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

Jue Wang[1], Karl Anderson[1], Ellen Yang[1], Lian He[1], Mary E. Lidstrom[1,2,*]

[1]Department of Chemical Engineering, University of Washington, USA

[2]Department of Microbiology, University of Washington, USA

* Address correspondence to Jue Wang: 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 *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.

Keywords: formate assimilation; one-carbon; directed evolution; metabolic engineering; methylotrophy

16 ## Introduction

17 Population growth and climate change have created an urgent need for processes to produce
18 more food, fuel, and chemicals while reducing CO_2 emissions. Engineered microbes have the
19 potential to renewably produce many useful chemicals (1). However, most commercial
20 bioproduction uses expensive sugar feedstocks that compete with the food supply. Carbon
21 dioxide, as a ubiquitous industrial waste and greenhouse gas, is an attractive feedstock, but
22 CO_2-fixing organisms are technically challenging to adapt to industrial scale. These problems
23 can potentially be solved by bio-inorganic hybrid systems, where electricity drives catalytic
24 production of an energy-carrying molecule used by microbes to produce value-added
25 compounds (2,3). Coupled to advanced photovoltaics, these systems can achieve solar-to-
26 biomass conversion efficiencies approaching 10%, well beyond values of 3% for microalgae and
27 1% for plants (4).

28 Formate is an attractive energy carrier for a bio-inorganic system because it can be produced
29 efficiently by electrocatalysis (5), is highly soluble in water, and provides both carbon and
30 reducing power to microbes (2,6). Formate can also be derived from waste biomass and fossil
31 carbon, making it a flexible feedstock for bridging existing and future carbon economies (2).
32 Unfortunately, organisms that naturally consume formate are poorly suited to industrial use, and
33 moreover, natural formate-assimilation pathways are theoretically less efficient in their
34 consumption of ATP and reducing equivalents than rationally designed alternatives (6–8).
35 Recently, the first synthetic formate-assimilation pathway, the reductive glycine pathway
36 (rGlyP), was successfully introduced into *E. coli* to support growth on formate and CO_2 as sole
37 carbon sources (9,10). Although the rGlyP is energy-efficient and has great biotechnological

38 potential, it involves a CO_2-fixation step that requires a high ambient CO_2 concentration in order
39 to operate, potentially limiting its range of applications.

40 Several alternative formate-assimilation pathways have been proposed that could rival the
41 efficiency of the rGlyP while not requiring CO_2 fixation (7). These pathways all have an initial
42 step where formate is reduced to formaldehyde, which could potentially be achieved in two
43 enzymatic reactions via a formyl-CoA intermediate (8). For example, the ribulose
44 monophosphate (RuMP) pathway naturally occurs in methylotrophic bacteria and assimilates
45 formaldehyde derived from methanol oxidation (11). Assuming formate could be reduced to
46 formaldehyde, a bacterium utilizing the RuMP pathway could easily assimilate formate as well.
47 A second option is the rationally designed homoserine cycle, in which formaldehyde is
48 assimilated by aldolases to generate serine or threonine, which are then assimilated by native
49 enzymes (12). Although this pathway is not naturally occurring, its reactions can be catalyzed
50 relatively efficiently by pre-existing *E. coli* enzymes. A few other pathways could assimilate
51 formate in theory but will require substantial enzyme engineering to support biomass production
52 in practice. For example, the formolase enzyme can convert formaldehyde into
53 dihydroxyacetone (8) or glycolaldehyde (13), which can be assimilated by either natural or
54 engineered enzymes (14). An engineered enzyme can convert formyl-CoA and formaldehyde
55 into glycolyl-CoA and then glycolate, which can be assimilated naturally (15,16). The common
56 advance needed to enable all of these pathways is the reduction of formate to formaldehyde.
57 Therefore, we sought to improve the two enzymes known to catalyze formate reduction.

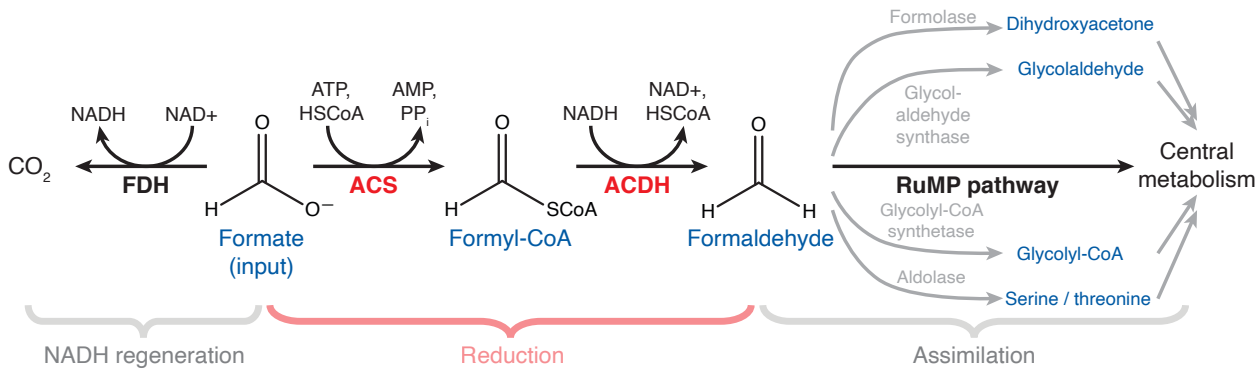


Figure 1. Schematic of formate reduction pathway and associated reactions. The proposed pathway reduces formate to formaldehyde via the enzymes ACS and ACDH, highlighted in red. A portion of the formate is oxidized by formate dehydrogenase (FDH) to generate the NADH needed for formyl-CoA reduction. To 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, glycolyl-CoA synthase, or a serine/threonine aldolase.

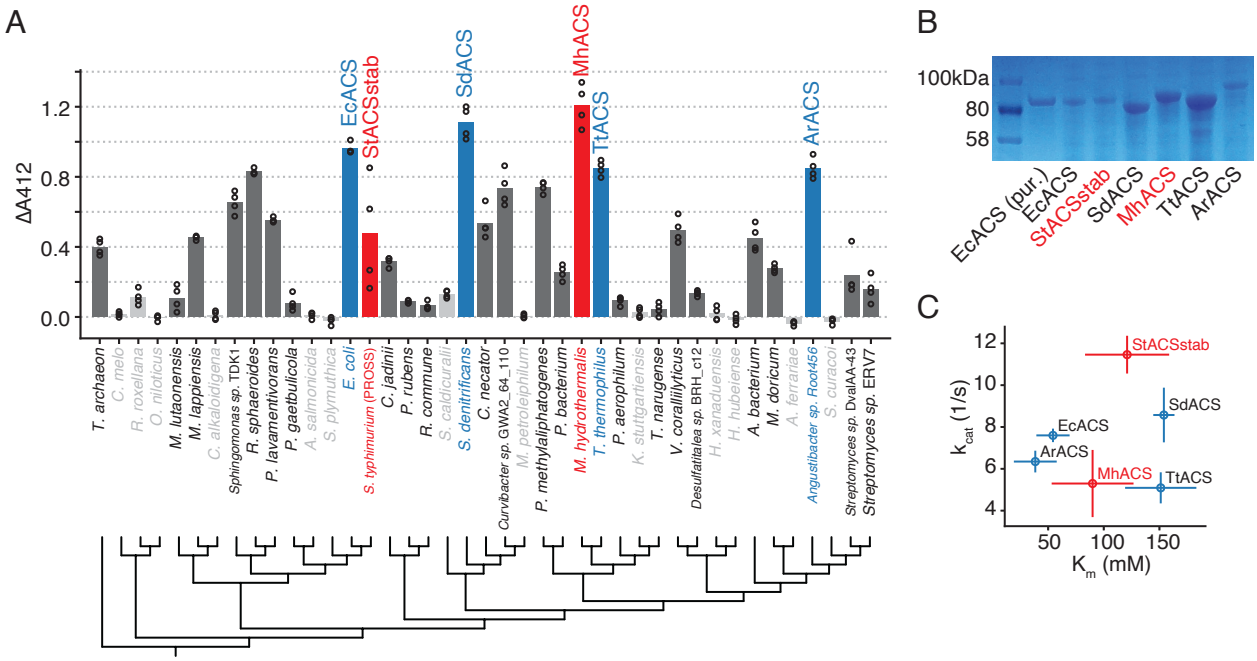
Results

Discovery of active natural ACS variants

Previous work showed that *E. coli* acetyl-CoA synthetase (EcACS) and *L. monocytogenes* acylating acetaldehyde dehydrogenase (LmACDH) can reduce formate to formaldehyde (8). However, the wildtype 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 8,911 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 *S. typhimurium* ACS well-suited to directed evolution (17). We also included ACSs from *P. aerophilum* and *K. stuttgartiensis*, which

are reported to have relatively high formate activities of 27% to 65% of their acetate activities, respectively (18,19).

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 50mM 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 (Figure 2A; 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 (20), we chose to include it along with the top five ACS homologs for further analysis.



91 **Figure 2. Screening natural ACS homologs identifies enzymes with formate activity.**
92 A) Lysate activity in *E. coli* for 41 ACS homologs versus their phylogeny. Activity is shown as absorbance
93 at 412nm from the DTNB-based discontinuous assay after subtracting background (See also Figure S1;
94 raw data in Table S1A). Circles show replicates and bars show the mean. Highlighted in color are 6
95 homologs chosen for purification and kinetic characterization; in red are 2 homologs chosen for directed
96 evolution. Enzymes with statistically significant activity compared to empty vector (FDR=0.05, 2-sample t-
97 test with Benjamini-Hochberg correction) are shown in dark gray or colored bars and black font; non-
98 significant activity is indicated by light gray bars and font. Phylogenetic tree is a maximum-likelihood tree
99 calculated via FastTree2 (Materials and Methods). B) SDS-PAGE on clarified lysates from 6 chosen ACS
100 homologs. Each lane contains lysate from equal biomass. Lane 2 contains purified EcACS. C) Scatterplot
101 of kinetic parameters k_{cat} versus K_m on formate of 6 chosen ACS homologs (see Figure S2 for raw
102 kinetics data).

103 To determine whether high lysate activity of top homologs was due to increased soluble
104 expression, we analyzed clarified lysates by SDS-PAGE. This showed that 3 of the enzymes
105 had much higher soluble expression than the others (SdACS, MhACS, and TtACS in Figure
106 2B). We then used a myokinase-coupled continuous assay to determine the kinetic parameters
107 of purified enzymes (Figure 2C, S2; Materials and Methods). We assayed these homologs using
108 formate as well as acetate, the likely native substrate, to determine whether any of these ACSs
109 are already naturally biased toward one-carbon substrates. Despite its relatively low lysate
110 activity, StACSstab had the highest k_{cat} of the 6 enzymes, on both formate and acetate (11.4 ±
111 0.9 s^{-1} and 50.9 ± 3.1 s^{-1}, respectively; Figure S4A). On the other hand, EcACS and ArACS had
112 the lowest K_ms (54.4 ± 14.9 mM and 38.5 ± 19.3 mM; Figure 2C) and highest catalytic
113 efficiencies (k_{cat}/K_m) on formate (148 ± 48 $M^{-1}s^{-1}$ and 204 ± 119 $M^{-1}s^{-1}$; Figure S4A). In general,
114 K_ms on formate were about 3 orders of magnitude higher than on acetate. As a result, all
115 enzymes had much lower catalytic efficiencies on formate (k_{cat}/K_m between 50 and 200 $M^{-1}s^{-1}$)
116 than on acetate (between 2×10^5 and 5×10^5 $M^{-1}s^{-1}$) (Figure 3A), although there is some variation
117 in this specificity ratio (Figure S4B). The measured k_{cat} values were also generally lower for

118 formate than for acetate, although usually by less than one order of magnitude. In one case,
119 SdACS actually has higher formate k_{cat} (8.6 ± 1.3 s^{-1}) than acetate k_{cat} (7.0 ± 0.4 s^{-1}).

120 **Directed evolution of ACS**

121 Given that even the most active of these ACS homologs is still 2-3 orders of magnitude less
122 efficient on formate than on acetate, we performed directed evolution to increase the formate
123 activity of ACS. No single homolog simultaneously had the highest activity, expression, and
124 specificity, so we chose two parent enzymes: StACSstab because it had the highest k_{cat}, and
125 MhACS (from *M. hydrothermalis*) because it had the highest lysate activity. Both are also likely
126 tolerant to mutation, since StACSstab is computationally stabilized and MhACS comes from a
127 thermophile (21).

128 We took a semi-rational approach to engineer StACSstab using a published crystal structure of
129 the wildtype StACS (PDB: 2p2f). StACSstab has 46 mutations relative to StACS (93% amino-
130 acid identity), but these mutations were designed to avoid perturbing the structure of the active
131 site (17). Therefore, we used the StACS structure and previous mutagenesis studies to choose
132 a set of 18 residues lining the active-site pocket near the acyl moiety for mutagenesis (Figure
133 3A, Table 1) (Materials and Methods) (20,22,23). We screened lysates of single-site-saturating
134 libraries of these 18 positions in StACS using a continuous assay in 50mM formate, followed by
135 a secondary screen with plate-based purification (Figure S5; Materials and Methods). We
136 identified 2 mutations, N521V (StACSstab1) and N521L (StACSstab2) that increased lysate
137 activity by almost 3-fold (Figure 3B). Then, using StACSstab1 as a parent, we mutated positions
138 that were beneficial in round 1 as well as new positions that were structurally proximal to N521,
139 screening using the discontinuous assay. We isolated N521V W414F, N521V F260W, and
140 N521V G524A (StACSstab3, StACSstab4, and StACSstab5, respectively) as variants with

141 further improved lysate activity (Figure 3B). Previous work found that V310 and V387, which line
142 the acyl-binding pocket in ACS, play a strong role in controlling substrate specificity (20,22).
143 Therefore, in a third round of evolution, we combinatorially mutated these 2 positions to each of
144 3 larger hydrophobic amino acids. However, this failed to generate any improved variants (Table
145 1). In a final round, we combined the mutations discovered in round 2 and identified N521V
146 W414F G524A (StACSstab6) as the most-improved candidates (Figure 3B). From StACSstab to
147 StACSstab6, lysate activity increased by 5.8-fold (ratio of median of 4 replicates in Figure 3B).

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 (F262W W416F N524S)
	2	MhACS1	W414X, F423X, Y499X	MhACS2 (F262W W416F Y499V N524S)
	3	MhACS2	L523X, N524X, G527X, E560X	MhACS3 (F262W W416F Y499V N524R)
LmACDH	1	LmACDH	I250X, C251X, A252X, L418X, L420X	LmACDH1 (A252S)
	2	LmACDH1	I250X, C251X, S253X, T374, A404X, G407X	LmACDH2 (A252S S253C)

148

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

All mutant libraries were from single-site-saturating mutagenesis (e.g. “F260X, W309X” denotes 2 libraries, of which one contains F260 mutated to all 20 substitutions, 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 the supplemental figures.

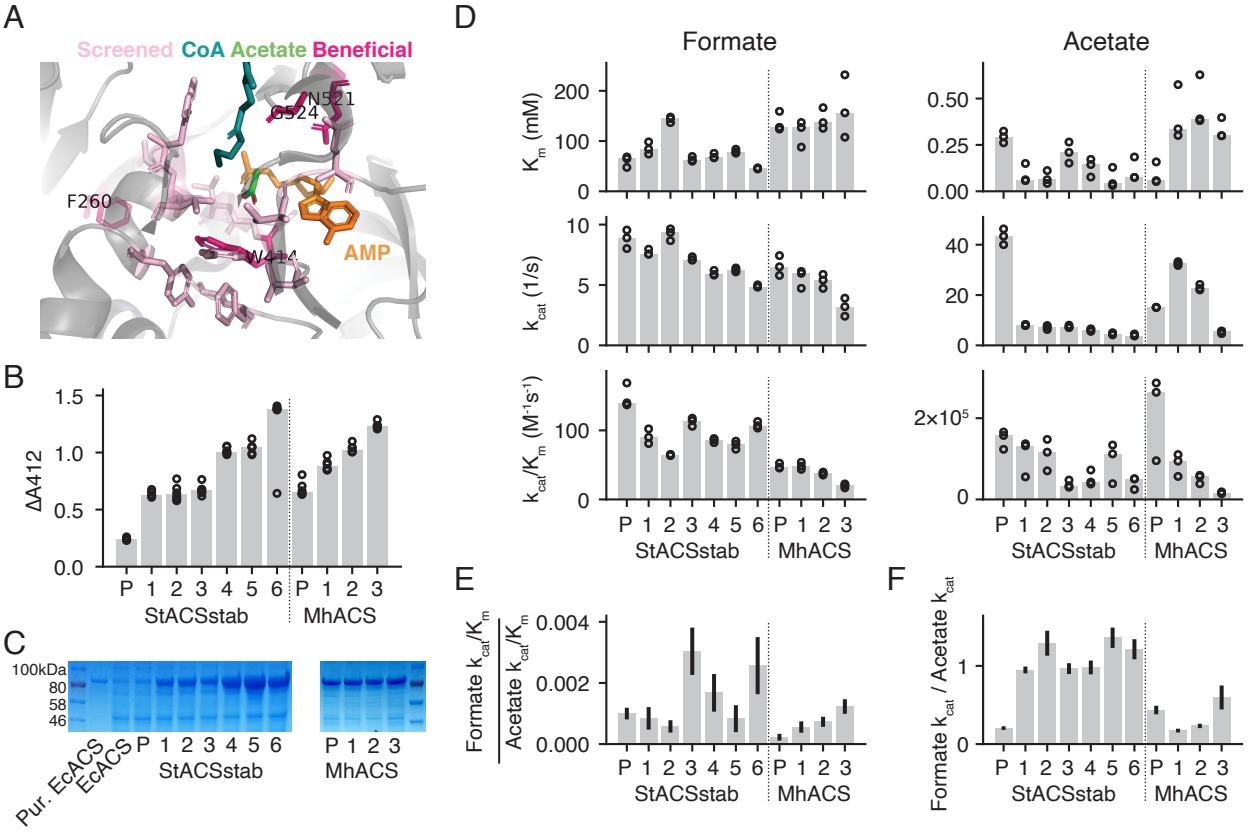


Figure 3. Directed evolution of ACS improves expression and specificity, but not specific activity. A) Structure of *S. typhimurium* ACS (PDB: 2p2f) with residues chosen for site-saturating mutagenesis highlighted in light pink. Residues that yielded beneficial mutations that were kept in the evolved 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 3 replicates and bars show the median. “P” indicates parental or wildtype 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 3 replicates and bars show

166 the median. These parameters are also listed in Table S2. E) Ratios of formate k_{cat}/K_m to acetate k_{cat}/K_m.
167 F) Ratios of formate k_{cat} to acetate k_{cat}. Error bars represent s.d. of the ratios estimated using the replicate
168 data.

169 To determine whether increases in lysate activity translated to increases in specific activity, we
170 purified the ACS variants and measured their kinetic parameters. The initial variants
171 StACSstab1 and StACSstab2 have similar k_{cat} on formate to the parent enzyme (7-10 s^{-1}), but
172 have 5.3-fold and 6.0-fold lower k_{cat} on acetate (Figure 3D), respectively. This is accompanied
173 by higher soluble expression (Figure 3C), suggesting that high levels of native (acetate) activity
174 may be toxic and prevents high expression of parental StACSstab. Subsequent variants
175 StACSstab2 through StACSstab6 continued to increase in soluble expression as well as
176 formate specificity. The final variant StACSstab6 had a ratio of formate to acetate k_{cat}/K_m 2.6-
177 fold higher than that of StACSstab. The formate to acetate k_{cat} ratio increased even more, by
178 5.9-fold, between these enzymes.

179 Although directed evolution increased expression and formate specificity, it did not increase
180 formate activity. In fact, catalytic activity decreased modestly over the course of evolution. The
181 final mutant, StACSstab6, has a formate k_{cat} of 4.9 ± 0.1 s^{-1} and k_{cat}/K_m of 108 ± 5 $M^{-1}s^{-1}$, 45%
182 and 28% lower, respectively, than StACSstab's k_{cat} of 8.8 ± 0.8 s^{-1} and k_{cat}/K_m of 149 ± 18 $M^{-1}s^{-1}$
183 Figure 3D, Table S2). Normalizing the relative change in lysate activity to that of k_{cat}/K_m, we
184 estimate the change in functional expression to be 8-fold, consistent with the qualitative
185 increase in band intensity on the SDS-PAGE gel.

186 To engineer MhACS, we mutated a small set of positions corresponding to those in StACSstab
187 that yielded beneficial mutations (Table 1). We first screened a combinatorial library with
188 mutations at positions F262, W416, N524, and G527 (corresponding to StACSstab F260, W414,
189 N521, and G524), which yielded a mutant F262Y W416F N524S (MhACS1) with improved

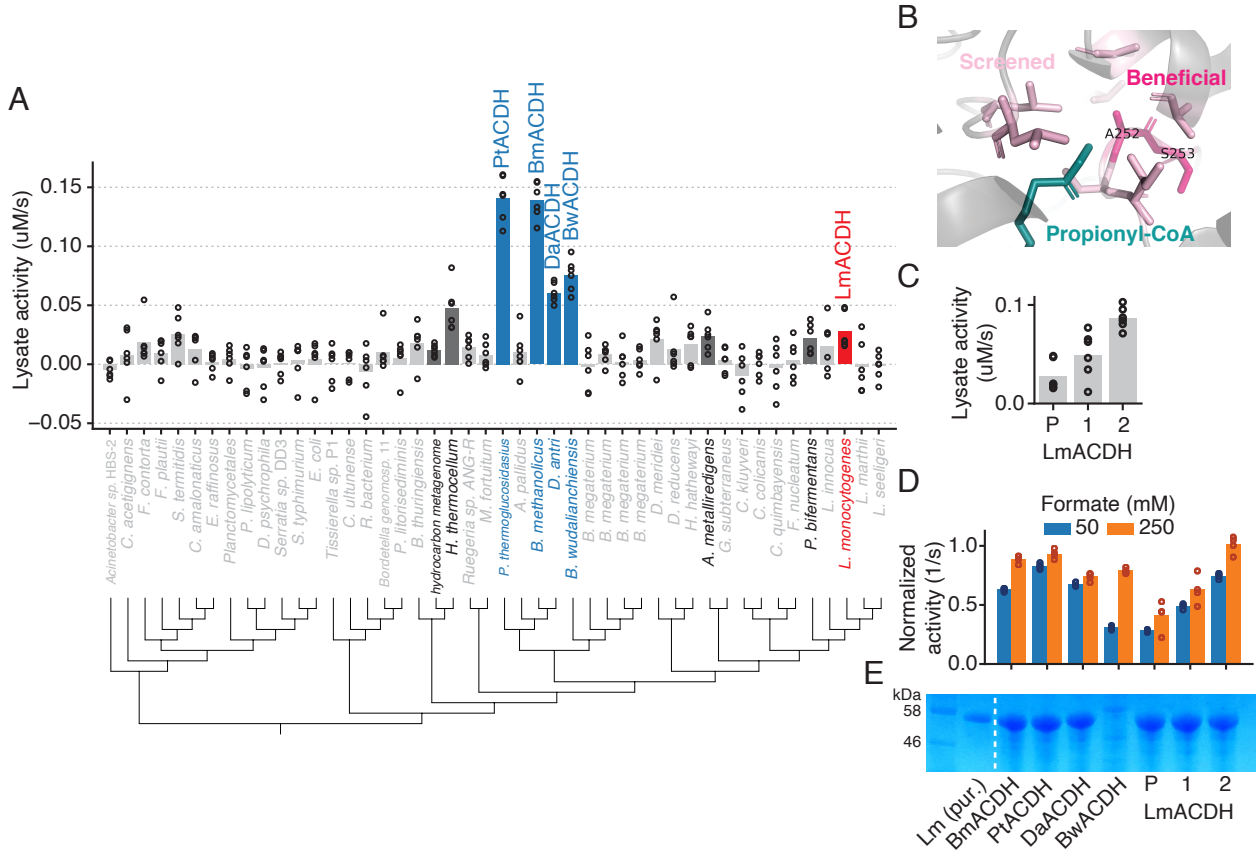
190 lysate activity. Using MhACS1 as a parent, we then screened single-site-saturating libraries and
191 discovered an improved variant with additional mutation Y499V (MhACS2). At this point,
192 screening additional site-saturating libraries MhACS2 failed to yield improved mutants. Given
193 the correlation between higher StACSstab expression and lower ACS activity, we hypothesized
194 that an *E. coli* host strain with lower basal expression may increase our chances of isolating
195 more improved mutants. Therefore, we switched from the BL21*(DE3) host strain to
196 NovaBlue(DE3), which has the stronger $LacI^q$ repressor (Figure S6). Starting with MhACS2 and
197 screening site-saturating libraries, we obtained a mutant with S524R (MhACS3; N524R relative
198 to MhACS) with 1.9-fold higher lysate activity than the MhACS parent (Figure 3B).

199 As with StACSstab, increased lysate activity of MhACS mutants did not translate to increased
200 specific activity, but rather a decrease in formate k_{cat} and k_{cat}/K_m over the course of evolution
201 (Figure 3D). Formate K_m also did not change appreciably, staying around 150 mM in all variants
202 (Figure 3D, Table S2). Like StACSstab, the k_{cat}/K_m for acetate decreased in successive rounds
203 of mutation. Thus, the ratio of formate to acetate k_{cat}/K_m increased by 5.5-fold from MhACS to
204 MhACS3, although its absolute value is lower for MhACS3 than for StACSstab6. Interestingly,
205 the decrease in acetate k_{cat}/K_m in MhACS3 was due to a combination of increased K_m and
206 decreased k_{cat} contributed by different mutations. By contrast, in the StACSstab evolutionary
207 trajectory, the major change was a drastic decrease in k_{cat} caused by the initial N521V mutation,
208 which was actually negated somewhat by a decrease in K_m. Normalizing the increase in lysate
209 activity from MhACS to MhACS3 by the 59% decrease in formate k_{cat}/K_m, we find that functional
210 expression of MhACS3 is 4.6-fold higher than that of MhACS. This is surprising given the
211 roughly similar apparent expression of all MhACS variants, perhaps suggesting a change in the
212 active fraction of expressed protein.

213 **Discovery and improvement of ACDH**

214 Previous work used *Listeria monocytogenes* (LmACDH) for formate reduction because it was
215 the most active of 5 homologs tested (8). We sought to identify additional active variants by a
216 two-pronged strategy of homolog screening and directed evolution. Although formyl-CoA is the
217 desired substrate of ACDH, it is not commercially available and has a very short half-life(15).
218 Therefore, we screened ACDHs using formate as a substrate instead, including ACS as a
219 coupling enzyme to generate formyl-CoA in the reaction (Figure S7A). This does not allow
220 quantitative estimation of the K_m of ACDH for formyl-CoA but is sufficient to determine the
221 relative activities of ACDH variants.

222 We analyzed all available ACDH homologs in UniProt and BRENDA, and chose 46 for gene
223 synthesis and testing. Alignment and clustering of ACDHs revealed two divergent clades with
224 roughly equal numbers of sequences (Figure 4). One clade contained members such as *E. coli*
225 MhpF, which natively operates as a complex with an aldolase (24). EcMhpF was previously
226 shown to have low activity compared to LmACDH(8), suggesting difficulties in expressing the
227 monomer form. Therefore, we avoided members of the MhpF-like clade and focused instead on
228 the clade containing *E. coli* AdhE, LmACDH, and bacterial-microcompartment-associated
229 enzymes such as EutE (25).



230

231 **Figure 4. Homolog screening and directed evolution of ACDH enzymes.**

232 A) Lysate activity of ACDH homologs. Dots show technical replicates from 2 independent expression
233 strain transformants; bars show mean of all replicates. Enzymes with statistically significant activity
234 compared to empty vector (FDR=0.05, 2-sample t-test with Benjamini-Hochberg correction) are shown in
235 dark gray or colored bars and black font; non-significant activity is indicated by light gray bars and font.
236 Blue indicates enzymes chosen for followup characterization and red indicates enzyme used for directed
237 evolution. Raw data in Table S1B. B) Structure of *L. monocytogenes* ACDH (3k9d, propionyl-CoA from
238 5jfn) with residues chosen for site-saturating mutagenesis shown in light pink. Residues that yielded
239 beneficial mutations that were kept in the evolved variants are highlighted in dark pink. C) Lysate activity
240 of evolved LmACDH variants. "P" indicates parental or wildtype enzyme, and numbered variants
241 correspond to mutants listed in Table 1. Circles represent 3 technical replicates and bars show the mean.
242 D) Activity of selected homologs and evolved LmACDH variants, normalized to enzyme concentration.
243 Shares x-axis labels with panel E. E) SDS-PAGE of clarified lysates of *E. coli* strains expressing ACDH
244 variants. Each lane contains lysate from equal biomass. Lane 2 contains purified LmACDH.

245

246 We screened clarified lysates of the ACDH homologs for the ability to oxidize NADH in the
247 presence of ACS, ATP, CoA, and 50mM formate. We initially screened 9 ACDH homologs and
248 then used those results to choose 37 more (Figure S1). From the full set of 46 homologs, we
249 found 9 with activity significantly higher than empty vector at a 5% false discovery rate (Figure
250 4A, S8A; t-test with Benjamini-Hochberg correction). Five homologs had higher activity than
251 LmACDH (Figure 4A, colored bars). We then purified the 4 homologs with highest lysate activity
252 as well as LmACDH and assayed them in 50mM and 250mM formate with excess ACS. Unlike
253 in our ACS screen, all the ACDH homologs with higher lysate activity than LmACDH also had
254 higher activity after normalizing by enzyme concentration, with PtACDH having the highest
255 activity in both assays (normalized activity of 0.93 ± 0.04 s^{-1} in 250mM formate; Figure 4D,
256 S8B). All homologs were more active in 250mM formate than in 50mM. Their relative rankings
257 were unchanged by formate concentration, except for BwACDH, which is the least active in
258 50mM formate but among the most active homologs at 250mM. This suggests that it has a
259 higher Km for formyl-CoA than other homologs.

260 To engineer LmACDH, we used its crystal structure 3k9d along with a propionyl-CoA substrate
261 superimposed from a related structure 5jfn to choose positions close to the acyl moiety for
262 mutagenesis. We screened site-saturating libraries at 5 positions and identified a mutant A252S
263 (LmACDH1) with increased activity (Table 1). We then screened some of the same positions on
264 top of the LmACDH1 background, as well as additional residues close to A252 in the structure,
265 and found A252S S253C (LmACDH2) to have even higher activity (1.00 ± 0.07 s^{-1} at 250mM
266 formate, or 2.5-fold higher than LmACDH; Figure 4D). In fact, LmACDH2 has slightly higher
267 activity than PtACDH, the best homolog we discovered.

268 **Expression of pathway in *M. flagellatus* KT**

269 Having identified ACSs with improved expression and ACDHs with increased activity, we next
270 asked whether these enzymes could support formate reduction *in vivo*. Methylotrophic bacteria
271 are able to assimilate formaldehyde as an intermediate of methanol, and those that do this via
272 the RuMP pathway cannot natively assimilate formate. Therefore, if we introduced ACS and
273 ACDH activities into an RuMP methylotroph (which also had an NADH-producing FDH), this
274 would in principle confer partial or complete formatotrophy. In practice, our enzymes are likely
275 too inefficient to support growth, but even a low flux from formate into biomass could potentially
276 be used to select for further enzyme improvements.

277 We chose the betaproteobacterium *Methylobacillus flagellatus* KT to express the pathway
278 because it assimilates methanol via the RuMP pathway, grows robustly under standard
279 laboratory conditions, and is amenable to genetic manipulation (26). We first tested a panel of
280 promoters for their ability to drive high constitutive expression of a red fluorescent protein
281 reporter from a IncP-based broad-host-range plasmid in *M. flagellatus* KT (27) (Figure S9).
282 Based on this, we chose to use the native promoters Phps and PmxaF to drive ACS and ACDH
283 expression, respectively, from the plasmid. We cloned a panel of expression vectors containing
284 different ACSs coexpressed with the same ACDH, or vice versa, and conjugated them into *M.*
285 *flagellatus* KT (Figure 5A; Materials and Methods).

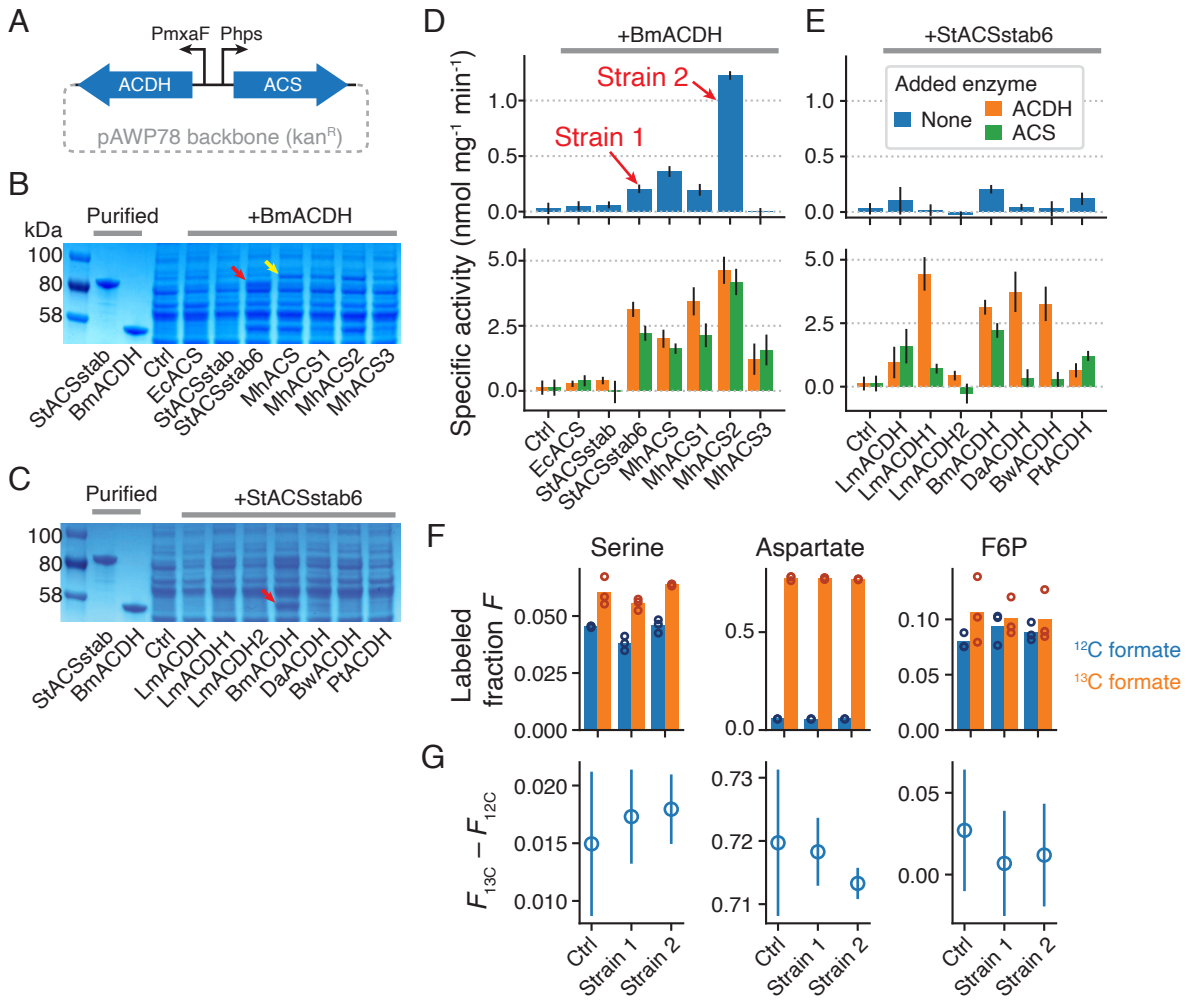


Figure 5. ACS and ACDH are expressed and active in *M. flagellatus* KT lysates but do not assimilate formate *in vivo*.

A) Schematic of vector used to express ACS and ACDH variants in *M. flagellatus* KT. PmxaF 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 co-expressed with StACSstab6. Red arrow: gel band for BmACDH. D) Specific activity of *M. flagellatus* KT lysates for formaldehyde production from 300mM formate, for various ACSs co-expressed with BmACDH (mean and s.d. from 3 technical replicates; see Materials and Methods, Figure S10). Results for cell lysates only (blue bars) or with 4μM purified ACDH (orange bars) or ACS (green bars) added. E) Same as (D), but for various ACDHs co-expressed with StACSstab6. "Strain 1" and "Strain 2" were chosen for ^{13}C labeling. F) Fraction of ^{13}C-labeled proteinogenic serine, aspartate, or fructose-6-

300 phosphate (F6P) from cells grown in methanol + 200mM ^{12}C or ^{13}C formate. 3 biological replicates of the
301 control strain and 2 different pathway variants ("Strain 1" and "Strain 2" from (D)) were assayed. Bars
302 show the mean. G) The difference in ^{13}C-labeled fraction between ^{13}C-formate-grown cells and ^{12}C-
303 formate-grown cells (mean and s.d., n=3). If ^{13}C formate were being assimilated by the pathway, then the
304 pathway-containing strains should have a higher value of this difference than the control strain.

305 To test for expression and activity of our enzymes in *M. flagellatus* KT, we analyzed clarified
306 lysates by SDS-PAGE and Nash assay (28), which measures formaldehyde production (Figure
307 S10; Materials and methods). Formaldehyde is only produced from formate if both ACS and
308 ACDH are active, so we assayed lysates with and without an added excess of purified ACS or
309 ACDH to detect activity of each enzyme individually (Figure 5D,E). This also has a side benefit
310 of boosting the sensitivity of the assay. Independent *M. flagellatus* KT transconjugants varied in
311 phenotype (growth rates, enzyme activities), so for each vector we screened multiple
312 transconjugants and picked the one with the highest enzyme activity for further characterization.

313 We found that StACSstab6 and all MhACS variants, but not EcACS or StACSstab, had a visible
314 band in SDS-PAGE when expressed in *M. flagellatus* KT. MhACS has a higher-molecular-
315 weight band than StACSstab6, reflecting their predicted molecular weights (74 and 72 kDa,
316 respectively). Consistent with the SDS-PAGE, the Nash assay only showed ACS activity in
317 StACSstab6 and MhACS variants (orange bars in Figure 5D, lower panel). Across ACDHs, only
318 BmACDH had visible SDS-PAGE expression. It also had the highest ACDH activity by Nash
319 assay, although LmACDH, LmACDH1, and PtACDH also had above-background activities
320 (green bars in Figure 5E, bottom panel). The strains with the highest lysate activity without any
321 added enzymes are those containing BmACDH and either StACSstab6, MhACS, MhACS1, or
322 MhACS2. Notably, neither EcACS nor the LmACDH variants, which were used in the previous
323 version of the pathway (8), were well-expressed in *M. flagellatus* KT. We chose the strains
324 containing StACSstab6/BmACDH ("Strain 1"), which had the 3rd-highest activity (0.20 $\pm$ 0.04

325 nmol mg^{-1} min^{-1}; Figure 5D), and MhACS2/BmACDH ("Strain 2"), which had the highest activity
326 (1.22 ± 0.04 nmol mg^{-1} min^{-1}), for further characterization.

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

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

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

344 First, we analyzed proteinogenic amino acids as an indicator of overall incorporation of formate
345 into biomass. We inoculated control and pathway strains into MM2 medium with 0.2% ^{12}C
346 methanol and 200mM ^{12}C or ^{13}C-formate, harvested the saturated cultures, and acid-hydrolyzed
347 the biomass for LC-MS. We found almost no labeling of more than one carbon atom across the

348 amino acids examined, so we used total ^{13}C-labeled fraction, or 1 – unlabeled fraction, as a
349 simple metric for the degree of labeling (Figure 5F, S11). Serine, whose carbon atoms are
350 derived from pyruvate and thus immediately downstream of the RuMP pathway, displayed a
351 background ^{13}C-labeled fraction of about 4% in ^{12}C-formate across all strains, but a 1.5 - 2%
352 increase in labeling in ^{13}C-formate (Figure 5F). However, this increase occurred in both control
353 and pathway strains and was similar in magnitude (Figure 5G), indicating that the extra labeling
354 is not due to formate assimilation via our pathway. Aspartate had much higher labeling in ^{13}C-
355 formate than in ^{12}C-formate, although the differential labeling was again the same in all strains
356 (Figure 5F,G, middle panel). Since aspartate is derived from oxaloacetate, the high labeled
357 fraction in ^{13}C-formate is possibly due to re-assimilation of $^{13}CO_2$ by pyruvate carboxylase after
358 formate oxidation by formate dehydrogenases (29). A similar jump in labeling in ^{13}C-formate
359 was observed for threonine and glutamate, which can both be derived from aspartate (Figure
360 S11B,C). Alanine, on the other hand, had <5% labeling in ^{13}C-formate like serine, consistent
361 with also being derived from pyruvate (Figure S11B,C). Overall, no amino acid examined had
362 labeling indicative of formate assimilation by the introduced pathway.

363 The analysis above requires a sufficiently high formate reduction flux to result in labeled
364 proteins. For a more sensitive test of formate assimilation, we monitored fructose-6-phosphate
365 (F6P), a metabolite immediately downstream of formaldehyde assimilation into the RuMP
366 pathway. We added 200mM ^{12}C- or ^{13}C-formate to mid-exponential-phase cultures of control
367 and pathway strains, continued incubating the cultures for 2 hours, and then harvested and
368 extracted metabolites for LC-MS. We saw 8-10% labeling of F6P in ^{12}C-formate, close to the
369 expected background rate of 6% (Figure 5F, right panel). Labeling in ^{13}C-formate was higher, at
370 around 10% in all 3 strains. However, as in the case of the amino acids, there was no increase

371 in the difference in labeling between labeled and unlabeled formate conditions (Figure 5G, right
372 panel). Therefore, we were unable to detect evidence of our pathway assimilating formate into
373 biomass *in vivo*.

374 It is possible that our ACS and ACDHs still do not have the activity needed to supply even
375 detectable formaldehyde flux through the RuMP pathway. To test this, we used flux balance
376 analysis (FBA) to calculate the theoretical growth rate that could be supported by the measured
377 rate of formate reduction in the Nash assay. A genome-scale model of *M. flagellatus* KT
378 metabolism does not exist, but we used a model developed for another RuMP-pathway
379 methylotroph, *Methylotuvimicrobium buryatense* 5GB1C (30). We assumed that the flux through
380 the methanol dehydrogenase reaction, which provides all the formaldehyde (and reduced
381 carbon) for biomass production, is the same as the highest specific activity we measured in *M.*
382 *flagellatus* KT lysates, or 1.2 nmol/min/mg (Figure 5D). We found that this would support a
383 theoretical growth rate of 0.00048 h^{-1}, or a doubling time of 8.6 weeks, even without ATP
384 maintenance (with ATP maintenance, growth was infeasible). This is much slower than even the
385 55-hour growth supported by an unoptimized reductive glycine pathway (9), indicating that
386 further improvements to activity and/or expression are needed.

387 Discussion

388 **Utility of phylogenetically diverse enzymes**

389 Previous work showed that EcACS and LmACDH have formate-reduction activity and that the
390 enzymes are functional when expressed in *E. coli*. We extend that work to identify a panel of
391 natural and engineered ACS and ACDH variants with improved expression and lysate activity
392 and show that StACSstab6, MhACS2, and BmACDH, are well-expressed and active in the
393 methylotroph *M. flagellatus* KT. None of these three enzymes had the highest ACS or ACDH

394 activities *in vitro*, showing that expression in the host cytosolic environment is an equally if not
395 more important factor than catalytic properties in practice. The computational design of
396 StACSstab and the thermophilic source organisms of MhACS and BmACDH may have played a
397 role in their greater expression and host range. By contrast, neither EcACS and LmACDH, the
398 previous best enzymes for this pathway, were expressed in *M. flagellatus* KT, despite EcACS
399 having the 2nd-highest formate k_{cat}/K_m of the wildtype ACSs and LmACDH2 being the most
400 active ACDH we found. This highlights a key advantage of screening phylogenetic diversity in
401 that this approach offers not only the chance to discover high activity, but also high expression
402 and evolvability (20,31,32).

403 **Improving expression versus activity**

404 Since we performed directed evolution on ACSs using a lysate-based screen, it is reassuring
405 that we obtained increased lysate activity (5.8-fold for StACSstab and 1.9-fold for MhACS).
406 However, this was entirely due to increases in functional expression (8-fold for StACSstab and
407 4.6-fold for MhACS) and not catalytic efficiency. In fact, k_{cat}/K_m decreased by 28% for
408 StACSstab6 on formate, although it decreased by 72% on acetate, leading to an overall
409 increase in the formate specificity from the parent enzyme. Despite this lack of improvement in
410 catalytic activity, the increased soluble expression proved crucial to functionality in *M. flagellatus*
411 KT, where StACSstab6, but not the StACSstab parent, was expressed and active. Interestingly,
412 even wildtype ACS homologs differed widely in soluble expression in *E. coli* as well as in *M.*
413 *flagellatus*. By contrast, there were no obvious differences in expression between the various
414 ACDH homologs or evolved variants in *E. coli*, and our directed evolution of ACDH using a
415 lysate-based assay led to increases in both lysate and specific activity.

416 Why did our lysate assays select for increased specific activity in ACDH but not in ACS? The
417 ACS parents we chose perhaps started with poor stability or expression, but this is unlikely
418 given their origins . Moreover, stability usually *decreases* while evolving for activity (33). A more
419 likely possibility is that ACS activity is toxic. This has been observed previously (34), and thus
420 decreasing it may allow cells to tolerate increased expression. This is consistent with the
421 expression gain concomitant with a sudden reduction of acetate activity from StACSstab to
422 StACSstab1/2 (N521V/L). Despite having a much lower acetate activity, however, even
423 StACSstab6 still appears to be toxic, frequently leading to *E. coli* colonies with spontaneously
424 decreased activity (one such colony can be seen as a replicate in Figure 3B). This problem can
425 be mitigated in future rounds of evolution by using a low-background expression host and/or
426 reducing induced expression level.

427 **Challenges of one-carbon substrates**

428 A more fundamental problem is the possibility of biophysical limits on the formate activity of
429 ACS. We chose ACS for synthesizing formyl-CoA because formate is structurally similar to
430 ACS's native substrate acetate. However, formate is less electrophilic than acetate, which could
431 make it challenging to achieve a high k_{cat}. Indeed, formate k_{cat}s among our natural and evolved
432 ACSs never exceeded 12 s^{-1}, while the highest acetate k_{cat} was 43.2 $\pm$ 3.1 s^{-1}. One-carbon
433 compounds also have relatively few functional groups for interacting with a substrate-binding
434 pocket, leading to higher K_ms (35) and potentially explaining why even our lowest ACS K_m for
435 formate is greater than 40mM. As a result, our highest formate k_{cat}/K_m values are between 100-
436 200 $M^{-1}s^{-1}$, 2-3 orders of magnitude lower than many natural enzymes, including the acetate
437 activity for native ACSs. Encouragingly, however, natural enzymes find formate equally
438 challenging. The *M. extorquens* formate-tetrahydrofolate ligase (FTL), which activates formate

439 for assimilation, has a K_m of 22 mM and a k_{cat} of ~100 s^{-1}, for a k_{cat}/K_m of ~5000 $M^{-1}s^{-1}$, about 30-
440 fold higher than our best ACSs (36). Despite being 2 orders of magnitude lower than the median
441 enzyme (35), this activity can support fully formatotrophic growth in natural and engineered
442 pathways. Therefore, a physiologically relevant activity of the formate reduction module may be
443 within reach given further enzyme engineering.

444 ACDH is not expected to be as challenging an engineering target as ACS, because most of the
445 substrate binding affinity is contributed by the CoA group. We did not directly measure the K_m of
446 ACDH for formyl-CoA, but the ACS-coupled assays show it is at most 7 mM (Figure S7). In
447 reality it is probably much lower; the K_m of ACDH for acetyl-CoA can be <100μM (37), and the
448 K_m of 2-hydroxyacyl-CoA lyase for formyl-CoA is 200μM (15), despite this not being its native
449 substrate. However, a potential problem with ACDH is k_{cat}. Even though the fastest ACDH
450 homolog in the literature has a k_{cat} of ~60 s^{-1} on acetate (38), our best ACSs were almost 2
451 orders of magnitude slower on formate. However, given that we were able to increase this value
452 by ~2-fold in 2 rounds of directed evolution, further engineering will likely result in additional
453 gains.

454 **Effects of mutations on ACS and ACDH**

455 Previous work found that ACS substrate specificity can be changed from acetate to larger or
456 more polar substrates by mutating V310, T311, V386, or W414 in StACS, which are all with 4Å
457 of the acetyl moiety (20,22,23). However, for formate, we did not find increases in lysate activity
458 when mutating V310, T311, or V386 in isolation or V310 and V386 combinatorially (Table 1).
459 Instead, we found that the previously unexplored N521V/L mutations cause a large decrease in
460 acetate activity in StACSstab. In the 2p2f structure of StACS, acetate is modeled in the active
461 site with its acetyl carbon pointed toward N521, 5.7 Å away from the asparagine's sidechain

462 carbonyl (23). This orientation is consistent with the ability of large hydrophobic substitutions at
463 N521 to favor a smaller acyl moiety in the substrate.

464 Several other mutations increased ACS lysate activity while decreasing acetate k_{cat}/K_m.
465 StACSstab F260W and W414F increased acetate K_m, as can be seen by comparing
466 StACSstab3 and StACSstab4 to StACSstab1. In the StACS structure, W414 is within 4 Å of the
467 acyl substrate, while F260 is 10.9 Å away but in contact with the first-shell V310 sidechain
468 (Figure 3A). MhACS Y499V decreased acetate k_{cat}; it is 11.1 Å from the acyl substrate but
469 contacts the backbone of W414. The homologous StACSstab Y496 had mutants among the top
470 hits in the first round of directed evolution (Figure S5B), but was not pursued in favor of N521V.
471 StACSstab G524 lines the CoA binding site in StACSstab and as a result, G524S/L is known to
472 block CoA addition to the acyl group(23). Our results show that G524A, on top of N521V
473 W414F, maintains formate activity while increasing acetate K_m. Additional residues (e.g. F421)
474 showed evidence of improved activity in our initial screens on StACSstab (Figure S5), but were
475 not pursued fully. They are prime candidates for mutagenesis in future rounds of evolution.

476 Over two rounds of site-saturating mutagenesis and screening at 9 residues comprising the
477 acyl-binding tunnel in LmACDH, we found a double-mutant LmACDH2 (A252S S253C) that
478 contributed to improved activity on formyl-CoA. These positions partially overlap with those
479 mutagenized in a recent effort to engineer an ACDH to reduce glycolyl-CoA to glycolate (20).
480 Our evolved isolates were comparable in activity to the best natural homologs BmACDH and
481 PtACDH, but ultimately only BmACDH expressed well in *M. flagellatus* KT. BmACDH has the
482 same sequence as LmACDH at positions homologous to A252 and S253, so these are clear
483 candidates for mutagenesis in future directed evolution efforts.

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

485 We chose to use *M. flagellatus* KT for testing *in vivo* activity of our pathway because it could
486 potentially gain formatotrophy with only the expression of ACS and ACDH. This is the first
487 published instance, to our knowledge, of metabolic engineering in this organism. Recently, an *E.*
488 *coli* strain was engineered to grow on methanol as a sole carbon source via the RuMP pathway
489 (39). This provides the alternative option of engineering the ACS/ACDH pathway in *E. coli*
490 instead, which would allow access to a wider range of genetic tools and pathway manipulations
491 (40). Most importantly, it would allow the improvements we obtain from directed evolution in *E.*
492 *coli* lysates to be directed translated into *in vivo* activity or expression. However, even in the fully
493 methylotrophic *E. coli*, formaldehyde toxicity is still a major problem, and perhaps as a result, its
494 doubling time on methanol is more than 8 hours. By contrast, *M. flagellatus* KT has a doubling
495 time of 2 hours on methanol, indicating its naturally evolved robustness against formaldehyde
496 toxicity. Perhaps the best approach in future work is to use a combination of strains – *E. coli* for
497 initial troubleshooting and improvement of enzymes, followed by *M. flagellatus* KT for fine-tuning
498 for maximal flux.

499 Conclusion

500 Through phylogenetic homolog screening and directed evolution, we identified a panel of highly
501 expressed and active ACS and ACDH enzymes and gained insight into the genetic
502 determinants of their acyl-substrate specificity. We established a plasmid-based expression
503 system in the RuMP-pathway methylotroph *M. flagellatus* KT and used it to introduce a formate-
504 reduction pathway in an attempt to confer synthetic formatotrophy. Although we ultimately did
505 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 CO_2 into value-added chemicals.

509 # Materials and Methods

510 **Bioinformatics and enzyme homolog selection**

511 All bioinformatics and analysis/visualization of experimental data was performed in
512 Python/Jupyter. Phylogenetic trees for figures were computed by FastTree (41) and visualized
513 using iTOL (42).

514 To identify ACS homologs for testing, an initial candidate list of 8,911 sequences was compiled
515 that included: 6,104 sequences from the "Acetate-CoA ligase" Interpro family (IPR011904)(43)
516 with the same 3 domains as EcACS downclustered to 90% identity using CD-HIT (44); 2,790
517 sequences from RefProt based on a pHMMER search (E-value < 10^{-200}) with query EcACS
518 (P27550) (45); 17 experimentally characterized ACS homologs from BRENDA (EC 6.2.1.1) (46).
519 From the initial candidate list, an alignment and distance matrix was generated using Clustal
520 Omega (47), and a hierarchical clustering of the distance matrix was used to guide manual
521 selection of a diverse final set of ACSs. Initially 11 ACSs were chosen for testing (see below for
522 details). Then, given the results of the 1st round of testing, 30 additional ACSs were chosen to
523 further sample clades with high activity while also exploring new areas of sequence space.

524 For ACDHs, 7,646 sequences were obtained from UniProt via a pHMMER search (E-value <
525 $7x10^{-17}$) with query LmACDH (Q8Y7U1) or enzyme commission number search (EC 1.2.1.10),
526 or from BRENDA (EC 1.2.1.10). After alignment and clustering of ACDHs, mhpF-like sequences
527 were removed, leaving 4,037 adhE-like sequences from which the final selection was made.
528 Initially 9 ACDHs were chosen for testing; then these results were used to choose 37 additional
529 ACDHs to test.

530 The full phylogeny of all ACSs and ACDHs considered for homolog discovery would be too
531 large to visualize as a tree, so Figure S1 shows only untested sequences that have less than

532 50% (ACSs) or 40% (ACDHs) amino-acid identity to each other and to the tested homologs.
533 Some of the tested homologs are more similar to each other than this because they were
534 chosen for reasons other than diversity; these were all included in the trees.

535 **DNA synthesis and *E. coli* strain construction**

536 ACS and ACDH sequences were codon optimized for *E. coli* expression using Integrated DNA
537 Technologies' online tool (https://www.idtdna.com/CodonOpt; accessed September 2018). DNA
538 synthesis was performed at Twist or the Joint Genome Institute of the U.S. Department of
539 Energy. Genes were cloned into vector pET29b+ between *Nde*I and *Xho*I such that expressed
540 enzymes have a C-terminal 6xHis tag. Expression vectors with ACDH genes were
541 electroporated into *E. coli* expression strain BL21*(DE3), propagated on lysogeny broth (LB) +
542 50ug/mL kanamycin, and stored at -80C in 25% glycerol.

543 ACS is known to be repressed under standard physiological conditions by acetylation at K609,
544 but can be derepressed by a point mutation L641P (48). We found that a simpler method of
545 knocking out the patZ deacetylase leads to comparable EcACS activity (Figure S2A), so we
546 used a BL21*(DE3) ΔpatZ host strain for all ACS experiments. The *patZ* gene was deleted from
547 BL21*(DE3) using lambdaRed recombinase (49). For the round of directed evolution from
548 MhACS2 to MhACS3, the alternate expression strain NovaBlue(DE3) ΔpatZ was constructed
549 and used.

550 **Protein expression, lysis, and SDS-PAGE**

551 To express proteins, strains were inoculated directly from -80C stocks into auto-induction
552 medium (50) at 1:500 to 1:50,000 dilution and incubated at 37C for 24 hours. For screening,
553 500μL cultures were grown in 2mL 96-well microtiter plates (Axygen P-DW-20-C) with shaking
554 at 1000rpm on a benchtop shaker (Heidolph Titramax 1000) in a temperature-controlled room.

555 For SDS-PAGE and Nash assays, 5mL cultures were grown in round-bottom glass tubes in a
556 rotary shaker incubator at 250rpm. For purification, 50mL or 500mL cultures were grown in
557 Erlenmeyer flasks and shaken at 250rpm.

558 To prepare lysates for screening, 96-well plate cultures were pelleted at 2200g for 10 min,
559 washed once in 4°C water, and resuspended by vortexing after adding 300μL/well of lysis buffer
560 (50mM HEPES pH 7.5, 50mM NaCl, 2 mM MgCl2) with 0.6 mg/ml lysozyme, 0.1mg/mL
561 polymyxin B, and 1:50,000 Sigma benzonase nuclease. Plates were incubated at 37°C for 50
562 min without shaking followed by 10 min shaking at 1000rpm. Then, lysates were pelleted at
563 2200g for 10 min, and the supernatant was used for downstream assays.

564 To prepare lysates for SDS-PAGE, Nash assay, or protein purification, 5mL, 50mL, or 500mL
565 cultures were pelleted at 4000rpm for 10min and resuspended, respectively, in 0.5mL, 8mL, or
566 30mL lysis buffer with 0.1mM DTT, 1:500 Sigma protease inhibitor cocktail, 1:50,000 Sigma
567 benzonase nuclease. Cell suspensions were sonicated on ice (Branson SLPt) for 3 repeats of
568 10 seconds on and 10 seconds off at 30% amplitude for 5mL cultures, or 6 repeats of 30
569 seconds on and off at 70% amplitude for 50mL cultures, or 12 repeats of 30 seconds on and off
570 at 70% amplitude for 500mL cultures. Lysates were pelleted at 4000rpm for 15 minutes and the
571 supernatant used for analysis or purification.

572 To analyze lysates and purified enzymes by SDS-PAGE, samples were mixed 1:1 with 2x
573 Laemmli sample buffer with 2-mercaptoethanol (Bio-Rad) and boiled for 10min. A sample
574 containing 1-20μg of protein was loaded into a 4-15% precast gel (Bio-Rad Mini-ProTEAN) and
575 run at 60V for 20min followed by 160V for 1 hour. To compare expression levels across lysates,
576 protein concentration was determined by BCA and equal μg of protein were loaded in each

577 lane. Gels were stained by Coomassie blue and imaged using a digital camera. Minor contrast
578 adjustments were made to the images to improve visibility of bands.

579 **Assays for enzyme activity in lysates**

580 ACS was assayed in lysates using a discontinuous assay with DTNB (5,5-dithio-bis-(2-
581 nitrobenzoic acid)), which reacts with CoA to yield absorbance at 412nm (20,51). In a microtiter
582 plate (Corning Costar 3370), 100μL of reaction buffer (10 μL of expression-induced *E. coli*
583 lysate, 50mM HEPES pH 7.5, 2mM $MgCl_2$, 5mM ATP, 0.5mM CoA, and 50mM sodium formate)
584 was aliquoted. The formate was added last to start the reaction, everything was incubated for
585 10min at 37°C, then stopped by adding 100μL DTNB reagent (50mM HEPES pH 7.5, 2mM
586 DTNB). Absorbance at 412nm was read on a plate reader (Molecular Devices Spectramax 190).
587 Empty vector control lysates were used to establish the background signal, and the metric
588 $\Delta A412 = A412_{Empty\ vector} - A412$ was used to quantify lysate ACS activity. Note that higher
589 ACS activity corresponds to lower $A412$ but higher $\Delta A412$. For the first round of ACS directed
590 evolution, a continuous assay was used (see "Protein purification and enzyme kinetics" below),
591 but subsequent rounds of evolution used the discontinuous assay described above.

592 ACDH lysates were assayed in a continuous assay by coupling to ACS, all steps at 37°C.
593 Reactions were performed in a microtiter plate with a total volume 200μL containing 2μL
594 clarified lysate, 2μM StACSstab1, 50mM HEPES pH 7.5, 5mM $MgCl_2$, 1mM DTT, 2.5mM ATP,
595 0.5mM CoA, 0.6mM NADH, and 50mM formate. Reactions were prepared in 100μL at 2x
596 concentration and then 100μL of 2x formate was added to start the reaction. Absorbance at
597 340nm was monitored and NADH concentration was calculated as $[NADH] = \frac{A340}{\varepsilon \cdot l}$, where $\varepsilon =$
598 $6.22\ mM^{-1}cm^{-1}$ is the extinction coefficient of NADH and $l = 0.56\ cm$ is the path length of
599 200μL of reaction mixture in the microtiter plate. Initial velocities were calculated from least-

600 square linear fits to the first 3-10 datapoints. The amount of ACS to use for coupling was
601 determined by titrating ACS for every new batch of purified ACS or round of lysate screening
602 (Figure S5). For assaying purified ACDHs, we used 30x molar excess of coupling ACS.

603 **Directed evolution to improve ACS and ACDH**

604 To engineer StACSstab and MhACS, residues were selected for site-saturating mutagenesis
605 based on proximity to the acetate molecule in the crystal structure of the *Salmonella*
606 *typhimurium* ACS (PDB: 2p2f). Mutant libraries at single positions were constructed using
607 "inside-out" PCR from NNK or "22c" (52) degenerate primers and multi-site combinatorial
608 libraries were made by overlap-extension PCR. Libraries were electroporated into expression
609 host strains (see above). One 96-well plate of mutant clones (plus control strains) was screened
610 for each single-site library. Eight plates were screened for the 4-site library in MhACS evolution
611 round 1. The best 10-20 mutants were restreaked on LB + kan plates and 4 colonies of each
612 mutant were screened again. The best mutant from the secondary screen was used as the
613 parent for the next round of evolution. The DTNB assay with 50mM formate was used for all
614 screening all ACS mutants, except in round 1 of StACSstab evolution, when the myokinase
615 coupled assay was used. To engineer LmACDH, its crystal structure (3k9d) was superimposed
616 on the *R. palustris* ACDH (5jfn) (53), and the position of the substrate propionyl-CoA from 5jfn
617 was used to choose residues in 3k9d for saturation mutagenisis. Screening was done using the
618 ACS-coupled assay described above on clarified lysates at 50mM formate.

619 **Protein purification and enzyme kinetics**

620 All steps were done at 4°C. 1mL of Ni-NTA superflow resin (Qiagen) was placed in a gravity-
621 flow column (GE Healthcare PD-10) and equilibrated by flowing through 10mL of lysis buffer.
622 Then 8mL of clarified lysate was added and the column was sealed and placed on ice and

623 nutated (VWR 12620-916) for 10 minutes. Then the lysate was flowed through the column,
624 10mL of wash buffer (50mL HEPES pH 7.5, 300mM NaCl, 35mM imidazole) was applied, and
625 protein was eluted in 10mL of elution buffer (50mM HEPES pH 7.5, 50mM NaCl, 150mM
626 imidazole). Eluate was exchanged to lysis buffer by spinning at 4000g for 15min in Amicon
627 Ultra-15 30kDa (for ACS) or 10kDa (for ACDH) centrifugal filters. Glycerol was added to 10%
628 and protein concentration was determined by BCA assay (Pierce 23227). Kinetic assays were
629 performed immediately after purification. Additional purified enzyme was split into aliquots and
630 stored at -20°C.

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

638 ACS kinetics were determined by a continuous assay using coupling enzymes (23,51). All steps
639 were performed at 37°C. All coupling enzymes were from Sigma. A reaction buffer was
640 prepared with 0.05 – 0.2μM of ACS, 15 U/mL pyruvate kinase, 23 U/mL lactate dehydrogenase,
641 and 25 U/mL myokinase in 50mM HEPES pH 7.5, 5mM $MgCl_2$, 1mM DTT, 0.6mM NADH,
642 2.5mM phosphoenolpyruvate, 2.5mM ATP, and 0.5mM CoA. 100μL of a 2x portion of the
643 reaction mixture was aliquoted into a microtiter plate and the reaction started by adding 100μL
644 of 2x sodium formate or acetate. Absorbance at 340nm was monitored for 10 minutes on the
645 plate reader and initial velocities of NADH oxidation were calculated as above. Kinetic

646 parameters k_{cat} and K_m were extracted from plots of initial velocities versus substrate

647 concentration by fitting $v = k_{cat}\frac{[S]}{[S]+K_m} + b$, where v is the per-enzyme initial velocity, $[S]$ is the

648 substrate concentration, and b is a background rate. Kinetic curves were fit using

649 `scipy.optimize.curve_fit`.

650 Purified ACDHs were assayed similarly to ACDH lysates as described above, except rates of

651 NADH oxidation were normalized to enzyme concentration as determined by BCA.

652 **Cloning and *M. flagellatus* KT strain construction**

653 All genetic constructs for *M. flagellatus* KT expression were maintained on a IncP-based broad-

654 host-range plasmid, whose backbone was derived from pAWP87 / pCM66 (27). Expression

655 vectors for *M. flagellatus* KT were constructed using PCR amplified backbone fragments,

656 promoters from the *M. flagellatus* KT genome, and ACS/ACDH coding regions from DNA

657 synthesis. PCR primers were designed with 20-25bp of overlap and Gibson assembled (New

658 England Biolabs) and electroporated into *E. coli* strain TOP10. Electrotransformants were

659 propagated in LB + 50ug/mL kanamycin at 20°C to avoid toxicity of the constructs.

660 A rifamycin-resistant isolate of *M. flagellatus* KT was used for triparental conjugation. Briefly, *M.*

661 *flagellatus* KT was patched, along with plasmid donor strain and helper strain pRK2073 (54),

662 onto plates containing MM2 medium (14.5 mM K_2HPO_4, 18.8 mM NaH_2PO_4 (monohydrous), 0.8

663 mM $MgSO_4$ (heptahydrous), 3.8mM Na2SO4, 9.9mM KNO3, and 1x Vishniac trace elements

664 (55)) with 5% LB and 2% methanol and incubated 16-20 hours at 37°C. Cell mixtures were then

665 plated on plates with MM2 + 0.1% pyruvate + 2% methanol + 50ug/mL rifamycin + 50ug/mL

666 kanamycin and incubated for 2-3 days at 37°C, colonies were restreaked onto the same

667 medium and incubated for another 2-3 days at 37°C. For each conjugation, multiple colonies

668 were screened and by SDS-PAGE and Nash assay and the colony with the best phenotype was

669 used for downstream assays. Strains were stored by culturing in liquid MM2 medium and then
670 freezing at -80°C with 10% DMSO.

671 Promoter reporter constructs were constructed by Gibson assembly of various promoters
672 upstream of dTomato in pAWP87. These were conjugated into *M. flagellatus* KT, colonies were
673 inoculated into seed cultures in MM2 + 2% methanol + kanamycin, grown for 24-48 hours at
674 37°C with shaking, diluted 1:50 and grown 24 hours, and then measured at 535nm excitation /
675 590nm emission on a plate reader (Tecan Infinite 500). Four promoters (Phps, PmxaF, Ptrc,
676 Ptac) were chosen for driving ACS/ACDH, but intact plasmids with Ptrc and Ptac could not be
677 isolated in the *E. coli* TOP10 cloning strain and were omitted from further experiments.

678 **Nash assay**

679 *M. flagellatus* KT strains were patched from -80°C stocks onto MM2 + 2% methanol + 50ug/mL
680 kanamycin ("MM2 Me2 kan") plates and incubated at 37°C for 2 days, then inoculated into 5mL
681 liquid MM2 Me2 kan medium in a round-bottom glass tube and incubated at 37°C with 250rpm
682 shaking for 24 hours. Cultures were lysed as described above and the soluble fraction was used
683 for the assay. In a microtiter plate, 8μL lysate was added to a reaction mixture (50mM HEPES
684 pH 7.5, 5mM $MgCl_2$, 1mM DTT, 4mM ATP, 6mM NADH, and 1mM CoA, either 4μM ACDH, 2μM
685 ACS, or no additional enzyme, and either 0mM or 300mM formate) for 100μL total volume.
686 Reactions were incubated for 1 hour at 37°C. Then 100μL of Nash reagent (0.1M ammonium
687 acetate, 0.2% acetic acid, 3.89M acetylacetone) is added and reactions are incubated at 65°C
688 for 30 min. Then, to precipitate proteins, 20μL of 100% w/v trichloroacetic acid is added and the
689 reactions are placed on ice for 5 min. The plate is spun down at 2200g for 10min and 100μL of
690 supernatant is transferred to a new plate and A412nm is measured. The difference between

691 absorbance at 300mM and 0mM formate, $\Delta A412 = A412_{300mM} - A412_{0mM}$, was used to
692 quantify lysate activity.

693 **^{13}C formate labeling and analysis by LC-MS**

694 To analyze proteinogenic amino acids, *M. flagellatus* KT strains were revived from -80°C and
695 grown in 5mL liquid MM2 Me2 kan seed cultures as described above. 200μL of seed culture
696 was transferred to 5mL of MM2 + kan + 0.05% methanol + 200mM ^{12}C or ^{13}C formate. Initial
697 experiments used cultures with one growth cycle in MM2 + 2% methanol + 200mM formate.
698 Later experiments used multiple growth cycles as follows. Cultures were grown for 24 hours,
699 pelleted at 2200g for 10min, and resuspended in fresh medium with methanol and formate. This
700 iterative re-feeding was done for 3 days, and the final cell pellets were resuspended in 1mL 6N
701 hydrochloric acid and boiled in glass vials for 24 hours (56). The vials were uncapped and left to
702 dry for another 24 hours. The biomass was then resuspended in 1mL of water, dried for 24
703 hours, and then resuspended in 0.5mL of water and centrifuge-filtered (Costar Spin-X 0.22 µm,
704 Sigma). Samples were stored at -20°C until LC-MS.

705 LC-MS was performed as in (57). A Waters Xevo mass spectrometry triple quad (Xevo, Waters,
706 Milford, MA) with UPLC system equipped with a Zic-pHILIC column (SeQuant, PEEK 150 mm
707 length × 2.1 mm metal free, with 5 µm polymeric film thickness, EMD Millipore) was used for
708 detection of MID of metabolites with following LC condition. Mobile phase A is 20 mM
709 bicarbonate in water (OPTIMA grade, Thermo Fisher Scientific), mobile phase B is 100%
710 acetonitrile (OPTIMA grade, Thermo Fisher Scientific). The LC condition starts with 0.15 ml/min
711 flow rate with initial gradient A = 15% for 0.5 min, the increased to 80% A in 20 min, at 21 min, A
712 = 90%, and hold for 5 min, at 26.5 min, mobile phase A switched to 15% and then re-equilibrate
713 the column for 5.5 min. Multiple reaction monitors (MRM) were set up for each metabolite of

714 interest. For each metabolite, 12C chemical standards were used to set up the mass channel for
715 unlabeled isotopomers. The predicted mass fragments were then used to predict the MRM for
716 labeled isotopomers for each metabolite. The MassLynx software (Waters) was used to
717 integrate ion peak intensities, with subsequent analysis in Python.

718 **Availability of Materials**

719 All strains and plasmids used in this study are available upon request from the authors, and all
720 plasmids other than the homolog screening libraries have been deposited in Addgene.

721 Acknowledgements

722 We thank Yakov Kipnis, Devin Trudeau, Joseph Groom, Julius Palme, and Dan Tawfik for
723 valuable technical advice and feedback on the manuscript.

724 Data availability

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

728 Funding

729 This work was supported by a Washington Research Foundation Postdoctoral Fellowship to
730 J.W. and funding from the University of Washington to M.E.L. Gene synthesis was provided by
731 the Synthetic Biology program at the Joint Genome Institute of the Department of Energy.

732 Conflicts of interest

733 The authors declare no conflicts of interest.

References

1 Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B. and Keasling, J.D. (2012) Microbial Engineering for the Production of Advanced Biofuels. *Nature*, **488**, 320–328.

2 Yishai, O., Lindner, S.N., Gonzalez de la Cruz, J., Tenenboim, H. and Bar-Even, A. (2016) The Formate Bio-Economy. *Current Opinion in Chemical Biology*, Elsevier Ltd, **35**, 1–9. https://doi.org/10.1016/j.cbpa.2016.07.005.

3 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.

4 Liu, C., Colón, B.C., Ziesack, M., Silver, P.A. and Nocera, D.G. (2016) Water Splitting–Biosynthetic System with CO_2 Reduction Efficiencies Exceeding Photosynthesis. *Science*, **352**, 1210–1213. https://doi.org/10.1126/science.aaf5039.

5 Jouny, M., Luc, W. and Jiao, F. (2018) General Techno-Economic Analysis of CO2 Electrolysis Systems. *Industrial & Engineering Chemistry Research*, American Chemical Society, **57**, 2165–2177. https://doi.org/10.1021/acs.iecr.7b03514.

6 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.

7 Bar-Even, A. (2016) Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry*, **55**, 3851–3863. https://doi.org/10.1021/acs.biochem.6b00495.

8 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., Gallaher, J.L., Bale, J., Yoshikuni, Y., Gelb, M.H., Keasling, J.D., Stoddard, B.L., Lidstrom, M.E. and Baker, D. (2015) Computational Protein Design Enables a Novel One-Carbon Assimilation Pathway. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 3704–3709.

9 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. *Nature Chemical Biology*, Nature Publishing Group, **16**, 538–545. https://doi.org/10.1038/s41589-020-0473-5.

10 Bang, J., Hwang, C.H., Ahn, J.H., Lee, J.A. and Lee, S.Y. (2020) Escherichia Coli Is Engineered to Grow on CO 2 and Formic Acid. *Nature Microbiology*, Nature Publishing Group, 1–5. https://doi.org/10.1038/s41564-020-00793-9.

11 KATO, N., YURIMOTO, H. and THAUER, R.K. (2006) The Physiological Role of the Ribulose Monophosphate Pathway in Bacteria and Archaea. *Bioscience, Biotechnology, and Biochemistry*, **70**, 10–21. https://doi.org/10.1271/bbb.70.10.

12 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. *Metabolic Engineering*, **60**, 1–13. https://doi.org/10.1016/j.ymben.2020.03.002.

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. https://doi.org/10.1002/cbic.201500228.

14 Lu, X., Liu, Y., Yang, Y., Wang, S., Wang, Q., Wang, X., Yan, Z., Cheng, J., Liu, C., Yang, X., Luo, H., Yang, S., Gou, J., Ye, L., Lu, L., Zhang, Z., Guo, Y., Nie, Y., Lin, J., Li, S., Tian,

C., Cai, T., Zhuo, B., Ma, H., Wang, W., Ma, Y., Liu, Y., Li, Y. and Jiang, H. (2019) Constructing a Synthetic Pathway for Acetyl-Coenzyme A from One-Carbon through Enzyme Design. *Nature Communications*, **10**. https://doi.org/10.1038/s41467-019-09095-z.

15 Chou, A., Clomburg, J.M., Qian, S. and Gonzalez, R. (2019) 2-Hydroxyacyl-CoA Lyase Catalyzes Acyloin Condensation for One-Carbon Bioconversion. *Nature Chemical Biology*, **15**, 900–906. https://doi.org/10.1038/s41589-019-0328-0.

16 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. *Angewandte Chemie International Edition*, **59**, 5526–5530. https://doi.org/10.1002/anie.201915155.

17 Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R.L., Aharoni, A., Silman, I., Sussman, J.L., Tawfik, D.S. and Fleishman, S.J. (2016) Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell*, **63**, 337–346. https://doi.org/10.1016/j.molcel.2016.06.012.

18 Bräsen, C., Urbanke, C. and Schönheit, P. (2005) A Novel Octameric AMP-Forming Acetyl-CoA Synthetase from the Hyperthermophilic Crenarchaeon Pyrobaculum Aerophilum. *FEBS Letters*, **579**, 477–482. https://doi.org/10.1016/j.febslet.2004.12.016.

19 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. *Archives of Microbiology*, **194**, 943–948. https://doi.org/10.1007/s00203-012-0829-7.

20 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., Tawfik, D.S. and Bar-Even, A. (2018) Design and in Vitro Realization of Carbon-Conserving Photorespiration. *Proceedings of the National Academy of Sciences*, 201812605. https://doi.org/10.1073/pnas.1812605115.

21 Copeland, A., Gu, W., Yasawong, M., Lapidus, A., Lucas, S., Deshpande, S., Pagani, I., Tapia, R., Cheng, J.-F., Goodwin, L.A., Pitluck, S., Liolios, K., Ivanova, N., Mavromatis, K., Mikhailova, N., Pati, A., Chen, A., Palaniappan, K., Land, M., Pan, C., Brambilla, E.-M., Rohde, M., Tindall, B.J., Sikorski, J., Göker, M., Detter, J.C., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., Klenk, H.-P. and Woyke, T. (2012) Complete Genome Sequence of the Aerobic, Heterotroph Marinithermus Hydrothermalis Type Strain (T1T) from a Deep-Sea Hydrothermal Vent Chimney. *Standards in Genomic Sciences*, **6**, 21–30. https://doi.org/10.4056/sigs.2435521.

22 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. https://doi.org/10.1021/bi061023e.

23 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. https://doi.org/10.1021/bi6026506.

24 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. *Chemico-Biological Interactions*, **202**, 70–77. https://doi.org/10.1016/j.cbi.2012.11.006.

25 Zhu, H., Gonzalez, R. and Bobik, T.A. (2011) Coproduction of Acetaldehyde and Hydrogen during Glucose Fermentation by Escherichia Coli. *Applied and Environmental Microbiology*, **77**, 6441–6450. https://doi.org/10.1128/AEM.05358-11.

26 Chistoserdova, L., Lapidus, A., Han, C., Goodwin, L., Saunders, L., Brettin, T., Tapia, R., Gilna, P., Lucas, S., Richardson, P.M. and Lidstrom, M.E. (2007) Genome of Methylobacillus Flagellatus, Molecular Basis for Obligate Methylotrophy, and Polyphyletic Origin of Methylotrophy. *Journal of Bacteriology*, **189**, 4020–4027. https://doi.org/10.1128/JB.00045-07.

27 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 Methylomicrobium Buryatense. *Appl. Environ. Microbiol.*, **81**, 1775–1781. https://doi.org/10.1128/AEM.03795-14.

28 Kleeberg, U. and Klinger, W. (1982) Sensitive Formaldehyde Determination with NASH's Reagent and a 'Tryptophan Reaction.' *Journal of Pharmacological Methods*, **8**, 19–31. https://doi.org/10.1016/0160-5402(82)90004-3.

29 He, L., Fu, Y. and Lidstrom, M.E. (2019) Quantifying Methane and Methanol Metabolism of " *Methylotuvimicrobium Buryatense* " 5GB1C under Substrate Limitation. Mackelprang, R., Ed., *mSystems*, **4**, e00748-19, /msystems/4/6/msys.00748-19.atom. https://doi.org/10.1128/mSystems.00748-19.

30 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 Methylomicrobium Buryatense 5GB1. *mBio*, American Society for Microbiology, **10**. https://doi.org/10.1128/mBio.00406-19.

31 Davidi, D., Shamshoum, M., Guo, Z., Bar-On, Y.M., Prywes, N., Oz, A., Jablonska, J., Flamholz, A., Wernick, D.G., Antonovsky, N., de Pins, B., Shachar, L., Hochhauser, D., Peleg, Y., Albeck, S., Sharon, I., Mueller-Cajar, O. and Milo, R. (2020) Highly Active Rubiscos Discovered by Systematic Interrogation of Natural Sequence Diversity. *The EMBO Journal*, John Wiley & Sons, Ltd, **n/a**, e104081. https://doi.org/10.15252/embj.2019104081.

32 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**, 10005.

33 Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D.S. (2008) How Protein Stability and New Functions Trade Off. *PLoS Computational Biology*, **4**, 35–37. https://doi.org/10.1371/journal.pcbi.1000002.

34 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. *Molecular Microbiology*, **80**, 168–183. https://doi.org/10.1111/j.1365-2958.2011.07566.x.

35 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*, 4402–4410.

36 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. *Journal of Bacteriology*, **185**, 7169–7175. https://doi.org/10.1128/JB.185.24.7169-7175.2003.

37 I. Lan, E., Y. Ro, S. and C. Liao, J. (2013) Oxygen-Tolerant Coenzyme A-Acylating Aldehyde Dehydrogenase Facilitates Efficient Photosynthetic n -Butanol Biosynthesis in Cyanobacteria. *Energy & Environmental Science*, Royal Society of Chemistry, **6**, 2672–2681. https://doi.org/10.1039/C3EE41405A.

38 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. *Annals of Microbiology*, **59**, 97–104. https://doi.org/10.1007/BF03175605.

39 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*, Elsevier, **182**, 933-946.e14. https://doi.org/10.1016/j.cell.2020.07.010.

40 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 for Growth of E . Coli. https://doi.org/10.1021/acssynbio.8b00093.

41 Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*, Public Library of Science, **5**, e9490. https://doi.org/10.1371/journal.pone.0009490.

42 Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Research*, Oxford Academic, **47**, W256–W259. https://doi.org/10.1093/nar/gkz239.

43 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., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.-Y. and Finn, R.D. (2019) InterPro in 2019: Improving Coverage, Classification and Access to Protein Sequence Annotations. *Nucleic Acids Research*, Oxford Academic, **47**, D351–D360. https://doi.org/10.1093/nar/gky1100.

44 Li, W. and Godzik, A. (2006) Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics*, Oxford Academic, **22**, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

45 Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Research*, Oxford Academic, **39**, W29–W37. https://doi.org/10.1093/nar/gkr367.

46 Jeske, L., Placzek, S., Schomburg, I., Chang, A. and Schomburg, D. (2019) BRENDA in 2019: A European ELIXIR Core Data Resource. *Nucleic Acids Research*, Oxford Academic, **47**, D542–D549. https://doi.org/10.1093/nar/gky1048.

47 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G. (2011) Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Molecular Systems Biology*, John Wiley & Sons, Ltd, **7**, 539. https://doi.org/10.1038/msb.2011.75.

48 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. *Journal of Biological Chemistry*, **280**, 26200–26205. https://doi.org/10.1074/jbc.M504863200.

49 Datsenko, K.A. and Wanner, B.L. (2000) One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6640–5. https://doi.org/10.1073/pnas.120163297.

50 Studier, F.W. (2005) Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expression and Purification*, **41**, 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

51 Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D.L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D.S. and Fleishman, S.J. (2018) Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2018.08.033.

52 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 Synthetic Biology*, **2**, 83–92. https://doi.org/10.1021/sb300037w.

53 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 Microcompartment. *Scientific Reports*, Nature Publishing Group, **7**, 1–12. https://doi.org/10.1038/srep42757.

54 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.

55 Vishniac, W. and Santer, M. (1957) THE THIOBACILLI, 12. *Bacteriological Reviews*, **21**, 195–213.

56 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 Synthetic Biology*, acssynbio.7b00086. https://doi.org/10.1021/acssynbio.7b00086.

57 Fu, Y., Li, Y. and Lidstrom, M. (2017) The Oxidative TCA Cycle Operates during Methanotrophic Growth of the Type I Methanotroph Methylomicrobium Buryatense 5GB1. *Metabolic Engineering*, **42**, 43–51. https://doi.org/10.1016/j.ymben.2017.05.003.