nature portfolio

https://doi.org/10.1038/s44160-022-00095-4

Supplementary information

Acetyl-CoA synthesis through a bicyclic carbon-fixing pathway in gas-fermenting bacteria

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION

Acetyl-CoA synthesis through a bicyclic carbon fixing pathway in gas-fermenting bacteria

Chao Wu[1]†, Jonathan Lo[1]†, Chris Urban[1], Xiang Gao[1], Bin Yang[1], Jonathan Humphreys[1], Shrameeta Shinde[2], Xin Wang[2], Katherine J. Chou[1], PinChing Maness[1], Nicolas Tsesmetzis[3], David Parker[3], Wei Xiong[1]*

[1] Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA

[2] Department of Microbiology, Miami University, Oxford, OH 45056, USA

[3] Shell International Exploration and Production Inc, Houston, TX 77082, USA

* Correspondence to: wei.xiong@nrel.gov

† Denotes equal contribution

TABLE OF CONTENTS

Supplementary Methods

In silico analysis

Pathway thermodynamics and enzyme protein cost analysis

Thermodynamics and enzyme protein cost analysis were applied to assess the feasibility of a native or designed metabolic pathway as well as corresponding protein burden. Pathway feasibility was evaluated by solving a max-min driving force (MDF) problem which seeks to maximize the Gibbs energy change, ΔG', of the most thermodynamically unfavorable reaction by tuning the concentrations of intermediates, which is defined as[1, 2, 3]:

$$\begin{aligned} \max_{\ln(\mathbf{c})} \ & \min(-\Delta G_1^{'}, -\Delta G_2^{'}, ..., -\Delta G_m^{'}) \\ s.t. \ & \mathbf{\Delta G}^{'} = \mathbf{\Delta G}^{'\mathbf{0}} + R \cdot T \cdot \mathbf{S} \cdot \ln(\mathbf{c}) \\ & \ln(\mathbf{c}_{\min}) \le \ln(\mathbf{c}) \le \ln(\mathbf{c}_{\max}) \end{aligned}$$

where m is the number of pathway enzymes, **c** is the vector of involved metabolite concentrations with upper bounds and lower bounds as $\mathbf{c}_{min}$ and $\mathbf{c}_{max}$, respectively, and **S** is the stoichiometric matrix of the pathway reactions.

In contrast, enzyme protein cost analysis aims to estimate the minimal protein mass required to support a unit pathway flux by solving a nonlinear optimization problem[1, 2]:

$$\begin{aligned} \min_{\ln(\mathbf{c})} \ & \Lambda = \frac{\sum_{i}^{m} (MW_i \cdot E_i)}{v_{in}} \\ s.t. \ & \ln(\mathbf{c}_{\min}) \le \ln(\mathbf{c}) \le \ln(\mathbf{c}_{\max}) \end{aligned}$$

where v_{in} denotes influx to a pathway, MW is enzyme molecular weight, and E is the expression of reaction enzyme level derived from a common modular rate law[4] and Haldane relationship[5].

The calculation was performed using our previously developed pathway analysis tool, PathParser[1]. The thermodynamic and kinetic parameters of enzymes in the WLP and the designed reductive acetyl-CoA Bi-cycle pathway (with CO_2 and formate fixation) were listed in Supplementary Table 2, 3 and 4. The tandard Gibbs free energies ($\Delta G'^{m}$) were searched in eQuilibrator database[6]. Michaelis constants (K_m), catalytic rate constants (k_{cat}) and enzyme molecular weights (MW) of *C. ljungdahlii*, *M. thermoacetica*, *Acetobacterium woodii*, *Clostridium formicoaceticum*[7] and *E. coli* were preferentially chosen from Brenda[8]. If no data are available, default values of 200 s^{-1}, 0.2 mM and 40 kDa were used, respectively. During optimization, concentrations of involved metabolites were constrained to vary between 1 μM and 10mM[1]. Optimization was performed using LP constructor applying the "cvxopt_lp" solver for MDF analysis from an open-source optimization package openopt, and the NLP constructor applying the "ralg" solver for protein cost estimation from the same package. To evaluate the effect of uncertainties of the kinetic parameters on estimated enzyme protein costs, we performed a Monte Carlo simulation of

500 runs with kinetic parameters log-uniformly sampled in their feasible regions defined by the available data from database.

Metabolic robustness analysis

Pathway robustness was evaluated using an ensemble modeling approach combined with the continuation method[9]. First, an ensemble of models was generated with log-uniformly sampled kinetic parameters from the feasible spaces, meanwhile subjecting to the same flux distribution of reference state[10, 11]. Then the continuation method was used to simulate the system response to enzyme expression perturbations. The dynamic system of metabolite concentrations can be expressed as:

$$\frac{d\mathbf{c}}{dt} = \mathbf{f}(\mathbf{c}, \mathbf{E}) = \mathbf{S}^T \cdot \mathbf{v}(\mathbf{c}, \mathbf{E})$$

where **f** denotes the derivative of the concentrations with respect to time which is a function of metabolite concentrations **C** and enzyme levels **E**. At metabolic steady state, **C** is constant, and the derivative of **C** with respect to t equals zero. Accordingly, the derivative of **f**(**c**, **E**) with respect to **E** equals zero too, which yields[1, 9]:

$$\frac{d\mathbf{c}_{SS}}{d\mathbf{E}} = -\left(\frac{\partial \mathbf{f}}{\partial \mathbf{c}_{SS}}\right)^{-1} \cdot \frac{\partial \mathbf{f}}{\partial \mathbf{E}}$$

where $\mathbf{c}_{SS}$ denotes metabolite concentrations at steady state. The Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{c}_{SS}}$ determines the metabolic robustness of pathway in which a system failure occurs if real part of any of the Jacobian eigenvalues pass through zero.

The above differential equations are solved along with up- and downregulation of enzyme levels until system failure happens, and the number of models is counted to calculate probability of system failure for each enzyme at varied enzyme levels. When solving the equations, metabolite concentrations will be updated, and pathway fluxes can be re-estimated accordingly responding to enzyme perturbations. Robustness analysis of native WLP and the reductive acetyl-CoA Bi-cycle pathway was also performed using PathParser[1] with an ensemble of 100 generated models.

Genome scale flux balance analysis

The synergy of reductive acetyl-CoA bi-cycle and the WLP for ethanol and acetate production under different growth mode was evaluated by flux balance analysis using the metabolic modeling package COBRApy[12]. We adapted the genome-scale metabolic model (iHN637) which consists of 785 reactions and 698 metabolites including all the major central metabolic and biosynthetic pathways in *C. ljungdahlii*[13] by adding the phosphoketolase reaction into the model. In autotrophic model, uptake of CO, CO_2 and H_2 were allowed as carbon and electron donor, whereas fructose was consumed in heterotrophic mode with CO_2 emission. Both fructose and CO were used as substrate in mixotrophic mode of *C. ljungdahlii*.

Visualization of proteomic data

Proteomic data was visualized with an online illustration tool Proteomaps which shows the quantitative composition of proteomes according to protein function[14]. Specifically, a proteomap template was first generate for *C. ljungdahlii* which contains KEGG pathways gene classification and color assignment for each classification. Proteomaps were plotted in the online server once proteomic data with normalized spectral abundance factor (NSAF) was uploaded. Details of description and usage of the tool can be found at: https://www.proteomaps.net/index.html.

Simulation of acetate labeling pattern

To simulate the labeling pattern of acetate from [1-^{13}C] pyruvate, we adopted a small reaction network (Supplementary Table 5) which consists of primary pathways in acb*/acs*$^{-}$ including gluconeogenesis, the NOG and the PFOR. The network was decomposed to generate the set of Elementary Metabolite Units (EMU) relevant to the carbon labeling of acetate and establish the isotopomer balance equations for simulation[15, 16]. The labeling pattern of acetate can be simulated as the function of pathway fluxes. For reversible reactions, the bidirectional fluxes were calculated according to[17]:

$$\begin{cases} v_{for} = \dfrac{\beta \cdot v_{exch}}{1 - v_{exch}} - \min(-v_{net}, 0) \\ v_{back} = \dfrac{\beta \cdot v_{exch}}{1 - v_{exch}} - \min(v_{net}, 0) \end{cases}$$

where v_{net} denotes the predefined pathway flux, v_{exch} denotes exchange flux which represents the extent of reversibility of reaction, and β denotes the scaler which was set to the input flux to the network. The network decomposition and labeling pattern simulation were performed using EMUlator[16] we developed previously.

Phylogenetic analysis

Protein sequences of phosphoketolase and the four subunits of PFOR (por A-D) were aligned against the genomes of acetogenic bacteria [18] using BLASTP [19] with top five hits retained which have the highest scores for the query sequences. Phylogenetic tree of the acetogenic bacteria was generated with a Python package ete3 and visualized by a R package phytools.

Experimental analysis

Strains, media, and chemicals

C. ljungdahlii DSM 13528 was purchased from the Leibniz Institute DSMZ (Braunschweig, Germany). The *C. ljungdahlii* ACS deficient strain OTA1 was a gift from Dr. Amy M Grunden at NC State University[20]. The YTF rich medium consisting of 10 g L^{-1} Bacto yeast extract, 16 g L^{-1} Bacto tryptone, 4 g L^{-1} NaCl, 5 g L^{-1} fructose and 0.5 g L^{-1} cysteine-HCl and the defined PETC medium (ATCC 1754, American Type Culture Collection, Manassas, VA)

were utilized for growing *C. ljungdahlii*. Cell growth was monitored at 600nm with a DU 800 spectrophotometer (Beckman-Coulter, Brea, CA). All chemical reagents used in growth studies were purchased from Sigma-Aldrich, except Bacto yeast extract and tryptone which were purchased from Becton Dickinson.

Phosphoketolase expression in *E. coli*, purification, and activity assay

The phosphoketolase (CA_C1343) gene from *C. acetobutylicum* ATCC 824 was synthesized from Genscript and cloned into plasmid pET28a. The successful transformants were confirmed by DNA sequencing. *E. coli* (DE3) containing the expression plasmid for N-terminal His6-tagged protein was grown on LB medium to an OD of 0.8 at 37°C, induced by 0.2 mM IPTG and harvested after overnight shaking at room temperature. Protein purification was performed according to a nickel-nitrilotriacetic (Ni-NTA) agarose minicolumn protocol[21]. The purified protein was run on a sodium dodecyl sulfate-polyacrylamide gel to monitor its size and purity. Phosphoketolase activity was measured using enzyme-linked assay[22]. Briefly, the assay included 50 mM Tri-HCl pH7.5, 5 mM $MgCl_2$, 5 mM K_3PO_4, 1 mM ADP, 2 mM glucose, 0.2 mM NADP, 0.5 U glucose kinase, 0.5 U glucose-6-phosphate 1-dehydrogenase, 1 mM thiamine pyrophosphate (TPP), 1U acetate kinase, 10 mM of ribose 5-phosphate, 2U ribulose 5-phosphate 3-epimerase, 2U ribose-5-phosphate isomerase and 2 μg purified phosphoketolase. The formation of NADPH was recorded at 340 nm by plate reader.

Construction of acb strains in *C. ljungdahlii*

Plasmid pMTL82151 was from Chain Biotech (Nottingham, UK) and used to generate the constructs transformed into *C. ljungdahlii*. The *C. ljungdahlii pta* promoter was amplified from gDNA and linked with the phosphoketolase to drive expression. To construct the plasmid, pMTL82151 was linearized with SmaI and ligated with the *pta* promoter and phosphoketolase gene using Gibson assembly from New England Biolabs (Ipswich, MA). Supplementary Methods Table 1 listed the primers used for cloning. The plasmid map and sequence of pMTL82151-pkt are provided in Supplementary Figure 10.

Supplementary Methods Table 1. Primers used for the PCR amplification and cloning.

Primer	Sequence (5'-3')	Note
XJL.579.f	TGATTACGAATTCGAGCTCGGTACCCCATTTGTCAACTATAGATGGTG	Pta promoter into SmaI site of pMTL plasmid
XJL.580.r	GTTCATTTCCTCCCTTTAAATTTAACAC	For Pta promoter linking to Ca.pkt
XJL.581.f	ATTTAAAGGGAGGAAATGAACATGCAATCTATAATTGGAAAGCATAAG	Ca.pkt linking to Pta promoter
XJL.391.r	GTCGACTCTAGAGGATCCCCTTATACGTGCCATTGCCAATTAGTTATTT	Ca.pkt into SmaI site of pMTL plasmid

Transformation was based on previously reported protocols[23]. Briefly, cultures were grown in 40 mM DL-Threonine to mid log phase (OD 0.4-0.8), then harvested by centrifugation. Cells were washed twice with ice cold SMP buffer (270 mM sucrose, 1 mM $MgCl_2$, 7 mM sodium phosphate, pH 6), then resuspended in SMP buffer with 10% dimethyl sulfoxide (DMSO) and stored at -80°C until used for transformation. Electrotransformation was performed in a COY anaerobic chamber with Gene Pulser Xcell Bio-Rad electroporator (Hercules, CA) with the following settings: in a 1 mm cuvette, 25 µL cells were mixed with 2-5 µg of DNA, pulsed 625 kV with resistance at 600 Ω and a capacitance of 25 µF. Cells were then resuspended in 5 mL of YTF and recovered overnight at 37°C. Cells were plated in 1.5% agar YTF with thiamphenicol (Tm) at a concentration of 10 µg/mL.

Activity assay of phosphoketolase in *C. ljungdahlii*

Cells in mid log phase (OD 0.4-0.8) were harvested by centrifugation. Cell pellets were resuspended in 500 µL 50 mM histidine-HCl buffer (pH 7.0) containing 20 mM KH_2PO_4-Na_2HPO_4, 2 mM dithiothreitol, 1 mM $MgSO_4$ and 5 µg Read-Lyse lysozyme (Lucigen, Middleton, WI). Resuspended cells were added to a screwcap tube containing acid washed glass beads (Sigma-Aldrich, St. Louis, MO) and lysed using a Tissuelyser II (Qiagen, Hilden, Germany) at a frequency of 30 s^{-1} for 2 min. Cell free extract was made by spinning lyse cells at 15,000g for 3 minutes and taking the supernatant.

Phosphoketolase activity assay was adapted from previously reported protocols[21, 24]. Briefly, cell free extracts of the wild-type and transformed phosphoketolase strain were added to buffer containing 25 mM xylulose 5-phosphate to generate acetyl-P, which was then converted to acetate by adding 1 µl of 1 M $MgCl_2$, 1 µl of 30 mM ADP, and 0.2 U of acetate kinase to 75 µl of the assay mixture and incubating at 30°C for 30 minutes. The acetate was then determined enzymatically with the Acetic Acid Assay Kit (Megazyme, Bray, Ireland), using an assay mixture without xylulose 5-phosphate as a control.

Proteome measurement and analysis

The proteomic analysis of wild type and acb strains were conducted following the same method in our previous report[1]. 10 µg of trypsin digested peptides from each sample were loaded onto a C18 capillary column coupled to a Thermo LTQ Orbitrap mass spectrometry (Thermo-Scientific, Rockford, IL). The peptide identity was analyzed at the resolution of 30,000. Dynamic exclusion was enabled in this case with the setup of repeat count of 1, repeat duration of 30 seconds, and exclusion duration of 90 seconds. The peptide identify was obtained by searching the tandem MS spectra using Patternlab for Proteomics[25].

Protein cost comparison was conducted by the relative quantification method used in label-free proteomics [26]. Briefly, each MS2 spectrum represents a peptide, which can be identified by matching the experimental MS spectrum with the theoretical spectrum based on constraints set in the data searching algorithm. The abundance of each protein in a

sample is represented by the sum of spectral counts from all its identified peptides. To normalize the protein size differences and protein number differences among replicates, the normalized spectral abundance factor (NSAF) [27] was used to compare between the control group and the experimental group.

^{13}C-labeling experiment and metabolite measurement by GC-MS

20 mM [U-$^{13}C_2$] acetate and 50 mM [1-^{13}C] pyruvate was separately fed to PETC media for tracking modular activities. In acetate labeling experiment, autotrophic *C. ljungdahlii* cells were grown in sealed bottles with a 4:1 v/v blend of CO/CO_2 in the headspace, and heterotrophic cultures were supplied with no gas substrates but 5 g/L fructose in the media. In pyruvate labeling experiment, pyruvate was fed as the sole carbon source. Cells in mid log phase (OD 0.4-0.8) were harvested and pretreated as described previously[28]. Briefly, 3 mL of cultures was centrifuged at 10,000g for 1 min, and the cell pellets were digested with 500 µL of 6 M HCl at 105 °C for 12 h. The hydrolysate was dried under nitrogen gas flow at 65 °C. For GC-MS measurement, the dried hydrolysate was dissolved in 50 µL of water-free dimethylformamide and derivatized by TBDMS with 1% tert-butyl-dimethylchlorosilane at 85 °C for 60 min. Labeling patterns of proteinogenic amino acids from cell mass and acetate from culture supernatant were then analyzed by an Agilent 6890 GC equipped with a 5975 MS Detector (Santa Clara, CA) as previously reported[16, 29]. The oven temperature was initially held at 50°C for 2 min which was then raised to 150°C at 5°C min^{-1} and held for 2 min. Finally, it was raised to 320°C at 7°C min^{-1} and held for 2 min.

Analysis of *C. ljungdahlii* strains grown under heterotrophic, mixotrophic and autotrophic conditions

C. ljungdahlii was first grown in liquid YTF medium, then centrifuged (Beckman-Coulter, Brea, CA) at 4000g for 15 min. The pellet was transferred into PETC medium with 5 g L^{-1} fructose and grown on a MaxQ 5000 shaker (Thermo Fisher Scientific, Waltham, MA) at 37°C. Cells in mid log phase (OD 0.4-0.8) were harvested again by centrifugation at 4000g for 15 min, and the pellet was transferred to Duran Pressure Plus laboratory bottles (DWK Life Sciences, Mainz, Germany) containing PETC media with 5 g L^{-1} fructose for heterotrophic and mixotrophic growth or no fructose for autotrophic growth at an initial OD of 0.1.

The bottles for autotrophic and mixotrophic growth were degassed with a 4:1 v/v blend of CO/CO_2 then pressurized to 2 psi. The cultures were grown at 37°C, 220rpm on the MaxQ 5000 shaker. Cell growth was monitored every 24 or 48 hours. Headspace pressure was measured with an DG25 Digital Pressure Gauge (Ashcroft, Stratford, CT), and metabolites including acetate and ethanol were analyzed by an Agilent 1200 high-pressure liquid chromatography (Santa Clara, CA) with an Aminex HPX-87H column using 4 mM H_2SO_4 as the eluent. Liquid samples were filtered prior to analysis.

Analysis of *acs*- and acb/*acs*- strains for H_2 and CO_2 fermentation

The PETC medium and YTF medium without fructose (also denoted as YT medium) were used for this experiment. 15mM sodium nitrate were supplemented into the medium before fermentation. The bottles for H_2 and CO_2 fermentation were degassed with a 1:1 v/v blend of H_2/CO_2 then pressurized to 7 psi. The cultures were grown at 37°C, 220rpm on the MaxQ 5000 shaker. Headspace pressure was measured daily. When pressure was lower than 7 psi, the headspace was refilled with H_2 and CO_2 to 7psi.

Supplementary Table 1. Enzymes involved in the designed bicyclic acetyl-CoA pathway for syngas-fermenting in *Clostridium ljungdahlii*. Of the 16 enzymes, 15 are native in as being found in the genome of *C. ljungdahlii* DSM 13528 except phosphoketolase.

Enzyme ID	Enzyme	Gene Number
pta	Phosphotransacetylase	CLJU_c12770
pfor[a]	Pyruvate:ferredoxin oxidoreductase	CLJU_c29340
pfl[b]	Pyruvate formate lyase	CLJU_c25980
pc	Pyruvate carboxylase	CLJU_c37390
pepck	Phosphoenolpyruvate carboxykinase	CLJU_c06210
end	Endolase	CLJU_c39110
pgm	Phosphoglycerate mutase	CLJU_c26320
pgk	Phosphoglycerate kinase	CLJU_c39140
gapdh	Glyceraldehyde-3-phosphate dehydrogenase	CLJU_c13400
tpi	Triose phosphate isomerase	CLJU_c39103
fba	Fructose 1,6-bisphosphate aldolase	CLJU_c02810
fbp	Fructose 1,6-bisphosphatase	CLJU_c29050
tal	Transaldolase	CLJU_c25960 CLJU_c39640
tkt	Transketolase	CLJU_c03050 CLJU_c03051 CLJU_c25830 CLJU_c25820
rpi	Ribose-5-phosphate isomerase	CLJU_c02310 CLJU_c02190
rpe	Ribulose-5-phosphate epimerase	CLJU_c01210
pkt	Phosphoketolase	-

[a]: Variant of the designed pathway using CO as substrate.

[b]: Variant of the designed pathway using formate as substrate.

Supplementary Table 2. Parameters for thermodynamics and kinetics analysis of the designed reductive acetyl-CoA Bi-cycle pathway (CO_2 fixation) in *C. ljungdahlii*. Standard Gibbs free energies ($\Delta G'^m$) were searched in eQuilibrator database, Michaelis constants (K_m), catalytic rate constants (k_{cat}) and enzyme molecular weight (MW) were collected from references available or chosen from Brenda. Data from Clostridia was preferential used, then that from *Escherichia coli*. If multiple data are available, they are presented as geometric mean(lower bound, upper bound). If no data are available, default values of 200 s^{-1}, 0.2 mM and 40 kDa were assigned for k_{cat}, K_m and MW, respectively.

Enzyme ID	Reversibility	$\Delta G'^m$ (kJ mol^{-1})	Substrates	Products	Substrate K_m (mM)	Product K_m (mM)	k_{cat} (s^{-1})	Enzyme MW (kDa)
[a]tkt1	1	9.4	F6P;G3P	E4P;X5P	1.1;2.1	0.09;0.16	200	32.423
pkt	0	-58.2	X5P;Pi	G3P;AcP	0.2;0.2	0.2;0.2	200	40
pta	1	-11.3	CoA;AcP	AcCoA;Pi	0.56;0.66	0.02(0.0095,0.045);2.1	135.2	38
pfor	1	13.2	CO2;AcCoA;2rFdx	Pyr;CoA;2oFdx	2;0.009[30];0.2	0.2;0.2;0.2	3.2[30]	129.94
pc	0	-4.2	Pyr;ATP;CO2	OAA;ADP;Pi	0.2;0.2;0.2	0.2;0.2;0.2	200	128.5
pepck	0	-7.5	OAA;Pi	PEP;CO2	0.2;0.2	0.34;0.19	540	58.769
end	1	4.1	PEP	G2P	0.2	0.2	200	46.94
pgm	1	-4.7	G2P	PGA	0.2	4.7(4.2,5.2)	200	56.3
pgk	1	17.7	PGA;ATP	Go3P;ADP	0.2;0.2	0.2;0.2	200	42.956
gapdh	1	-31.3	Go3P;NADPH	G3P;NADP;Pi	0.2;0.2	0.89;0.2;0.53	200	38.575
tpi	1	-5.1	G3P	DHAP	0.2	0.2	9000	27.172
fba	1	-2.9	G3P;DHAP	FBP	0.2;0.2	0.17(0.12,0.3)	10.8	33.579
fbp	0	-29.6	FBP	F6P;Pi	0.009(0.0012,1)	0.2;0.2	12.1	77.46
tal	1	0.4	E4P;F6P	G3P;S7P	0.16(0.078,0.555); 1.15(0.6,4.9)	0.27(0.038,1.9);0.285	13	24.446
[a]tkt2	1	4.1	G3P;S7P	R5P;X5P	2.1;4	1.4;0.16	200	32.423
rpi	1	1.5	R5P	Ru5P	2.2(0.83,4.4)	0.2	2100	16.385
rpe	1	-3.4	Ru5P	X5P	0.2	0.2	200	23.776
ak	1	-12.6	ADP;AcP	ATP;Ac	3.355(0.71,6);0.58	1.435(0.37,2.5); 116.5(73,160)	200	88

[a]: tkt involved in reactions with different reactants was assigned different IDs.

Supplementary Table 3. Parameters for thermodynamics and kinetics analysis of the designed reductive acetyl-CoA Bi-cycle pathway (formate fixation) in *C. ljungdahlii*. Standard Gibbs free energies ($\Delta G'^{m}$) were searched in eQuilibrator database, Michaelis constants (K_m), catalytic rate constants (k_{cat}) and enzyme molecular weight (MW) were collected from references available or chosen from Brenda. Data from Clostridia was preferential used, then that from *Escherichia coli*. If multiple data are available, they are presented as geometric mean(lower bound, upper bound). If no data are available, default values of 200 s^{-1}, 0.2 mM and 40 kDa were assigned for k_{cat}, K_m and MW, respectively.

Enzyme ID	Reversibility	$\Delta G'^{m}$ (kJ mol^{-1})	Substrates	Products	Substrate K_m (mM)	Product K_m (mM)	k_{cat} (s^{-1})	Enzyme MW (kDa)
[a]tkt1	1	9.4	F6P;G3P	E4P;X5P	1.1;2.1	0.09;0.16	200	32.423
pkt	0	-58.2	X5P;Pi	G3P;AcP	0.2;0.2	0.2;0.2	200	40
pta	1	-11.3	CoA;AcP	AcCoA;Pi	0.56;0.66	0.02(0.0095,0.045);2.1	135.2	38
pfl	1	13.2	Fm;AcCoA	Pyr;CoA	14(10,20);0.26	1(0.6,1.6);0.012	200	17
pc	0	-4.2	Pyr;ATP;CO2	OAA;ADP;Pi	0.2;0.2;0.2	0.2;0.2;0.2	200	128.5
pepck	0	-7.5	OAA;Pi	PEP;CO2	0.2;0.2	0.34;0.19	540	58.769
end	1	4.1	PEP	G2P	0.2	0.2	200	46.94
pgm	1	-4.7	G2P	PGA	0.2	4.7(4.2,5.2)	200	56.3
pgk	1	17.7	PGA;ATP	Go3P;ADP	0.2;0.2	0.2;0.2	200	42.956
gapdh	1	-31.3	Go3P;NADPH	G3P;NADP;Pi	0.2;0.2	0.89;0.2;0.53	200	38.575
tpi	1	-5.1	G3P	DHAP	0.2	0.2	9000	27.172
fba	1	-2.9	G3P;DHAP	FBP	0.2;0.2	0.17(0.12,0.3)	10.8	33.579
fbp	0	-29.6	FBP	F6P;Pi	0.009(0.0012,1)	0.2;0.2	12.1	77.46
tal	1	0.4	E4P;F6P	G3P;S7P	0.16(0.078,0.555); 1.15(0.6,4.9)	0.27(0.038,1.9);0.285	13	24.446
[a]tkt2	1	4.1	G3P;S7P	R5P;X5P	2.1;4	1.4;0.16	200	32.423
rpi	1	1.5	R5P	Ru5P	2.2(0.83,4.4)	0.2	2100	16.385
rpe	1	-3.4	Ru5P	X5P	0.2	0.2	200	23.776
ak	1	-12.6	ADP;AcP	ATP;Ac	3.355(0.71,6);0.58	1.435(0.37,2.5); 116.5(73,160)	200	88

[a]: tkt involved in reactions with different reactants was assigned different IDs.

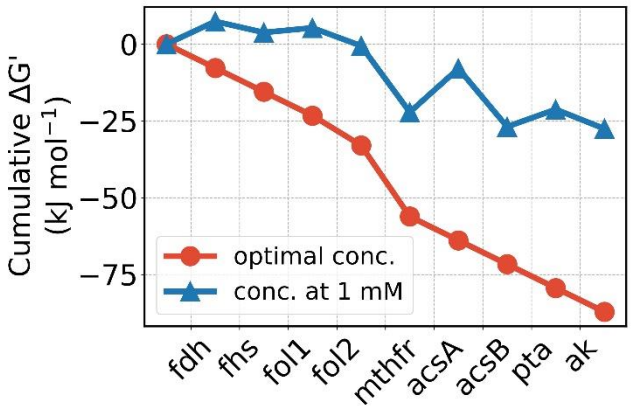
Supplementary Table 4. Parameters for thermodynamics and kinetics analysis of the Wood-Ljungdahl pathway in *C. ljungdahlii*. Standard Gibbs free energies ($\Delta G'^{m}$) were searched in eQuilibrator database, Michaelis constants (K_m), catalytic rate constants (k_{cat}) and enzyme molecular weight (MW) were collected from references available or chosen from Brenda. Data from Clostridia was preferential used, then that from *Escherichia coli*. If multiple data are available, they are presented as geometric mean(lower bound, upper bound). If no data are available, default values of 200 s^{-1}, 0.2 mM and 40 kDa were assigned for k_{cat}, K_m and MW, respectively.

Enzyme ID	Reversibility	$\Delta G'^{m}$ (kJ mol^{-1})	Substrates	Products	Substrate K_m (mM)	Product K_m (mM)	k_{cat} (s^{-1})	Enzyme MW (kDa)
fdh	1	14.8	CO2;NADH	Fm;NAD	0.2;0.05	0.2;0.2	6.2[7]	80.7
fhs	1	-7.2	THF;Fm;ATP	FormylTHF;ADP;Pi	0.22(0.055,0.69); 8.2(3,20.2); 0.1555(0.091,0.22)	10;0.06(0.032,0.13);5	420.6[31, 32, 33]	240
fol1	1	3	FormylTHF	MethenylTHF	0.2	0.19	200	41
fol2	1	-11.7	MethenylTHF;NADPH	MethyleneTHF;NADP	0.057(0.026,0.228); 0.029(0.0004,0.69)	0.127(0.026,0.228); 0.0595(0.034,0.085)	1600[7]	70
mthfr	1	-43.2	MethyleneTHF;NADH	MethylTHF;NAD	0.001(0.0005,0.0039); 0.026(0.017,0.066)	0.12;0.2	324[7]	124
acsA	1	28.4	CO2;rFdx;0.5NADH	CO;oFdx;0.5NAD	0.2;0.2;0.2	0.2;0.2;0.2	10[34]	67.955
acsB	1	-37.9	MethylTHF;CoA;CO	THF;AcCoA	0.2;0.2;0.2	0.2;0.2	2.1[35]	83.554
pta	1	11.3	AcCoA;Pi	CoA;AcP	0.02(0.0095,0.045);2.1	0.56;0.66	135.2	38
ak	1	-12.6	ADP;AcP	ATP;Ac	3.355;0.58	1.435(0.37,2.5); 116.5(73,160)	200	88

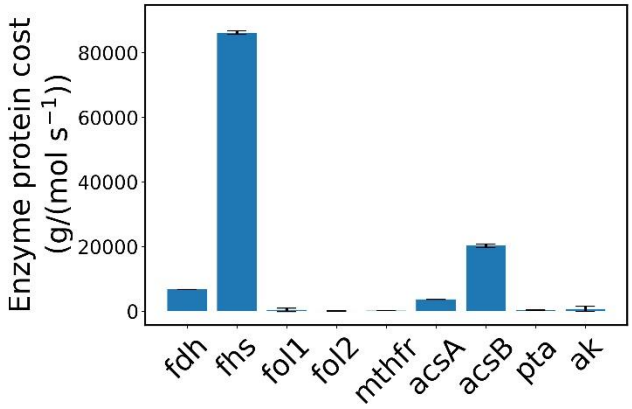
Supplementary Table 5. Reaction list with atom transitions for simulation of the labeling pattern of acetate using [1-^{13}C] pyruvate as substrate.

Reaction ID	Substrates(atom)	Products(atom)	Reversibility
pt	Pyr.ex(abc)	Pyr(abc)	0
pc	Pyr(abc)+CO_2(d)	OAA(abcd)	0
pepck	OAA(abcd)	PEP(abc)+CO_2(d)	0
end	PEP(abc)	PGA(abc)	1
pgk	PGA(abc)	G3P(abc)	1
tpi	G3P(abc)	DHAP(abc)	1
fba	DHAP(cba)+G3P(def)	FBP(abcdef)	1
fbp	FBP(abcdef)	F6P(abcdef)	0
tal	E4P(abcd)+F6P(efghjk)	G3P(hjk)+S7P(efgabcd)	1
tkt2	G3P(abc)+S7P(defghjk)	X5P(deabc)+R5P(fghjk)	1
rpi	R5P(abcde)	Ru5P(abcde)	1
rpe	Ru5P(abcde)	X5P(abcde)	1
tkt1	G3P(abc)+F6P(defghi)	X5P(deabc)+E4P(fghi)	1
pkt	X5P(abcde)	AcP(ba)+G3P(cde)	0
pfor	Pyr(abc)	AcCoA(bc)+CO_2(a)	0
pta	AcCoA(ab)	AcP(ab)	1
ak	AcP(ab)	Ac(ab)	0

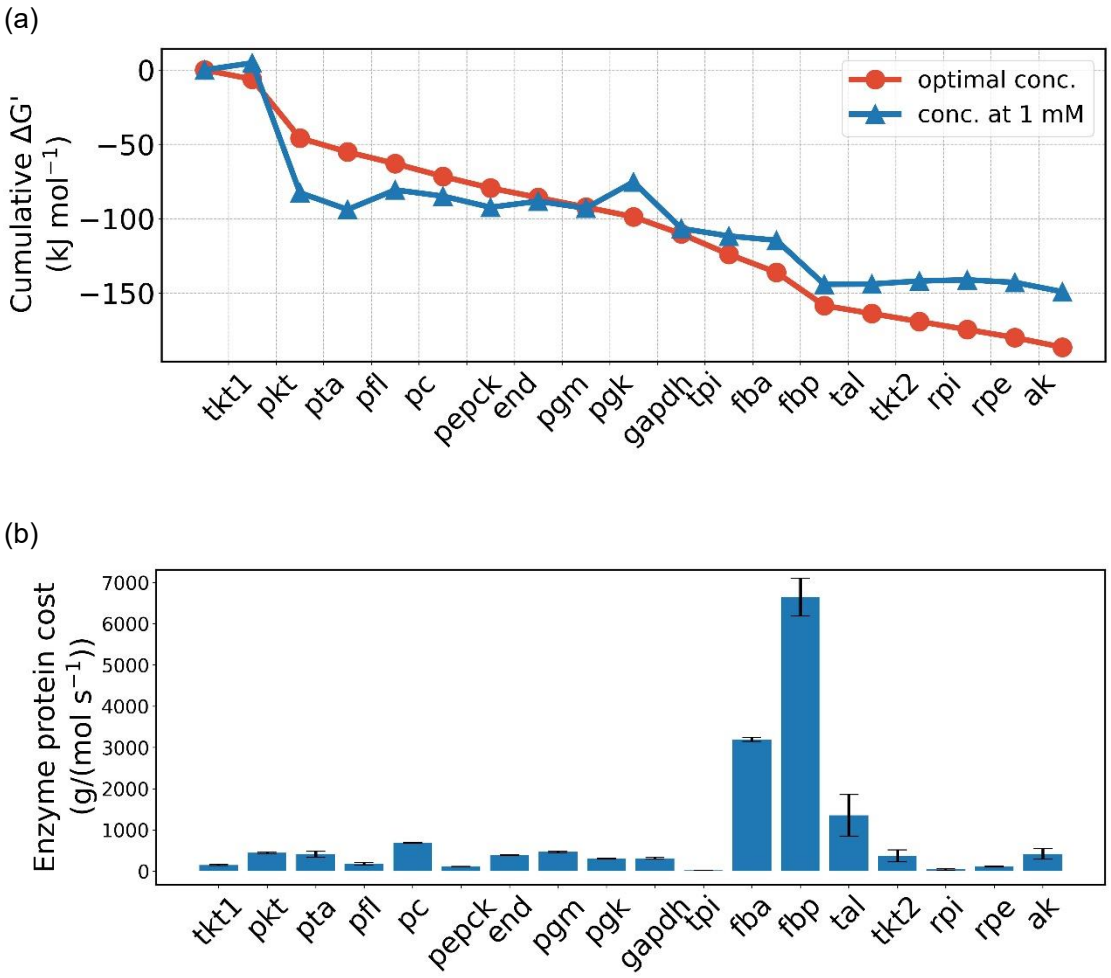
(a)



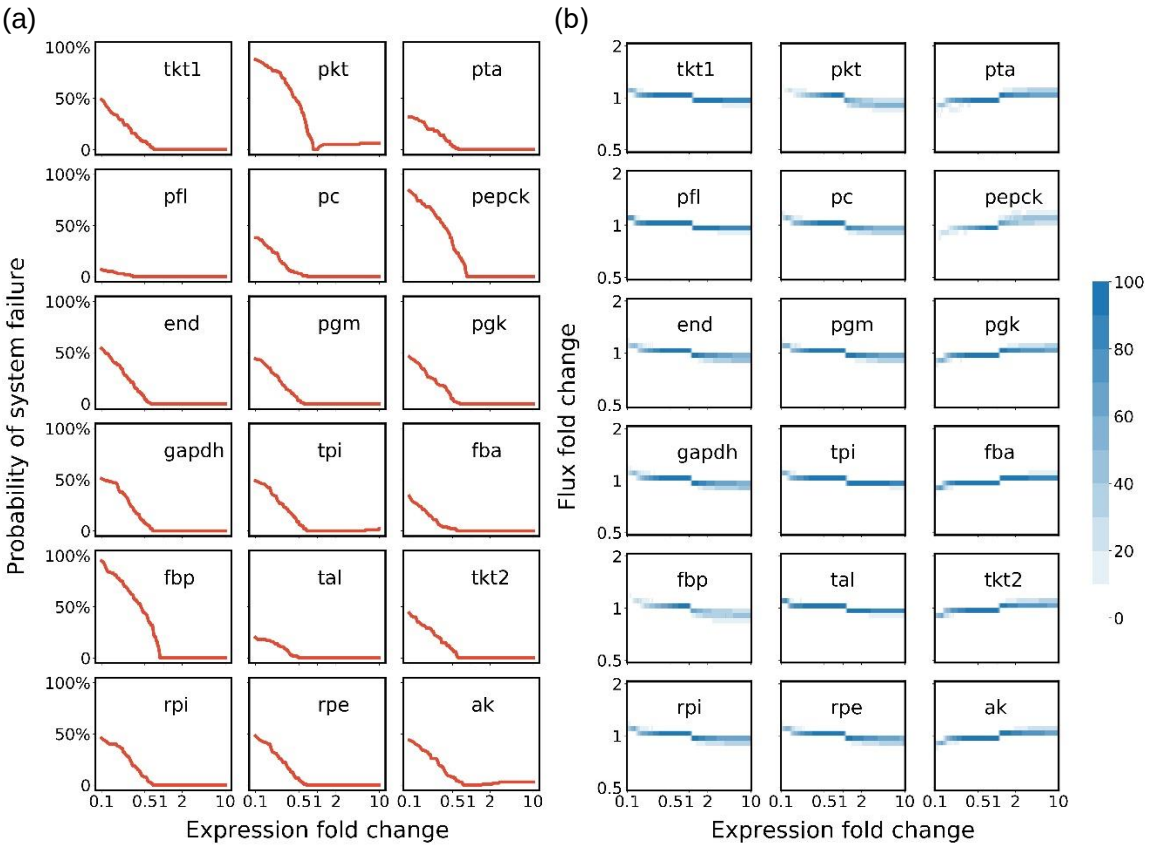
(b)



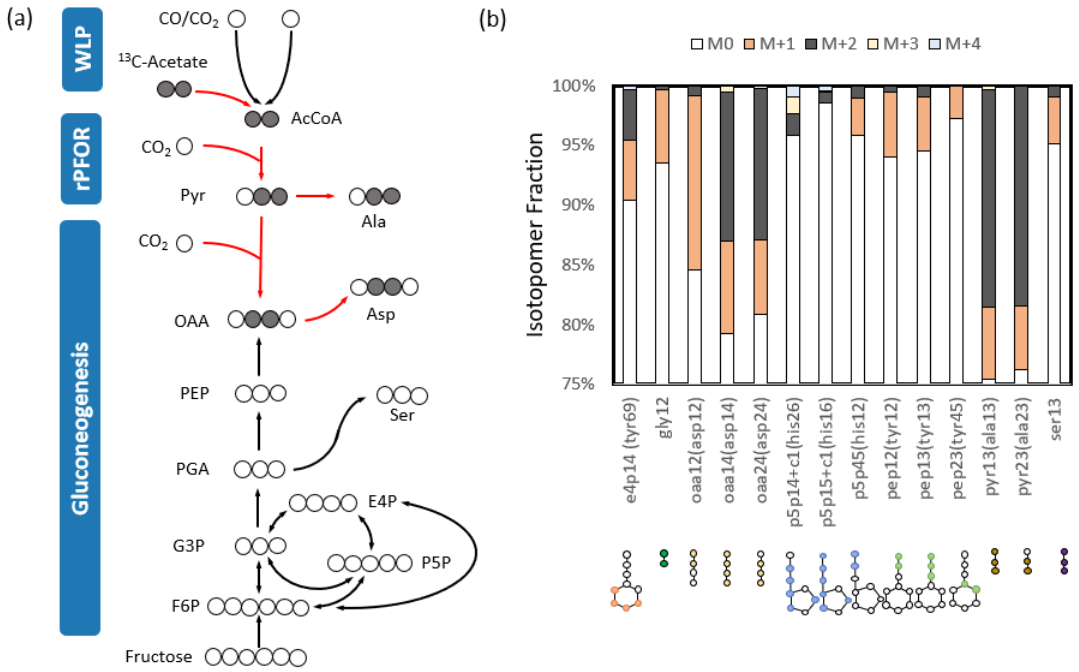
Supplementary Figure 1. Thermodynamics (a) and enzyme protein cost (b) analysis of the Wood-Ljungdahl pathway. Thermodynamic driving force of pathway is presented as the cumulative sum of reaction Gibbs energies, ΔG'. Blue line denotes standard Gibbs energies with all metabolite concentrations fixed at 1 mM, and red line denotes Gibbs energies when the highest reaction ΔG′ is iteratively maximized in the negative direction. Enzyme protein costs ($1.2\times10^{5} \pm 1.8\times10^{3}$ g/(mol s^{-1})) are estimated by minimizing the total enzyme mass required to support unit metabolic flux through the pathway. The uncertainties were obtained by performing the estimation 500 times with kinetic parameters log-uniformly sampled in their feasible regions defined by the available data from database. All optimizations were performed with metabolite concentrations constrained to range from 1 μM to 10 mM. The estimations were made under the assumption that all enzymes are working at their optimal temperature. Enzyme abbreviations: acsA, Carbon monoxide dehydrogenase; acsB, CO-methylating acetyl-CoA synthase; ak, Acetate kinase; fdh, Formate dehydrogenase; fhs, Formyltetrahydrofolate synthetase; fol1, Methenyltetrahydrofolate cyclohydrolase; fol2, Methylenetetrahydrofolate dehydrogenase; mthfr, Methylenetetrahydrofolate reductase; pta, Phosphotransacetylase.



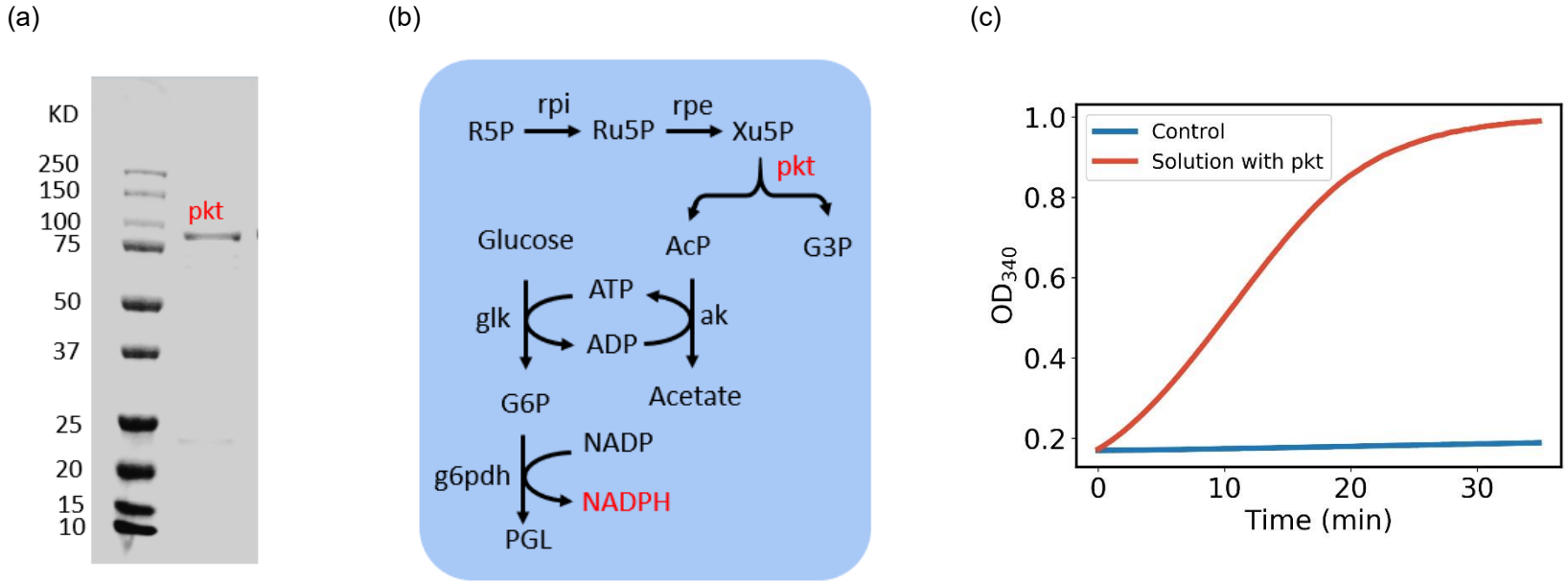
Supplementary Figure 2. Thermodynamics (a) and enzyme protein cost (b) analysis of the reductive acetyl-CoA Bi-cycle for formate fixation. Pathway's ability to fix formate is of interest because formate is a soluble C1 substrate and could serve as a driving force for new C1 bioeconomy [36]. Thermodynamic driving force of pathway is presented as the cumulative sum of reaction Gibbs energies, ΔG'. Blue line denotes standard Gibbs energies with all metabolite concentrations fixed at 1 mM, and red line denotes Gibbs energies when the highest reaction ΔG′ is iteratively maximized in the negative direction. Enzyme protein costs ($1.6\times10^4 \pm 6.8\times10^2$ g/(mol s^{-1})) are estimated by minimizing the total enzyme mass required to support unit metabolic flux through the pathway. The uncertainties were obtained by performing the estimation 500 times with kinetic parameters log-uniformly sampled in their feasible regions defined by the available data from database. All optimizations were performed with metabolite concentrations constrained to range from 1 μM to 10 mM. The estimations were made under the assumption that all enzymes are working at their optimal temperature. Enzyme abbreviations: ak, Acetate kinase; end, Endolase; fba, Fructose-1,6-bisphosphate aldolase; fbp, Fructose-1,6-bisphosphatase; gapdh, Glyceraldehyde-3-phosphate dehydrogenase; pc, Pyruvate carboxylase ; pepck, Phosphoenolpyruvate carboxykinase; pfl, pyruvate formate lyase; pgk, Phosphoglycerate kinase; pgm, Phosphoglycerate mutase; pkt, Phosphoketolase; pta, Phosphotransacetylase; rpe, Ribulose-5-phosphate epimerase; rpi, Ribose-5-phosphate isomerase; tal, Transaldolase; tkt, Transketolase; tpi, Triose phosphate isomerase. Tkt involved in reactions with different reactants was marked with numbers.



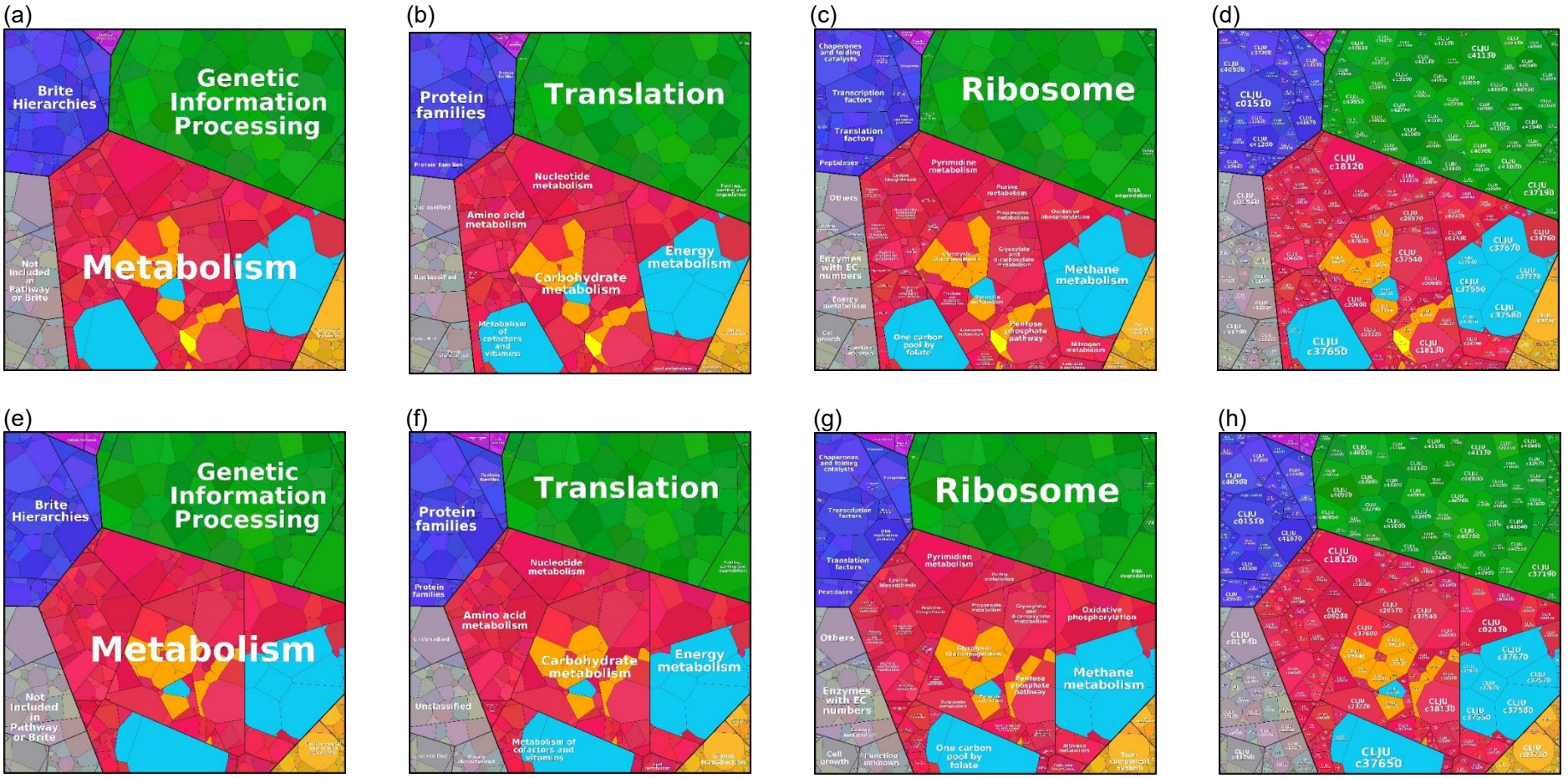
Supplementary Figure 3. Pathway robustness (a) and flux change (b) in response to enzyme level perturbations in the reductive acetyl- CoA Bi-cycle for formate fixation. Pathway robustness is represented by the probability of system failure at varied fold change of enzyme expression levels over reference state. A pathway is considered as entering system failure when any intermediate is depleted or over accumulated over time, and the probability of system failure is calculated as counts in an ensemble of 100 models which were generated with log-uniformly sampled kinetic parameters from the feasible spaces, meanwhile subjecting to the same flux distribution of reference state. The flux change plot on the right is a set of heatmaps, and the color indicates the number of models of corresponding flux fold change at some enzyme expression level for each enzyme.



Supplementary Figure 4. Functionality of only carbon fixation module proved by labeling experiment using [U-$^{13}C_2$] acetate in heterotrophic mode. (a) Reactions and metabolites involved in carbon fixation and gluconeogenesis modules. Red arrows denote active reactions verified by the formation of labeled downstream metabolites. (b) Isotopomer fraction of metabolite fragments inferred from measurable amino acid fragments. M0 denotes fraction of fragments with all ^{12}C carbon atoms, and M+i denotes fraction of fragments with i labeled carbon atoms. The isotopomer fraction denotes mean value of biological duplicates. Labeled metabolites are presented as “precursor fragment (amino acid fragment)” and contributing carbon atoms from the precursor are colored.



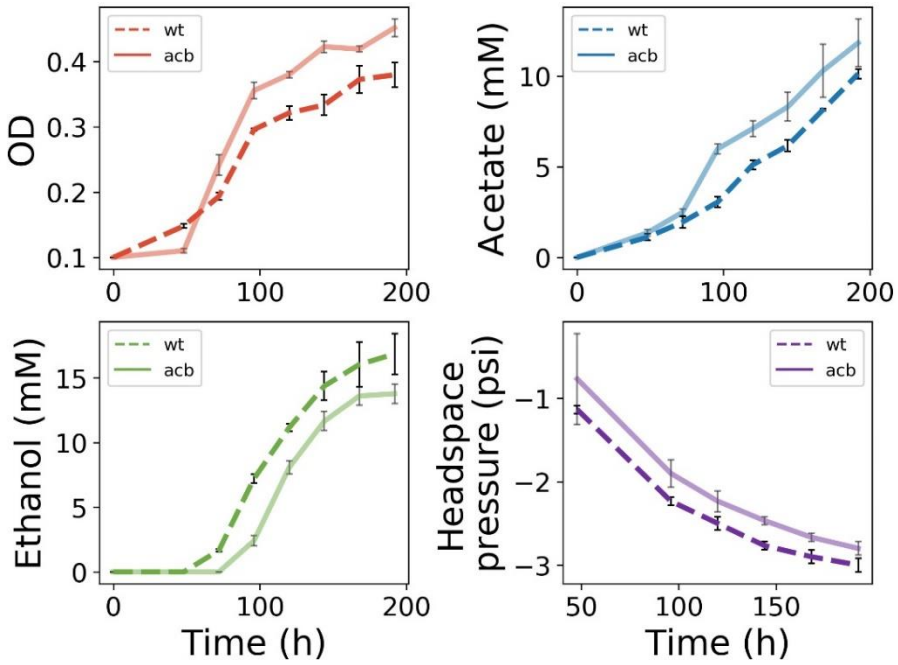
Supplementary Figure 5. Purification and activity measurement of heterologous phosphoketolase. (a) SDS-PAGE gel of purified phosphoketolase; (b) Principle of in vitro enzyme assay of phosphoketolase. The formation of NADPH was recorded as the indicator of enzyme activity; (c) Solution with phosphoketolase showed a significant higher absorbance at 340 nm than the control. The absorbance curve denotes the representative of biological triplicates. Abbreviations: G6P, glucose 6-phosphate; g6pdh, glucose-6-phosphate 1-dehydrogenase; glk, glucose kinase; PGL, 6-phospho-D-glucono-1,5-lactone; Xu5P, xylulose 5-phosphate.



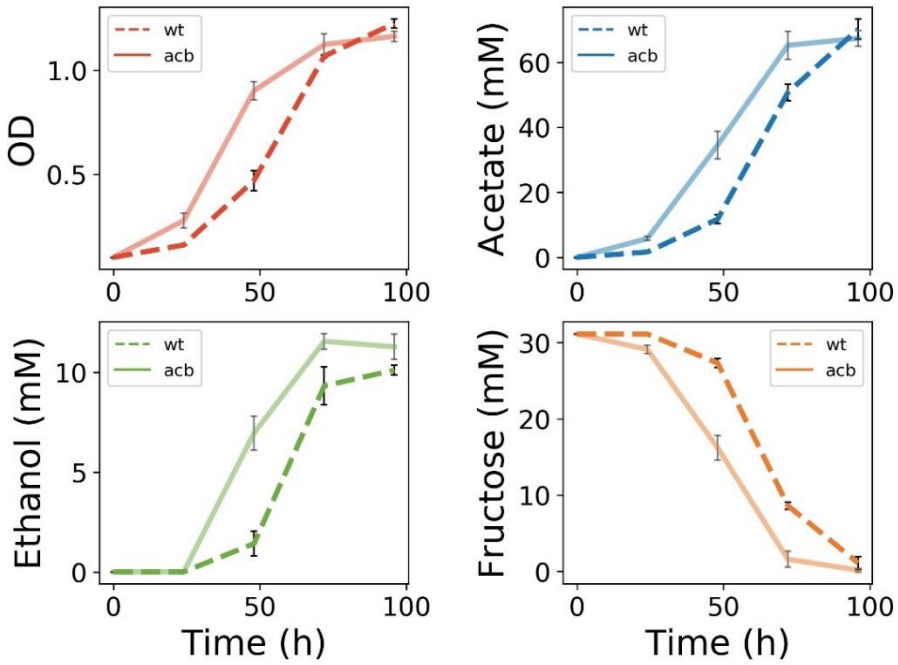
Supplementary Figure 6. Pathway enzyme allocation in the proteome of *C. ljungdahlii* acb strain (a - d) and wild type (wt) strain (e - h) grown on CO and CO_2. Proteins are illustrated in gradually detailed categorization. Pathway enzymes for carbon fixation are under the Metabolism category. Every tile (small polygon) represents one type of protein. Tile sizes denote the mass fractions of proteins from the representative of biological duplicates.

The enzymes in the Wood-Ljungdahl pathway are shown in cyan and enzymes in the reductive acetyl-CoA bi-cycle are shown in ocher. Engineered phosphoketolase are shown in bright yellow.

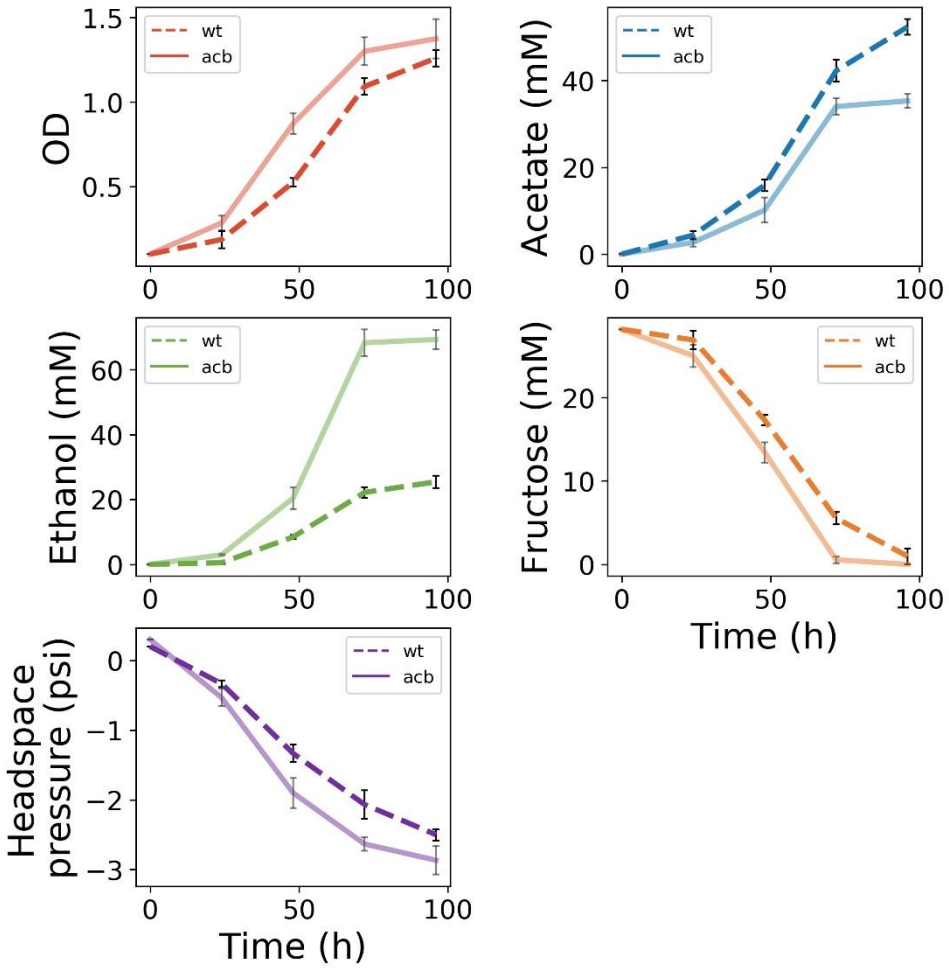
(a)


wt
acb
OD
Acetate (mM)
Ethanol (mM)
Headspace pressure (psi)
Time (h)
Time (h)

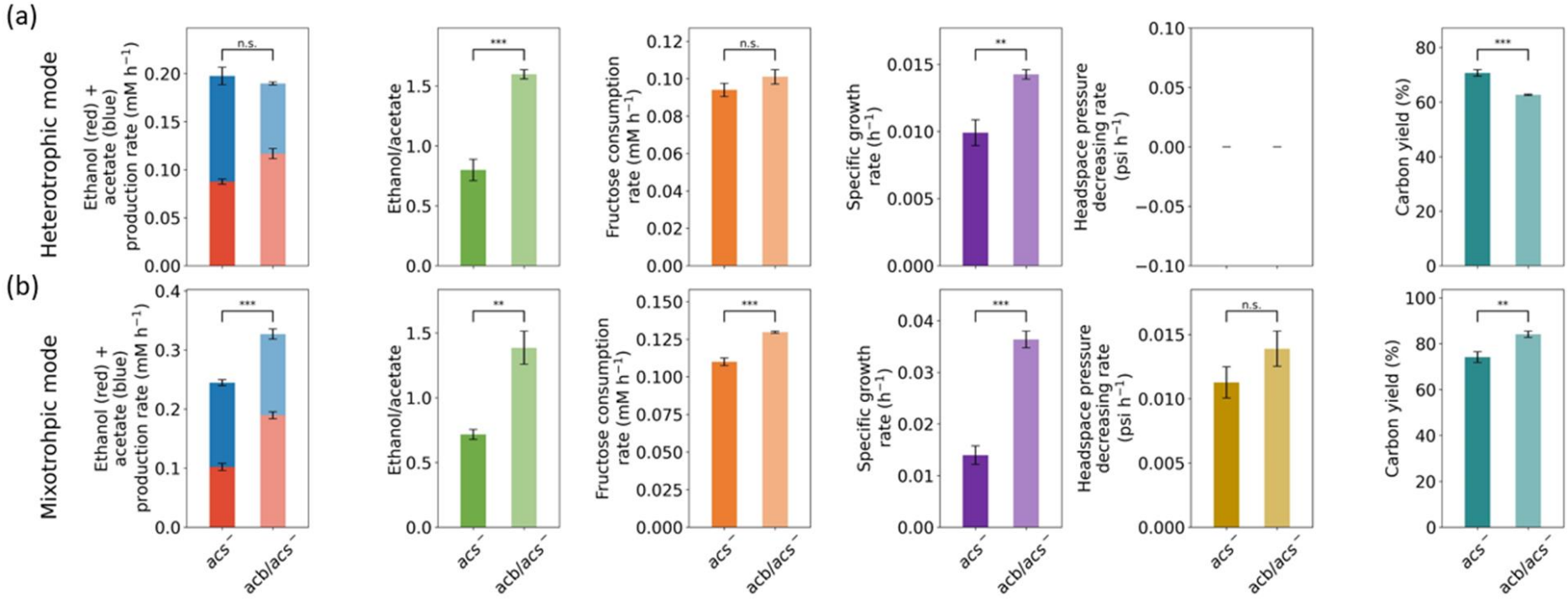
(b)


wt
acb
OD
Acetate (mM)
Ethanol (mM)
Fructose (mM)
Time (h)
Time (h)

(c)

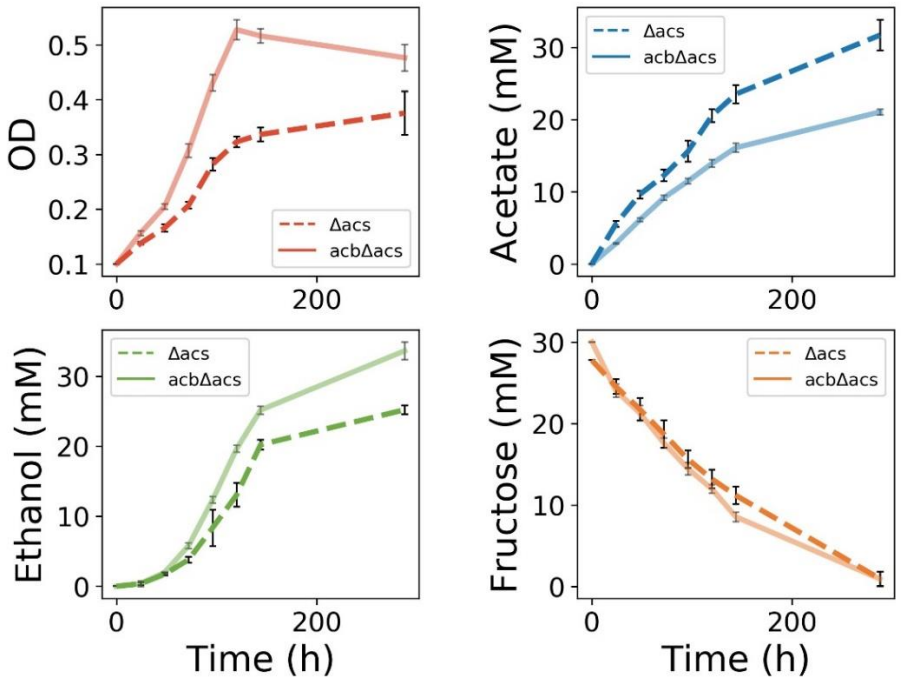


Supplementary Figure 7. Growth, product formation and substrate consumption of the wt and heterologous phosphoketolase expressed strain (acb) in autotrophic (a), heterotrophic (b) and mixotrophic (c) modes. In autotrophic mode, CO, CO_2 were fed as carbon source, whereas fructose and fructose/CO were used as substrates in heterotrophic mode and mixotrophic mode, respectively. Error bars represent standard deviations from biological triplicates.

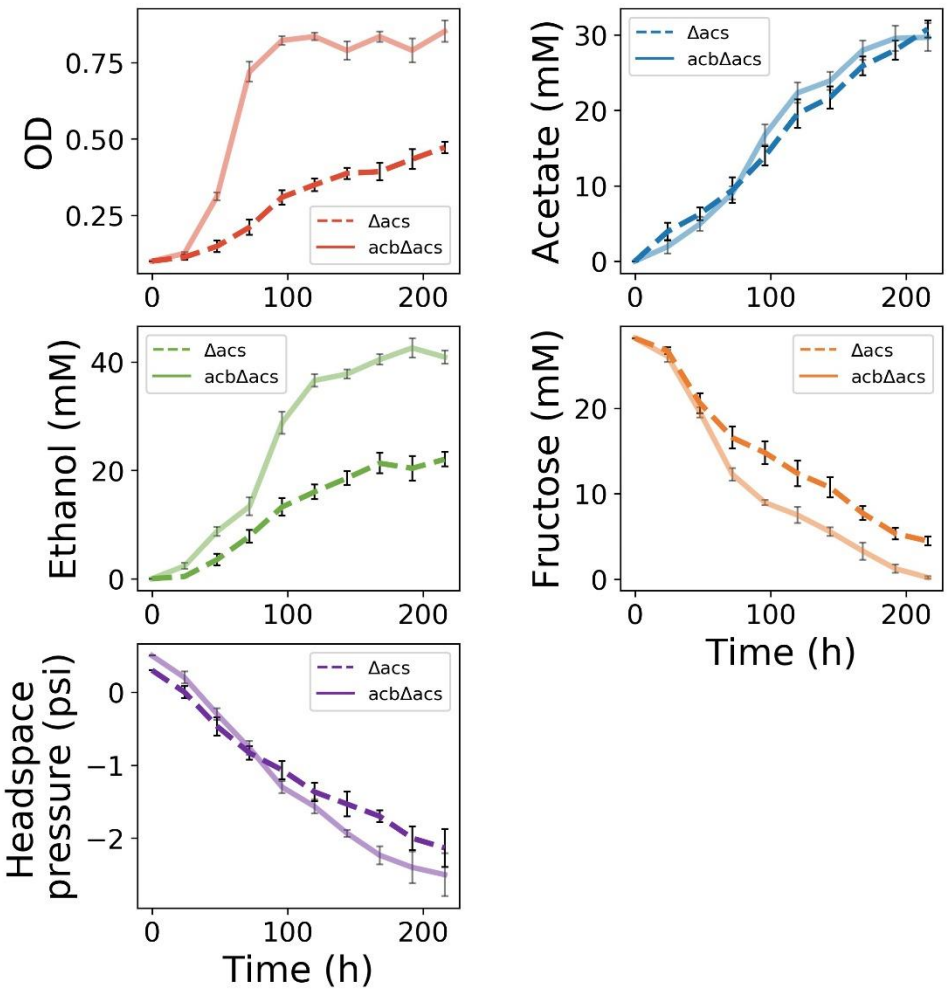


Supplementary Figure 8. Growth and productivities of the CO-methylating acetyl-CoA synthase knockout strain (*acs*$^-$) and heterologous phosphoketolase expressed *acs*$^-$ strain (acb/*acs*$^-$) in heterotrophic (a) and mixotrophic (b) modes. Fructose and fructose/CO were used as substrates in heterotrophic mode and mixotrophic mode, respectively. In the calculation of carbon yield, only organic carbon source, i.e., fructose was considered in the denominator. Error bars represent standard deviations from biological triplicates. Student's t-test was performed to compare the difference of each specie. '*' denotes pvalue < 0.05, '**' denotes pvalue < 0.005, '***' denotes pvalue < 0.5×10^{-3}, '******' denotes pvalue < 0.5×10^{-6} and 'n.s.' denotes no significance.

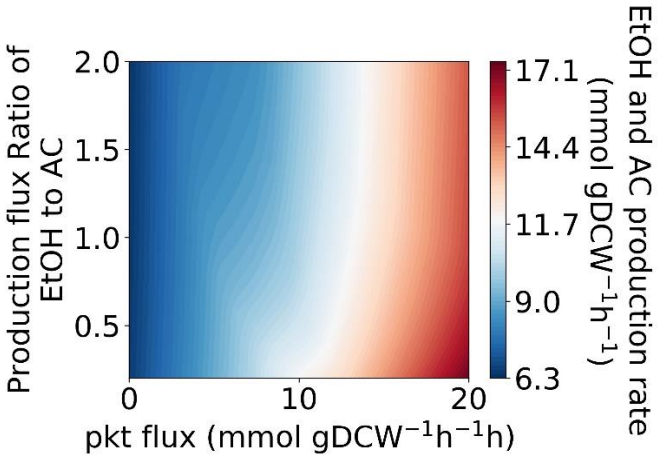
(a)


OD
Acetate (mM)
Ethanol (mM)
Fructose (mM)
Time (h)
Time (h)
Δacs
acbΔacs

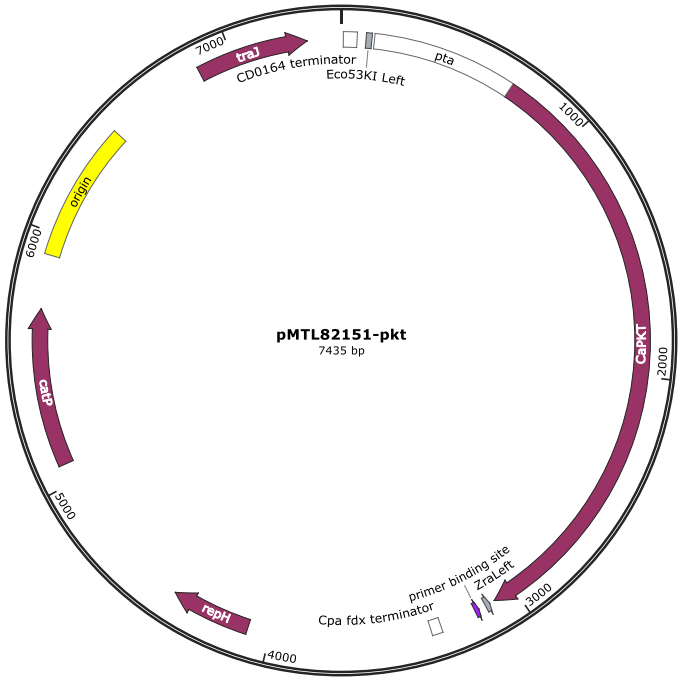
(b)


OD
Acetate (mM)
Ethanol (mM)
Fructose (mM)
Headspace pressure (psi)
Time (h)
Time (h)
Δacs
acbΔacs

Supplementary Figure 9. Growth, product formation and substrate consumption of the CO-methylating acetyl-CoA synthase knockout strain (*acs*$^-$) and heterologous phosphoketolase expressed *acs*$^-$ strain (acb/*acs*$^-$) in heterotrophic (a) and mixotrophic (b) modes. Fructose and fructose/CO were used as substrates in heterotrophic mode and mixotrophic mode, respectively. Error bars represent standard deviations from biological triplicates.



Supplementary Figure 10. Contour profiling of C2 metabolites productivity as a function of the ratio of production rate of ethanol to acetate and the reductive acetyl-CoA bi-cycle activity in heterotrophic mode. The contour was profiled by using a flux balance mode of *C. ljungdahlii*[13] with the addition of the phosphoketolase reaction setting specific growth rate as the objective function. Fructose and fructose/CO were consumed as substrates in heterotrophic mode. X axis denotes the activity of acetyl-CoA bi-cycle represented by phosphoketolase (pkt) flux. Y axis denotes acetyl-CoA synthase (acs) flux as a proxy of the WLP.



Supplementary Figure 11. The plasmid map and sequence for pMTL82151-pkt.

Sequence (5'-3'):

CCTGCAGGATAAAAAAATTGTAGATAAATTTTATAAAATAGTTTTATCTACAATTTTTTTATCAGGAAACA
GCTATGACCGCGGCCGCTGTATCCATATGACCATGATTACGAATTCGAGCTCGGTACCCCATTTGTC
AACTATAGATGGTGAACCTCTTTGCACGAAGTTACCACTCATAACTGCTATTACAGCATCACATTTTGT
AGCAGAACGAGCACTTTCAATATGATATTTATGTCCATTGTGAAAGGGATTATATTCAACTATTATTCCA
GTTACGTTCATAGAAATTTTCCTTTCTAAAATATTTTATTCCATGTCAAGAACTCTGTTTATTTCATTAAA
GAACTATAAGTACAAAGTATAGGGCATTTGAAAAAATAGGCTAGTATATTGATTGATTATTTATTTTAAAA
TGCCTAAGTGAAATATATACATATTATAACAATAAAATAAGTATTAGTGTAGGATTTTTAAATAGAGTATCT
ATTTTCAGATTAAATTTTTGCTTATTTGATTTACATTATATAATATTGAGTAAAGTATTGACTAGCAAAATT
TTTTGATACTTTAATTTGTGAAATTTCTTATCAAAAGTTATATTTTTGAATGATTTTTATTGAAAAATACAA
CTAAAAAGGATTATAGTATAAGTGTGTGTAATTTTGTGTTAAATTTAAAGGGAGGAAATGAACATGCAA
TCTATAATTGGAAAGCATAAGGACGAAGGTAAGATTACACCTGAGTACTTAAAGAAAATTGACGCTTA
CTGGAGAGCAGCAAATTTTATAAGTGTGGGACAACTTTATCTACTAGATAACCCACTTTTAAGAGAGC
CATTAAAACCTGAACATCTTAAAAGGAAGGTGGTGGGACATTGGGGAACTATTCCTGGTCAAAATTTC
ATTTATGCTCATCTAAACAGAGTTATAAAAAAGTATGATTTAGATATGATTTACGTATCAGGTCCTGGTC
ACGGTGGACAGGTTATGGTTTCAAACTCTTATTTAGACGGAACATACTCTGAGGTATACCCAAATGTG
TCTAGGGATCTTAATGGATTGAAGAAATTGTGCAAACAATTCAGCTTTCCTGGAGGTATATCTAGTCA
CATGGCTCCTGAAACTCCAGGATCTATAAACGAAGGAGGTGAGCTTGGTTATAGTCTAGCTCACAGT
TTTGGAGCTGTGTTCGATAATCCAGATCTAATTACAGCATGCGTTGTGGGTGATGGAGAGGCTGAGA
CAGGACCTCTTGCTACTTCTTGGCAGGCAAATAAATTCCTTAATCCAGTGACTGATGGAGCAGTTCTA
CCAATACTACACCTTAATGGTTACAAGATATCTAACCCAACTGTTTTGTCAAGAATACCTAAAGACGAG
CTAGAGAAGTTTTTTGAAGGAAATGGATGGAAGCCTTACTTCGTTGAAGGTGAAGATCCTGAGACAA
TGCACAAATTGATGGCAGAAACATTAGATATTGTAACAGAGGAGATACTTAACATACAGAAAAACGCA
AGAGAAAACAATGACTGTTCTAGACCTAAATGGCCTATGATTGTACTAAGAACACCAAAGGGATGGA
CAGGTCCTAAATTCGTTGACGGAGTGCCAAACGAGGGTTCATTCAGGGCACATCAGGTACCACTTG
CAGTAGACAGATACCACACTGAGAATTTAGACCAGCTAGAGGAGTGGCTTAAGTCATATAAACCAGA
GGAATTATTTGATGAAAACTACAGATTGATACCTGAACTAGAGGAGCTTACACCTAAAGGAAATAAGA
GAATGGCTGCAAACTTGCACGCAAATGGAGGTCTACTACTAAGAGAACTTAGGACTCCAGATTTCAG
GGACTACGCTGTGGATGTTCCTACTCCAGGAAGTACAGTGAAGCAAGATATGATAGAGCTAGGTAAG
TACGTAAGGGATGTGGTGAAGCTAAATGAGGACACTAGAAACTTTAGGATTTTCGGACCTGATGAGA
CTATGTCAAACAGATTGTGGGCTGTATTCGAGGGTACTAAAAGGCAATGGCTTAGCGAGATTAAGGA
ACCTAATGACGAATTTCTATCTAACGATGGTAGAATTGTTGACAGTATGTTATCTGAGCACCTATGCGA
AGGTTGGCTAGAAGGTTACTTGCTTACTGGAAGACACGGATTCTTCGCTTCTTACGAAGCATTTCTTA
GAATTGTAGATAGCATGATAACTCAGCACGGAAAATGGTTAAAGGTAACATCTCAGCTTCCTTGGAGG
AAGGACATTGCTTCTTTAAACCTAATAGCAACATCAAATGTATGGCAGCAGGATCACAATGGTTATACA
CACCAAGATCCTGGTCTTCTTGGTCATATTGTTGACAAAAAACCTGAAATTGTGAGGGCATACTTGCC

AGCAGACGCTAACACATTACTTGCAGTTTTTGACAAATGTCTTCATACAAAGCATAAAATAAACTTACT
AGTGACTTCTAAGCATCCAAGACAACAATGGCTTACTATGGACCAAGCAGTAAAGCATGTGGAACAA
GGTATTAGTATTTGGGATTGGGCTTCAAACGATAAAGGTCAAGAACCTGATGTGGTGATTGCTTCATG
CGGTGACACACCTACTCTAGAGGCTTTAGCAGCAGTGACAATTTTGCATGAACATTTGCCAGAGCTT
AAAGTGAGATTCGTTAATGTAGTGGACATGATGAAGCTTTTGCCTGAGAATGAGCACCCTCATGGTC
TTTCAGATAAGGACTATAACGCATTATTCACAACAGACAAGCCAGTGATATTTGCATTTCATGGTTTTG
CACATCTAATAAATCAACTAACTTATCACAGAGAAAACAGAAACTTGCATGTTCATGGTTACATGGAAG
AGGGTACAATAACAACACCTTTTGACATGAGAGTTCAAAACAAATTGGATAGATTCAATCTTGTAAAA
GACGTAGTGGAGAACCTACCACAATTAGGTAACAGAGGTGCACACTTGGTACAGTTGATGAACGAC
AAATTAGTGGAGCATAACCAATATATAAGAGAAGTAGGAGAGGATTTGCCTGAAATAACTAATTGGCAA
TGGCACGTATAAGGGGATCCTCTAGAGTCGACGTCACGCGTCCATGGAGATCTCGAGGCCTGCAGA
CATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAA
CTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGAT
CGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTAGCATAAAAATAAGAAGCCTGCA
TTTGCAGGCTTCTTATTTTTATGGCGCGCCGCCATTATTTTTTTGAACAATTGACAATTCATTTCTTATT
TTTTATTAAGTGATAGTCAAAAGGCATAACAGTGCTGAATAGAAAGAAATTTACAGAAAAGAAAATTATA
GAATTTAGTATGATTAATTATACTCATTTATGAATGTTTAATTGAATACAAAAAAAAATACTTGTTATGTAT
TCAATTACGGGTTAAAATATAGACAAGTTGAAAAATTTAATAAAAAAATAAGTCCTCAGCTCTTATATATT
AAGCTACCAACTTAGTATATAAGCCAAAACTTAAATGTGCTACCAACACATCAAGCCGTTAGAGAACT
CTATCTATAGCAATATTTCAAATGTACCGACATACAAGAGAAACATTAACTATATATATTCAATTTATGAG
ATTATCTTAACAGATATAAATGTAAATTGCAATAAGTAAGATTTAGAAGTTTATAGCCTTTGTGTATTGGA
AGCAGTACGCAAAGGCTTTTTTATTTGATAAAAATTAGAAGTATATTTATTTTTTCATAATTAATTTATGAA
AATGAAAGGGGGTGAGCAAAGTGACAGAGGAAAGCAGTATCTTATCAAATAACAAGGTATTAGCAATA
TCATTATTGACTTTAGCAGTAAACATTATGACTTTTATAGTGCTTGTAGCTAAGTAGTACGAAAGGGGG
AGCTTTAAAAAGCTCCTTGGAATACATAGAATTCATAAATTAATTTATGAAAAGAAGGGCGTATATGAAA
ACTTGTAAAAATTGCAAAGAGTTTATTAAAGATACTGAAATATGCAAAATACATTCGTTGATGATTCATG
ATAAAACAGTAGCAACCTATTGCAGTAAATACAATGAGTCAAGATGTTTACATAAAGGGAAAGTCCAAT
GTATTAATTGTTCAAAGATGAACCGATATGGATGGTGTGCCATAAAAATGAGATGTTTTACAGAGGAA
GAACAGAAAAAAGAACGTACATGCATTAAATATTATGCAAGGAGCTTTAAAAAAGCTCATGTAAAGAA
GAGTAAAAAGAAAAAATAATTTATTTATTAATTTAATATTGAGAGTGCCGACACAGTATGCACTAAAAAA
TATATCTGTGGTGTAGTGAGCCGATACAAAAGGATAGTCACTCGCATTTTCATAATACATCTTATGTTAT
GATTATGTGTCGGTGGGACTTCACGACGAAAACCCACAATAAAAAAAGAGTTCGGGGTAGGGTTAA
GCATAGTTGAGGCAACTAAACAATCAAGCTAGGATATGCAGTAGCAGACCGTAAGGTCGTTGTTTAG
GTGTGTTGTAATACATACGCTATTAAGATGTAAAAATACGGATACCAATGAAGGGAAAAGTATAATTTTT
GGATGTAGTTTGTTTGTTCATCTATGGGCAAACTACGTCCAAAGCCGTTTCCAAATCTGCTAAAAAGT
ATATCCTTTCTAAAATCAAAGTCAAGTATGAAATCATAAATAAAGTTTAATTTTGAAGTTATTATGATATTA
TGTTTTTCTATTAAAATAAATTAAGTATATAGAATAGTTTAATAATAGTATATACTTAATGTGATAAGTGTCT
GACAGTGTCACAGAAAGGATGATTGTTATGGATTATAAGCGGCCGGCCAGTGGGCAAGTTGAAAAAT
TCACAAAAATGTGGTATAATATCTTTGTTCATTAGAGCGATAAACTTGAATTTGAGAGGGAACTTAGAT
GGTATTTGAAAAAATTGATAAAAATAGTTGGAACAGAAAAGAGTATTTTGACCACTACTTTGCAAGTGT
ACCTTGTACCTACAGCATGACCGTTAAAGTGGATATCACACAAATAAAGGAAAAGGGAATGAAACTAT
ATCCTGCAATGCTTTATTATATTGCAATGATTGTAAACCGCCATTCAGAGTTTAGGACGGCAATCAATC
AAGATGGTGAATTGGGGATATATGATGAGATGATACCAAGCTATACAATATTTCACAATGATACTGAAA
CATTTTCCAGCCTTTGGACTGAGTGTAAGTCTGACTTTAAATCATTTTTAGCAGATTATGAAAGTGATA
CGCAACGGTATGGAAACAATCATAGAATGGAAGGAAAGCCAAATGCTCCGGAAAACATTTTTAATGTA
TCTATGATACCGTGGTCAACCTTCGATGGCTTTAATCTGAATTTGCAGAAAGGATATGATTATTTGATT
CCTATTTTTACTATGGGGAAATATTATAAAGAAGATAACAAAATTATACTTCCTTTGGCAATTCAAGTTC
ATCACGCAGTATGTGACGGATTTCACATTTGCCGTTTTGTAAACGAATTGCAGGAATTGATAAATAGTT
AACTTCAGGTTTGTCTGTAACTAAAAACAAGTATTTAAGCAAAAACATCGTAGAAATACGGTGTTTTTT
GTTACCCTAAGTTTAAACTCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCC
ACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAAT
CTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACC
AACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAG
CCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGT
TACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTAC
CGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAAC
GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAG
AAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCA
GGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTT
TGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC
CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCG
TATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAG
TGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAGGGCCCCCTGCTTCGGGGTCATTATAGCGATT
TTTTCGGTATATCCATCCTTTTTCGCACGATATACAGGATTTTGCCAAAGGGTTCGTGTAGACTTTCCT
TGGTGTATCCAACGGCGTCAGCCGGGCAGGATAGGTGAAGTAGGCCCACCCGCGAGCGGGTGTTC
CTTCTTCACTGTCCCTTATTCGCACCTGGCGGTGCTCAACGGGAATCCTGCTCTGCGAGGCTGGCC
GGCTACCGCCGGCGTAACAGATGAGGGCAAGCGGATGGCTGATGAAACCAAGCCAACCAGGAAGG

GCAGCCCACCTATCAAGGTGTACTGCCTTCCAGACGAACGAAGAGCGATTGAGGAAAAGGCGGCG
GCGGCCGGCATGAGCCTGTCGGCCTACCTGCTGGCCGTCGGCCAGGGCTACAAAATCACGGGCG
TCGTGGACTATGAGCACGTCCGCGAGCTGGCCCGCATCAATGGCGACCTGGGCCGCCTGGGCGG
CCTGCTGAAACTCTGGCTCACCGACGACCCGCGCACGGCGCGGTTCGGTGATGCCACGATCCTCG
CCCTGCTGGCGAAGATCGAAGAGAAGCAGGACGAGCTTGGCAAGGTCATGATGGGCGTGGTCCG
CCCGAGGGCAGAGCCATGACTTTTTTAGCCGCTAAAACGGCCGGGGGGTGCGCGTGATTGCCAAG
CACGTCCCCATGCGCTCCATCAAGAAGAGCGACTTCGCGGAGCTGGTGAAGTACATCACCGACGA
GCAAGGCAAGACCGATCGGGCCC

References

1. Wu C*, et al.* A generalized computational framework to streamline thermodynamics and kinetics analysis of metabolic pathways. *Metab Eng* **57**, 140-150 (2020).

2. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences* **110**, 10039-10044 (2013).

3. Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol* **10**, e1003483 (2014).

4. Liebermeister W, Uhlendorf J, Klipp E. Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* **26**, 1528-1534 (2010).

5. Alberty RA. Relations between biochemical thermodynamics and biochemical kinetics. *Biophysical Chemistry* **124**, 11-17 (2006).

6. Flamholz A, Noor E, Bar-Even A, Milo R. eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res* **40**, D770-775 (2012).

7. Ragsdale SW, Pierce E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2) fixation. *Biochim Biophys Acta* **1784**, 1873-1898 (2008).

8. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* **47**, D542-D549 (2019).

9. Lee Y, Lafontaine Rivera JG, Liao JC. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metab Eng* **25**, 63-71 (2014).

10. Tan Y, Rivera JG, Contador CA, Asenjo JA, Liao JC. Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metab Eng* **13**, 60-75 (2011).

11. Tran LM, Rizk ML, Liao JC. Ensemble modeling of metabolic networks. *Biophys J* **95**, 5606-5617 (2008).

12. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* **7**, 74 (2013).

13. Nagarajan H*, et al.* Characterizing acetogenic metabolism using a genome-scale

metabolic reconstruction of Clostridium ljungdahlii. *Microb Cell Fact* **12**, 118 (2013).

14. Liebermeister W, Noor E, Flamholz A, Davidi D, Bernhardt J, Milo R. Visual account of protein investment in cellular functions. *Proc Natl Acad Sci U S A* **111**, 8488-8493 (2014).

15. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab Eng* **9**, 68-86 (2007).

16. Wu C, Chen CH, Lo J, Michener W, Maness P, Xiong W. EMUlator: An Elementary Metabolite Unit (EMU) Based Isotope Simulator Enabled by Adjacency Matrix. *Front Microbiol* **10**, 922 (2019).

17. Wiechert W, de Graaf AA. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnol Bioeng* **55**, 101-117 (1997).

18. Bengelsdorf FR*, et al.* Bacterial Anaerobic Synthesis Gas (Syngas) and CO2+H2 Fermentation. *Adv Appl Microbiol* **103**, 143-221 (2018).

19. Sayers EW*, et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res* **50**, D20-D26 (2022).

20. Whitham JM*, et al.* Characterization of Clostridium ljungdahlii OTA1: a non-autotrophic hyper ethanol-producing strain. *Appl Microbiol Biotechnol* **101**, 1615-1630 (2017).

21. Liu L*, et al.* Phosphoketolase pathway for xylose catabolism in Clostridium acetobutylicum revealed by 13C metabolic flux analysis. *J Bacteriol* **194**, 5413-5422 (2012).

22. Bogorad IW, Lin TS, Liao JC. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **502**, 693-697 (2013).

23. Leang C, Ueki T, Nevin KP, Lovley DR. A genetic system for Clostridium ljungdahlii: a chassis for autotrophic production of biocommodities and a model homoacetogen. *Appl Environ Microbiol* **79**, 1102-1109 (2013).

24. Sonderegger M, Schumperli M, Sauer U. Metabolic engineering of a phosphoketolase pathway for pentose catabolism in Saccharomyces cerevisiae. *Appl Environ Microbiol* **70**, 2892-2897 (2004).

25. Carvalho PC*, et al.* Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. *Nat Protoc* **11**, 102-117 (2016).

26. Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, 840518 (2010).

27. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J Proteome Res* **5**, 2339-2347 (2006).

28. Xiong W*, et al.* CO2-fixing one-carbon metabolism in a cellulose-degrading bacterium Clostridium thermocellum. *Proc Natl Acad Sci U S A* **113**, 13180-13185 (2016).

29. Xiong W, *et al.* Isotope-Assisted Metabolite Analysis Sheds Light on Central Carbon Metabolism of a Model Cellulolytic Bacterium Clostridium thermocellum. *Front Microbiol* **9**, 1947 (2018).

30. Furdui C, Ragsdale SW. The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. *J Biol Chem* **275**, 28494-28499 (2000).

31. Elliott JI, Ljungdahl LG. Chemical modification of cysteine and tyrosine residues in formyltetrahydrofolate synthetase from Clostridium thermoaceticum. *Arch Biochem Biophys* **215**, 245-252 (1982).

32. Ljungdahl L, Brewer JM, Neece SH, Fairwell T. Purification, stability, and composition of formyltetrahydrofolate synthetase from Clostridium thermoaceticum. *J Biol Chem* **245**, 4791-4797 (1970).

33. Lovell CR, Przybyla A, Ljungdahl LG. Cloning and expression in Escherichia coli of the Clostridium thermoaceticum gene encoding thermostable formyltetrahydrofolate synthetase. *Arch Microbiol* **149**, 280-285 (1988).

34. Can M, Armstrong FA, Ragsdale SW. Structure, function, and mechanism of the nickel metalloenzymes, CO dehydrogenase, and acetyl-CoA synthase. *Chem Rev* **114**, 4149-4174 (2014).

35. Roberts JR, Lu WP, Ragsdale SW. Acetyl-coenzyme A synthesis from methyltetrahydrofolate, CO, and coenzyme A by enzymes purified from Clostridium thermoaceticum: attainment of in vivo rates and identification of rate-limiting steps. *J Bacteriol* **174**, 4667-4676 (1992).

36. Yishai O, Lindner SN, Gonzalez de la Cruz J, Tenenboim H, Bar-Even A. The formate bio-economy. *Curr Opin Chem Biol* **35**, 1-9 (2016).