

Enzyme engineering and *in vivo* testing of a formate reduction pathway

Jue Wang¹, Karl Anderson¹, Ellen Yang¹, Lian He¹, and Mary E. Lidstrom^{1,2,*}

¹Department of Chemical Engineering, University of Washington, Seattle, DC, USA

²Department of Microbiology, University of Washington, Seattle, DC, USA

*Corresponding author: E-mail: jue@uw.edu

Abstract

Formate is an attractive feedstock for sustainable microbial production of fuels and chemicals, but its potential is limited by the lack of efficient assimilation pathways. The reduction of formate to formaldehyde would allow efficient downstream assimilation, but no efficient enzymes are known for this transformation. To develop a 2-step formate reduction pathway, we screened natural variants of acyl-CoA synthetase (ACS) and acylating aldehyde dehydrogenase (ACDH) for activity on one-carbon substrates and identified active and highly expressed homologs of both enzymes. We then performed directed evolution, increasing ACDH-specific activity by 2.5-fold and ACS lysate activity by 5-fold. To test for the *in vivo* activity of our pathway, we expressed it in a methylotroph which can natively assimilate formaldehyde. Although the enzymes were active in cell extracts, we could not detect formate assimilation into biomass, indicating that further improvement will be required for formatotrophy. Our work provides a foundation for further development of a versatile pathway for formate assimilation.

Key words: formate assimilation; one carbon; directed evolution; metabolic engineering; enzyme engineering

1. Introduction

Population growth and climate change have created an urgent need for processes to produce more food, fuel and chemicals while reducing CO₂ emissions. Engineered microbes have the potential to renewably produce many useful chemicals (1). However, most commercial bioproduction uses expensive sugar feedstocks that compete with the food supply. Carbon dioxide, as a ubiquitous industrial waste and greenhouse gas, is an attractive feedstock, but CO₂-fixing organisms are technically challenging to adapt to industrial scale. These problems can potentially be solved by bio-inorganic hybrid systems, where electricity drives catalytic production of an energy-carrying molecule used by microbes to produce value-added compounds (2, 3). Coupled to advanced photovoltaics, these systems can achieve solar-to-biomass conversion efficiencies approaching 10%, well beyond values of 3% for microalgae and 1% for plants (4).

Formate is an attractive energy carrier for a bio-inorganic system because it can be produced efficiently by electrocatalysis (5), is highly soluble in water, and provides both carbon and reducing power to microbes (2, 6). Formate can also be derived from waste biomass and fossil carbon, making it a flexible feedstock for bridging existing and future carbon economies (2). Unfortunately, organisms that naturally consume formate are poorly suited to industrial use, and moreover, natural formate assimilation pathways are theoretically less efficient in their consumption of adenosine triphosphate (ATP) and reducing equivalents than rationally

designed alternatives (6–8). Recently, the first synthetic formate assimilation pathway, the reductive glycine pathway (rGlyP), was successfully introduced into *Escherichia coli* to support growth on formate and CO₂ as sole carbon sources (9, 10). Although the rGlyP is energy-efficient and has great biotechnological potential, it involves a CO₂-fixation step that requires a high ambient CO₂ concentration in order to operate, potentially limiting its range of applications.

Several alternative formate assimilation pathways have been proposed that could rival the efficiency of the rGlyP while not requiring CO₂ fixation (Figure 1) (7). These pathways all have an initial step where formate is reduced to formaldehyde, which could potentially be achieved in two enzymatic reactions via a formyl-CoA intermediate (8). For example, the ribulose monophosphate (RuMP) pathway naturally occurs in methylotrophic bacteria and assimilates formaldehyde derived from methanol oxidation (11). If formate could be reduced to formaldehyde, a bacterium with the RuMP pathway and a formate reduction pathway would be able to assimilate formate as well. A second option is the homoserine cycle, which although not naturally occurring, can be implemented using preexisting aldolases in *E. coli* (12). Another option is to use the engineered formolase enzyme to convert formaldehyde into dihydroxyacetone or glycoaldehyde (8, 13), which can then be assimilated by natural enzymes. Finally, an engineered enzyme can convert formyl-CoA and formaldehyde into glycolyl-CoA and then glycolate, which can

Submitted: 14 February 2021; Received (in revised form): 8 July 2021; Accepted: 4 August 2021

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

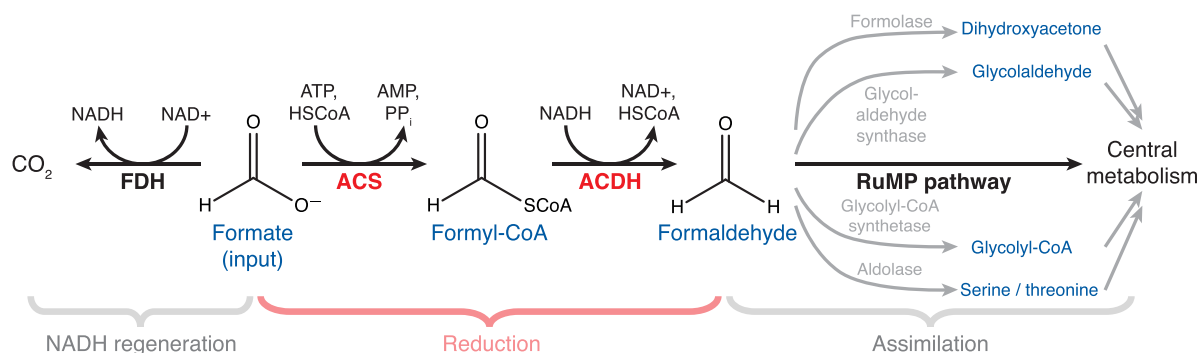


Figure 1. Schematic of formate reduction pathway and associated reactions. The proposed pathway reduces formate to formaldehyde via the enzymes ACS and ACDH, highlighted in red. A portion of the formate is oxidized by FDH to generate the NADH needed for formyl-CoA reduction. To assimilate the formaldehyde into central metabolism and thereby support growth, the pathway is integrated into an organism which natively contains the RuMP pathway (as well as FDH). Formaldehyde could also in principle be assimilated via other pathways, such as those starting with formolase, glycolaldehyde synthase, glycyl-CoA synthase or a serine/threonine aldolase.

be assimilated naturally (14–16). The common advance needed to enable all of these pathways is the reduction of formate to formaldehyde. Therefore, we sought to improve the two enzymes known to catalyze formate reduction.

2. Materials and methods

2.1 Bioinformatics and enzyme homolog selection

All bioinformatics and analysis/visualization of experimental data were performed in Python/Jupyter. Phylogenetic trees for figures were computed by FastTree (17) and visualized using iTOL (18).

To identify acyl-CoA synthetase (ACS) homologs for testing, an initial candidate list of 8911 sequences was compiled that included 6104 sequences from the ‘Acetate-CoA ligase’ Interpro family (IPR011904) (19) with the same three domains as EcACS downclustered to 90% identity using CD-HIT (20); 2790 sequences from RefProt based on a pHMMER search (E -value $< 10^{-200}$) with query EcACS (P27550) (21); and 17 experimentally characterized ACS homologs from BRENDA (EC 6.2.1.1) (22). From the initial candidate list, an alignment and distance matrix was generated using Clustal Omega (23), and a hierarchical clustering of the distance matrix was used to guide manual selection of a diverse final set of ACSs. Initially 11 ACSs were chosen for testing (see below for details). Then, given the results of the first round of testing, 30 additional ACSs were chosen to further sample clades with high activity while also exploring new areas of sequence space.

For acylating aldehyde dehydrogenases (ACDHs), 7646 sequences were obtained from UniProt via a pHMMER search (E -value $< 7 \times 10^{-17}$) with query LmACDH (Q8Y7U1) or enzyme commission number search (EC 1.2.1.10), or from BRENDA (EC 1.2.1.10). After alignment and clustering of ACDHs, mhpF-like sequences were removed, leaving 4037 adhE-like sequences from which the final selection was made. Initially, nine ACDHs were chosen for testing and then these results were used to choose 37 additional ACDHs to test.

The full phylogeny of all ACSs and ACDHs considered for homolog discovery would be too large to visualize as a tree, so Figure S1 shows only untested sequences that have less than 50% (ACSs) or 40% (ACDHs) amino acid identity to each other and to the tested homologs. Some of the tested homologs are more similar to each other than this because they were chosen for reasons other than diversity; these were all included in the trees.

2.2 DNA synthesis and *E. coli* strain construction

ACS and ACDH sequences were codon optimized for *E. coli* expression using Integrated DNA Technologies’ online tool (<https://www.idtdna.com/CodonOpt>; accessed September 2018). DNA synthesis was performed at Twist or the Joint Genome Institute of the U.S. Department of Energy. Genes were cloned into vector pET29b+ between NdeI and XhoI such that expressed enzymes have a C-terminal 6xHis tag. Expression vectors with ACDH genes were electroporated into *E. coli* expression strain BL21*(DE3), propagated on lysogeny broth (LB) + 50 μ g/ml kanamycin, and stored at -80°C in 25% glycerol.

ACS is known to be repressed under standard physiological conditions by acetylation at K609, but can be derepressed by a point mutation L641P (24). We found that a simpler method of knocking out the *patZ* deacetylase leads to comparable EcACS activity (Figure S2A), so we used a BL21*(DE3) Δ *patZ* host strain for all ACS experiments. The *patZ* gene was deleted from BL21*(DE3) using lambdaRed recombinase (25). For the round of directed evolution from MhACS2 to MhACS3, the alternate expression strain NovaBlue(DE3) Δ *patZ* was constructed and used.

2.3 Protein expression, lysis and SDS-PAGE

To express proteins, strains were inoculated directly from -80°C stocks into auto-induction medium (26) at 1:500–1:50 000 dilution and incubated at 37°C for 24 h. For screening, 500- μ l cultures were grown in 2 ml 96-well microtiter plates (Axygen P-DW-20-C) with shaking at 1000 rpm on a benchtop shaker (Heidolph Titramax 1000) in a temperature-controlled room. For sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Nash assays, 5-ml cultures were grown in round-bottom glass tubes in a rotary shaker incubator at 250 rpm. For purification, 50 ml or 500 ml cultures were grown in Erlenmeyer flasks and shaken at 250 rpm.

To prepare lysates for screening, ninety-six-well plate cultures were pelleted at 2200g for 10 min, washed once in 4°C water, and resuspended by vortexing after adding 300 μ l/well of lysis buffer (50 mM HEPES pH 7.5, 50 mM NaCl, and 2 mM MgCl_2) with 0.6 mg/ml lysozyme, 0.1 mg/ml polymyxin B and 1:50 000 Sigma benzonase nuclease. Plates were incubated at 37°C for 50 min without shaking followed by 10-min shaking at 1000 rpm. Then, lysates were pelleted at 2200g for 10 min, and the supernatant was used for downstream assays.

To prepare lysates for SDS-PAGE, Nash assay or protein purification, 5 ml, 50 ml or 500 ml cultures were pelleted at 4000 rpm

for 10 min and resuspended, respectively, in 0.5 ml, 8 ml or 30 ml lysis buffer with 0.1 mM dithiothreitol (DTT), 1:500 Sigma protease inhibitor cocktail and 1:50 000 Sigma benzonase nuclease. Cell suspensions were sonicated on ice (Branson SLPt) for 3 repeats of 10 s on and 10 s off at 30% amplitude for 5-ml cultures, or 6 repeats of 30 s on and off at 70% amplitude for 50-ml cultures, or 12 repeats of 30 s on and off at 70% amplitude for 500-ml cultures. Lysates were pelleted at 4000 rpm for 15 min and the supernatant was used for analysis or purification.

To analyze lysates and purified enzymes by SDS-PAGE, samples were mixed 1:1 with 2× Laemmli sample buffer with 2-mercaptoethanol (Bio-Rad) and boiled for 10 min. A sample containing 1–20 μg of protein was loaded into a 4–15% precast gel (Bio-Rad Mini-ProTEAN) and run at 60 V for 20 min followed by 160 V for 1 h. To compare expression levels across lysates, protein concentration was determined by bicinchoninic acid (BCA) and equal micrograms of protein were loaded in each lane. Gels were stained by Coomassie blue and imaged using a digital camera. Minor contrast adjustments were made to the images to improve visibility of bands.

2.4 Assays for enzyme activity in lysates

ACS was assayed in lysates using a discontinuous assay with 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB), which reacts with CoA to yield absorbance at 412 nm (27, 28). In a microtiter plate (Corning Costar 3370), 100 μl of reaction buffer (10 μl of expression-induced *E. coli* lysate, 50 mM HEPES pH 7.5, 2 mM MgCl₂, 5 mM ATP, 0.5 mM CoA and 50 mM sodium formate) was aliquoted. The formate was added last to start the reaction; everything was incubated for 10 min at 37°C and then stopped by adding 100 μl DTNB reagent (50 mM HEPES pH 7.5 and 2 mM DTNB). Absorbance at 412 nm was read on a plate reader (Molecular Devices Spectramax 190). Empty vector control lysates were used to establish the background signal, and the metric $\Delta A_{412} = A_{412}^{\text{Emptyvector}} - A_{412}$ was used to quantify lysate ACS activity. Note that higher ACS activity corresponds to lower A_{412} but higher ΔA_{412} . For the first round of ACS directed evolution, a continuous assay was used (see ‘Protein purification and enzyme kinetics’), but subsequent rounds of evolution used the discontinuous assay described above.

ACDH lysates were assayed in a continuous assay by coupling to ACS; all steps were performed at 37°C. Reactions were performed in a microtiter plate with a total volume 200 μl containing 2 μl of clarified lysate, 2 μM StACSstab1, 50 mM HEPES pH 7.5, 5 mM MgCl₂, 1 mM DTT, 2.5 mM ATP, 0.5 mM CoA, 0.6 mM nicotinamide adenine dinucleotide hydride (NADH) and 50 mM formate. Reactions were prepared in 100 μl at 2× concentration and then 100 μl of 2× formate was added to start the reaction. Absorbance at 340 nm was monitored and NADH concentration was calculated as $[\text{NADH}] = \frac{A_{340}}{\epsilon \cdot l}$, where $\epsilon = 6.22 \text{ mM}^{-1} \text{ cm}^{-1}$ is the extinction coefficient of NADH and $l = 0.56 \text{ cm}$ is the path length of 200 μl of reaction mixture in the microtiter plate. Initial velocities were calculated from least-square linear fits to the first 3–10 datapoints. The amount of ACS to use for coupling was determined by titrating ACS for every new batch of purified ACS or round of lysate screening (Figure S5). For assaying purified ACDHs, we used 30× molar excess of coupling ACS.

2.5 Directed evolution to improve ACS and ACDH

To engineer StACSstab and MhACS, residues were selected for site-saturating mutagenesis based on proximity to the acetate molecule in the crystal structure of the *Salmonella typhimurium* ACS (Protein Data Bank (PDB): 2p2f). Mutant libraries at single positions were constructed using ‘inside-out’ polymerase chain

reaction (PCR) from NNK or ‘22c’ (29) degenerate primers and multisite combinatorial libraries were made by overlap extension PCR. Libraries were electroporated into expression host strains (see above). One 96-well plate of mutant clones (plus control strains) was screened for each single-site library. Eight plates were screened for the 4-site library in MhACS evolution Round 1. The best 10–20 mutants were restreaked on LB + kan plates and four colonies of each mutant were screened again. The best mutant from the secondary screen was used as the parent for the next round of evolution. The DTNB assay with 50 mM formate was used for screening all ACS mutants, except in Round 1 of StACSstab evolution, when the myokinase-coupled assay was used. To engineer LmACDH, its crystal structure (3k9d) was superimposed on the *Rhodospseudomonas palustris* ACDH (5jfn) (30), and the position of the substrate propionyl-CoA from 5jfn was used to choose residues in 3k9d for saturation mutagenesis. Screening was done using the ACS-coupled assay described above on clarified lysates at 50 mM formate.

2.6 Protein purification and enzyme kinetics

All steps were done at 4°C. 1 ml of Ni-NTA superflow resin (Qiagen) was placed in a gravity-flow column (GE Healthcare PD-10) and equilibrated by flowing through 10 ml of lysis buffer. Then 8 ml of clarified lysate was added and the column was sealed and placed on ice and rocked to mix (VWR 12620-916) for 10 min. Then the lysate was passed through the column, 10 ml of wash buffer (50 mM HEPES pH 7.5, 300 mM NaCl and 35 mM imidazole) was added, and protein was eluted in 10 ml of elution buffer (50 mM HEPES pH 7.5, 50 mM NaCl and 150 mM imidazole). Eluate was exchanged to lysis buffer by spinning at 4000 g for 15 min in Amicon Ultra-15 30 kDa (for ACS) or 10 kDa (for ACDH) centrifugal filters. Glycerol was added to 10% and protein concentration was determined by bicinchoninic acid (BCA) assay (Pierce 23227). Kinetic assays were performed immediately after purification. Additional purified enzyme was split into aliquots and stored at –20°C.

For the first round of ACS directed evolution, a secondary screen was performed with high-throughput purification. 1-ml cultures of *E. coli* expression strains were lysed in 300 μl lysis buffer + 1:50 000 benzonase + 0.6 mg/ml lysozyme + 0.1 mg/ml polymyxin B and clarified lysates were passed over 50 μl Ni-NTA superflow resin in each well of a 96-well filter plate (Pall) and by centrifugation at 2200 g for 10 min. The resin was washed with 200 μl of wash buffer and eluted in 100 μl of elution buffer. Eluates were used immediately without buffer exchange and protein was quantified by bicinchoninic acid (BCA).

ACS kinetics was determined by a continuous assay using coupling enzymes (27, 31). All steps were performed at 37°C. All coupling enzymes were from Sigma. A reaction buffer was prepared with 0.05–0.2 μM ACS, 15 U/ml pyruvate kinase, 23 U/ml lactate dehydrogenase and 25 U/ml myokinase in 50 mM HEPES pH 7.5, 5 mM MgCl₂, 1 mM DTT, 0.6 mM NADH, 2.5 mM phosphoenolpyruvate, 2.5 mM ATP and 0.5 mM CoA. 100 μl of a 2× portion of the reaction mixture was aliquoted into a microtiter plate and the reaction was started by adding 100 μl of 2× sodium formate or acetate. Absorbance at 340 nm was monitored for 10 min on the plate reader and initial velocities of NADH oxidation were calculated as above. Kinetic parameters k_{cat} and K_m were extracted from plots of initial velocities versus substrate concentration by fitting $v = k_{\text{cat}} \frac{[S]}{[S] + K_m} + b$, where v is the per-enzyme initial velocity, $[S]$ is the substrate concentration and b is a background rate. Kinetic curves were fit using `scipy.optimize.curve_fit`.

Purified ACDHs were assayed similarly to ACDH lysates as described above, except rates of NADH oxidation were normalized to enzyme concentration as determined by bicinchoninic acid (BCA).

2.7 Cloning and *M. flagellatus* KT strain construction

All genetic constructs for *Methylobacillus flagellatus* KT expression were maintained on an IncP-based broad-host-range plasmid, whose backbone was derived from pAWP87/pCM66 (32). Expression vectors for *M. flagellatus* KT were constructed using PCR-amplified backbone fragments, promoters from the *M. flagellatus* KT genome, and ACS/ACDH coding regions from DNA synthesis. PCR primers were designed with 20–25 bp of overlap and Gibson assembled (New England Biolabs) and electroporated into *E. coli* strain TOP10. Electrotransformants were propagated in LB + 50 µg/ml kanamycin at 20°C to avoid toxicity of the constructs.

A rifamycin-resistant isolate of *M. flagellatus* KT was used for triparental conjugation. Briefly, *M. flagellatus* KT was patched, along with plasmid donor strain and helper strain pRK2073 (33), onto plates containing MM2 medium (14.5 mM K₂HPO₄, 18.8 mM NaH₂PO₄ (monohydrated), 0.8 mM MgSO₄ (heptahydrated), 3.8 mM Na₂SO₄, 9.9 mM KNO₃ and 1× Vishniac trace elements (34)) with 5% LB and 2% methanol and incubated for 16–20 h at 37°C. Cell mixtures were then plated on plates with MM2 + 0.1% pyruvate + 2% methanol + 50 µg/ml rifamycin + 50 µg/ml kanamycin and incubated for 2–3 days at 37°C, and colonies were restreaked onto the same medium and incubated for another 2–3 days at 37°C. For each conjugation, multiple colonies were screened by SDS-PAGE and Nash assay, and the colony with the best phenotype was used for downstream assays. Strains were stored by culturing in liquid MM2 medium and then freezing at –80°C with 10% dimethylsulfoxide.

Promoter reporter constructs were constructed by Gibson assembly of various promoters upstream of dTomato in pAWP87. These were conjugated into *M. flagellatus* KT, colonies were inoculated into seed cultures in MM2 + 2% methanol + kanamycin, grown for 24–48 h at 37°C with shaking, diluted 1:50 and grown 24 h, and then measured at 535 nm excitation/590 nm emission on a plate reader (Tecan Infinite 500). Four promoters (Phps, PmxαF, P_{trc} and P_{tac}) were chosen for driving ACS/ACDH, but intact plasmids with P_{trc} and P_{tac} could not be isolated in the *E. coli* TOP10 cloning strain and were omitted from further experiments. The expression cassettes for ACS and ACDH (i.e. promoter + gene) were oriented divergently to avoid the possibility of transcriptional interference.

2.8 Nash assay

M. flagellatus KT strains were patched from –80°C stocks onto MM2 + 2% methanol + 50 µg/ml kanamycin ('MM2 Me2 kan') plates and incubated at 37°C for 2 days, then inoculated into 5 ml liquid MM2 Me2 kan medium in a round-bottom glass tube and incubated at 37°C with 250 rpm shaking for 24 h. Cultures were lysed as described above and the soluble fraction was used for the assay. In a microtiter plate, 8 µl of lysate was added to a reaction mixture (50 mM HEPES pH 7.5, 5 mM MgCl₂, 1 mM DTT, 4 mM ATP, 6 mM NADH and 1 mM CoA, either 4 µM ACDH, 2 µM ACS or no additional enzyme, and either 0 mM or 300 mM formate) to a total volume of 100 µl. This concentration of formate is high enough that it could potentially have inhibitory effects on the enzymes; however, reaction rates increased monotonically with formate concentration until at least 500 mM (Figure S3C). Reactions were

incubated for 1 h at 37°C. Then 100 µl of Nash reagent (0.1 M ammonium acetate, 0.2% acetic acid and 3.89 M acetylacetone) is added and reactions are incubated at 65°C for 30 min. Then, to precipitate proteins, 20 µl of 100% w/v trichloroacetic acid is added and the reactions are placed on ice for 5 min. The plate is spun down at 2200 g for 10 min and 100 µl of supernatant is transferred to a new plate and A412nm is measured. The difference between absorbance at 300 mM and 0 mM formate, $\Delta A_{412} = A_{412_{300\text{ mM}}} - A_{412_{0\text{ mM}}}$, was used to quantify lysate activity.

2.9 ¹³C formate labeling and analysis by LC-MS

To analyze proteinogenic amino acids, *M. flagellatus* KT strains were revived from –80°C and grown in 5 ml liquid MM2 Me2 kan seed cultures as described above. 200 µl of seed culture was transferred to 5 ml of MM2 + kan + 0.05% methanol + 200 mM ¹²C or ¹³C formate. Initial experiments used cultures with one growth cycle in MM2 + 2% methanol + 200 mM formate. Later experiments used multiple growth cycles as follows. Cultures were grown for 24 h, pelleted at 2200 g for 10 min and resuspended in fresh medium with methanol and formate. This iterative re-feeding was done for 3 days, and the final cell pellets were resuspended in 1 ml 6N hydrochloric acid and boiled in glass vials for 24 h (35). The vials were uncapped and left to dry for another 24 h. The biomass was then resuspended in 1 ml of water, dried for 24 h and then resuspended in 0.5 ml of water and centrifuge-filtered (Costar Spin-X 0.22 µm, Sigma). Samples were stored at –20°C until liquid chromatography-mass spectrometry (LC-MS).

LC-MS was performed as described previously (36). A Waters Xevo mass spectrometry triple quad (Xevo, Waters, Milford, MA) with ultra-performance liquid chromatography system equipped with a Zic-pHILIC column (SeQuant, poly(etherether ketone) (PEEK) 150 mm length × 2.1 mm metal free, with 5 µm polymeric film thickness, EMD Millipore) was used for detection of mass ID (MID) of metabolites with the following LC condition. Mobile Phase A is 20 mM bicarbonate in water (Optima grade, Thermo Fisher Scientific) and Mobile Phase B is 100% acetonitrile (Optima grade, Thermo Fisher Scientific). The LC condition starts with 0.15 ml/min flow rate with initial gradient A = 15% for 0.5 min, then increased to 80% A linearly up to 20 min; at 21 min, A is set to 90% and held for 5 min; at 26.5 min, mobile phase A is switched to 15% and then the column is re-equilibrated for 5.5 min. Multiple reaction monitors were set up for each metabolite of interest. For each metabolite, ¹²C chemical standards were used to set up the mass channel for unlabeled isotopomers. The predicted mass fragments were then used to predict the multiple reaction monitors for labeled isotopomers for each metabolite. The MassLynx software (Waters) was used to integrate ion peak intensities, with subsequent analysis in Python.

3. Results

3.1 Discovery of active natural ACS variants

Previous work showed that *E. coli* acetyl-CoA synthetase (EcACS) and *Listeria monocytogenes* acylating acetaldehyde dehydrogenase (LmACDH) can reduce formate to formaldehyde (8). However, the wild-type enzymes tested had poor activity on the one-carbon substrates and failed to support formate reduction *in vivo* as part of the formolase pathway. To identify potential homologs of ACS with increased formyl-CoA synthetase activity, we collected 8911 ACS sequences from UniProt and chose 41 phylogenetically diverse homologs to test experimentally (Figure S1; 'Materials and Methods'). The chosen sequences include EcACS as well as StACSstab, a computationally stabilized variant of the

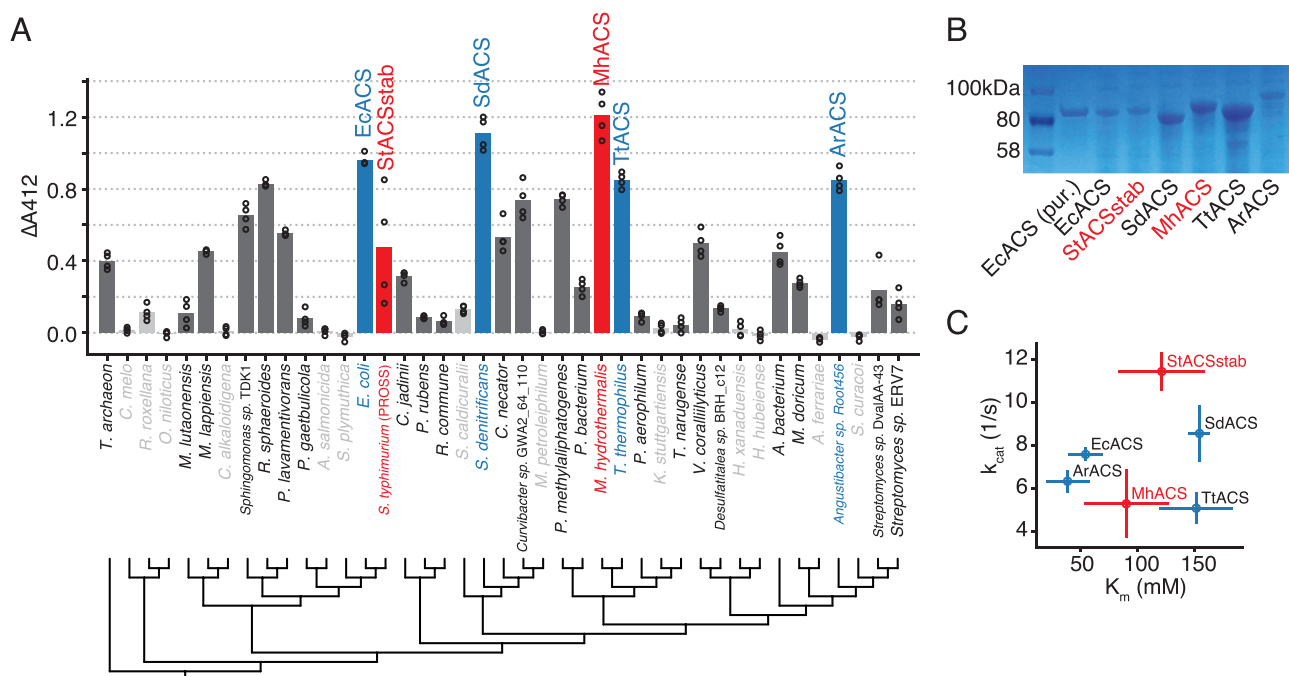


Figure 2. Screening natural ACS homologs identifies enzymes with formate activity. (A) Lysate activity in *E. coli* for 41 ACS homologs versus their phylogeny. Activity is shown as absorbance at 412 nm from the DTNB-based discontinuous assay after subtracting background (See also Figure S1; raw data in Table S1A). Circles show replicates and bars show the mean. Highlighted in color are six homologs chosen for purification and kinetic characterization; in red are two homologs chosen for directed evolution. Enzymes with statistically significant activity compared to empty vector (FDR = 0.05, 2-sample t-test with Benjamini–Hochberg correction) are shown in dark gray or colored bars and black font; non-significant activity is indicated by light gray bars and font. Phylogenetic tree is a maximum-likelihood tree calculated via FastTree2 (‘Materials and Methods’). (B) SDS-PAGE on clarified lysates from six chosen ACS homologs. Each lane contains lysate from equal biomass. ‘EcACS (pur.)’ means purified EcACS. (C) Scatterplot of kinetic parameters k_{cat} versus K_m on formate of six chosen ACS homologs (see Figure S2 for raw kinetics data).

S. typhimurium ACS with 47 mutations (92.7% identity to wild type) (37). StACSstab’s mutations are outside the active site, do not affect catalysis and contribute to a 100-fold higher expression level in *E. coli*, greatly facilitating directed evolution (28). We also included ACSs from *Pyrobaculum aerophilum* and *Kuenenia stuttgartiensis*, which are reported to have relatively high specific activities on formate relative to those on acetate, of 27% and 65%, respectively (38, 39).

We obtained the set of ACS homologs via DNA synthesis, expressed them in *E. coli*, and screened their activity in clarified *E. coli* lysates using a plate-based endpoint assay with the DTNB reagent (‘Materials and Methods’). We performed the screens in 50 mM formate, close to the k_m of EcACS from pilot experiments, to reveal variation in k_{cat}/k_m across homologs. Initially we tested 11 homologs (Figure S1); using these results to highlight clades containing active variants, we then chose 30 more homologs to test. From the full set of 41 homologs, 30 had significantly higher activity than the empty vector control at a 5% false discovery rate (FDR; Figure 2A and Table S1; t-test with Benjamini–Hochberg correction) and two had higher activity than EcACS. StACSstab had lower activity than EcACS in this lysate assay, but since StACSstab is well-characterized (28), we chose to include it along with the top five ACS homologs for further analysis.

To determine whether high lysate activity of top homologs was due to increased soluble expression, we analyzed clarified lysates by SDS-PAGE. This showed that three of the enzymes had much higher soluble expression than the others (SdACS, MhACS and TtACS in Figure 2B). We then used a myokinase-coupled continuous assay to determine the kinetic parameters of purified enzymes (Figure 2C and S2; ‘Materials and Methods’).

We assayed these homologs using formate as well as acetate, the likely native substrate, to determine whether any of these ACSs are already naturally biased toward one-carbon substrates. Despite its relatively low lysate activity, StACSstab had the highest k_{cat} of the six enzymes, on both formate and acetate ($11.4 \pm 0.9 \text{ s}^{-1}$ and $50.9 \pm 3.1 \text{ s}^{-1}$, respectively; Figure S4A). On the other hand, EcACS and ArACS had the lowest k_m values ($54.4 \pm 14.9 \text{ mM}$ and $38.5 \pm 19.3 \text{ mM}$; Figure 2C) and highest catalytic efficiencies (k_{cat}/k_m) on formate ($148 \pm 48 \text{ M}^{-1} \text{ s}^{-1}$ and $204 \pm 119 \text{ M}^{-1} \text{ s}^{-1}$; Figure S4A). In general, k_m values on formate were about three orders of magnitude higher than those on acetate. As a result, all enzymes had much lower catalytic efficiencies on formate (k_{cat}/k_m between 50 and $200 \text{ M}^{-1} \text{ s}^{-1}$) than on acetate (between 2×10^5 and $5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) (Figure 3A), although there is some variation in this specificity ratio (Figure S4B). The measured k_{cat} values were also generally lower for formate than for acetate, although usually by less than one order of magnitude. In one case, SdACS actually has higher formate k_{cat} ($8.6 \pm 1.3 \text{ s}^{-1}$) than acetate k_{cat} ($7.0 \pm 0.4 \text{ s}^{-1}$).

3.2 Directed evolution of ACS

Given that even the most active of these ACS homologs are still 2–3 orders of magnitude less efficient on formate than on acetate, we performed directed evolution to increase the formate activity of ACS. No single homolog simultaneously had the highest activity, expression, and specificity, so we chose two parent enzymes: StACSstab because it had the highest k_{cat} , and MhACS (from *Marinithermus hydrothermalis*) because it had the highest lysate activity. Both are also likely tolerant to mutation, since StACSstab is computationally stabilized and MhACS comes from a thermophile (40).

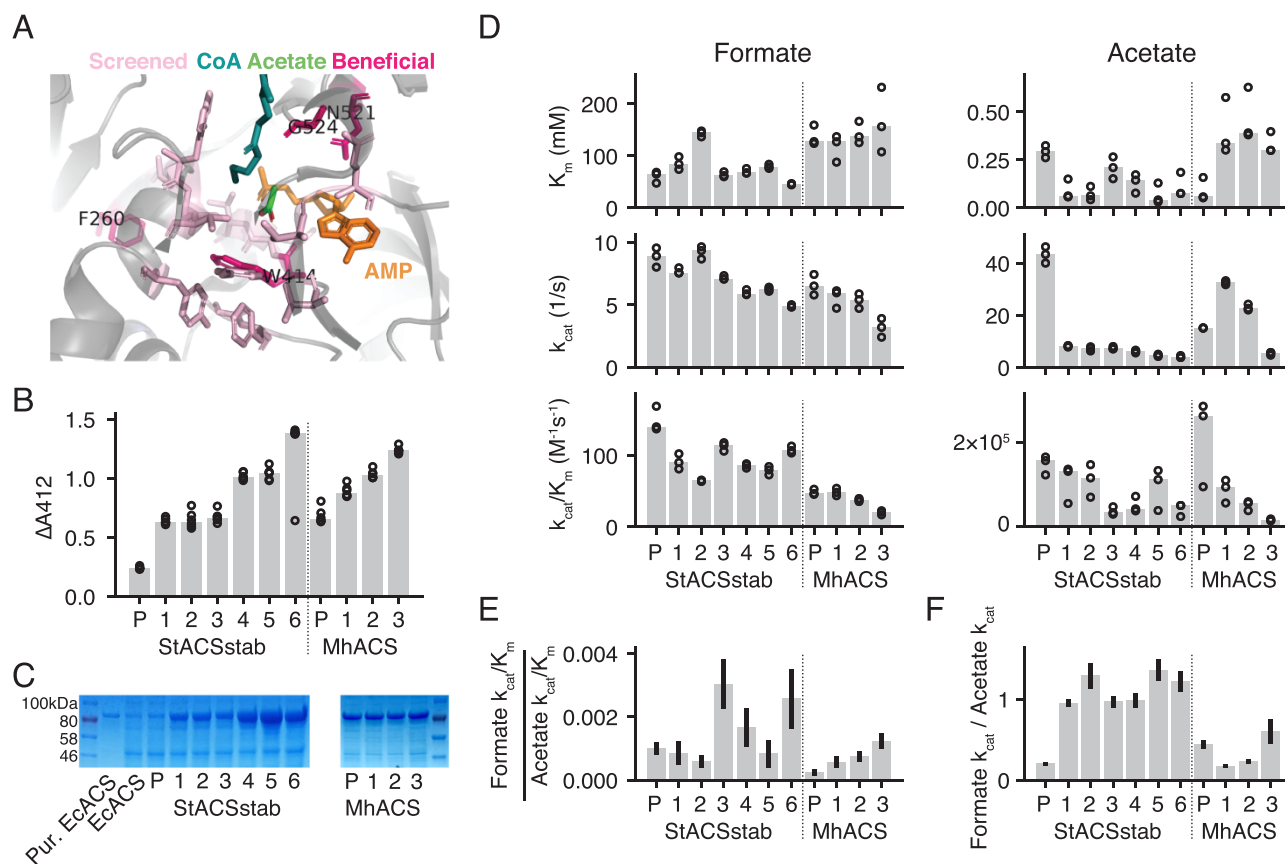


Figure 3. Directed evolution of ACS improves expression and specificity, but not specific activity. (A) Structure of *S. typhimurium* ACS (PDB: 2p2f) with residues chosen for site-saturating mutagenesis highlighted in light pink. Residues that yielded beneficial mutations that were kept in the evolved variants are highlighted in dark pink. (B) Background-subtracted lysate activity in *E. coli* using the discontinuous assay ('Materials and Methods') on evolved variants of StACSstab and MhACS. All enzyme variants shown were isolated in host strain BL21*(DE3), except MhACS3, which was isolated in NovaBlue(DE3) (Figure S6). Circles show three replicates and bars show the median. 'P' indicates parental or wild-type enzyme, and numbered variants correspond to mutants listed in Table 1. (C) SDS-PAGE of clarified lysates of expression cultures of each evolved variant. Each lane contains lysates from equal biomass. (D) k_m , k_{cat} , and k_{cat}/k_m for formate and acetate of evolved variants. Circles show three replicates and bars show the median. These parameters are also listed in Table S2. (E) Ratios of formate k_{cat}/k_m to acetate k_{cat}/k_m . (F) Ratios of formate k_{cat} to acetate k_{cat} . Error bars represent standard deviation of the ratios estimated using the replicate data.

We took a semi-rational approach to engineer StACSstab using a published crystal structure of the wild-type StACS (PDB: 2p2f). StACSstab has 46 mutations relative to StACS (93% amino acid identity), but these mutations were designed to avoid perturbing the structure of the active site (37). Therefore, we used the StACS structure and previous mutagenesis studies to choose a set of 18 residues lining the active-site pocket near the acyl moiety for mutagenesis (Figure 3A and Table 1) ('Materials and Methods') (28, 31, 41). We screened lysates of single-site-saturating libraries of these 18 positions in StACS using a continuous assay in 50 mM formate, followed by a secondary screen with plate-based purification (Figure S5; 'Materials and Methods'). We identified two mutations, N521V (StACSstab1) and N521L (StACSstab2), that increased lysate activity by almost 3-fold (Figure 3B). Then, using StACSstab1 as a parent, we mutated positions that were beneficial in Round 1 as well as new positions that were structurally proximal to N521, screening using the discontinuous assay. We isolated N521V W414F, N521V F260W and N521V G524A (StACSstab3, StACSstab4 and StACSstab5, respectively) as variants with further improved lysate activity (Figure 3B). Previous work found that V310 and V387, which line the acyl-binding pocket in ACS, play a strong role in controlling substrate specificity (28, 41). Therefore, in a third round of evolution, we combinatorially mutated

these two positions to each of three larger hydrophobic amino acids. However, this failed to generate any improved variants (Table 1). In a final round, we combined the mutations discovered in Round 2 and identified N521V W414F G524A (StACSstab6) as the most-improved candidates (Figure 3B). From StACSstab to StACSstab6, lysate activity increased by 5.8-fold (ratio of median of four replicates in Figure 3B).

To determine whether increases in lysate activity translated to increases in specific activity, we purified the ACS variants and measured their kinetic parameters. The initial variants StACSstab1 and StACSstab2 have similar k_{cat} on formate to the parent enzyme (7–10 s^{-1}), but have 5.3-fold and 6.0-fold lower k_{cat} on acetate (Figure 3D), respectively. This is accompanied by higher soluble expression (Figure 3C), suggesting that high levels of native (acetate) activity may be toxic and prevent high expression of parental StACSstab. Subsequent variants StACSstab2–StACSstab6 continued to increase in soluble expression as well as formate specificity. The final variant StACSstab6 had a ratio of formate to acetate k_{cat}/k_m 2.6-fold higher than that of StACSstab. The formate to acetate k_{cat} ratio increased even more, by 5.9-fold, between these enzymes.

Although directed evolution increased expression and formate specificity, it did not increase formate activity. In fact, catalytic

Table 1. Directed evolution campaigns on StACSstab, MhACS and LmACDH

Campaign parent	Round	Round parent(s)	Mutations made	Highest lysate activity
StACSstab	1	StACSstab	F260X, W309X, V310X, T311X, Y315X, Y355X, V386X, G387X, E388X, T412X, W414X, Q415X, T416X, G420X, F421X, Y496X, N521X, G524X	StACSstab1 (N521V), StACSstab2 (N521L)
	2	StACSstab1	F260X, L262X, V310X, V386X, W414X, F421X, R515X, L520X, V522X, S523X, G524X, G556X, Q557X	StACSstab3 (N521V W414F), StACSstab4 (N521V F260W), StACSstab5 (N521V G524A)
	3	StACSstab, StACSstab1-5	V310VILM + V386VILM	None
	4	StACSstab2-5	F260X, L262X, W414X, G524X	StACSstab6 (N521V W414F G524A)
MhACS	1	MhACS	F262X + W414X + N521NKIVLSDQ	MhACS1 (F262Y W416F N524S)
	2	MhACS1	W414X, F423X, Y499X	MhACS2 (F262Y W416F Y499V N524S)
	3	MhACS2	L523X, N524X, G527X, E560X	MhACS3 (F262Y W416F Y499V N524R)
LmACDH	1	LmACDH	I250X, C251X, A252X, L418X, L420X	LmACDH1 (A252S)
	2	LmACDH1	I250X, C251X, S253X, T374, A404X, G407X	LmACDH2 (A252S S253C)

All mutant libraries were from single-site-saturating mutagenesis (e.g. 'F260X, W309X' denotes two libraries, of which one contains F260 mutated to all 20 substitutions and the other W309 mutated to all 20 substitutions), unless specific subsets of substitutions are indicated. Plus sign indicates a combinatorial library at multiple positions. Screening data for a subset of evolution rounds is shown in supplemental figures.

activity decreased modestly over the course of evolution. The final mutant, StACSstab6, has a formate k_{cat} of $4.9 \pm 0.1 \text{ s}^{-1}$ and k_{cat}/k_m of $108 \pm 5 \text{ M}^{-1} \text{ s}^{-1}$, 45% and 28% lower, respectively, than StACSstab's k_{cat} of $8.8 \pm 0.8 \text{ s}^{-1}$ and k_{cat}/k_m of $149 \pm 18 \text{ M}^{-1} \text{ s}^{-1}$ (Figure 3D and Table S2). Normalizing the relative change in lysate activity to that of k_{cat}/k_m , we estimate the change in functional expression to be 8-fold, consistent with the qualitative increase in band intensity on the SDS-PAGE gel.

To engineer MhACS, we mutated a small set of positions corresponding to those in StACSstab that yielded beneficial mutations (Table 1). We first screened a combinatorial library with mutations at positions F262, W416, N524 and G527 (corresponding to StACSstab F260, W414, N521 and G524), which yielded a mutant F262Y W416F N524S (MhACS1) with improved lysate activity. Using MhACS1 as a parent, we then screened single-site-saturating libraries and discovered an improved variant with additional mutation Y499V (MhACS2). At this point, screening additional site-saturating libraries MhACS2 failed to yield improved mutants. We hypothesized that an *E. coli* host strain with lower basal expression may reduce any potential toxicity of the heterologous enzyme and increase our chances of isolating additional improved mutants. Therefore, we switched from the BL21*(DE3) host strain to NovaBlue(DE3), which has the stronger LacI^q repressor (Figure S6). Starting with MhACS2 and screening site-saturating libraries, we obtained a mutant with S524R (MhACS3; N524R relative to MhACS) with 1.9-fold higher lysate activity than the MhACS parent (Figure 3B).

As with StACSstab, increased lysate activity of MhACS mutants did not translate to increased specific activity. In fact, the k_{cat} and k_{cat}/k_m on formate decreased over the course of evolution (Figure 3D). However, the k_{cat}/k_m for acetate decreased even more, so that the formate specificity of the enzyme, as seen by the ratio of formate to acetate k_{cat}/k_m , increased by 5.5-fold from MhACS to MhACS3. Interestingly, the decrease in acetate k_{cat}/k_m in MhACS3 was due to a combination of increased k_m and decreased k_{cat}

contributed by different mutations. By contrast, during StACSstab directed evolution the major change was a large decrease in k_{cat} . The k_m on formate did not change appreciably, staying around 150 mM in all variants (Figure 3D and Table S2). Normalizing the increase in lysate activity from MhACS to MhACS3 by the 59% decrease in formate k_{cat}/k_m , we find that functional expression of MhACS3 is 4.6-fold higher than that of the original parent enzyme.

3.3 Discovery and improvement of ACDH

Previous work used *L. monocytogenes* (LmACDH) for formyl-CoA reduction because it was the most active of five homologs tested (8). We sought to identify additional active variants by a two-pronged strategy of homolog screening and directed evolution. Although formyl-CoA is the desired substrate of ACDH, it is not commercially available and has a very short half-life (14). Therefore, we screened ACDHs using formate as a substrate instead, including ACS as a coupling enzyme to generate formyl-CoA in the reaction (Figure S7A). This does not allow quantitative estimation of the k_m of ACDH for formyl-CoA but is sufficient to determine the relative activities of ACDH variants.

We analyzed all available ACDH homologs in UniProt and BRENDA and chose 46 for gene synthesis and testing. Alignment and clustering of ACDHs revealed two divergent clades with roughly equal numbers of sequences (Figure 4). One clade contained members such as *E. coli* MhpF, which natively operates as a complex with an aldolase (42). EcMhpF was previously shown to have low activity compared to LmACDH (8), suggesting difficulties in expressing the monomer form. Therefore, we avoided members of the MhpF-like clade and focused instead on the clade containing *E. coli* AdhE, LmACDH and bacterial-microcompartment-associated enzymes such as EutE (43).

We screened clarified lysates of the ACDH homologs for the ability to oxidize NADH in the presence of ACS, ATP, CoA and 50 mM formate. We initially screened nine ACDH homologs and then used those results to choose 37 more (Figure S1). From

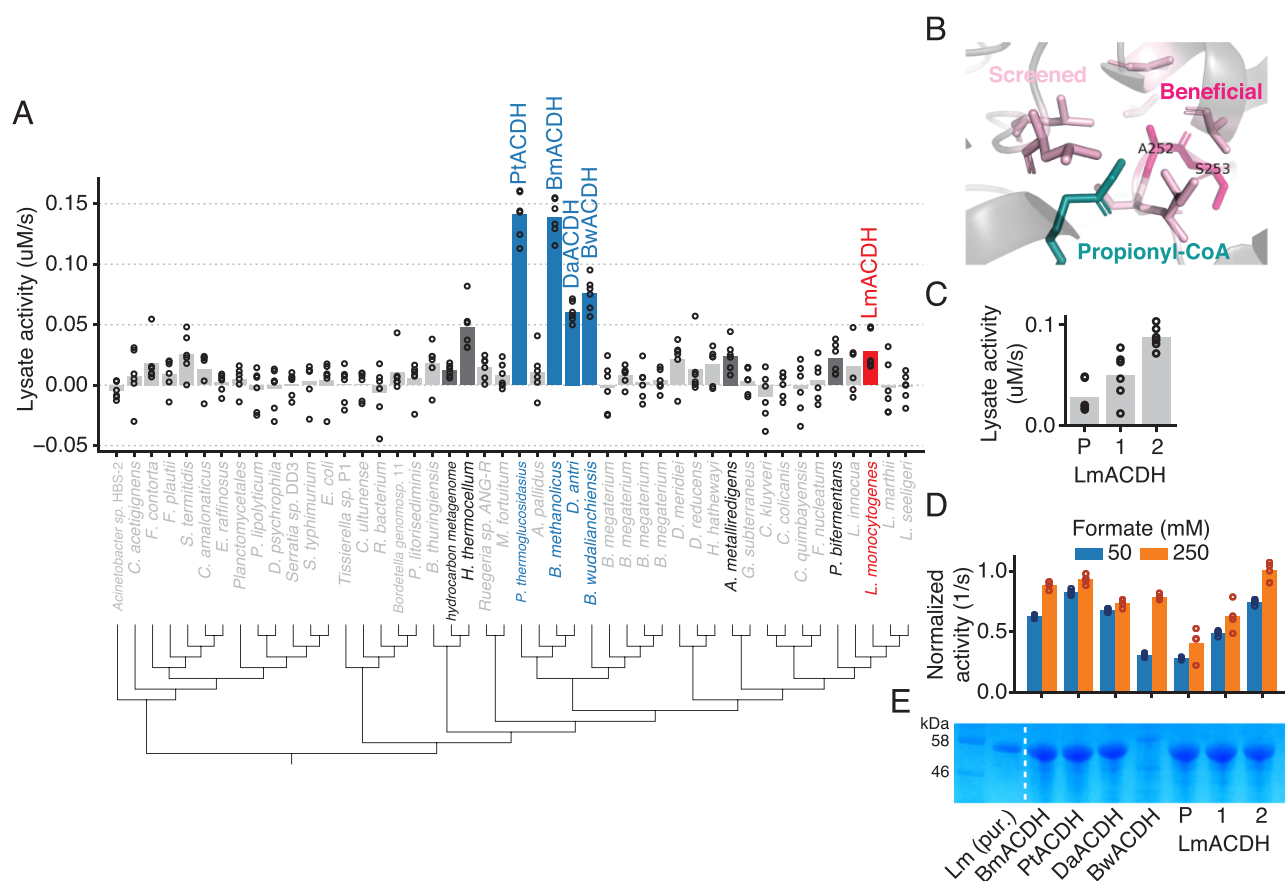


Figure 4. Homolog screening and directed evolution of ACDH enzymes. (A) Lysate activity of ACDH homologs. Dots show technical replicates from two independent expression strain transformants; bars show mean of all replicates. Enzymes with statistically significant activity compared to empty vector (FDR = 0.05, 2-sample t-test with Benjamini–Hochberg correction) are shown in dark gray and black font; non-significant activity is indicated by light gray bars and font. Blue indicates enzymes chosen for follow-up characterization and red indicates enzyme used for directed evolution. Raw data are in Table S1B. (B) Structure of *L. monocytogenes* ACDH (3k9d, propionyl-CoA from 5jfn) with residues chosen for site-saturating mutagenesis shown in light pink. Residues that yielded beneficial mutations that were kept in the evolved variants are highlighted in dark pink. (C) Lysate activity of evolved LmACDH variants. ‘P’ indicates parental or wild-type enzyme, and numbered variants correspond to mutants listed in Table 1. Circles represent three technical replicates and bars show the mean. (D) Activity of selected homologs and evolved LmACDH variants, normalized to enzyme concentration. Shares x-axis labels with Panel E. (E) SDS-PAGE of clarified lysates of *E. coli* strains expressing ACDH variants. Each lane contains lysate from equal biomass. Lane 2 contains purified LmACDH.

the full set of 46 homologs, we found nine with activity significantly higher than empty vector at a 5% FDR (Figure 4A and S8A; t-test with Benjamini–Hochberg correction). Five homologs had higher activity than LmACDH (Figure 4A, colored bars). We then purified the four homologs with highest lysate activity as well as LmACDH and assayed them in 50 mM and 250 mM formate with excess ACS. Unlike in our ACS screen, all the ACDH homologs with higher lysate activity than LmACDH also had higher activity after normalizing by enzyme concentration, with PtACDH having the highest activity in both assays (normalized activity of $0.93 \pm 0.04 \text{ s}^{-1}$ in 250 mM formate; Figure 4D and S8B). All homologs were more active in 250 mM formate than in 50 mM formate. Their relative rankings were unchanged by formate concentration, except for BwACDH, which is the least active in 50 mM formate but among the most active homologs at 250 mM. This suggests that it has a higher k_m for formyl-CoA than other homologs.

To engineer LmACDH, we used its crystal structure 3k9d along with a propionyl-CoA substrate superimposed from a related structure 5jfn to choose positions close to the acyl moiety for mutagenesis. We screened site-saturating libraries at five positions and identified a mutant A252S (LmACDH1) with increased

activity (Table 1). We then screened some of the same positions on top of the LmACDH1 background, as well as additional residues close to A252 in the structure, and found A252S S253C (LmACDH2) to have even higher activity ($1.00 \pm 0.07 \text{ s}^{-1}$ at 250 mM formate, or 2.5-fold higher than LmACDH; Figure 4D). In fact, LmACDH2 has slightly higher activity than PtACDH, the best homolog we discovered.

3.4 Expression of pathway in *M. flagellatus* KT

Having identified ACSs with improved expression and ACDHs with increased activity, we next asked whether these enzymes could support formate reduction *in vivo*. Methylophilic bacteria are able to assimilate formaldehyde as an intermediate of methanol, and those that do this via the RuMP pathway cannot natively assimilate formate (unlike serine-pathway methylotrophs, which can assimilate both formate and formaldehyde). Therefore, if we introduced ACS and ACDH activities into an RuMP methylotroph (which also had an NADH-producing formate dehydrogenase (FDH)), this would in principle confer partial or complete formatotrophy. In practice, our enzymes are likely too inefficient to support growth, but even a low flux from formate into

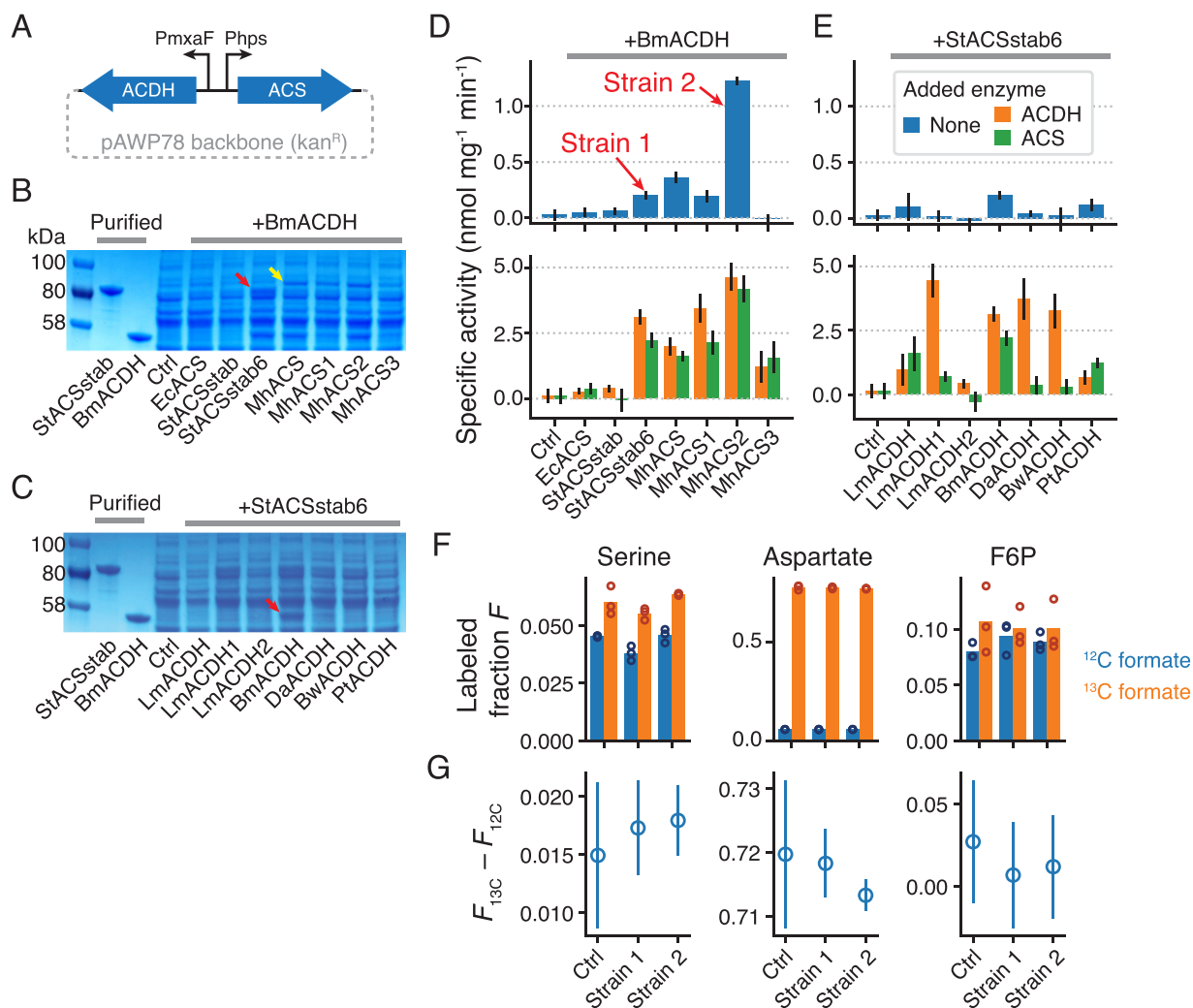


Figure 5. Nash assay and ¹³C labeling to determine pathway activity in cell extracts and *in vivo*. (A) Schematic of vector used to express ACS and ACDH variants in *M. flagellatus* KT. PmxαF and Phps are native promoters (see also Figure S9). (B) SDS-PAGE of clarified lysates from *M. flagellatus* KT strains expressing different ACSs, as well as BmACDH, from the expression vector. Control strain ('Ctrl') contains the same vector, but with a pJ23101-dTomato insert instead of ACS/ACDH enzymes. Red arrow: gel band for StACSstab6; yellow arrow: gel band for MhACS. (C) Same as (B), but for different ACDHs coexpressed with StACSstab6. Red arrow: gel band for BmACDH. (D) Specific activity of *M. flagellatus* KT lysates for formaldehyde production from 300 mM formate, for various ACSs coexpressed with BmACDH (mean and standard deviation from three technical replicates; see 'Materials and Methods', Figure S10). Results for cell lysates only (blue bars) or with 4 μM purified ACDH (orange bars) or ACS (green bars) added. (E) Same as (D), but for various ACDHs coexpressed with StACSstab6. 'Strain 1' and 'Strain 2' were chosen for ¹³C labeling. (F) Fraction of ¹³C-labeled proteinogenic serine, aspartate or F6P from cells grown in methanol + 200mM ¹²C or ¹³C formate. Three biological replicates of the control strain and two different pathway variants ('Strain 1' and 'Strain 2' from (D)) were assayed. Bars show the mean. (G) The difference in ¹³C-labeled fraction between ¹³C-formate-grown cells and ¹²C-formate-grown cells (mean and standard deviation, n = 3). If ¹³C formate were being assimilated by the pathway, then the pathway-containing strains should have a higher value of this difference than the control strain.

biomass could potentially be used to select for further enzyme improvements.

We chose the betaproteobacterium *M. flagellatus* KT to express the pathway because it assimilates methanol via the RuMP pathway, grows robustly under standard laboratory conditions and is amenable to genetic manipulation (44). We first tested a panel of promoters for their ability to drive high constitutive expression of a red fluorescent protein reporter from an IncP-based broad-host-range plasmid in *M. flagellatus* KT (32) (Figure S9). Based on this, we chose to use the native promoters Phps and PmxαF to drive ACS and ACDH expression, respectively, from the plasmid. We cloned a panel of expression vectors containing different ACSs coexpressed with the same ACDH, or vice versa, and conjugated them into *M. flagellatus* KT (Figure 5A; 'Materials and Methods').

To test for expression and activity of our enzymes in *M. flagellatus* KT, we analyzed clarified lysates by SDS-PAGE and Nash assay (45), which measures formaldehyde production (Figure S10; 'Materials and methods'). Formaldehyde is only produced from formate if both ACS and ACDH are active, so we assayed lysates with and without an added excess of purified ACS or ACDH to detect activity of each enzyme individually (Figure 5D and E). This also has a side benefit of boosting the sensitivity of the assay. Independent *M. flagellatus* KT transconjugants varied in phenotype (growth rates and enzyme activities), so for each vector we screened multiple transconjugants and picked the one with the highest enzyme activity for further characterization.

We found that StACSstab6 and all MhACS variants, but not EcACS or StACSstab, had a visible band in SDS-PAGE when

expressed in *M. flagellatus* KT. MhACS has a higher-molecular-weight band than StACSstab6, reflecting their predicted molecular weights (74 and 72 kDa, respectively). Consistent with the SDS-PAGE, the Nash assay only showed ACS activity in StACSstab6 and MhACS variants (orange bars in Figure 5D, lower panel). Across ACDHs, only BmACDH had visible SDS-PAGE expression. It also had the highest ACDH activity by Nash assay, although LmACDH, LmACDH1 and PtACDH also had above-background activities (green bars in Figure 5E, bottom panel). The strains with the highest lysate activity without any added enzymes are those containing BmACDH and either StACSstab6, MhACS, MhACS1 or MhACS2. Notably, neither EcACS nor the LmACDH variants, which were used in the previous version of the pathway (8), were well-expressed in *M. flagellatus* KT. We chose the strains containing StACSstab6/BmACDH ('Strain 1'), which had the third-highest activity ($0.20 \pm 0.04 \text{ nmol mg}^{-1} \text{ min}^{-1}$; Figure 5D), and MhACS2/BmACDH ('Strain 2'), which had the highest activity ($1.22 \pm 0.04 \text{ nmol mg}^{-1} \text{ min}^{-1}$), for further characterization.

Some strains with the same ACDH or ACS differed in their apparent activities for that enzyme in the Nash assay. This could be due to unintended genetic variation between transconjugants, or some interaction between the divergent Phps and PmxαF promoters. The latter might explain why, for example, across the MhACS variants in Figure 5D, ACDH activity seems to vary and correlate with ACS activity even though the ACDH enzyme is the same. We verified that the sequence of the promoters and enzyme genes on the expression vector are as expected in every strain. Therefore, any cryptic genetic variation would have to be in the genome of the host strain.

3.5 Test for formate assimilation in *M. flagellatus* KT

Given the observed activity of ACS and ACDH in *M. flagellatus* KT lysates, we next tested for assimilation of formate into biomass. To do this, we cultured Strain 1 and Strain 2 in ^{12}C - or ^{13}C -formate and monitored ^{13}C labeling of metabolites via LC-MS. Because *M. flagellatus* KT contains formate dehydrogenases capable of oxidizing formate to CO_2 , which can potentially be reassimilated via carboxylation, we also assayed a control strain containing a dTomato-expressing vector. If formate is being assimilated into biomass via our pathway, we should observe more ^{13}C labeling in central carbon metabolites and proteins in a pathway-containing strain than in a control strain, and only when labeled formate is provided.

First, we analyzed proteinogenic amino acids as an indicator of overall incorporation of formate into biomass. We inoculated control and pathway strains into MM2 medium with 0.2% ^{12}C methanol and 200 mM ^{12}C or ^{13}C -formate, harvested the saturated cultures, and acid-hydrolyzed the biomass for LC-MS. We found almost no labeling of more than one carbon atom across the amino acids examined, so we used total ^{13}C -labeled fraction, or 1—unlabeled fraction, as a simple metric for the degree of labeling (Figure 5F and S11).

Serine, whose carbon atoms are derived from pyruvate and thus immediately downstream of the RuMP pathway, displayed a background ^{13}C -labeled fraction of about 4% in ^{12}C -formate across all strains, but a 1.5–2% increase in labeling in ^{13}C -formate (Figure 5F). However, this increase occurred in both control and pathway strains and was similar in magnitude (Figure 5G), indicating that the extra labeling is not due to formate assimilation via our pathway. Aspartate had much higher labeling in ^{13}C -formate than in ^{12}C -formate, although the differential labeling was again the same in all strains (Figure 5F and G, middle panel). Since

aspartate is derived from oxaloacetate, the high labeled fraction in ^{13}C -formate is possibly due to re-assimilation of $^{13}\text{CO}_2$ by pyruvate carboxylase after formate oxidation by formate dehydrogenases (46). A similar jump in labeling in ^{13}C -formate was observed for threonine and glutamate, which can both be derived from aspartate (Figure S11B and C). Alanine, on the other hand, had <5% labeling in ^{13}C -formate like serine, consistent with also being derived from pyruvate (Figure S11B and C). Overall, no amino acid examined had labeling indicative of formate assimilation by the introduced pathway.

The analysis above requires a sufficiently high formate reduction flux to result in labeled proteins. For a more sensitive test of formate assimilation, we monitored fructose-6-phosphate (F6P), a metabolite immediately downstream of formaldehyde assimilation into the RuMP pathway. We added 200 mM ^{12}C - or ^{13}C -formate to mid-exponential-phase cultures of control and pathway strains, continued incubating the cultures for 2 h, and then harvested and extracted metabolites for LC-MS. We saw 8–10% labeling of F6P in ^{12}C -formate, close to the expected background rate of 6% (Figure 5F, right panel). Labeling in ^{13}C -formate was higher, at around 10% in all three strains. However, as in the case of the amino acids, there was no increase in the difference in labeling between labeled and unlabeled formate conditions (Figure 5G, right panel). Therefore, we were unable to detect evidence of our pathway assimilating formate into biomass *in vivo*.

It is possible that our ACS and ACDHs still do not have the activity needed to supply even detectable formaldehyde flux through the RuMP pathway. To test this, we used flux balance analysis to calculate the theoretical growth rate that could be supported by the measured rate of formate reduction in the Nash assay. A genome-scale model of *M. flagellatus* KT metabolism does not exist, but we used a model developed for another RuMP-pathway methylotroph, *Methylotuvimicrobium buryatense* 5GB1C (47). We assumed that the flux through the methanol dehydrogenase reaction, which provides all the formaldehyde (and reduced carbon) for biomass production, is the same as the highest specific activity we measured in *M. flagellatus* KT lysates, or 1.2 nmol/min/mg (Figure 5D). We found that this would support a theoretical growth rate of 0.00048 h^{-1} , or a doubling time of 8.6 weeks, even without ATP maintenance (with ATP maintenance, growth was infeasible). This is much slower than even the 55-h growth supported by an unoptimized reductive glycine pathway (9), indicating that further improvements to activity and/or expression are needed.

4. Discussion

4.1 Utility of phylogenetically diverse enzymes

Previous work showed that EcACS and LmACDH have formate reduction activity and that the enzymes are functional when expressed in *E. coli*. We extend that work to identify a panel of natural and engineered ACS and ACDH variants with improved expression and lysate activity and show that StACSstab6, MhACS2 and BmACDH are well-expressed and active in the methylotroph *M. flagellatus* KT. None of these three enzymes had the highest ACS or ACDH activities *in vitro*, showing that expression in the host cytosolic environment is an equally if not more important factor than catalytic properties in practice. The computational design of StACSstab and the thermophilic source organisms of MhACS and BmACDH may have played a role in their greater expression and host range. By contrast, neither EcACS nor LmACDH, the previous best enzymes for this pathway, was expressed in *M. flagellatus* KT, despite EcACS having the second-highest formate k_{cat}/K_m of

the wild-type ACSs and LmACDH2 being the most active ACDH we found. This highlights a key advantage of screening phylogenetic diversity in that this approach offers not only the chance to discover high activity, but also high expression and evolvability (28, 48, 49).

4.2 Improving expression versus activity

Since we performed directed evolution on ACSs using a lysate-based screen, it is reassuring that we obtained increased lysate activity (5.8-fold for StACSstab and 1.9-fold for MhACS). However, this was entirely due to increases in functional expression (8-fold for StACSstab and 4.6-fold for MhACS) and not catalytic efficiency. In fact, k_{cat}/k_m decreased by 28% for StACSstab6 on formate, although it decreased by 72% on acetate, leading to an overall increase in the formate specificity from the parent enzyme. Despite this lack of improvement in catalytic activity, the increased soluble expression proved crucial to functionality in *M. flagellatus* KT, where StACSstab6, but not the StACSstab parent, was expressed and active. Interestingly, even wild-type ACS homologs differed widely in soluble expression in *E. coli* as well as in *M. flagellatus*. By contrast, there were no obvious differences in expression between the various ACDH homologs or evolved variants in *E. coli*, and our directed evolution of ACDH using a lysate-based assay led to increases in both lysate and specific activity.

Why did our lysate assays select for increased specific activity in ACDH but not in ACS? The ACS parents we chose perhaps started with poor stability or expression, but this is unlikely given their origins. Moreover, stability usually 'decreases' while evolving for activity (50). A more likely possibility is that ACS activity is toxic. This has been observed previously (51), and thus decreasing it may allow cells to tolerate increased expression. This is consistent with the expression gain concomitant with a sudden reduction of acetate activity from StACSstab to StACSstab1/2 (N521V/L). Despite having a much lower acetate activity, however, even StACSstab6 still appears to be toxic, frequently leading to *E. coli* colonies with spontaneously decreased activity (one such colony can be seen as a replicate in Figure 3B). This problem can be mitigated in future rounds of evolution by using a low-background expression host and/or reducing induced expression level.

4.3 Challenges of one-carbon substrates

A more fundamental problem is the possibility of biophysical limits on the formate activity of ACS. We chose ACS for synthesizing formyl-CoA because formate is structurally similar to ACS's native substrate acetate. However, formate is less electrophilic than acetate, which could make it challenging to achieve a high k_{cat} . Indeed, formate k_{cat} values among our natural and evolved ACSs never exceeded 12 s^{-1} , while the highest acetate k_{cat} was $43.2 \pm 3.1 \text{ s}^{-1}$. One-carbon compounds also have relatively few functional groups for interacting with a substrate-binding pocket, leading to higher k_m values (52) and potentially explaining why even our lowest ACS k_m for formate is greater than 40 mM. As a result, our highest formate k_{cat}/k_m values are between 100 and $200 \text{ M}^{-1} \text{ s}^{-1}$, 2–3 orders of magnitude lower than many natural enzymes, including the acetate activity for native ACSs. Encouragingly, however, natural enzymes find formate equally challenging. The *Methylobacterium extorquens* formate-tetrahydrofolate ligase, which activates formate for assimilation, has a k_m of 22 mM and a k_{cat} of $\sim 100 \text{ s}^{-1}$, for a k_{cat}/k_m of $\sim 5000 \text{ M}^{-1} \text{ s}^{-1}$, about 30-fold higher than our best ACSs (53). Despite being two orders of magnitude lower than the median enzyme (52), this activity can support fully formatotrophic growth in natural and engineered

pathways. Therefore, a physiologically relevant activity of the formate reduction module may be within reach given further enzyme engineering.

ACDH is not expected to be as challenging an engineering target as ACS, because most of the substrate binding affinity is contributed by the CoA group. We did not directly measure the k_m of ACDH for formyl-CoA, but the ACS-coupled assays show it is at most 7 mM (Figure S7). In reality it is probably much lower; the k_m of ACDH for acetyl-CoA can be $<100 \mu\text{M}$ (54) and the k_m of 2-hydroxyacyl-CoA lyase for formyl-CoA is $200 \mu\text{M}$ (14), although this is not its native substrate. However, a potential problem with ACDH is k_{cat} . Even though the fastest ACDH homolog in the literature has a k_{cat} of $\sim 60 \text{ s}^{-1}$ on acetate (55), our best ACSs were almost two orders of magnitude slower on formate. However, given that we were able to increase this value by ~ 2 -fold in two rounds of directed evolution, further engineering will likely result in additional gains.

4.4 Effects of mutations on ACS and ACDH

Previous work found that ACS substrate specificity can be changed from acetate to larger or more polar substrates by mutating V310, T311, V386 or W414 in StACS, which are all with 4 Å of the acetyl moiety (28, 31, 41). However, for formate, we did not find increases in lysate activity when mutating V310, T311 or V386 in isolation or V310 and V386 combinatorially (Table 1). Instead, we found that the previously unexplored N521V/L mutations cause a large decrease in acetate activity in StACSstab. In the 2p2f structure of StACS, acetate is modeled in the active site with its acetyl carbon pointed toward N521, 5.7 Å away from the asparagine's side chain carbonyl (31). This orientation could explain why large hydrophobic substitutions at N521 favor a smaller acyl substrate.

Several other mutations increased ACS lysate activity while decreasing acetate k_{cat}/k_m . StACSstab F260W (StACSstab4) and W414F (StACSstab3) both increased acetate k_m . In the StACS structure, W414 is within 4 Å of the acyl substrate, while F260 is 10.9 Å away but in contact with the first-shell V310 side chain (Figure 3A). MhACS Y499V decreased acetate k_{cat} ; it is 11.1 Å from the acyl substrate but contacts the backbone of W414. StACSstab G524 lines the CoA binding site in StACSstab and, as a result, G524S/L is known to block CoA addition to the acyl group (31). Our results show that G524A, on top of N521V W414F, maintains formate activity while increasing acetate k_m . Additional residues (e.g. F421) showed evidence of improved activity in our initial screens on StACSstab (Figure S5), but were not pursued fully. They are prime candidates for mutagenesis in future rounds of evolution.

Over two rounds of site-saturating mutagenesis and screening at nine residues comprising the acyl-binding tunnel in LmACDH, we found a double-mutant LmACDH2 (A252S S253C) that contributed to improved activity on formyl-CoA. These positions partially overlap with those mutagenized in a recent effort to engineer an ACDH to reduce glycolyl-CoA to glycolate (28). Our evolved isolates were comparable in activity to the best natural homologs BmACDH and PtACDH, but ultimately only BmACDH expressed well in *M. flagellatus* KT. BmACDH has the same sequence as LmACDH at positions homologous to A252 and S253, so these are clear candidates for mutagenesis in future directed evolution efforts.

4.5 *M. flagellatus* KT as an *in vivo* pathway testing platform

We chose to use *M. flagellatus* KT for testing the *in vivo* activity of our pathway because it could potentially gain formatotrophy

with only the expression of ACS and ACDH. This is the first published instance, to our knowledge, of metabolic engineering in this organism. We did not observe *in vivo* assimilation of formate in *M. flagellatus* KT, which is most likely due to the low activity of the pathway. However, other factors may also have reduced the pathway flux, such as the presence of FDHs in *M. flagellatus* KT, which natively oxidize formate to CO₂ and generate NADH and thus could divert formate from our pathway (56). Knocking out FDH activity causes a large growth defect (56), but partial knockdown may make more formate available to our heterologous pathway. Additionally, *M. flagellatus* KT has a formaldehyde oxidation pathway that could divert flux from formaldehyde assimilation (57). However, this is a detoxification pathway and only carries appreciable flux at high formaldehyde concentrations (58), which our formate reduction module does not reach.

Recently, an *E. coli* strain was engineered to grow on methanol as a sole carbon source via the RuMP pathway (59). This provides the alternative option of engineering the ACS/ACDH pathway in *E. coli* instead, which would allow access to a wider range of genetic tools and pathway manipulations (60). Most importantly, it would allow the improvements we obtain from directed evolution in *E. coli* lysates to be directly translated into *in vivo* activity or expression. However, even in the fully methylotrophic *E. coli*, formaldehyde toxicity is still a major problem, and perhaps as a result, its doubling time on methanol is more than 8 h. By contrast, *M. flagellatus* KT has a doubling time of 2 h on methanol, indicating its naturally evolved robustness against formaldehyde toxicity. Perhaps the best approach in future work is to use a combination of strains—*E. coli* for initial troubleshooting and improvement of enzymes, followed by *M. flagellatus* KT for fine-tuning for maximal flux.

5. Conclusion

Through phylogenetic homolog screening and directed evolution, we identified a panel of highly expressed and active ACS and ACDH enzymes and gained insight into the genetic determinants of their acyl substrate specificity. We established a plasmid-based expression system in the RuMP-pathway methylotroph *M. flagellatus* KT and used it to introduce a formate reduction pathway in an attempt to confer synthetic formatotrophy. Although we ultimately did not observe *in vivo* formate assimilation via our pathway, the enzymes and insights from this work should enable continued improvement of this pathway toward the ultimate goal of efficient conversion of CO₂ into value-added chemicals.

Supplementary data

Supplementary data are available at SYNBO Online.

Data availability

All homolog protein sequences, plasmid sequences and processed data on lysate and purified enzyme activities are contained in the supplementary tables. Raw data on enzyme activities and Python code used for analysis are available from authors upon request.

Funding

Washington Research Foundation Postdoctoral Fellowship [to J.W.]; University of Washington [to M.E.L.].

Acknowledgment

Gene synthesis was provided by the Synthetic Biology program at the Joint Genome Institute of the Department of Energy. We thank Yakov Kipnis, Devin Trudeau, Joseph Groom, Julius Palme and Dan Tawfik for valuable technical advice and feedback on the manuscript.

Conflict of interest statement. None declared.

References

- Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B. and Keasling, J.D. (2012) Microbial engineering for the production of advanced bio-fuels. *Nature*, **488**, 320–328.
- Yishai, O., Lindner, S.N., Gonzalez de la Cruz, J., Tenenboim, H. and Bar-Even, A. (2016) The formate bio-economy. *Curr. Opin. Chem. Biol.*, **35**, 1–9.
- Li, H., Opgenorth, P.H., Wernick, D.G., Rogers, S., Wu, T.-Y., Higashide, W., Malati, P., Huo, Y.-X., Cho, K.M. and Liao, J.C. (2012) Integrated electromicrobial conversion of {CO₂} to higher alcohols. *Science*, **335**, 1596.
- Liu, C., Colón, B.C., Ziesack, M., Silver, P.A. and Nocera, D.G. (2016) Water splitting–biosynthetic system with CO₂ reduction efficiencies exceeding photosynthesis. *Science*, **352**, 1210–1213.
- Jouny, M., Luc, W. and Jiao, F. (2018) General techno-economic analysis of CO₂ electrolysis systems. *Ind. Eng. Chem. Res.*, **57**, 2165–2177.
- Bar-Even, A., Noor, E., Flamholz, A. and Milo, R. (2013) Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochim. Biophys. Acta*, **1827**, 1039–1047.
- Bar-Even, A. (2016) Formate assimilation: the metabolic architecture of natural and synthetic pathways. *Biochemistry*, **55**, 3851–3863.
- Siegel, J.B., Smith, A.L., Poust, S., Wargacki, A.J., Bar-Even, A., Louw, C., Shen, B.W., Eiben, C.B., Tran, H.M., Noor, E. et al. (2015) Computational protein design enables a novel one-carbon assimilation pathway. *Proc. Natl. Acad. Sci. USA*, **112**, 3704–3709.
- Kim, S., Lindner, S.N., Aslan, S., Yishai, O., Wenk, S., Schann, K. and Bar-Even, A. (2020) Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.*, **16**, 538–545.
- Bang, J., Hwang, C.H., Ahn, J.H., Lee, J.A. and Lee, S.Y. (2020) *Escherichia coli* is engineered to grow on CO₂ and formic acid. *Nat. Microbiol.*, Nature Publishing Group, **5**, 1–5.
- Kato, N., Yurimoto, H. and Thauer, R.K. (2006) The physiological role of the ribulose monophosphate pathway in bacteria and archaea. *Biosci. Biotechnol. Biochem.*, **70**, 10–21.
- He, H., Höper, R., Dodenhöft, M., Marlière, P. and Bar-Even, A. (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*. *Metab. Eng.*, **60**, 1–13.
- Poust, S., Piety, J., Bar-Even, A., Louw, C., Baker, D., Keasling, J.D. and Siegel, J.B. (2015) Mechanistic analysis of an engineered enzyme that catalyzes the formose reaction. *ChemBioChem*, **16**, 1950–1954.
- Chou, A., Clomburg, J.M., Qian, S. and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion. *Nat. Chem. Biol.*, **15**, 900–906.
- Burgener, S., Cortina, N.S. and Erb, T.J. (2020) Oxalyl-CoA decarboxylase enables nucleophilic one-carbon extension of aldehydes to chiral α -hydroxy acids. *Angew. Chem. Int. Ed.*, **59**, 5526–5530.
- Nattermann, M., Burgener, S., Pfister, P., Chou, A., Schulz, L., Lee, S.H., Paczia, N., Zarzycki, J., Gonzalez, R. and Erb, T.J. (2021)

- Engineering a highly efficient carbonylase for synthetic one-carbon metabolism. *ACS Catal.*, **11**, 5396–5404.
17. Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
 18. Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (ITOL) v4: recent updates and new developments. *Nucleic Acids Res.*, **47**, W256–W259.
 19. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.-Y., El-Gebali, S., Fraser, M.I. et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.*, **47**, D351–D360.
 20. Li, W. and Godzik, A. (2006) Cd-Hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
 21. Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–W37.
 22. Jeske, L., Placzek, S., Schomburg, I., Chang, A. and Schomburg, D. (2019) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.*, **47**, D542–D549.
 23. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.*, John Wiley & Sons, Ltd, **7**, 1–6.
 24. Starai, V.J., Gardner, J.G. and Escalante-Semerena, J.C. (2005) Residue Leu-641 of acetyl-CoA synthetase is critical for the acetylation of residue Lys-609 by the protein acetyltransferase enzyme of *Salmonella enterica*. *J. Biol. Chem.*, **280**, 26200–26205.
 25. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA*, **97**, 6640–6645.
 26. Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
 27. Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D.L., Prilusky, J. et al. (2018) Automated design of efficient and functionally diverse enzyme repertoires. *Mol. Cell*, **72**, 178–186.e5.
 28. Trudeau, D.L., Edlich-Muth, C., Zarzycki, J., Scheffen, M., Goldsmith, M., Khersonsky, O., Avizemer, Z., Fleishman, S.J., Cotton, C.A.R., Erb, T.J. et al. (2018) Design and in vitro realization of carbon-conserving photorespiration. *Proc. Natl. Acad. Sci. USA*, **115**, E11455–E11464.
 29. Kille, S., Acevedo-Rocha, C.G., Parra, L.P., Zhang, Z.-G., Opperman, D.J., Reetz, M.T. and Acevedo, J.P. (2013) Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.*, **2**, 83–92.
 30. Zarzycki, J., Sutter, M., Cortina, N.S., Erb, T.J. and Kerfeld, C.A. (2017) In vitro characterization and concerted function of three core enzymes of a glycyl radical enzyme - associated bacterial micro-compartment. *Sci. Rep.*, **7**, 1–12.
 31. Reger, A.S., Carney, J.M. and Gulick, A.M. (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase. *Biochemistry*, **46**, 6536–6546.
 32. Puri, A.W., Owen, S., Chu, F., Chavkin, T., Beck, D.A.C., Kalyuzhnaya, M.G. and Lidstrom, M.E. (2015) Genetic tools for the industrially promising methanotroph *Methylobacillus buryatense*. *Appl. Environ. Microbiol.*, **81**, 1775–1781.
 33. Marx, C.J. and Lidstrom, M.E. (2001) Development of improved versatile broad-host-range vectors for use in methylotrophs and other Gram-negative bacteria. *Microbiology*, **147**, 2065–2075.
 34. Vishniac, W. and Santer, M. (1957) The Thiobacilli, 12. *Bacteriol. Rev.*, **21**, 195–213.
 35. Yishai, O., Goldbach, L., Tenenboim, H., Lindner, S.N. and Bar-Even, A. (2017) Engineered assimilation of exogenous and endogenous formate in *Escherichia coli*. *ACS Synth. Biol.*, **6**, 1722–1731.
 36. Fu, Y., Li, Y. and Lidstrom, M. (2017) The oxidative TCA cycle operates during methanotrophic growth of the Type I methanotroph *Methylobacillus buryatense* 5GB1. *Metab. Eng.*, **42**, 43–51.
 37. Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J. et al. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell*, **63**, 337–346.
 38. Bräsen, C., Urbanke, C. and Schönheit, P. (2005) A novel octameric AMP-forming acetyl-CoA synthetase from the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *FEBS Lett.*, **579**, 477–482.
 39. Russ, L., Harhangi, H.R., Schellekens, J., Verdellen, B., Kartal, B., Op Den Camp, H.J.M. and Jetten, M.S.M. (2012) Genome analysis and heterologous expression of acetate-activating enzymes in the anammox bacterium *Kuenenia stuttgartiensis*. *Arch. Microbiol.*, **194**, 943–948.
 40. Copeland, A., Gu, W., Yasawong, M., Lapidus, A., Lucas, S., Deshpande, S., Pagani, I., Tapia, R., Cheng, J.-F., Goodwin, L.A. et al. (2012) Complete genome sequence of the aerobic, heterotroph *Marinithermus hydrothermalis* type strain (T1T) from a deep-sea hydrothermal vent chimney. *Stand. Genomic Sci.*, **6**, 21–30.
 41. Ingram-Smith, C., Woods, B.I. and Smith, K.S. (2006) Characterization of the acyl substrate binding pocket of acetyl-CoA synthetase. *Biochemistry*, **45**, 11482–11490.
 42. Fischer, B., Boutserin, S., Mazon, H., Collin, S., Branlant, G., Gruez, A. and Talfournier, F. (2013) Catalytic properties of a bacterial acylating acetaldehyde dehydrogenase: evidence for several active oligomeric states and coenzyme A activation upon binding. *Chem. Biol. Interact.*, **202**, 70–77.
 43. Zhu, H., Gonzalez, R. and Bobik, T.A. (2011) Coproduction of acetaldehyde and hydrogen during glucose fermentation by *Escherichia coli*. *Appl. Environ. Microbiol.*, **77**, 6441–6450.
 44. Chistoserdova, L., Lapidus, A., Han, C., Goodwin, L., Saunders, L., Brettin, T., Tapia, R., Gilna, P., Lucas, S., Richardson, P.M. et al. (2007) Genome of *Methylobacillus flagellatus*, molecular basis for obligate methylotrophy, and polyphyletic origin of methylotrophy. *J. Bacteriol.*, **189**, 4020–4027.
 45. Kleeberg, U. and Klinger, W. (1982) Sensitive formaldehyde determination with Nash's reagent and a 'tryptophan reaction'. *J. Pharmacol. Methods*, **8**, 19–31.
 46. He, L., Fu, Y. and Lidstrom, M.E. (2019) Quantifying methane and methanol metabolism of "*Methylobacillus buryatense*" 5GB1C under substrate limitation. Mackelprang, R., Ed., *mSystems*, **4**, 1–14.
 47. Fu, Y., He, L., Reeve, J., Beck, D.A.C. and Lidstrom, M.E. (2019) Core metabolism shifts during growth on methanol versus methane in the methanotroph *Methylobacillus buryatense* 5GB1. *mBio*, American Society for Microbiology, **10**, 1–14.
 48. Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y.M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D.G., Antonovsky, N. et al. (2020) Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J.*, John Wiley & Sons, Ltd, **n/a**, 1–11.

49. Mak, W.S., Tran, S., Marcheschi, R., Bertolani, S., Thompson, J., Baker, D., Liao, J.C. and Siegel, J.B. (2015) Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun.*, **6**, 1–9.
50. Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D.S. (2008) How protein stability and new functions trade off. *PLoS Comput. Biol.*, **4**, 35–37.
51. Chan, C.H., Garrity, J., Crosby, H.A. and Escalante-Semerena, J.C. (2011) In *Salmonella enterica*, the sirtuin-dependent protein acylation/deacylation system (SDPADS) maintains energy homeostasis during growth on low concentrations of acetate. *Mol. Microbiol.*, **80**, 168–183.
52. Bar-even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Taw, D.S. and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, **50**, 4402–4410.
53. Marx, C.J., Laukel, M., Vorholt, J.A. and Lidstrom, M.E. (2003) Purification of the formate-tetrahydrofolate ligase from *Methylobacterium extorquens* AM1 and demonstration of its requirement for methylotrophic growth. *J. Bacteriol.*, **185**, 7169–7175.
54. Lan, E.I., Ro, S.Y. and Liao, J.C. (2013) Oxygen-tolerant coenzyme A-acylating aldehyde dehydrogenase facilitates efficient photosynthetic n-butanol biosynthesis in cyanobacteria. *Energy Environ. Sci.*, **6**, 2672–2681.
55. Zhao, Y., Lei, M., Wu, Y., Wang, C., Zhang, Z., Deng, F. and Wang, H. (2009) Molecular cloning and expression of the complete DNA sequence encoding NAD⁺-dependent acetaldehyde dehydrogenase from *Acinetobacter* Sp. Strain HBS-2. *Ann. Microbiol.*, **59**, 97–104.
56. Hendrickson, E.L., Beck, D.A.C., Wang, T., Lidstrom, M.E., Hackett, M. and Chistoserdova, L. (2010) Expressed genome of *Methylobacillus flagellatus* as defined through comprehensive proteomics and new insights into methylotrophy. *J. Bacteriol.*, **192**, 4859–4867.
57. Chistoserdova, L., Gomelsky, L., Vorholt, J.A., Gomelsky, M., Tsygankov, Y.D. and Lidstrom, M.E. (2000) Analysis of two formaldehyde oxidation pathways in *Methylobacillus flagellatus* KT, a ribulose monophosphate cycle methylotroph. *Microbiology*, **146**, 233–238.
58. Vorholt, J.A., Marx, C.J., Lidstrom, M.E. and Thauer, R.K. (2000) Novel formaldehyde-activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J. Bacteriol.*, **182**, 6645–6650.
59. Chen, F.Y.-H., Jung, H.-W., Tsuei, C.-Y. and Liao, J.C. (2020) Converting *Escherichia coli* to a synthetic methylotroph growing solely on methanol. *Cell*, **182**, 933–946.e14.
60. He, H., Muth, C.E., Lindner, S.N. and Bar-even, A. (2018) Ribulose monophosphate shunt provides nearly all biomass and energy required for growth of *E. Coli*. *ACS Synth. Biol.*, **7**, 1601–1611.