmdB1 as alternative 4HB-CoA ligase (pTE3237) (for DmdB1 activity on 4HB, see Supplementary Fig. 15). Introducing module 2 (pTE3236 or pTE3237) and the C_4-acetyl-CoA shunt (pTE3217) into SL2 did not rescue acetyl-CoA auxotrophy on succinate. However, screening for colonies on plates identified two strains, 1736 and 1737 (SL2 carrying pTE3217 + pTE3236 and pTE3217 + pTE3237, respectively) that could grow on minimal medium with 1% glucose and 5 mM succinate after short-term evolution on plates (Fig. 4f and Supplementary Fig. 16; see details for this evolution method in Methods). We sequenced both strains, but found neither direct explanations for the observed phenotypes nor mutation overlaps between the two strains, beyond the interesting finding that both strains had turned off the production of type I fimbriae (Supplementary Table 4).

To confirm that growth on succinate in these strains was based on module 2, we performed ^{13}C-labelling experiments. We fed the cultures with [U-^{13}C]succinate and determined the labelling patterns of selected amino acids. Leucine, derived from one acetyl-CoA and two pyruvate, was predominantly double labelled, whereas alanine, derived from pyruvate only, was mainly unlabelled (Fig. 4g), demonstrating exquisite conversion of succinate into acetyl-CoA and thus successful integration of module 2 in vivo.

Implementation of the THETA cycle in vivo: module 3

Finally, we aimed to integrate module 3, which converts crotonyl-CoA to acetyl-CoA and pyruvate in vivo. We decided to use acetyl-CoA auxotroph SL2 (see above) as selection strain and generate acetyl-CoA through module 3 from externally added crotonate (Fig. 5a,b). Module 3, as prototyped in vitro, consists of seven genes. To lower metabolic burden, we integrated five genes (*epi*, *ecm*, *mct*, *meh* and *ccl*) into the chromosome of acetyl-CoA auxotroph SL2, resulting in strain HH61

(for enzyme activity verification, see Supplementary Fig. 17), and expressed the remaining two enzymes (Ccr and Mco), and crotonyl-CoA ligase (DmdB1) from plasmids (pTE3260 and pTE3265). However, this did not result in growth of HH61 on crotonate. To identify potential bottlenecks in module 3, we performed metabolomics and found that methylsuccinate accumulated in minimal medium containing glucose, acetate and crotonate (Fig. 5c). This indicated that the upper part of module 3 (Ccr, Epi and Ecm, reaction numbers 11–13) was functional and produced methylsuccinyl-CoA, which had been hydrolysed in vivo, diverting flux away from the downstream reactions of module 3.

To identify the enigmatic methylsuccinyl-CoA thioesterase activity, we tested the effects of two thioesterases of *E. coli*, TesB and YciA, which were both reported to have activity with methysuccinyl-CoA[37]. Knocking out *tesB* and *yciA* in HH61 resulted in strain HH169, which showed substantially decreased production titres (normalized by optical density (OD)) of methylsuccinate by 27-fold (Fig. 5c), indicating that these enzymes are mainly responsible for methylsuccinyl-CoA hydrolysis in vivo.

Much to our surprise, besides methylsuccinate, we also detected mesaconate in the medium of HH61, even when we did not overexpress *mco* (Fig. 5c), suggesting that *E. coli* possesses native enzymes that convert methylsuccinate into mesaconate. We suspected that succinate dehydrogenase (Sdh) might be responsible for the enigmatic methylsuccinate oxidation. Indeed, *sdhA* or *sdhB* knockout in *E. coli* reduced mesaconate titres by 80–90% (Supplementary Fig. 18), while activity assays with purified SdhAB showed that the enzyme natively oxidizes methylsuccinate at almost 70% of its specific activity with succinate (Supplementary Fig. 19).

Based on these findings, we re-drafted our strategy to integrate module 3 in vivo (Fig. 5b). Instead of a pure CoA-based route, we designed an acid bypass that relies on the native thioesterases of *E. coli* for the hydrolysis of (2*S*)-methylsuccinyl-CoA into methylsuccinate, which is oxidized by *E. coli* Sdh into mesaconate and subsequently activated to mesaconyl-C_4-CoA. For the last step, we tested two CoA transferases[38] and one CoA ligase (Supplementary Fig. 20) in wild-type *E. coli* BW25113 with ^{13}C-labelling experiments to identify itaconate CoA transferase (Ict) from *Yersinia pestis* as candidate for the mesaconate:CoA transfer (Supplementary Figs. 21 and 22). In combination with the lower part of module 3, Ict$_{Yp}$ was able to convert mesaconate into pyruvate in HH61 (Fig. 5d), demonstrating that this route is a viable option for realizing module 3 in vivo. Overexpression of *E. coli* native Sdh increased mesaconate production from methylsuccinate by seven-fold (Fig. 5e), demonstrating its crucial role in enhancing flux through module 3. Since Ict$_{Yp}$ converts mesaconate to mesaconyl-C_4-CoA (Supplementary Fig. 22), we decided to omit *mct* and keep only *epi* and *ecm* on the chromosome of acetyl-CoA auxotroph SL2 (generating strain SL3). We put the remaining genes on plasmids (pTE4501 containing *ict*, *ccr*, *ccl* and *meh*; and pTE4502 containing *sdhCDAB* and *dmdB1*) and introduced them into SL3, resulting in a notable increase in normalized extracellular pyruvate production from crotonate (Supplementary Fig. 23). However, due to background production of non-labelled pyruvate in SL3 (see Supplementary Fig. 23 for more details), we decided to confirm the result by testing module 3 in another strain without background pyruvate formation. We integrated *epi* and *ecm* in strain SIJ488, a derivative of wild-type *E.coli* MG1655. The resulting strain, HH422, when introduced with module 3, demonstrated non-labelled pyruvate production from crotonate on [U-^{13}C]glucose (Fig. 5f), confirming the functionality of module 3 in vivo. Interestingly, we observed that 2% (v/v) lysogeny broth (LB) was necessary to observe this phenotype, indicating that the strain requires additional nutritional support, which may be caused by metabolic burden or regulatory effects. These experiments showed that module 3 is functional in vivo, but also indicate that its metabolic flux must be further improved for full complementation of acetyl-CoA auxotrophy.
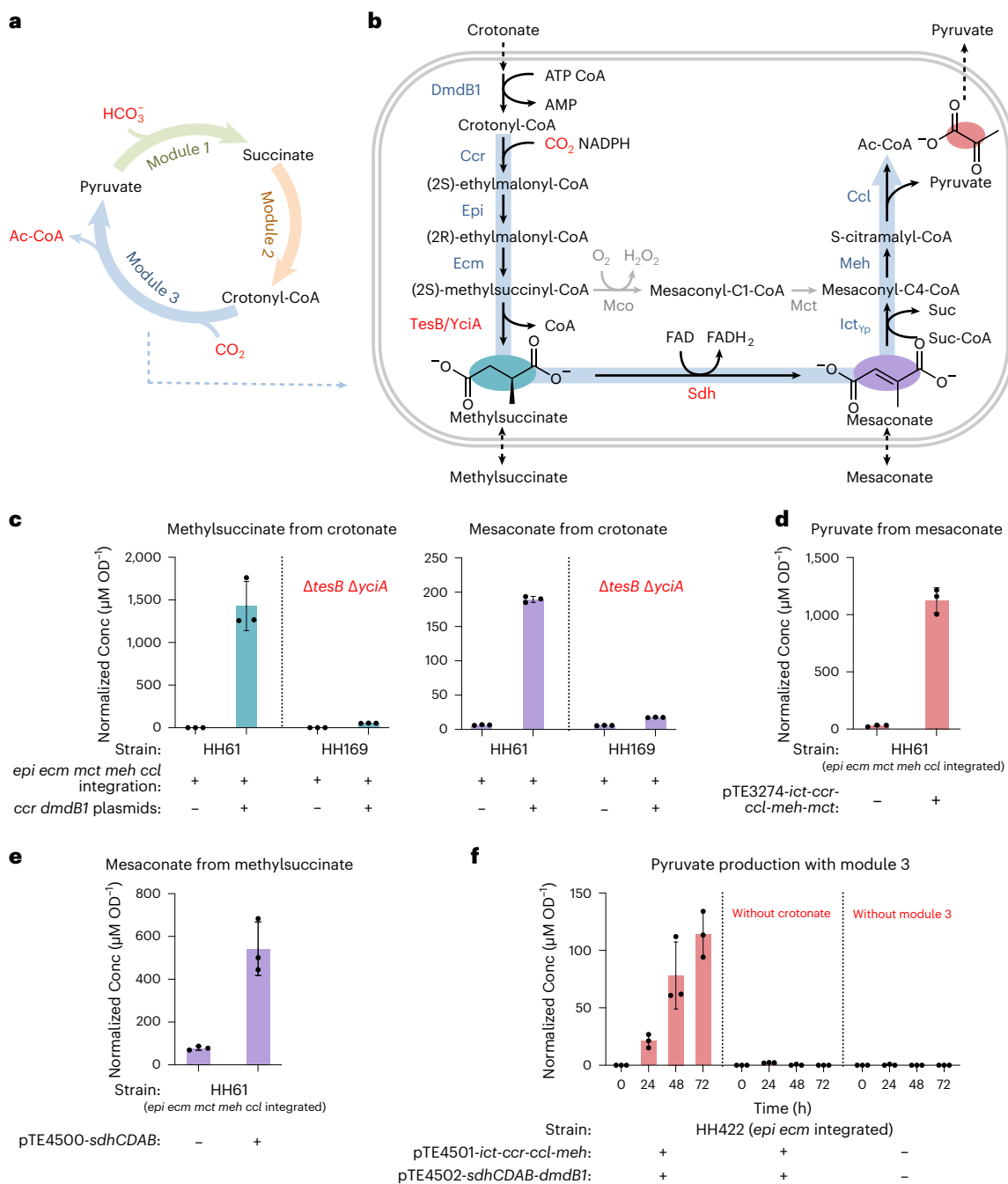
Fig. 5 | Implementation of module 3 of THETA in vivo. a, A simplified scheme of the THETA cycle, consisting of three modules. **b**, The scheme for implementing module 3 in vivo. The conversion of (2S)-methylsuccinyl-CoA to mesaconyl-C4-CoA, which was initially planned to take place through a CoA-based route (involving Mco or Mcd), was replaced with an acid bypass, in which native thioesterases (TesB and YciA) hydrolyse (2S)-methylsuccinyl-CoA into methylsuccinate, which is further converted to mesaconate and mesaconyl-C4-CoA using Sdh and Ict, respectively. **c**, Expression of Ccr and DmdB1 enabled HH61 to produce methylsuccinate (left) and mesaconate (right) in M9 minimal medium with 1% glucose, 10 mM acetate, 0.5 μM coenzyme B12 and 10 mM crotonate, indicating that the upper part of module 3 (Ccr, Epi and Ecm) worked in vivo so that *E. coli* converted methylsuccinate to mesaconate (through Sdh). Production of methylsuccinate was substantially decreased in a thioesterase double knockout (ΔtesB ΔyciA), indicating that these native enzymes are responsible for methylsuccinyl-CoA hydrolysis. **d**, Expression of Ict (*Y. pestis*) and the lower part of module 3 (Meh and Ccl steps) enabled HH61 to produce

non-labelled pyruvate in M9 minimal medium containing 1% [U-¹³C]glucose, 10 mM [U-¹³C]acetate and 10 mM mesaconate, indicating that the lower part from mesaconate to pyruvate and acetyl-CoA was functional in vivo. **e**, Over-expression of *E. coli* native Sdh resulted in a notable increase in mesaconate production in M9 minimal medium with 1% glucose, 10 mM acetate and 10 mM methylsuccinate. **f**, Expression of module 3 (integrated *epi* and *ecm*, pTE4501 and pTE4502) enabled HH422 to produce non-labelled pyruvate in M9 minimal medium containing 1% [U-¹³C]glucose, 2% LB, 0.5 μM coenzyme B12 and 10 mM crotonate, indicating that module 3 worked in vivo. To confirm that the non-labelled pyruvate originated from crotonate and not from LB, two negative controls were used: HH422 with module 3 in the same medium in the absence of crotonate, and HH422 without module 3 in the same medium containing crotonate. All media contained 0.5 mM IPTG and appropriate antibiotics. Metabolites measured in **c**–**f** are colour-coded according to **b**. The data represent the mean ± s.d determined from *n* = 3 independent experiments. Yp, *Yersinia pestis*.

Discussion

Here, we designed and realized the THETA cycle, a new-to-nature pathway for the conversion of CO_2 into acetyl-CoA in vitro, and integrated its reaction sequence as three modules into the metabolic network of *E. coli* (for summary of testing conditions for each module, see Supplementary Table 5). Our study expands the number of oxygen-insensitive in vitro CO_2-fixation pathways to three and paves a way for the implementation and evolution of these new-to-nature systems in the future. With a current in vitro CO_2-fixation rate of 2.7 nmol min^{-1} mg^{-1} (for the rate comparison of different THETA versions, see Supplementary Table 6), the THETA cycle is comparable to the efficiency of other synthetic CO_2-fixation pathways[10,29] and still holds the potential for further improvements, especially if the specific activity of the rate-limiting enzyme (Mco; Supplementary Table 2) can be further increased through enzyme engineering.

Rational engineering allowed us to improve the THETA cycle by 20 fold from version 1.0 to version 3.0. Using a recently developed machine learning-guided workflow allowed us to further increase the cycle's performance by an additional factor of five within only nine rounds of optimization, which is remarkable, given the potential space of approximately 10^{30} (that is, 8^{34}) different combinations. This underlines the power of combining machine learning and laboratory automation for exploring and improving complex in vitro systems[30]. Overall, the THETA cycle improved by almost two orders of magnitude from version 1.0 (final yield ~9 μM acetyl-CoA) to version 3.9.9 (~890 μM acetyl-CoA).

Our efforts to transfer the THETA cycle into *E. coli* showed that in vivo implementations can strongly differ from the in vitro solution. For instance, while we needed to establish a dihydroorotate-based bypass to overcome the insufficiencies of Frd in vitro, Frd proved to be sufficiently active in vivo. Moreover, the native metabolic network of *E. coli* can prevent the implementation of certain routes, but also provide alternative solutions. One example is the methylsuccinyl-CoA thioesterase activity of YciA and TesB, which posed a challenge for establishing methylsuccinyl-CoA oxidation in vivo. At the same time, the in vivo production of methylsuccinate opened the way for the discovery of the enigmatic methylsuccinate oxidation activity of Sdh, which we ultimately exploited for establishing module 3. Overall, these results highlight the opportunities and challenges that come with the integration of orthogonal pathways into the native metabolic network of *E. coli*, as was also noted and observed recently[39–41].

Demonstration of the three different pathway modules of the THETA cycle in *E. coli* through growth-dependent phenotypes and/or ^{13}C-labelling establishes the groundwork for the integration and experimental evolution of the complete THETA cycle in *E. coli* in the next step. Such engineering efforts might require further rational optimization and extensive adaptive laboratory evolution to achieve a finely balanced interplay between newly introduced metabolic fluxes, natural central metabolic networks, metabolic burden and genetic regulations. Moreover, additional reducing power and ATP will be required to drive the complete cycle in an autotrophic mode, which could be provided by electricity, hydrogen and/or sustainably generated formate or methanol[42], as suggested recently by several groups[15,17,43]. Demonstrating the THETA cycle in vivo will not only be important for showcasing the possibility of realizing highly orthogonal and complex metabolic pathways in the background of the native metabolism of living cells, but also for bringing one of the most promising designs for synthetic CO_2 fixation that was proposed recently, into life[9].

Methods

Materials

Chemicals used were purchased from Sigma-Aldrich, Carl Roth and Santa Cruz Biotechnology. NaH^{13}CO$_3$ was obtained from Cambridge Isotope Laboratories. Coenzyme A was purchased from Roche Diagnostics. Oligonucleotides were obtained from Eurofins Genomics Germany GmbH or Sigma-Aldrich. Synthesized genes were obtained from BaseClear B.V. or Twist Bioscience. Q5 Hot Start high-fidelity DNA polymerase was purchased from New England Biolabs. Prime-STAR GXL DNA polymerase was purchased from Takara Bio. Other materials for cloning and protein expression were obtained from Thermo Fisher Scientific, New England Biolabs and Macherey-Nagel. Materials and equipment for protein purification were obtained from GE Healthcare, Bio-Rad and Merck Millipore. PEP carboxylase, malic dehydrogenase, fumarase, glucose-6-phosphate dehydrogenase (*Leuconostoc mesenteroides*), creatine phosphokinase, pyruvate kinase/lactic dehydrogenase and carbonic anhydrase were purchased from Sigma-Aldrich.

Synthesis of CoA thioesters

Acetyl-CoA and malonyl-CoA were synthesized and purified according to ref. [44]. The concentration was quantified by determining the absorption at 260 nm ($\varepsilon_{260\,nm} = 16.4$ mM^{-1} cm^{-1}).

Culture medium and conditions

For general molecular biology purposes, *E. coli* strains were grown in LB with appropriate antibiotics at 37 °C. For growth rescue experiments, M9 minimal medium was used, which consists of M9 salts (8.9 g l^{-1} Na$_2$HPO$_4$·2H$_2$O, 3 g l^{-1} KH$_2$PO$_4$, 0.59 g l^{-1} NaCl and 1 g l^{-1} NH$_4$Cl; pH 7.2), 1 mM MgSO$_4$, 0.2 mM CaCl$_2$, 50 μM FeSO$_4$, 1 μg ml^{-1} thiamine hydrochloride, 1 μg ml^{-1} biotin, trace metals (68.2 μM MnCl$_2$, 3.7 μM ZnSO$_4$, 0.4 μM CoCl$_2$, 0.6 μM CuCl$_2$, 1.6 μM H$_3$BO$_3$, 2.1 μM NiCl$_2$, 2.1 μM Na$_2$MoO$_4$ and 1.9 μM Na$_2$SeO$_3$) and carbon sources as noted in the study. Antibiotics were used at the following concentrations: spectinomycin (Spec), 50 μg ml^{-1}; kanamycin (Kan), 50 μg ml^{-1}; and chloramphenicol (Cam), 34 μg ml^{-1}.

MDF analysis

MDF analysis[23] was applied to evaluate the thermodynamics feasibility of implementing the pathway in vivo with Python packages equilibrator_api (version 0.4.5) and equilibrator_pathway (version 0.4.4). The changes in Gibbs energy of the reactions were estimated using the component contribution method[45]. CO_2 was considered as the substrate for all carboxylation reactions as its concentration is pH independent, unlike that of bicarbonate, thus simplifying the calculations. Ambient CO_2 in solution is 10 μM, ubiquinone/ubiquinol is 1 μM (ref. [46]), and other cofactor and metabolite concentrations were constrained to the range 1–10 mM as described previously[23]. pH was assumed to be 7.0, ionic strength was assumed to be 0.25 M and $-\log[Mg^{2+}]$ (pMg) was assumed to be 3. The scripts and details can be found at ref. [47].

Plasmid construction

All plasmids constructed in this study were assembled using the Gibson isothermal DNA assembly method, while DNA fragments were amplified by Q5 Hot Start high-fidelity DNA polymerase or PrimeSTAR GXL DNA polymerase. Plasmids are listed in Supplementary Table 7. Primers used for strain construction are listed in Supplementary Table 8. All plasmids were constructed in *E. coli* strain DH5α for propagation and storage.

Assay of THETA 1.0

The assay was performed in a 500 μl (final volume) reaction mixture containing 100 mM MOPS pH 7.0, 5 mM MgCl$_2$, 20 mM NaH^{13}CO$_3$, 3 mM ATP, 10 mM creatine phosphate, 3 mM NADH, 4 mM NADPH, 1 mM CoA and all the enzymes in their corresponding amounts (Supplementary Table 9). The reaction was kept at 30 °C and started with the addition of 0.2 mM pyruvate. Note that Frd was added 3 min after adding pyruvate to alleviate its wasteful NADH oxidation, since Frd oxidizes NADH in the absence of fumarate. Samples (100 μl each) were withdrawn from the reaction mixture, quenched with 3% formic acid, and analysed with LC–MS for CoA esters.

Assay of THETA 2.0

The assay was performed in a 150 µl (final volume) reaction mixture containing 100 mM MOPS pH 7.0, 5 mM MgCl$_2$, 20 mM NaH^{13}CO$_3$, 3 mM ATP, 10 mM creatine phosphate, 0.2 mM NADH, 0.2 mM NADPH, 6 mM glucose-6-phosphate, 1 mM CoA and all the enzymes in the corresponding amounts (Supplementary Table 9). The reaction was kept at 30 °C and started with the addition of 0.2 mM pyruvate, fumarate or succinate. Samples (20 µl each) were withdrawn from the reaction mixture, quenched with 3% formic acid and analysed with LC–MS for CoA esters.

Assay of THETA 3.0

The assay was performed in a 150 µl (final volume) reaction mixture containing 100 mM MOPS pH 7.0, 5 mM MgCl$_2$, 20 mM NaH^{13}CO$_3$, 3 mM ATP, 10 mM creatine phosphate, 0.2 mM NADH, 0.2 mM NADPH, 6 mM glucose-6-phosphate, 1 mM CoA, 0.2 mM dihydroorotate and all the enzymes in their corresponding amounts (Supplementary Table 9). The reaction was kept at 30 °C and started with the addition of 0.2 mM pyruvate, fumarate or succinate. Samples (20 µl each) were withdrawn from the reaction mixture, quenched with 3% formic acid and analysed with LC–MS for CoA esters.

Machine learning-guided optimization of THETA assays in 96-well plates

The general workflow is outlined in ref. 30. The general parameters were set as follows: 30 conditions in triplicates per round, 9 iterative rounds of optimization and 10 µl reaction volume. The used concentrations of all components and their respective stock concentrations are listed in Supplementary Table 10. For model parameters we chose the default settings, whereas the exploration values for the different days were set to 1: 1.41, 2: 1.41, 3: 1.0, 4: 1.0, 5: 0.5, 6: 1.0, 7: 1.0, 8: 1.0 and 9: 1.0.

After starting the assays with 200 µM sodium pyruvate, we covered the 96-well plates (AB1058) with Axygen breathable sealing film (BF-400-S) to allow the transfer of oxygen. The reaction (10 µl volume) was carried out at 30 °C with mild shaking at 160 rpm in an Infors HT Ecotron shaker. The reactions were stopped after 3 h with 1.11 µl 50% formic acid. The plate was spun for 1 h at 2,272g and 4 °C to pellet the precipitated proteins.

For analysis by LC–MS, we used a multichannel pipette to transfer 1 µl of the supernatant into 19 µl of pre-cooled dH$_2$O in a new 96-well plate. The plate was sealed with a Corning microplate aluminium sealing tape (6570). The assay plate with the quenched reactions was sealed with a Corning microplate aluminium sealing tape too and stored at −80 °C.

Manual assays of optimized THETA versions

The assays were performed in triplicates in 120 µl (final volume) reaction mixture (all the corresponding amounts are listed in Supplementary Tables 9 and 11). The reaction was kept at 30 °C and started with the addition of 0.05 mM, 0.1 mM or 0.2 mM pyruvate or succinate. Samples (9 µl each) were withdrawn from the reaction mixture, quenched with 1 µl 50% formic acid (5% final concentration) and analysed with LC–MS for acetyl-CoA, malonyl-CoA or glycolate.

Strain construction

Strains used in this study are listed in Supplementary Table 12. Scarless Cas9 Assisted Recombineering was used for knocking out *sucAB* and *aceA* with the helping plasmid pCas9cr4 (ref. 48). Plasmid pKDsgRNA-sucAB and DNA oligonucleotides rec-SucAB were used for knocking out *sucAB*. And plasmid pKDsgRNA-ackA and DNA oligonucleotides rec-ackA were used for knocking out *aceA*. λ-Red recombineering was used iteratively for other gene knockouts and integrations with a helping plasmid pSIJ8 (ref. 49). For knockouts, Kan and Cam selection markers were polymerase chain reaction amplified with primers having 50 bp homologous arms[50], labelled with 'KO' in Supplementary Table 8, from pKD4 (GenBank: AY048743) and pKD3 (GenBank: AY048742)[51], respectively. The parent strain

was transformed with pSIJ8 plasmid, the procedures of deletion and antibiotic cassette removal were detailed in ref. 52 with the change of culturing temperature to 30 °C and all media containing 50 µg ml^{-1} ampicillin (Amp). The verification primers are listed with label 'Ver' or 'V' in Supplementary Table 8.

For genome integration of *ccl–meh–mct* and *epi–ecm* operons, the same recombineering method was used. The operons were first constructed to plasmids pTE3256 and pTE3257 with 600 bp homologous regions of safe sites SS7 and SS9 (ref. 53), under a constitutive strong promoter pgi-20 (ref. 54), a ribosome binding site rbsB[55] and a Kan cassette. The cassettes were polymerase chain reaction amplified with the P13_SS9_F and P14_SS9_R primer pair and the P15_SS7_F and P16_SS7_R primer pair from plasmid templates pTE3257 and pTE3256, respectively. In the end, the helping plasmid was cured by culturing at 37 °C.

Cell growth test

Single colonies of the strain transformed with the desired plasmids were cultured overnight in LB with 10 mM acetate and appropriate antibiotics (additional 10 mM succinate was added for strain SL1). On the next day, the overnight culture was inoculated at 2% into 3 ml M9 minimal medium with 1% glucose, 10 mM acetate and appropriate antibiotics (and additional 10 mM succinate was added for strain SL1). The cultures were grown at 37 °C to an OD$_{600}$ of 0.2–0.4, induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown at 30 °C for 18–20 h. Then cells were collected by centrifugation, washed twice with M9 minimal medium containing no carbon source, and inoculated into testing medium to an initial OD$_{600}$ of 0.02. Testing medium is M9 minimal medium with 0.5 mM IPTG, appropriate antibiotics and carbon sources as noted in the study. All growth experiments were performed in triplicate.

Short-term evolution on plates

To evolve the acetyl-CoA auxotroph strain SL2 to grow on succinate using the C$_4$-acetyl-CoA shunt and module 2, 32 colonies of SL2 transformed with plasmids containing genes of C$_4$-acetyl-CoA (pTE3217) and module 2 (pTE3236 or pTE3237) were picked from LB plates with 10 mM acetate and appropriate antibiotics, streaked on M9 relaxing plates (M9 minimal plates with 1% glucose, 10 mM acetate, 0.5 mM IPTG and appropriate antibiotics) and incubated at 37 °C. After 1–2 days, grown colonies on the M9 relaxing plates were re-streaked on M9 selection plates (M9 minimal plates with 1% glucose, 10 mM succinate, 0.5 mM IPTG and appropriate antibiotics), and incubated at 37 °C for 2–3 weeks. Colonies showed up after 10–14 days. Grown colonies were re-streaked on M9 selection plates, and also inoculated in M9 selection liquid medium (M9 minimal medium with 1% glucose, 10 mM succinate, 0.5 mM IPTG and appropriate antibiotics). Colonies showing consistent growth on both M9 selection plates and the M9 selection liquid medium were isolated. Strains 1736 (evolved from SL2 carrying pTE3217 and pTE3236) and 1737 (evolved from SL2 carrying pTE3217 and pTE3237) were isolated in this process.

^{13}C-labelling of proteinogenic amino acids

For labelling analysis of the acetyl-CoA auxotroph strain SL2 and evolved strains 1736 and 1737, cells were cultured in 3 ml M9 minimal medium supplemented with 1% glucose, 10 mM [U-^{13}C]succinate, 0.5 mM IPTG and appropriate antibiotics (additional 10 mM acetate was added for SL2). At the late exponential phase, cells with the amount equivalent to 1 ml OD$_{600}$ of 1 were collected, washed with water once and resuspended in 1 ml 6 M HCl. Biomass was hydrolysed at 95 °C overnight. Then, the samples were completely dried under an airstream at 95 °C, re-dissolved in 1 ml H$_2$O and centrifuged to remove insoluble particles.

Whole-genome sequencing

For Illumina sequencing, the NucleoSpin microbial DNA kit (Macherey-Nagel) was used for genomic DNA extraction following

the manufacturer's instructions. Library construction and genome sequencing were performed by Eurofins Genomics using the paired-end Illumina sequencing platform, NovaSeq 6000 S4 PE150 XP.

For Nanopore sequencing, genomic DNA of stationary phase cells was obtained using the NucleoBond HMW DNA kit (Macherey-Nagel) according to the manufacturer guidelines, and using lysozyme for cell lysis (final concentration 1 mg ml^{-1}) for 1 h at 37 °C in 2 ml of 10 mM Tris–HCl, pH 8.0. DNA quality and concentration were assessed via NanoDrop 8000 spectrophotometer and Qubit 3 fluorometer using double-stranded DNA broad range (BR) reagents. Library preparation was performed using the Ligation Sequencing Kit SQK-LSK109 in conjunction with the Native Barcoding Kit EXP-NBD104 and EXP-NBD105 (Oxford Nanopore Technologies), according to the manufacturers' guidelines, except the input DNA was increased five-fold to match the molarity expected in the protocol as no DNA shearing was applied. Sequencing was performed on a MinION Mk1B device using a MinION Flow Cell (FLO-MIN111, cell chemistry R10.1). Nanopore data were basecalled with ONT Guppy basecalling software (v 6.0.1+652ffd179). Raw reads have been deposited at the National Center for Biotechnology Information Sequence Read Archive and can be accessed under BioProject PRJNA884544.

For mutations and structural variants identification, we first constructed genomic sequences of the JCL301. With Illumina sequencing and Nanopore sequencing data, we modified the MG1655 genome (GenBank: NC_000913) by using gdtools from the breseq v0.35.7 (ref. 56), minimap2 v2.24 (r1122)[57] and canu v2.2 (ref. 58) to generate JCL301 genome. The genome has been deposited to GenBank under CP107281.

Using the reconstructed genome sequence and plasmid sequences, we mapped the Illumina sequencing data with breseq v0.35.7 (ref. 56), with a coverage parameter -l set to be 200, to identify mutations and knockouts. Structural variants were identified by ngmlr v0.2.7 and sniffles2 v2.0.3 (ref. 59) mapping Nanopore sequencing data to genome and plasmid sequences. All results were also manually checked in Tablet v1.21.02.08 (ref. 60).

Analysis of extracellular organic acids

Single colonies of the strain transformed with the desired plasmids were cultured overnight in LB with 10 mM acetate and appropriate antibiotics. On the next day, the overnight culture was inoculated at 2% into 3 ml M9 minimal medium with appropriate antibiotics and carbon sources as noted in the study. A total of 0.5 μM coenzyme B12 was also added, when crotonate was included in the medium. The cultures were grown at 37 °C to an OD$_{600}$ of 0.2–0.4, induced with 0.5 mM IPTG, and grown at 37 °C for 20–24 h. Then 0.5–1 ml cells were centrifuged at 20,817g for 2 min, and the supernatant was analysed with LC–MS for organic acids. Note that, when 1% [U-^{13}C]glucose and [U-^{13}C]acetate were used as carbon sources, overnight pre-cultures were prepared in M9 minimal medium with 1% [U-^{13}C]glucose, 10 mM [U-^{13}C]acetate and appropriate antibiotics.

Verification of module 1 of THETA in vivo with ^{13}C-labelling

Single colonies of strain SL1 transformed with pTE3245 or pTE3248 were cultured overnight in 3 ml LB with 10 mM acetate, 10 mM succinate and 50 μg ml^{-1} Kan. On the next day, the overnight culture was inoculated at 2% into 3 ml M9 minimal medium with 1% [U-^{13}C]glucose, 10 mM acetate, 10 mM succinate and 50 μg ml^{-1} Kan. The cultures were grown at 37 °C and 5% CO$_2$ for 19 h. Then cells were collected by centrifugation, washed twice with M9 minimal medium containing no carbon source, and inoculated into 3 ml M9 minimal medium with 1% [U-^{13}C]glucose, 10 mM acetate, 0.5 mM IPTG and 50 μg ml^{-1} Kan to an initial OD$_{600}$ of 0.01. Cells were grown at 37 °C and 5% CO$_2$, and were collected at an OD$_{600}$ of 0.4–0.6. The succinate in the medium was analysed with LC–MS. Note that 5% CO$_2$ was used to reduce the impact of intracellularly generated CO$_2$ on the labelling patterns.

Demonstration of module 3 of THETA in vivo with ^{13}C-labelling

Single colonies of strain SL3 or HH422 transformed with pTE4501 and pTE4502 were cultured overnight in LB supplemented with 10 mM acetate and appropriate antibiotics (acetate was omitted for HH422/HH422 with plasmids). On the following day, 1 ml of the overnight culture was collected by centrifugation, washed twice with LB, resuspended in 1 ml LB, and inoculated at 2% into 3 ml M9 minimal medium with 1% [U-^{13}C]glucose, 10 mM [U-^{13}C]acetate, 10 mM crotonate, 0.5 μM coenzyme B12, and appropriate antibiotics ([U-^{13}C]acetate was omitted for HH422/HH422 with plasmids). These cultures were grown at 37 °C to an OD$_{600}$ of 0.2–0.4, induced with 0.5 mM IPTG and grown at 37 °C. After 24 h, 48 h and 72 h, 0.3 ml cells were collected by centrifugation, and the supernatant was analysed with LC–MS for non-labelled pyruvate. Two negative controls were employed in this experiment: strain SL3 or HH422 without any plasmid cultured under the same conditions, as well as strain SL3 or HH422 containing pTE4501 and pTE4502 cultured in the same medium in the absence of crotonate.

Reporting summary

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

Data availability

The genome sequence of MG1655 used in this study is available at GenBank under accession no. NC_000913. The genome sequence of JCL301 has been deposited to GenBank under CP107281. Raw reads of Illumina and Nanopore sequencing have been deposited at the NCBI Sequence Read Archive and can be accessed under BioProject PRJNA884544. All other data are available in the main text or the Supplementary Materials. Data supporting the plots are available from the corresponding author upon request.

Code availability

The scripts for MDF analysis can be found at https://github.com/he-hai/PubSuppl/tree/master/2023_THETA. The METIS workflow (METIS_Optimization_Notebook.ipynb) used in this study is developed by Pandi et al.[30] and can be accessed at https://github.com/amirpandi/METIS. The parameters used in the workflow are described in the Methods.

References

1. Gong, F., Zhu, H., Zhang, Y. & Li, Y. Biological carbon fixation: from natural to synthetic. *J. CO2 Utilization* **28**, 221–227 (2018).
2. Rodríguez-Martínez, M. et al. Molecular biology for green recovery—a call for action. *PLoS Biol.* **20**, e3001623 (2022).
3. Fuchs, G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658 (2011).
4. Sánchez-Andrea, I. et al. The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. *Nat. Commun.* **11**, 1–12 (2020).
5. Berg, I. A. Ecological aspects of the distribution of different autotrophic CO$_2$ fixation pathways. *Appl. Environ. Microbiol.* **77**, 1925–1936 (2011).
6. Wurtzel, E. T. et al. Revolutionizing agriculture with synthetic biology. *Nat. Plants* **5**, 1207–1210 (2019).
7. Erb, T. J., Jones, P. R. & Bar-Even, A. Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol.* **37**, 56–62 (2017).
8. Bar-Even, A., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl Acad. Sci. USA* **107**, 8889–8894 (2010).
9. Löwe, H. & Kremling, A. In-depth computational analysis of natural and artificial carbon fixation pathways. *Biodes. Res.* https://doi.org/10.34133/2021/9898316 (2021).

10. Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354**, 900–904 (2016).

11. Luo, S. et al. A cell-free self-replenishing CO_2-fixing system. *Nat. Catal.* **5**, 154–162 (2022).

12. Sundaram, S. et al. A modular in vitro platform for the production of terpenes and polyketides from CO_2. *Angew. Chem. Int. Ed. Engl.* **60**, 16420–16425 (2021).

13. Diehl, C., Gerlinger, P. D., Paczia, N. & Erb, T. J. Synthetic anaplerotic modules for the direct synthesis of complex molecules from CO_2. *Nat. Chem. Biol.* **19**, 168–175 (2023).

14. Mattozzi, M. D., Ziesack, M., Voges, M. J., Silver, P. A. & Way, J. C. Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth. *Metab. Eng.* **16**, 130–139 (2013).

15. Gleizer, S. et al. Conversion of *Escherichia coli* to generate all biomass carbon from CO_2. *Cell* **179**, 1255–1263.e12 (2019).

16. Claassens, N. J. et al. Replacing the Calvin cycle with the reductive glycine pathway in *Cupriavidus necator*. *Metab. Eng.* **62**, 30–41 (2020).

17. Gassler, T. et al. The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO_2. *Nat. Biotechnol.* **38**, 210–216 (2020).

18. Schada, L. et al. An engineered Calvin–Benson–Bassham cycle for carbon dioxide fixation in *Methylobacterium extorquens* AM1. *Metab. Eng.* **47**, 423–433 (2018).

19. Erb, T. J. Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* **77**, 8466–8477 (2011).

20. Erb, T. J. et al. Synthesis of C_5-dicarboxylic acids from C_2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl Acad. Sci. USA* **104**, 10631–10636 (2007).

21. Liu, H. & Reynolds, K. A. Role of crotonyl coenzyme A reductase in determining the ratio of polyketides monensin A and monensin B produced by *Streptomyces cinnamonensis*. *J. Bacteriol.* **181**, 6806–6813 (1999).

22. Wilson, M. C. & Moore, B. S. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **29**, 72–86 (2012).

23. Noor, E. et al. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* **10**, e1003483 (2014).

24. Burgener, S., Schwander, T., Romero, E., Fraaije, M. W. & Erb, T. J. Molecular basis for converting (*2S*)-methylsuccinyl-CoA dehydrogenase into an oxidase. *Molecules* **23**, 68 (2017).

25. Mracek, J., Snyder, S. J., Chavez, U. B. & Turrens, J. F. A soluble fumarate reductase in *Trypanosoma brucei* procyclic trypomastigotes. *J. Protozool.* **38**, 554–558 (1991).

26. Nolte, J. C. et al. Novel characteristics of succinate coenzyme A (succinate-CoA) ligases: conversion of malate to malyl-CoA and CoA-thioester formation of succinate analogues in vitro. *Appl. Environ. Microbiol.* **80**, 166–176 (2014).

27. Takashima, E. et al. Characterization of the dihydroorotate dehydrogenase as a soluble fumarate reductase in *Trypanosoma cruzi*. *Mol. Biochem Parasitol.* **122**, 189–200 (2002).

28. Kahler, A. E., Nielsen, F. S. & Switzer, R. L. Biochemical characterization of the heteromeric *Bacillus subtilis* dihydroorotate dehydrogenase and its isolated subunits. *Arch. Biochem. Biophys.* **371**, 191–201 (1999).

29. Xiao, L. et al. A minimized synthetic carbon fixation cycle. *Am. Chem. Soc. Catal.* **12**, 799–808 (2022).

30. Pandi, A. et al. A versatile active learning workflow for optimization of genetic and metabolic networks. *Nat. Commun.* **13**, 3876 (2022).

31. Lee, S. Y. et al. A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2**, 18–33 (2019).

32. Barajas, J. F., Blake-Hedges, J. M., Bailey, C. B., Curran, S. & Keasling, J. D. Engineered polyketides: synergy between protein and host level engineering. *Synth. Syst. Biotechnol.* **2**, 147–166 (2017).

33. Steen, E. J. et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559–562 (2010).

34. Cruz-Morales, P. et al. Biosynthesis of polycyclopropanated high energy biofuels. *Joule* **6**, 1590–1605 (2022).

35. Schwander, T. *The Design and Realization of Synthetic Pathways for the Fixation of Carbon Dioxide In Vitro* (Philipps-Universität Marburg, 2018).

36. Duchoud, F. *Towards Improving Carbon Fixation in Plants: Cyanobacteria as a Model Organism* (University of California, Los Angeles, 2016).

37. Sonntag, F., Buchhaupt, M. & Schrader, J. Thioesterases for ethylmalonyl–CoA pathway derived dicarboxylic acid production in *Methylobacterium extorquens* AM1. *Appl. Microbiol. Biotechnol.* **98**, 4533–4544 (2014).

38. Sasikaran, J., Ziemski Michałand Zadora, P. K., Fleig, A. & Berg, I. A. Bacterial itaconate degradation promotes pathogenicity. *Nat. Chem. Biol.* **10**, 371–377 (2014).

39. Chen, F. Y. H., Jung, H. W., Tsuei, C. Y. & Liao, J. C. Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol. *Cell* **182**, 933–946.e14 (2020).

40. Claassens, N. J., Burgener, S., Vögeli, B., Erb, T. J. & Bar-Even, A. A critical comparison of cellular and cell-free bioproduction systems. *Curr. Opin. Biotechnol.* **60**, 221–229 (2019).

41. Herz, E. et al. The genetic basis for the adaptation of E. coli to sugar synthesis from CO_2. *Nat. Commun.* **8**, 1705 (2017).

42. Claassens, N. J., Cotton, C. A. R., Kopljar, D. & Bar-Even, A. Making quantitative sense of electromicrobial production. *Nat. Catal.* **2**, 437–447 (2019).

43. Kim, S. et al. Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* **16**, 538–545 (2020).

44. Peter, D. M., Vögeli, B., Cortina, N. S. & Erb, T. J. A chemo-enzymatic road map to the synthesis of CoA esters. *Molecules* **21**, 517 (2016).

45. Noor, E., Haraldsdóttir, H. S., Milo, R. & Fleming, R. M. T. Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* **9**, e1003098 (2013).

46. Nitzschke, A. & Bettenbrock, K. All three quinone species play distinct roles in ensuring optimal growth under aerobic and fermentative conditions in *E. coli* K12. *PLoS One* **13**, e0194699 (2018).

47. 2023_THETA. *GitHub* (2023) https://github.com/he-hai/PubSuppl/tree/master/2023_THETA

48. Reisch, C. R. & Prather, K. L. J. Scarless Cas9 assisted recom-bineering (no-SCAR) in *Escherichia coli*, an easy-to-use system for genome editing. *Curr. Protoc. Mol. Biol.* **117**, 31–38 (2017).

49. Jensen, S. I., Lennen, R. M., Herrgård, M. J. & Nielsen, A. T. Seven gene deletions in seven days: fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci. Rep.* **5**, 1–10 (2015).

50. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006–2008 (2006).

51. 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).

52. Wenk, S., Yishai, O., Lindner, S. N. & Bar-Even, A. in *Methods in Enzymology* (ed Scrutton N.) vol. 608 329–367 (Elsevier, 2018).

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

54. Braatsch, S., Helmark, S., Kranz, H., Koebmann, B. & Ruhdal Jensen, P. *Escherichia coli* strains with promoter libraries constructed by Red/ET recombination pave the way for transcriptional fine-tuning. *Biotechniques* **45**, 335–337 (2008).

55. Zelcbuch, L. et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **41**, e98 (2013).

56. Deatherage, D. E. & Barrick, J. E. in *Engineering and Analyzing Multicellular Systems* (ed Sun L., Shou W.) 165–188 (Springer, 2014).

57. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

58. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

59. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).

60. Milne, I. et al. Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**, 193–202 (2013).

Acknowledgements

Author contributions

S.L. and T.J.E. conceived the project. S.L. designed the THETA cycle, performed most of the experiments, and analysed the data. C.D. and Y.B. designed and performed in vitro optimization with machine learning, whereas C.D. performed all manual assays. H.H. performed genome editing and analysed genome sequencing results. M.K. assisted in plasmid construction and protein purifications. N.P., P.C. and N.S.C. performed mass spectrometry. D.S and A.A.R.R. performed NanoPore sequencing. C.A.F. assisted in constructing strain SL1. H.S.-M. constructed strain SL3. R.M. purified Sdh. S.L., C.D. and T.J.E. wrote the manuscript with contributions from all other authors.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41929-023-01079-z.

Correspondence and requests for materials should be addressed to Shanshan Luo or Tobias J. Erb.

Peer review information *Nature Catalysis* thanks Stephen Ragsdale, Michael Koepke, Wei Xiong, 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.

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

Corresponding author(s): Tobias J. Erb, Shanshan Luo

Last updated by author(s): Oct 31, 2023

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.

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Data were collected using Agilent MassHunter Workstation LC/MS Data Acquisition 10.1, Agilent Cary WinUV Kinetics Application V.5.0.0.999, Tecan i-control Version 3.9.1.0.
Data analysis	Data were analyzed using Microsoft Excel for Mac 2011 version 14.7.7, GraphPad Prism 8 for macOS version 8.4.0, Agilent MassHunter Workstation Quantitative Analysis (for Q-TOF & for QQQ) 10.0. MDF analysis used equilibrator_api v 0.4.5, equilibrator_pathway v 0.4.4. Sequencing data were analyzed using ONT Guppy base-calling software (v 6.0.1+652ffd179), breseq v0.35.7, minimap2 v2.24 (r1122), canu v2.2, ngmlr v0.2.7 and sniffles2 v2.0.3, Tablet v1.21.02.08.

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

> The genome sequence of MG1655 used in this study is available at GenBank under the accession number NC_000913. The genome sequence of JCL301 has been deposited to GenBank under CP107281. Raw reads of Illumina and Nanopore sequencing have been deposited at the NCBI Sequence Read Archive (SRA) and can be accessed under BioProject PRJNA884544. All other data are available in the main text or the supplementary materials. The source data that support the plots within this paper are available from the corresponding author upon request.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine the sample size. At least two or three independent experiments were performed in this work. Sample sizes were determined based on previous similar studies and considerations of biological and technical variabilities to ensure result confidence.
Data exclusions	No data were excluded in this study.
Replication	At least two or three independent experiments were performed to verify our results. All attempts of replication were successful.
Randomization	Randomization was not applicable in this study since it did not involve participants or samples requiring random allocation for statistical comparisons.
Blinding	Blinding was not relevant since the nature of the experiments in this study precludes any potential for subjective bias to influence the results.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

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