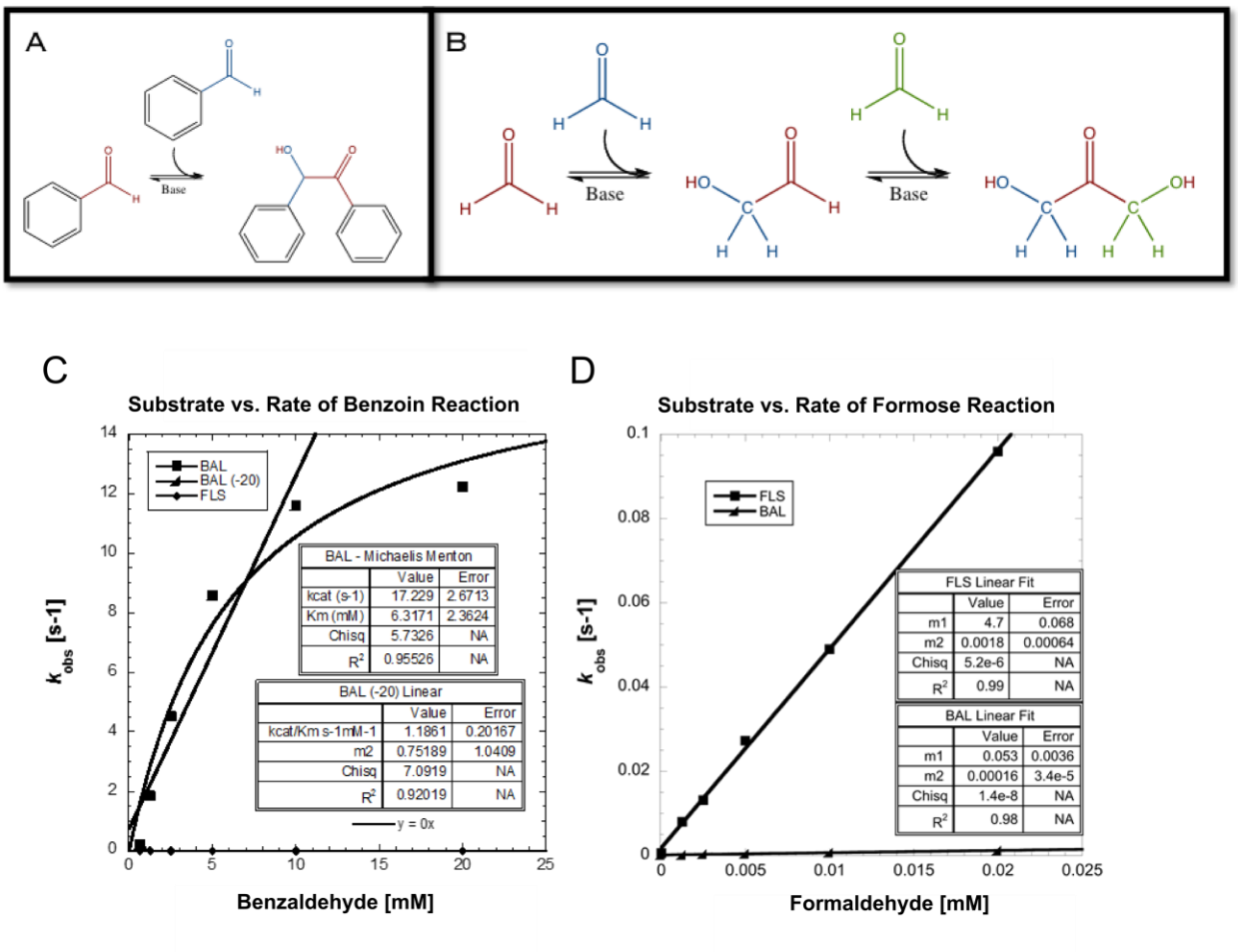


Figure-S1. Benzaldehye Lyase Catalyzed Reaction Descriptions and Kinetic Data. A) The native reaction catalyzed by BAL, in which two benzaldehydes are reversibly ligated together to form benzoin. The carbonyl carbon proton is shuttled to the newly formed hydroxyl of the aldehyde acceptor. B) The proposed formose reaction as catalyzed by BAL in which two consecutive reactions are catalyzed to reversibly convert formaldehyde into dihydroxyacetone. Substrate versus Velocity curve of the BAL catalyzed (C) benzoin reaction and (D) formose reaction**.** The assay was carried out as described in Materials and Methods. The steady state portion of the monitored reaction was used to determine the rate observed (k_{obs} s^{-1}) at each substrate concentration. For the benzoin reaction, our limit of detection was calculated to be 0.01 $M^{-1}s^{-1}$.

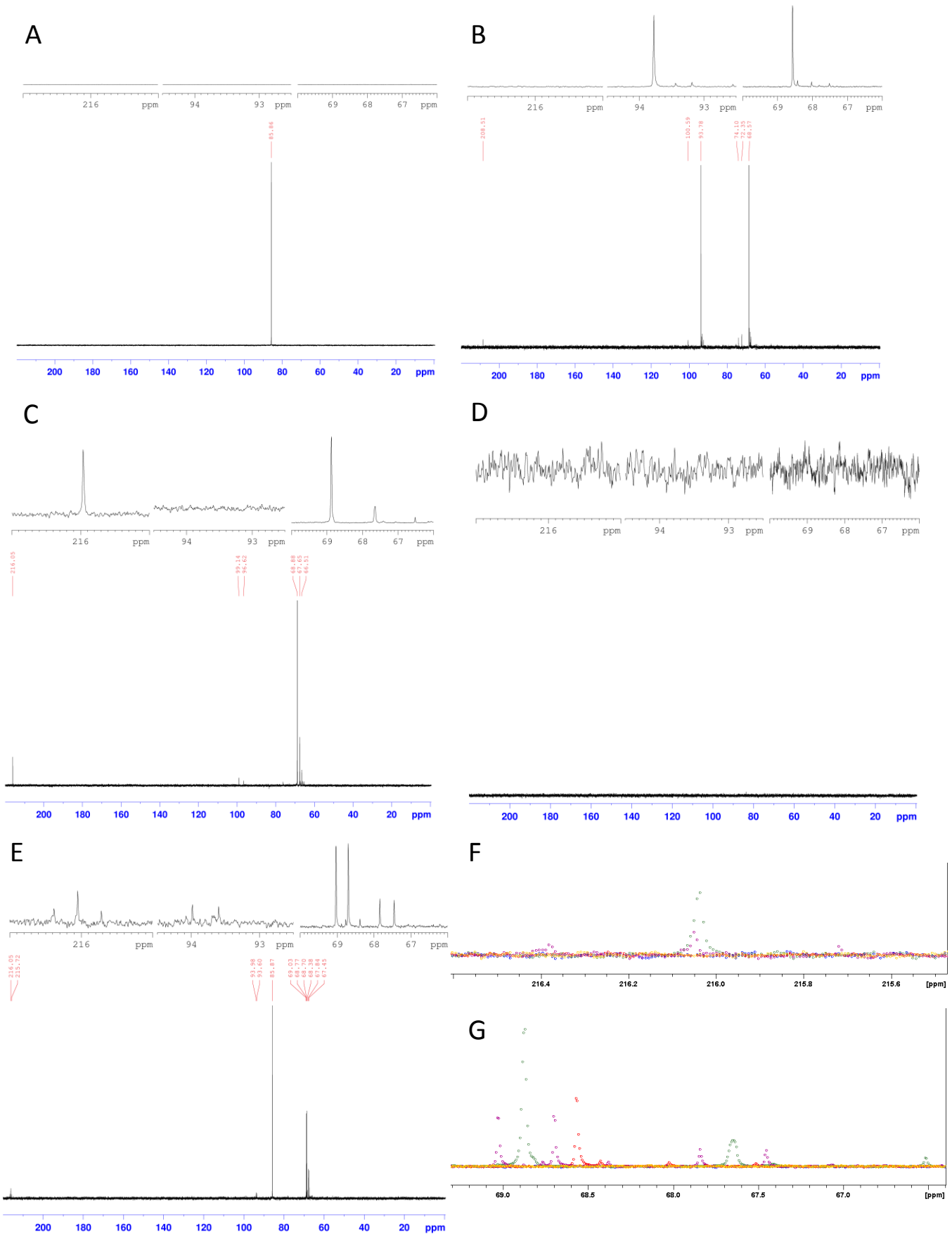

A
B
C
D
E
F
G
216
ppm
94
93
69
68
67
200
180
160
140
120
100
80
60
40
20
216.4
216.2
216.0
215.8
215.6
[ppm]
69.0
68.5
68.0
67.5
67.0

Figure-S2. NMR Analysis of the BAL catalyzed Formose Reaction. A) ^{13}C Formaldehyde in the absence of BAL. B) ^{12}C Glycolaldehyde scanned for natural abundance 13C, in the absence of BAL. C) ^{12}C Dihydroxyacetone scanned for natural abundance ^{13}C, in the absence of BAL. D) BAL in the absence of ^{13}C formaldehyde. E) Reaction mix with BAL (100 µM) and ^{13}C Formladehyde (10 mM) with 0.1 mM TPP and 2 mM $MgSO_4$ in pH 8.0 phosphate buffer. F) Zoom into the low field region 214-216 ppm. The reaction (purple) shows the expected J-coupling for middle carbon of 1,2,3-^{13}C-dihydroxyacetone (triplet, 1:2:1) around the natural abundance ^{13}C dihdyroxyacetone standard (green). G) Zoom into the low field region 66-70 ppm. The reaction (purple) shows the expected J-coupling for end carbons of 1,2,3-^{13}C -dihydroxyacetone (doublet, 1:1) around the natural abundance ^{13}C dihdyroxyacetone standard (green). Small peaks are also observed split around the natural abundance ^{13}C glycolaldehyde standard (red).

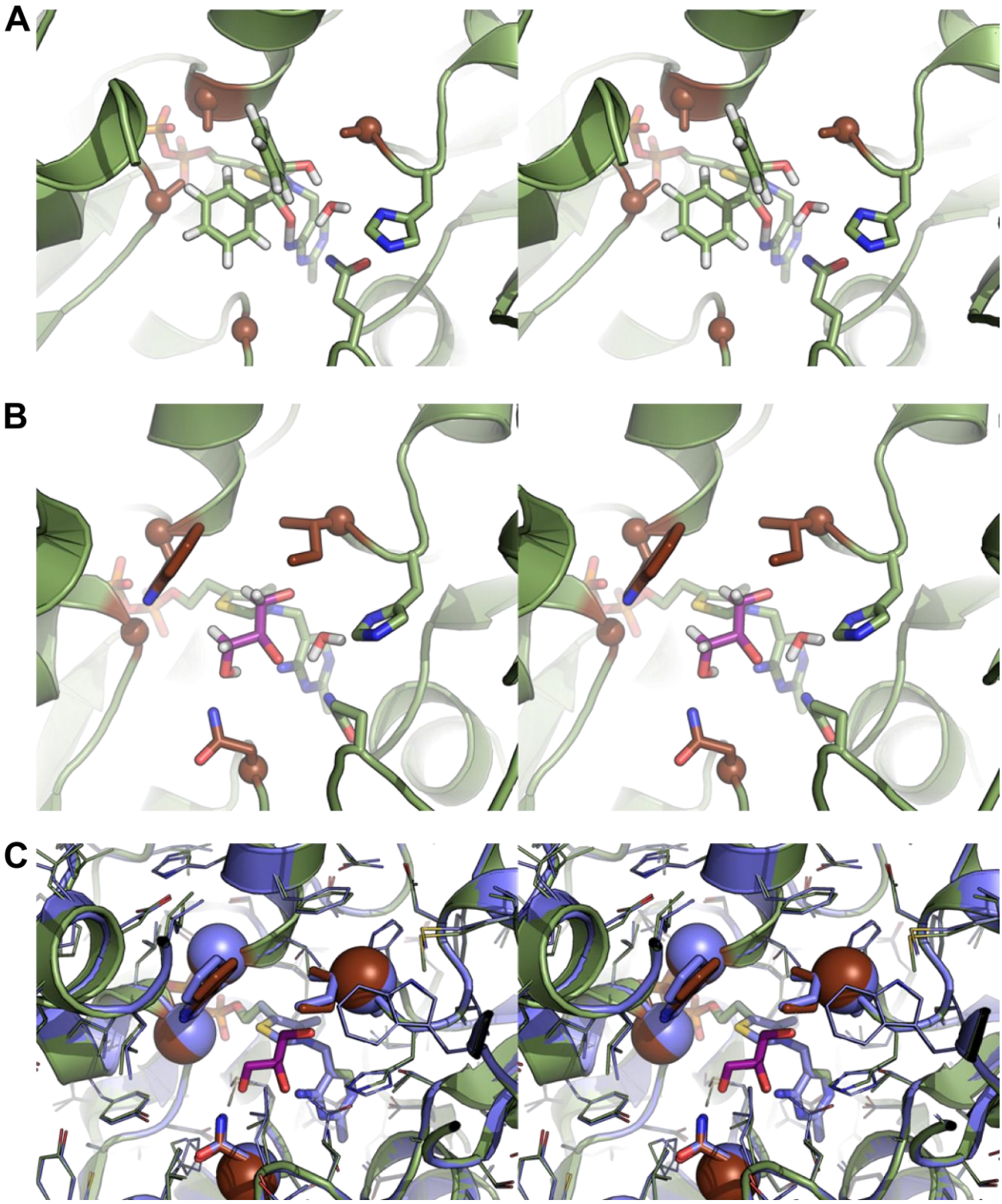
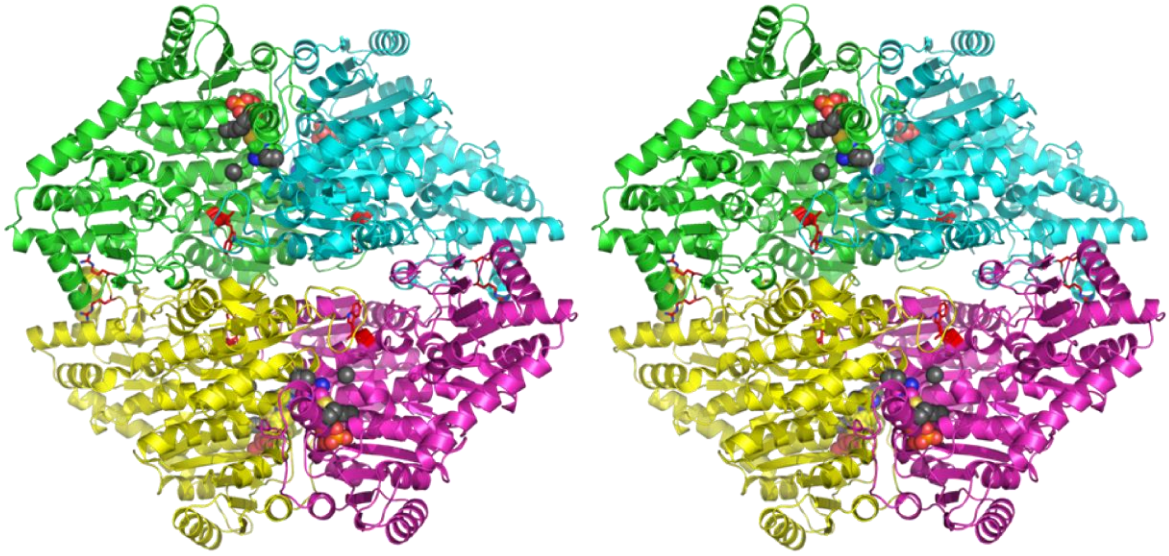


Figure-S3. Stereo representations of Figure 2.

A



B

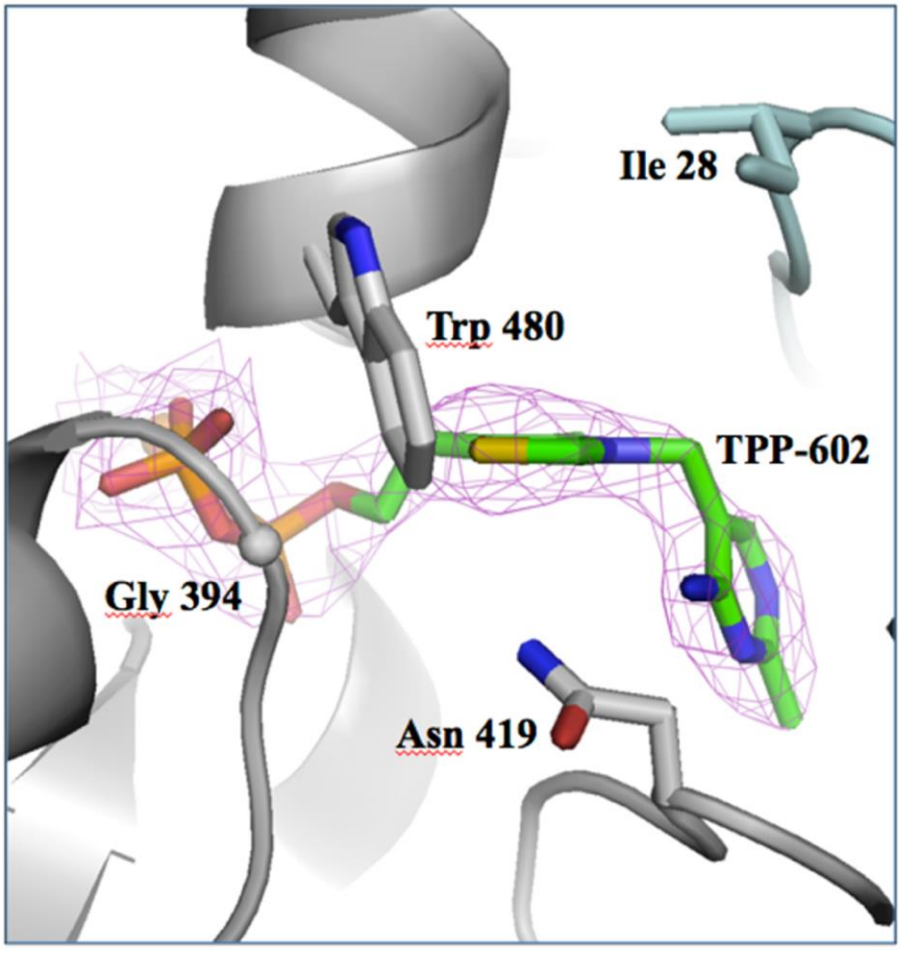


Figure-S4. FLS structure modifications. A) Overview image of the all mutations observed in BAL (PDB-ID 2AG0) in stereoview. The image depicts the relative locations for the 8 mutations that were engineered in the native homotetramer. Each subunit is presented as a cartoon with separate colors (green, cyan, pink, yellow) with TPP cofactor in spheres with carbons colored to match the chain. The four designed mutations are highlighted in gray spheres. The three mutations discovered through computationally guided site directed mutagenesis or error prone PCR are highlighted in Red and depicted in sticks. B) Close up view of the FLS structure around the TPP cofactor showing four of the seven mutated residues, A394G, G419N and A480W from chain A (colored grey) and A28I from chain B (colored in light blue) And a portion of a 2fo-fc omit map (magenta) covering the TPP cofactor.

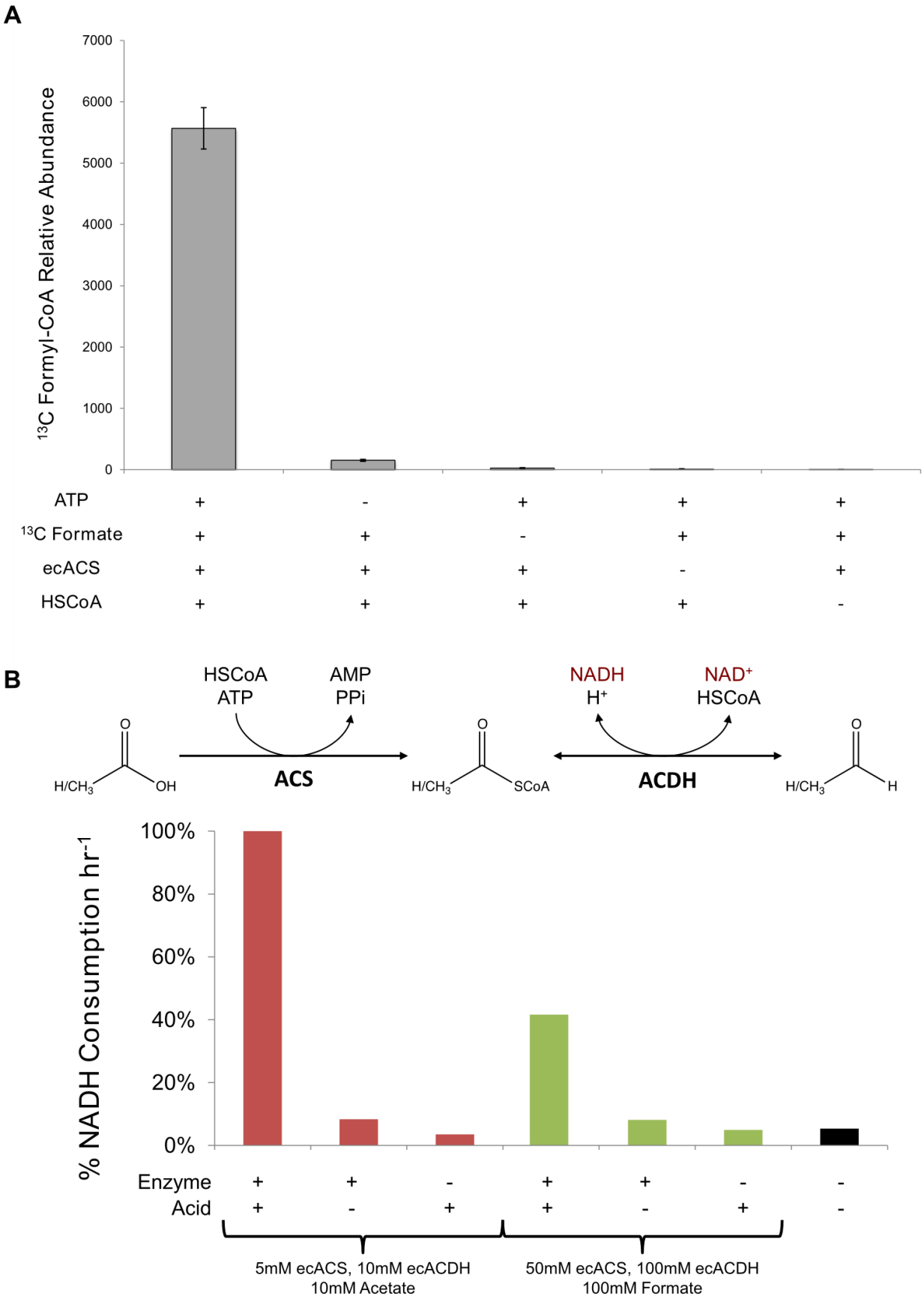

A
7000
6000
5000
4000
3000
2000
1000
0
13C Formyl-CoA Relative Abundance
ATP + - + + +
13C Formate + + - + +
ecACS + + + - +
HSCoA + + + + -
B
HSCoA
ATP
AMP
PPi
NADH
H+
NAD+
HSCoA
ACS
ACDH
H/CH3
OH
SCoA
H
100%
80%
60%
40%
20%
0%
% NADH Consumption hr-1
Enzyme + + - + + - -
Acid + - + + - + -
5mM ecACS, 10mM ecACDH
10mM Acetate
50mM ecACS, 100mM ecACDH
100mM Formate

Figure-S5. Confirmation of ecACS and ecACDH activity A**)** Confirmation of ecACS activity via formyl-CoA determination. Relative abundance of ^{13}C labeled formyl-CoA via LC-MS/MS was measured to confirm its production from ^{13}C formate by ecACS with HSCoA and ATP present. All components were required for significant ^{13}C formyl-CoA production. B) NADH oxidation by an ACS/ACDH coupled assay with formate and acetate as substrates**.** NADH consumption was largest with acetate as a substrate. Achieving over 40% of the NADH consumption with formate as a substrate required 10 times as much of each enzyme.

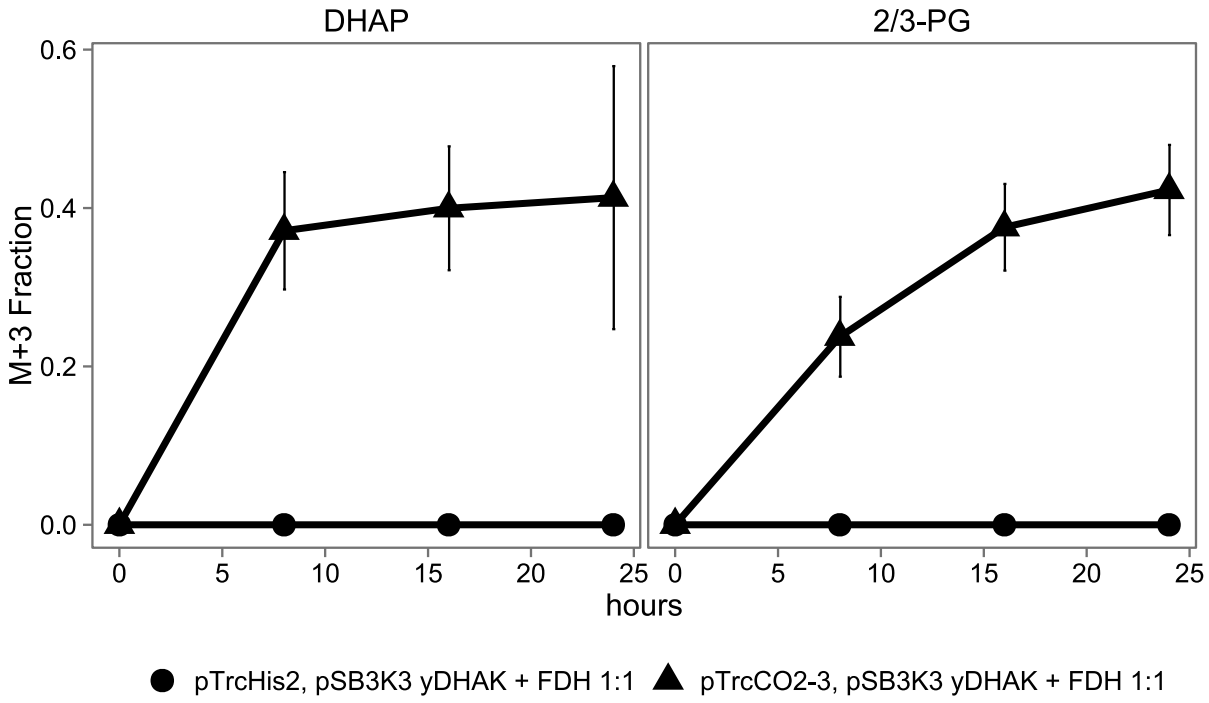


Figure-S6. ^{13}C intracellular metabolite pool labeling from clarified cell lysate reactions. Fraction of each metabolite pool, 2/3-PG and DHAP from clarified cell lysate assay, that is labelled M+3 after natural abundance contributions are removed. Lysates with the pathway genes, ecACS, lmACDH, FLS and yDHAK (pTrcCO2-3-pSB3K3 DHAK), or in the absence of the key formate assimilation enzymes with only yDHAK (pTrcHis2-pSB3K3 DHAK), after incubation with ^{13}C formate for 24 hours. Commercial FDH was added to balance the NADH oxidation rate of the test extracts in a 1:1 ratio based on measurements described in the Materials and Methods. Error bars represent the standard error of measurements for three biological replicates.

Table S1. Specific activities of candidate ACDH enzymes with both formaldehyde and acetaldehyde as substrates.

		Specific Activity (M^{-1} s^{-1})	
Enzyme Label	**UniProt Accession #**	**Formaldehyde**	**Acetaldehyde**
E. coli AdhE *Fe coordinating residues in alcohol dehydrogenase half mutated to alanine	H5NNB6	0.000	0.000
cbACDH	Q9X681	0.001	0.016
lmACDH	C8JVT0	5.262	40.599
psACDH	Q52060	0.027	0.049
E. coli ACDH	P77580	0.050	0.097

Table-S2. Formolase crystal structure data collection and refinement statistics.

Data set ID	DesI	FLS
Wavelength (Å)	1.0	1.5418
Data collection		
Space group	$P2_1$	$P4_32_12$
Cell parameters		
a (Å)	99.88	144.74
b (Å)	136.54	144.74
c (Å)	167.06	269.55
α (°)	90	90
β (°)	95.36	90
γ (°)	90	90
Resolution (Å)	166 - 3.0 (3.2 - 3.0)	45.77 - 2.88 (3.05 - 2.88)
Unique Reflections	89485	64091
Redundancy*	3.5 (3.3)	5.8 (5.2)
Completeness (%)*	91(86)	97.9 (95.1)
$I/\sigma I$*	10.1 (1.97)	9.9
(2.11)		
Rmerge[a] (%)*	9.7 (73.8)	13.2 (84.5)
B(iso)(Å^2)	31.46	47.85
Refinement		
Protein atoms	33229	16674
Metal ions	8 Mg^{++}	4 Mg^{++}
Cofactor	8 TPP	4 TPP
Solvent molecules		
R - factor [b] (%)*	20.52 (31.3)	17.15 (29.4)
R – free [b] (%)*	25.62 (37.4)	20.28 (31.4)
Rmsd		
Bond length (Å)	0.0132	
0.0114		
Bond angles (°)	1.713	1.479
Ramachandran (%)		
Core region	93.25	95.99
Allowed region	6.07	3.79
Outliers	0.67	0.22

* Highest resolution shell values in parentheses.

[a] $R_{merge} = \Sigma|I_{hi} - \langle I_h \rangle|/\Sigma I_h$, where I_{hi} is the i^{th} measurement of reflection h, and $\langle I_h \rangle$ is the average measured intensity of reflection h.

[b] R-factor/R-free = $\Sigma_h|F_{h(o)} - F_{h(c)}|/\Sigma_h|F_{h(o)}|$. R-free was calculated with 5% of the data excluded from refinement.

Table-S3. 3D Coordinates for the DHA intermediate used for molecular modeling in standard PDB format

```
HETATM    1  C   LG1  1109      38.923 -24.194   9.921  1.00 49.92           C
HETATM    2  C   LG1  1109      37.563 -23.987   9.205  1.00 49.86           C
HETATM    3  C   LG1  1109      38.650 -24.773  11.333  1.00 49.86           C
HETATM    4  O   LG1  1109      38.481 -20.937  10.144  1.00 49.38           O
HETATM    5  O   LG1  1109      40.037 -23.131  10.258  1.00 49.38           O
HETATM    6  O   LG1  1109      37.640 -22.781   8.517  1.00 49.38           O
HETATM    7  O   LG1  1109      39.890 -24.895  12.040  1.00 49.38           O
HETATM    8  H   LG1  1109      38.037 -20.097  10.440  1.00 50.47           H
HETATM    9  H   LG1  1109      39.011 -21.296  10.917  1.00 50.47           H
HETATM   10  H   LG1  1109      41.058 -23.258  10.202  1.00 50.47           H
HETATM   11  H   LG1  1109      37.307 -24.802   8.519  1.00 50.13           H
HETATM   12  H   LG1  1109      36.751 -23.889   9.931  1.00 50.13           H
HETATM   13  H   LG1  1109      37.961 -22.076   9.139  1.00 50.47           H
HETATM   14  H   LG1  1109      38.217 -25.778  11.289  1.00 50.13           H
HETATM   15  H   LG1  1109      37.984 -24.135  11.924  1.00 50.13           H
HETATM   16  H   LG1  1109      39.985 -24.076  12.570  1.00 50.47           H
HETATM   17  N   LG2  1110      42.607 -23.654  10.062  1.00 49.66           N
HETATM   18  N   LG2  1110      44.538 -23.018   8.942  1.00 49.60           N
HETATM   19  N   LG2  1110      44.495 -22.160   6.738  1.00 49.60           N
HETATM   20  N   LG2  1110      40.450 -24.309   7.500  1.00 49.66           N
HETATM   21  C   LG2  1110      46.679 -22.380   7.907  1.00 49.86           C
HETATM   22  C   LG2  1110      41.015 -24.711   5.251  1.00 50.01           C
HETATM   23  C   LG2  1110      40.891 -22.906   7.638  1.00 49.95           C
HETATM   24  C   LG2  1110      43.147 -22.268   6.655  1.00 50.01           C
HETATM   25  C   LG2  1110      42.411 -22.771   7.733  1.00 50.01           C
HETATM   26  C   LG2  1110      43.194 -23.158   8.951  1.00 50.01           C
HETATM   27  C   LG2  1110      45.181 -22.521   7.858  1.00 50.01           C
HETATM   28  C   LG2  1110      39.713 -24.867   8.638  1.00 49.95           C
HETATM   29  C   LG2  1110      40.653 -25.151   6.452  1.00 50.01           C
HETATM   30  C   LG2  1110      40.441 -26.448   6.753  1.00 50.01           C
HETATM   31  C   LG2  1110      40.835 -27.498   5.980  1.00 49.95           C
HETATM   32  C   LG2  1110      40.647 -28.978   6.285  1.00 49.95           C
HETATM   33  O   LG2  1110      40.878 -34.931   4.724  1.00 49.47           O
HETATM   34  O   LG2  1110      41.704 -31.891   3.844  1.00 49.37           O1-
HETATM   35  O   LG2  1110      39.529 -33.854   6.795  1.00 49.37           O
HETATM   36  O   LG2  1110      43.263 -31.023   5.662  1.00 49.37           O
HETATM   37  O   LG2  1110      38.625 -31.499   6.991  1.00 49.37           O1-
HETATM   38  O   LG2  1110      40.995 -31.992   6.303  1.00 49.47           O
HETATM   39  O   LG2  1110      39.127 -32.517   4.679  1.00 49.47           O
HETATM   40  O   LG2  1110      41.052 -29.770   5.153  1.00 49.47           O
HETATM   41  P   LG2  1110      41.834 -31.204   5.187  1.00 51.63           P
HETATM   42  P   LG2  1110      39.454 -32.476   6.170  1.00 51.63           P
HETATM   43 Mg2p LG2  1110      40.194 -33.381   3.206  1.00 52.13          Mg
HETATM   44  S   LG2  1110      39.663 -26.577   8.211  1.00 49.97           S
HETATM   45  H   LG2  1110      41.653 -35.059   5.278  1.00 50.56           H
HETATM   46  H   LG2  1110      40.546 -35.784   4.436  1.00 50.56           H
CONECT    1    2    3    5   28
CONECT    2    1    6   11   12
CONECT    3    1    7   14   15
CONECT    4    8    9
CONECT    5    1   10
CONECT    6    2   13
CONECT    7    3   16
CONECT    8    4
CONECT    9    4
CONECT   10    5
CONECT   11    2
CONECT   12    2
CONECT   13    6
CONECT   14    3
```

```
CONECT   15    3
CONECT   16    7
CONECT   17   26
CONECT   18   26
CONECT   18   26
CONECT   18   26
CONECT   18   26
CONECT   18   27
CONECT   18   27
CONECT   18   27
CONECT   18   27
CONECT   19   24
CONECT   19   24
CONECT   19   24
CONECT   19   24
CONECT   19   27
CONECT   19   27
CONECT   19   27
CONECT   19   27
CONECT   20   23
CONECT   20   28
CONECT   20   28
CONECT   20   28
CONECT   20   28
CONECT   20   29
CONECT   20   29
CONECT   20   29
CONECT   20   29
CONECT   21   27
CONECT   22   29
CONECT   23   20   25
CONECT   24   19
CONECT   24   19
CONECT   24   19
CONECT   24   19
CONECT   24   25
CONECT   24   25
CONECT   24   25
CONECT   24   25
CONECT   25   23
CONECT   25   24
CONECT   25   24
CONECT   25   24
CONECT   25   24
CONECT   25   26
CONECT   25   26
CONECT   25   26
CONECT   25   26
CONECT   26   17
CONECT   26   18
CONECT   26   18
CONECT   26   18
CONECT   26   18
CONECT   26   25
CONECT   26   25
CONECT   26   25
CONECT   26   25
CONECT   27   18
CONECT   27   18
CONECT   27   18
CONECT   27   18
CONECT   27   19
CONECT   27   19
```

```
CONECT   27   19
CONECT   27   19
CONECT   27   21
CONECT   28    1
CONECT   28   20
CONECT   28   20
CONECT   28   20
CONECT   28   20
CONECT   28   44
CONECT   28   44
CONECT   28   44
CONECT   28   44
CONECT   29   20
CONECT   29   20
CONECT   29   20
CONECT   29   20
CONECT   29   22
CONECT   29   30
CONECT   29   30
CONECT   29   30
CONECT   30   29
CONECT   30   29
CONECT   30   29
CONECT   30   29
CONECT   30   31
CONECT   30   44
CONECT   30   44
CONECT   30   44
CONECT   31   30   32
CONECT   32   31   40
CONECT   33   45   46
CONECT   34   41
CONECT   35   42
CONECT   36   41
CONECT   36   41
CONECT   37   42
CONECT   38   41   42
CONECT   39   42
CONECT   39   42
CONECT   40   32   41
CONECT   41   34
CONECT   41   36
CONECT   41   36
CONECT   41   38   40
CONECT   42   35   37   38
CONECT   42   39
CONECT   42   39
CONECT   44   28
CONECT   44   28
CONECT   44   28
CONECT   44   28
CONECT   44   30
CONECT   44   30
CONECT   44   30
CONECT   44   30
CONECT   45   33
CONECT   46   33
END
```

SI Materials and Methods

Cloning

Codon-optimized versions of BAL, FLS, lmACDH, cmFDH and yDHAK were obtained from Genescript cloned into pET29b+ vectors (Novagen). ecACDH and ecACS were amplified from genomic material and cloned into the pET29b+ expression vector between *Nco*I and *Bam*HI.

pTrcCO_2-3 was designed for simultaneous expression of the three critical enzymes in the formolase pathway. All three genes (ecACS, Des1 and lmACDH) are under control of a single IPTG inducible *trc* promoter. A Shine-Delgarno sequence precedes each open reading frame.

ACS was further amplified from this pET29b+ based template using the following primers:

Trc-ACS	CGATTAAATAAGGAGGAATAAACCATGGGAAGCCAGATCCACAAGCATACCAT
RBS2-R	CATATGACTTTCCTTTATTAGTGGTGGTGGTGGTGGTGCT

FLS was amplified from pET29b+ FLS using the following primers:

RBS2-FLS-F	CACCACTAATAAAGGAAAGTCATATGGCTATGATTACTGGTGGTGAACTGG
RBS3-R	CATTTGACTTTCCTTTACTAGTGGTGGTGGTGGTGGTGCT

lmACDH was amplified from pET29b+ lmACDHusing the following primers:

RBS3-ADH-F	CACCACTAGTAAAGGAAAGTCAAATGAGCCTGGAAGATAAGGACCTGCGTA
Trc2-term-R	CGTAAGCTTCGAATTCCCATATGGTACCTTATTAGTGGTGGTGGTGGTGGTGCT

The expression vector pTrcHis2C (Invitrogen), was subjected to restriction digest by *Nco*I, *BamHI* and CIP, then gel purified. All PCR products were gel purified (Qiagen). All three gene fragments were assembled into the linearized vector in one step via the method described by Gibson et al.(1). Ten ng of backbone DNA was used per reaction, along with equi-molar amounts of each insert. NEB 5-alpha cells were transformed with 1 µL of the Gibson product by the heat-shock method. Transformants were selected on LB/Agar carbenicillin plates and sequenced for the correct assemblage of inserts.

An alternate system of expressing ecACS, ecACDH, yDHAK and Des1 was also constructed. Two biscistronic operons were synthesized by DNA2.0 and subcloned by digestion with *EcoRI* and *PstI* (NEB) and ligation with T4 DNA ligase (Promega) into Biobrick vectors to give pSB1A3 Des1-yDHAK (constitutive promoter) and pSB3K3 ecACS_ecACDH (lac inducible promoter)(2).

yDHAK and cmFDH were combined in additional constructs along with formate transporter, *focA* and putative formate transporter, *focB*. Genes for yDHAK and cmFDH were amplified from pET29b+ yDHAK and pET29b+ cmFDH.

yDHAK XbaI Fwd-new	GCTGGTCTAGAAAAAATGAGCGCTAA
yDHAK BBSuffix Rev-new	GAAAGTATAGGAACTTCACTTCATTTTCTGC

cmFDH29b_XbaIadd_Fwd	GCTGGTCTAGATGAAAATCGTCCTGGTTCTGTATG
cmFDH29b_BBSuffixadd_Rev	AGTTGACTGCAGCGGCCGCTACTAGTATCATTAGGAACCGCCCTTTTTATCGT

focA and *focB* were amplified from *E. coli* BL21 genomic DNA.

F_FocA	gtttcttcgaattcgcggccgcttctagaggtgaaagctgacaacccttttgatcttttac
R_FocA	gtttcttcctgcagcggccgctactagtattaatggtggtcgttttcacgcagg
F_FocB	gtttcttcgaattcgcggccgcttctagatgagaaacaaactctctttcgacttgc
R_FocB	gtttcttcctgcagcggccgctactagtatcagggttcctgacgtaaataaatagcc

These PCR products had the necessary cut sites to function in the Biobrick system and were cloned using that associated method into vector backbone pSB3K3 with high constitutive expression cassette (BBa_K314100) or low constitutive expression cassette (BBa_K314101) in various combinations(2–4).

pSB3K3 J23100 DHAK and a set of four constructs, pSB4C5 J23100/114 cmFDH J23114 yDHAK J23114 FocA/B (pSB4C5 FDF3-6), were created.

Strain Manipulations

E. coli K-12 MG1655 with pSB1A3 FLS-DHAK and pSB3K3 ACS-ACDH, were serial passaged daily with a minimal medium containing formate and glycerol. Over 85 passages, formate levels were increased from 0 mM to 60 mM and glycerol levels were decreased from 40 mM to 5 mM. Cultures were grown at 37 ˚C with 225 rpm shaking. Resulting strain was cured of plasmids by plating on antibiotic-free LB agar until replica plating with both kanamycin (Kan) and ampicillin confirmed sensitivity. The resulting strain was named ALA2.

A Keio collection knockout strain for *fdoG,* which encodes FDH-O, CGSC 10790, was obtained from the Coli Genome Stock Center(5). The Kan-marked mutation was transferred to ALA2 via P1 phage transduction to give ALA2.1(6). The Kan marker cassette was removed as described by Datsenko and Wanner(7).

Enzyme Engineering

Computational Design

The transition state for dihydroxyacetone was built into the crystal structure of benzaldehyde lyase (2AG0) based on the geometric orientation observed for the related reaction intermediate observed in the 3FSJ PDB using ChemDraw 3D (8, 9). This model was used as input for RosettaDesign and Foldit-based optimization using previously described methods (10, 11). Briefly, Rosetta was used to evaluate the overall energy of the system as well as the interface energy between the protein and intermediate. Either Monte Carlo combinatorial optimization of amino acid identity and conformation was carried out using RosettaDesign, or substitutions were manually introduced using FoldIt. During design, neighboring side chains were allowed to change conformations and the overall protein backbone and the protein-ligand rigid body interface were subjected to gradient based minimization. Selected sequences were experimentally characterized as described in the text. At each round, the model with the lowest overall system energy was selected for analysis of the interactions with the reaction intermediate, and as starting points for the RosettaDesign and FoldIt calculations.

Design Cycle Summary Des1

All designs were conducted using the Foldit interface to the Rosetta Molecular Modeling Suite. The input ligand of the DHA intermediate and TPP cofactor, positioned within the active site based on the crystal structure cofactor as described in Materials and Methods, was fixed in space during modeling. All designs were produced and purified as described in the Materials and Methods. The specific activity of all mutants were determined by comparing specific activity at 20 mM formaldehyde under the conditions described in Kinetic Constant determination in Materials and Methods.

Round #	General Goal/Target	Mutations	Relative Activity to Input (1 = equivlent)	Input Scaffold	Mutations to Keep	Design Name
1	Reduce pocket size	A394R	no activity detected	BAL		-
2	Reduce pocket size	A394G+A480W	9.3	BAL	*	-
3	Introduce Hydrogen Bonding	A394G+A480W+G419N+A28S	13.0	BAL	*	Des0
4	Explore active site	L112F	0.74	Des0		
		L112Y	no activity detected			

	point mutations predicted to be favorable or neutral	L112W	no activity detected			
		L112K	0.16			
		L112H	0.34			
		L112Q	0.23			
		L112E	0.05			
		L112M	no activity detected			
		F484Y	no activity detected			
		F484W	0.35			
		F484K	no activity detected			
		F484R	0.09			
		W163R	0.08			
		S28L	no activity detected			
		S28I	1.95		*	Des1
		S28M	no activity detected			
		S28F	0.15			
		S28Y	0.21			
		S28W	1.12			
		S28R	0.21			
		S28K	0.04			
		S28D	0.04			
		S28E	0.10			
		S28H	no activity detected			
		S28N	1.25			
		S28C	0.03			
		N419Q	0.19			
		N419D	0.01			
		N419E	0.31			
		N419K	0.19			
		N419R	no activity detected			
		N419M	0.13			
		N419I	no activity detected			
		N419L	0.16			
		N419F	0.15			
		N419W	0.09			
		N419Y	0.09			
		N419H	no activity detected			
		Y397W	0.06			
		Y397M	0.52			
		Y397R	0.12			
		Y397E	no activity detected			

Mutations

Error-prone PCR library

The Agilent Genemorph II Kit was used according the manufacturer's protocol to introduce on average two mutations per kilobase into the Des1 gene. After amplification, the mutant library was gel purified, digested for 1 hour at 37˚C with *Nde*I and *Xho*I (NEB), desalted using a Qiagen PCR purification kit, ligated into a *Nde*I and *Xho*I digested pET29b+ vector, and electroporated into *E. coli* DH5α. Mutation frequency was verified by sequencing 24 isolated clones.

Computationally directed Kunkel library construction

Using the Rosetta Molecular Modeling Suite, every point mutation within 15 Å of the Des1 enzyme active site was systematically mutated into each of the twenty amino acids, which equated to roughly 2000 unique mutants. Two thousand instances of EnzdesFixBB.linuxiccrelease (version 40019M) each systematically specifying a single position to be mutated to one of the 20 amino acids were run using Des1 as the protein structure input with the following flags:

```
-run:version
-correct
-corrections::score:no_his_his_pairE
#Add ch_bond_bb_bb 0.5 to weights file in database
-corrections::score:hbond_His_Phil_fix
# Add fa_cust_pair_dist 1.0 to weights file in database
-run::preserve_header
-extra_res_fa LG.params #The Rosetta parameterized ligand, coordinates provided in S12
-enzdes::detect_design_interface
-enzdes::cut1 0.0
-enzdes::cut2 0.0
-enzdes::cut3 20.0
-enzdes::cut4 30.0
-enzdes::cst_design
-enzdes::cst_min
-enzdes::bb_min
-enzdes::bb_min_allowed_dev 0.05
-enzdes::chi_min
-enzdes::design_min_cycles 1
-enzdes::cstfile S28I.enzdes.cst #distance csts from crystal structure between TPP and protein to ensure it maintains
original position
-mute core.io.database
-score:weights enzdes.wts
-packing::use_input_sc
-packing::flip_HNQ
-linmem_ig 10
-docking::ligand::old_estat
#Add hack_elec 0.25 to weights file
-enzdes::lig_packer_weight 1.8
-nstruct 10
-nblist_autoupdate
```

After generating the point mutations, all residues within 20 Å of the active site were repacked and minimized in the context of the modeled formolase transition state. Mutants with Rosetta energy calculated to be greater than the un-mutated protein were discarded (~50% of all mutants). The remaining mutants were then sorted by energy and a total of 380 mutants (up to five at each position) were selected for experimental characterization. Each point mutation in the following list was generated using standard Kunkel mutagenesis techniques as previously reported(12).

Described oligos for the following mutations were ordered (residue number and encoded amino acid): 25A, 25N, 25S, 25T, 25V, 26A, 26F, 26S, 26T, 26V, 28E, 28H, 28R, 28T, 28Y, 30A,

30D, 30S, 30T, 30V, 31A, 31L, 31S, 31T, 31V, 32A, 32D, 32M, 32R, 32V, 33A, 33L, 33Q, 33S, 33V, 34A, 34H, 34T, 34V, 34Y, 35I, 35K, 35L, 35T, 35V, 46A, 46E, 46S, 46T, 46V, 48D, 48I, 48L, 48S, 48T, 49A, 49D, 49Q, 49S, 49T, 50A, 50H, 50N, 50Q, 50S, 71A, 71S, 71V, 72A, 72H, 72N, 72S, 72T, 73A, 73D, 73N, 73S, 74S, 78A, 78M, 78S, 78T, 78Y, 79A, 79D, 79M, 79N, 79S, 80A, 80D, 80S, 80T, 80Y, 81S, 83D, 83S, 86S, 87D, 87K, 87L, 87M, 87Q, 90A, 90E, 90K, 90M, 90T, 99A, 99S, 101D, 101H, 101T, 101W, 101Y, 103K, 103N, 103Q, 103R, 103S, 110D, 110E, 110M, 110Q, 110S, 111D, 111K, 111Q, 111R, 111S, 112A, 112M, 112R, 112S, 112T, 114H, 114K, 114N, 114S, 114T, 115M, 115N, 115S, 116A, 116D, 116L, 116S, 116V, 161E, 161N, 161S, 161T, 162E, 162M, 163A, 163H, 163L, 163M, 163Q, 164E, 164H, 164K, 164R, 164S, 166A, 166D, 166M, 166Q, 166S, 167A, 167R, 167S, 167T, 167V, 282A, 282D, 282S, 282T, 282Y, 283A, 283H, 283Q, 283S, 283V, 391S, 391T, 392A, 395A, 395I, 395K, 395Q, 395R, 396A, 396S, 397A, 397F, 397M, 397R, 397S, 398A, 398D, 398H, 398N, 398Y, 399A, 399F, 399K, 399T, 399V, 401A, 401N, 401T, 414A, 414D, 414N, 414S, 414T, 415D, 415E, 415L, 415N, 415S, 417A, 417D, 417I, 417L, 417S, 418A, 418E, 418H, 418I, 418T, 419A, 419D, 419L, 419S, 419T, 420A, 420D, 420N, 421A, 421E, 421N, 421R, 421S, 423A, 423L, 423N, 423S, 423T, 446A, 446I, 446S, 448A, 448S, 450A, 451A, 451I, 451N, 451S, 451T, 453A, 453D, 453F, 453S, 453V, 477A, 477T, 478A, 478F, 478H, 478Q, 478Y, 480K, 480Q, 480R, 480S, 480T, 481A, 481N, 481S, 482A, 482M, 482Q, 482R, 482T, 483A, 483I, 483M, 483S, 483V, 484A, 484D, 484E, 484S, 484Y, 485A, 485N, 485T, 486H, 486I, 486K, 486L, 486V, 487H, 487M, 487Q, 487S, 487Y, 488E, 488I, 488K, 488R, 488V, 489E, 489H, 489K, 489M, 489Q, 493D, 493H, 493K, 493Q, 493Y, 494D, 494I, 494K, 494M, 494N, 495K, 495N, 495S, 495V, 495Y, 496S, 497A, 497D, 497N, 497S, 497V, 498F, 498I, 498K, 498L, 498M, 546D, 546R, 546S, 546T, 546Y, 549A, 549K, 549L, 549N, 549V, 550A, 551A, 551D, 551E, 551I, 551Q, 552A, 552D, 552L, 552M, 552N, 553A, 553D, 553K, 553L, 553N, 554I, 554K, 554M, 554Q, 554S, 555H, 555N, 555Q, 555S, 555T

Structure Determination

Purified formolase enzymes were exchanged into 25 mM Tris-HCl pH 7.5 and 150 mM NaCl, concentrated to approximately 10 mg/mL, flash frozen in 50 µL aliquots and stored at -80° C. Initial crystallization conditions were obtained by screening against commercially available sparse matrices using a 'Mosquito' crystallization robot (TTP Inc.) that dispenses 50 nL sitting drops. Potential crystallization hits were then expanded/improved with using the hanging-drop vapor diffusion geometry, with 1 µL protein drops mixed with 1 µL drops of the corresponding reservoir solutions, using NEXTAL crystallization trays (Qiagen, Inc). All crystallization trials were performed at 18° C.

The one and only diffracting Des1 crystal was grown in the presence of 15% PEG3350 in 100 mM succinic Acid, pH 7.0. FLS crystals were grown repeatedly in the presence of 14 to 16% PEG3000 (w/v) in 100 mM Tris-HCl pH 7.0 plus 150 mM $Ca(OAc)_2$. All crystals were transferred to an artificial mother liquor corresponding to the final crystallization reservoir solution augmented with 25% (v/v) ethylene glycol, then suspended in nylon loops and flash frozen by submersion in liquid nitrogen prior to data collection.

All data were collected at 100 K, and processed with the software package HKL2000. The Des1 dataset was collected on beamline 5.0.1 at the Advanced Light Source (ALS) synchrotron facilitate at Lawrence Berkeley National Laboratories (LNBL). Multiple datasets of FLS were collected either in-house on a copper rotating anode x-ray generator equipped with a Rigaku Four^{++} image plate detector, or at beamline 5.0.1 at the ALS; several were used at different stages of refinement and model building. All data was processed using the HKL2000 program suite(13). Only the final data and refinement statistics, corresponding to the dataset that gave the highest resolution and final refined model, are reported in Supplementary Table 1. All stages of both structure determinations and refinements were performed using programs in the CCP4 program suite with the 'ccp4i' interface(14). A single polypeptide coordinates was extracted from the RCSB entry 2AGO and used as template for molecular replacement using PHASER(15). Models were refined using Refmac5 after each cycle of model rebuilding using the molecular graphic program COOT(16, 17).

Structure data deposition

X-ray crystallographic coordinate data has been deposited in the RCSB under PDB IDs 4QPZ (Des1) and 4QQ8 (FLS).

Protein Purification

Constructs, pET29b+ versions, were transformed into chemical competent E. coli strain BL21 (DE3). A single colony was picked and cultured overnight in 3 mL terrific broth (TB) medium at 37˚C. In the morning, this culture was decanted into 0.5 L of TB medium, incubated and shaken at 37°C until an optical density (OD600) of 0.6-0.8 was achieved. Then, expression of the gene of interest was induced with 1 mM IPTG. After 30 hours, the cells were pelleted by centrifugation, resuspended in PBS, pH 7.5, 1 mM β-mercaptoethanol (βME), 1 mM $MgSO_4$, 0.2 mM thiamine pyrophosphate (TPP) and 10 mM imidazole and lysed by sonication in the presence of 0.1 mg/mL chicken egg lysozyme. Lysates were cleared of cell debris by centrifugation and purified using Co-NTA IMAC resin. 30mL of cleared lysate was applied to 1mL of Talon resin in a gravity fed column. The resin was washed three times with 20 mL PBS pH 7.5, 1 mM βME, 1 mM $MgSO_4$, 0.2 mM TPP and 10 mM imidazole. Eluted was done with 15 mL PBS, pH 7.5, 1 mM βME, 1 mM $MgSO_4$, 0.2 mM TPP and 200 mM imidazole. Eluate was concentrated to approximately 1mL prior to dialysis against PBS pH 7.5, 0.5 mM βME, 1 mM $MgSO_4$ and 0.2 mM TPP.

Enzyme Assays

Unless otherwise noted, assay reagents used were sourced from Sigma.

Kinetic Constant Determination

Kinetic constants for BAL and FLS were measured over a one hour period in 100 mM potassium phosphate buffer pH 8.0, 1.0 mM $MgSO_4$, and 0.1 mM TPP. Enzyme concentrations ranged from 1 to 40 µM, and substrate (either formaldehyde or benzaldehyde) concentrations ranged from 20 mM to 1.3 mM. Formaldehyde to dihydroxyacetone was measured using the coupled enzyme assay described earlier. Benzoin formation was measured by quenching the reaction in acetonitrile every 10 minutes and measuring benzoin through liquid

chromatography. Briefly, after incubation of benzaldehyde in the presence or absence of protein, 5 µL aliquots were taken at various time points and quenched in a solution of 80:20 acetonitrile (ACN): water with 0.1% formic acid. This solution was incubated for 5 minutes and then the precipitated protein was filtered using a Millipore multiscreen solvinert filter plate (Product Number: MSRLN0450). Twenty µL of the filtered quench solution was subsequently injected and analyzed using liquid chromatography. The column used for the chromatography run was a Hypersil Gold C18 (100 mm x 2.1 mm, 1.9 µm particle size, Thermo), product was monitored at an absorbance of 250 nm. The following gradient was performed for the chromatography run at a flow of 500 µL/min: 95:5 Water:ACN (0.1% formic acid) for 30 seconds, followed by a gradient over 4.5 minutes ending at 5:95 Water:ACN (0.1% formic acid), ending by switching back to 95:5 Water:ACN (0.1% formic acid) and letting the column re-equilibrate for 1 min before the next injection. A purchased product standard of benzoin was used to establish elution time and quantitate the observed signal in the experiments. Kinetic constants for each enzyme using either formaldehyde or benzoin as a substrate were calculated using non-linear regression fitting of rate of product production as a function of substrate concentration using the Michaelis-Menten equation, or to a linear equation if no slope was apparent. Benzaldehyde was visibly cloudy above 5mM and therefore a linear slope to concentrations below 5mM was used since the apparent "saturation" is likely due to the limited solubility of the substrate, as opposed to Km.

NMR Analysis

Experiments were done with the University of Washington Chemistry Facilities Staff. The sample (20 µM BAL, 20 mM ^{13}C formaldehyde (Cambridge Isotope Labs), 0.1 mM TPP, 1 mM $MgSO_4$, 50 mM potassium phosphate, pH 8) was incubated for three hours after which the NMR spectra of each sample using standard methods for ^{13}C detection was measured. Samples with and without enzyme and or formaldehyde present were carried out. Samples of 0.5 M ^{12}C glycolaldehyde or dihydroxyacetone under equivalent conditions, but without enzyme or formaldehyde, were used to determine the chemical shifts of the expected products. In this case, natural abundance ^{13}C was measured.

ACS formyl-CoA measurement

EcACS, 40 µM, was added to an assay mixture with 63 mM ^{13}C formate, 12.5 mM ATP, 2.5 mM HSCoA, 1 mM $MgSO_4$, 1 mM DTT, 150 mM potassium phosphate, pH 8 and incubated for 24 hours at room temperature. LC-MS/MS detection of formyl-CoA was performed as in Buescher et al. adapting the given parameters for acetyl-CoA for formyl-CoA, specifically changing the parent ion mass to reflect one -CH_2 unit while retaining the product ion mass, which represents the CoA group and SRM transition settings(18).

ACDH activity measurement

For each ACDH candidate, a mixture of 1.8 µM protein, 10 mMa, 0.5 mM NAD, 0.5 mM HSCoA, 0.5 mM DTT, 10 µM ZnSO4, 1x PBS and 3 mM imidizole were monitored for NADH formation at 340 nm using a Spectramax M5e. NADH concentrations were calculated using the extinction coefficient 6200 $M^{-1}cm^{-1}$ for NADH at 340 nm.

Coupled ACS-ACDH Enzyme Assay

Purified proteins ecACS and ACDH are combined with an assay mix of 1 mM NADH, 0.2 mM HSCoA, 0.5 mM DTT, 2 mM ATP, 2 mM $MgSO_4$, 0.2 mM TPP, 0.1 mg/mL glycerokinase, 250 mM potassium phosphate pH 8.0 and 50 mM formate. Reactions were monitored on a Spectramax M5e at 340 nm. NADH concentrations were calculated using the extinction coefficient 6200 $M^{-1}cm^{-1}$ for NADH at 340 nm.

Coupled FLS Enzyme Assay.

Lysate-containing or purified FLS was combined with an assay mix of 100 mM $NaPO_4$ buffer pH 8.0, 2 mM $MgSO_4$, 50 µg/mL glycerol dehydrogenase, 0.8 mM NADH, 134 mM FALD, 0.1 mM TPP. NADH concentrations were monitored as in the ACDH assay.

Talon-bead based High-throughput screening

Single colonies were picked and incubated overnight in 0.5 mL of LB media with 50 mg/L Kan in a 96-well deep-well plate while shaking at 37˚C on a Heidolph Brinkmann Titramax 1000 at 1200 rpm. Twenty µL were used to inoculate a fresh culture of Terrific Broth (TB) with 50 mg/L Kan, which was then incubated under the same conditions for 3 hours before adding 50 µL of 10 mM IPTG (0.5 mM final). The cultures were then transferred to a shaker at 18˚C and grown for 30 hours. Plates were spun down, and resuspended in 500 µL of lysis buffer (50 mM potassium phosphate, pH 8.0, 2 mM $MgSO_4$, 0.1 mM TPP, 15 mM imidazole, 1mM PMSF, 1.5x Novagen Bug Buster , 2 mg/ml lysozyme, 0.2 mg/ml DNAse). Plates were incubated while shaking for 1 hour at 4˚C, after which they were spun down for 40 minutes at 4000 rpm. In Millipore Multiscreen-HTS DV 1.2 micron hydrophilic low protein binding filter plate, 200 µL lysate was applied to 50 microliters of TALON beads. The plates were spun at 1500 rpm for 2 minutes where the flow-through was collected and discarded. Three rounds of washing in which 200 µL of wash buffer (50 mM potassium phosphate, pH 8.0, 2 mM $MgSO_4$, 0.1 mM TPP, 15 mM imidazole) was applied to the plate, allowed to incubate while shaking for 5 minutes, and then spun to remove the wash media. Finally 125 µL of elution buffer (50 mM potassium phosphate, pH 8.0, 2 mM $MgSO_4$, 0.1 mM TPP, 150 mM imidazole) was added to each well, incubated for 10 minutes, and then the plate spun to collect flow through. Protein content of the flow through was measured using A280 and the previously described coupled enzyme assay used to detect formolase activity.

Lysate-based Robotic high-throughput screening

The computationally directed Kunkel library was transformed into BLR electrocompetent cells (Novagen) and plated to LB-Kan. Single colonies were picked using a QPIX2 Colony Picker (Genetix) into 0.18 ml of LB with 50 mg/L Kan and 10% glycerol into 20 Costar 3795 shallow well plates. Shallow well plates were then incubated overnight at 37 ˚C at 250 rpm on a Kunher Rack Shaker. Using a Biomex FX robot (Beckman Coulter), 20 µL from each well of the shallow well starter plates was used to inoculate 1.0 ml of TB with 50 mg/L Kan in a Costar 3961 deep well plate. The shallow well plates were then stored at -80 ˚C for later hit picking. Inoculated deep well plates were grown at 37 ˚C at 250 rpm for 5 hours. Using the FX robot, 50 µL of 10mM IPTG (0.5 mM final) was added to each well and incubated overnight at 18 ˚C and 250 rpm. Following the overnight induction, deep well plates were centrifuged at 4000 rpm for 5 min and supernatant was poured off. The cell pellet containing plates were then stored at -80 ˚C until screening. Screening was performed using a Biomex FXP robot with SAMI® Workstation EX software (Beckman Coulter). Briefly, cells were lysed by the re-suspension of the cell pellets in 400 µL of lysis buffer (50 mM HEPES, pH 7.4, 0.1% trition X-100, 2 mM $MgCl_2$, 150 mM NaCl, 10 mg/mL Lysozyme, 25 U/mL Benzonase), followed by a 20 minute incubation at room temperature. Plates were then centrifuged by the robot for 10 min at 4000 rpm. 50 µL supernatant from each well was then combined with 20 µL assay buffer (100 mM phosphate buffer pH 8, 2 mM $MgSO_4$, 50 µg/mL glycerol dehydrogenase, 0.8 mM NADH, 134 mM formaldehyde, 0.1 mM TPP) in a Costar 3631 clear bottom plate and absorbance at 340 nm was monitored for 30 min using a Paradigm MultiMode reader (Beckman Coulter). The rate of decrease in absorbance at 340 nm for each well was normalized by internal positive controls on each plate (colonies expressing unmutated FLS), and wells with the highest normalized slopes were sequenced. Unique variants were then purified further for detailed kinetic analysis, as described above.

In vitro crude cell lysate assay

Cell Preparation

ALA2.1 cells were made chemically competent via CCMB and transformed with appropriate constructs. Overnight LB cultures of each strain were inoculated with single colonies from a LB plate containing the necessary antibiotics. The overnight cultures were used to inoculate 50 mL TB with 100 mg/ml carbenicillin and 100 mg/ml Kan final

concentration. Growth at 37˚C, 225 rpm was conducted to mid-log phase (OD600 ~0.6) when cells were induced using 0.5 mM IPTG final concentration and transferred to 18˚C (225 rpm). After 24 hours at 18˚C, samples of 30 µL were taken to test protein expression via SDS-PAGE, then cell pellets were harvested by centrifugation (Thermo Scientific), 5,000 rpm and 20 min at 4°C, flash frozen with liquid nitrogen and stored at -80°C until use.

Extract Preparation

Cell pellets were resuspended in 3 mL protein buffer (10 mM Potassium phosphate, pH 8.0, 1 mM $MgSO_4$, 0.1 mM TPP and 0.025 mM DTT). Cell lysates were prepared by two passages through a French Press (SIM-Aminco, Spectronic Instruments) with mini-cell (1000 psig). Ultracentrifugation (Beckman) was conducted at 4˚C and 45,000 rpm (Beckman MLS 50) for 1 hour. High-speed extracts (2.5 mL each) were desalted with PD-10 columns (GE Healthcare) eluting in 3.5 mL of the protein buffer to get a clarified cell lysate.

NADH oxidation rates of the clarified cell lysates (CCLs) were measured using a variant of the coupled ACS-ACDH activity assay with assay mixture described in the "Assay reaction" section below, with 1 mM NADH replacing the NAD. NAD reduction rates by a 10 mg/mL solution of FDH from *Candida boidinii* were also measured in the same assay mixture with the NAD instead of the NADH and no HSCoA using that same method as the ACDH assay. The volume of FDH solution to add per volume of CCL was determined by calculating the average NADH oxidation rate of the full pathway CCLs, then finding amount of FDH needed to balance it out in a 1:1 ratio. This volume ratio was applied to all CCLs to give CCL + FDH 1:1 mixtures, which served as the final CCL preparations.

Assay reaction

The following components were combined in an assay mixture with the given final concentrations, vortexed to mix and incubated for various lengths of time: 150 mM Potassium phosphate, 0.2 mM TPP, 2 mM $MgSO_4$, 11 mM ATP, 0.5 mM NAD, 0.2 mM HSCoA, 50 mM ^{13}C sodium formate (Cambridge Isotope Labs), 0.1 mM DTT and 50% (v/v) CCL + FDH 1:1 preparations. Every eight hours, additional ATP, ^{13}C formate and DTT were added to boost concentrations by 2.5 mM, 10 mM and 10 µM, respectively. At each time point (eight hour intervals), 150 µL of reaction was flash frozen in liquid nitrogen and stored at -80˚C.

Extraction

The fast centrifugation method from Kleijn et al. was adapted for use with cell extracts(19). Timepoint samples were thawed on ice. Hot buffered ethanol (Decon Labs) was added to timepoint samples to a final concentration of 60% (v/v). Internal standard, glutaric acid, 7.5 nmol, was added immediately following addition of the hot buffered ethanol. Samples were then heated for 1 min at 78°C with occasional mixing, placed on ice and centrifuged at 14,000 rpm for 10 min at 4°C. Supernatant was transferred to new Eppendorf tubes and dried to completion in a Labconco CentriVap with cold trap connected to a Savant Gel pump (GP110). Samples were stored at -80°C until within a day of LC-MS/MS analysis, when they were re-suspended in 100 µL of Millipore water and vortexed to mix.

Mass Spec Analysis

Autosampler vial inserts were loaded with 35 µL of each sample. LC-MS/MS analysis was performed as in Buescher et al.(18). An extra minute of equilibration time was added to the end of each run and the mass spectrometer used was a Thermo

Scientific Quantum Access. Transitions used for SRM are below. Adjustments to the parent and product ion masses were done to monitor the M+x channels for each targeted compound, where x is the number of carbons in the compound backbone. Internal standard curves made with ^{12}C compounds were used for quantification. Isotope correction for natural abundance was done using the Isocor software (20).

Compound	Parent Ion	Product Ion	Tube Lens	Collision Energy
2/3-Phosphoglyceric acid	185.013	79.204	36	134
Dihydroxyacetone phosphate	169.004	97.120	12	46
Glutaric Acid	131.084	87.247	15	146

Growth on formate

Growth experiments were completed using a set of four M9 minimal media with all combinations of -/+ formate (normally 40 mM) and -/+ glycerol (0.1%) (21). Temperatures tested were room temperature (~22˚C) and 18˚C. Cultures were grown in Bioscreen plates (0.210 mL total volume) in the Bioscreen automated growth curve machine (Growth Curves USA) or glass tubes (5 mL total volume). In both cases, cultures were kept oxygenated using the high shaking setting. Growth experiments were carried out for up to 15 days. The cultures tested included ALA2.1 with pTrcCO_2-3 (ecACS, lmACDH, FLS) pathway core and associated controls with a combination of secondary plasmids pSB4C5 FDF3-6 (cmFDH, yDHAK, FocA/B). Room temperature growth on M9 minimal medium agar plates with a range of formate concentrations from 0-50 mM was also tested.

Computation of Biomass Yield and Max-min Driving Force

Biomass yield, in gram cellular dry weight (gCDW) per mole of formate consumed was calculated using flux balance analysis and the core metabolic model of *E. coli* supplemented with pathway enzymes and without considering ATP maintenance(22). The Max-min Driving Force (MDF) is the lowest value of -Δ_rG' in the pathway (*i.e.* the reaction(s) with the smallest chemical driving force), after optimizing reactant concentrations within a physiological range (1µM-10mM for non-cofactors). The co-factor constraints used: [ATP]/[ADP]=10, [ADP]/[AMP]=1, [NADH]/[NAD^+] = 0.1, [NADPH]/[$NADP^+$]=10, [$Ferredoxin_{red}$]/[$Ferredoxin_{ox}$] = 1, [orthophosphate]= 10 mM, [pyrophosphate]=1 mM, [CoA] = 1 mM, [CO2(aq)] = 10 mM (ambient conditions) See ref. [PMID 24586134] for further details.

Coding sequences for Formolase pathway genes

ecACS -

ATGGGAAGCCAGATCCACAAGCATACCATTCCTGCCAACATCGCAGACCGTTGCCTGATAAACCCTC
AGCAGTACGAGGCGATGTATCAACAATCTATTAACGTACCTGATACCTTCTGGGGCGAACAGGGAAA
AATTCTTGACTGGATCAAACCTTACCAGAAGGTGAAAAACACCTCCTTTGCCCCCGGTAATGTGTCCA
TTAAATGGTACGAGGACGGCACGCTGAATCTGGCGGCAAACTGCCTTGACCGCCATCTGCAAGAAA
ACGGCGATCGTACCGCCATCATCTGGGAAGGCGACGACGCCAGCCAGAGCAAACATATCAGCTATA
AAGAGCTGCACCGCGACGTCTGCCGCTTCGCCAATACCCTGCTCGAGCTGGGCATTAAAAAAGGTG
ATGTGGTGGCGATTTATATGCCGATGGTGCCGGAAGCCGCGGTTGCGATGCTGGCCTGCGCCCGC
ATTGGCGCGGTGCATTCGGTGATTTTCGGCGGCTTCTCGCCGGAAGCCGTTGCCGGGCGCATTATT
GATTCCAACTCACGACTGGTGATCACTTCCGACGAAGGTGTGCGTGCCGGGCGCAGTATTCCGCTG
AAGAAAAACGTTGATGACGCGCTGAAAAACCCGAACGTCACCAGCGTAGAGCATGTGGTGGTACTG
AAGCGTACTGGCGGGAAAATTGACTGGCAGGAAGGGCGCGACCTGTGGTGGCACGACCTGGTTGA
GCAAGCGAGCGATCAGCACCAGGCGGAAGAGATGAACGCCGAAGATCCGCTGTTTATTCTCTACAC
CTCCGGTTCTACCGGTAAGCCAAAAGGTGTGCTGCATACTACCGGCGGTTATCTGGTGTACGCGGC
GCTGACCTTTAAATATGTCTTTGATTATCATCCGGGTGATATCTACTGGTGCACCGCCGATGTGGGCT

GGGTGACCGGACACAGTTACTTGCTGTACGGCCCGCTGGCCTGCGGTGCGACCACGCTGATGTTTG
AAGGCGTACCCAACTGGCCGACGCCTGCCCGTATGGCGCAGGTGGTGGACAAGCATCAGGTCAAT
ATTCTCTATACCGCACCCACGGCGATCCGCGCGCTGATGGCGGAAGGCGATAAAGCGATCGAAGGC
ACCGACCGTTCGTCGCTGCGCATTCTCGGTTCCGTGGGCGAGCCAATTAACCCGGAAGCGTGGGA
GTGGTACTGGAAAAAAATCGGCAACGAGAAATGTCCGGTGGTCGATACCTGGTGGCAGACCGAAAC
CGGCGGTTTCATGATCACCCCGCTGCCTGGCGCTACCGAGCTGAAAGCCGGTTCGGCAACACGTCC
GTTCTTCGGCGTGCAACCGGCGCTGGTCGATAACGAAGGTAACCCGCTGGAGGGGGCCACCGAAG
GTAGCCTGGTAATCACCGACTCCTGGCCGGGTCAGGCGCGTACGCTGTTTGGCGATCACGAACGTT
TTGAACAGACCTACTTCTCCACCTTCAAAAATATGTATTTCAGCGGCGACGGCGCGCGTCGCGATGA
AGATGGCTATTACTGGATAACCGGGCGTGTGGACGACGTGCTGAACGTCTCCGGTCACCGTCTGGG
GACGGCAGAGATTGAGTCGGCGCTGGTGGCGCATCCGAAGATTGCCGAAGCCGCCGTAGTAGGTA
TTCCGCACAATATTAAAGGTCAGGCGATCTACGCCTACGTCACGCTTAATCACGGGGAGGAACCGTC
ACCAGAACTGTACGCAGAAGTCCGCAACTGGGTGCGTAAAGAGATTGGCCCGCTGGCGACGCCAG
ACGTGCTGCACTGGACCGACTCCCTGCCTAAAACCCGCTCCGGCAAAATTATGCGCCGTATTCTGC
GCAAAATTGCGGCGGGCGATACCAGCAACCTGGGCGATACCTCGACGCTTGCCGATCCTGGCGTA
GTCGAGAAGCTGCTTGAAGAGAAGCAGGCTATCGCGATGCCATCGGGGGATCCGAATTCGAGCTCC
GTCGACAAGCTTGCGGCCGCA

FLS -
ATGGCTATGATTACTGGTGGTGAACTGGTTGTTCGTACCCTGATTAAAGCTGGCGTAGAACATCTGTT
TGGCCTGCATGGCATTCATATTGACACCATTTTTCAGGCTTGCCTGGACCACGACGTCCCAATCATT
GATACTCGCCACGAAGCGGCGGCAGGCCACGCTGCGGAAGGTTATGCCCGCGCGGGCGCTAAACT
GGGTGTTGCCCTGGTGACCGCTGGCGGTGGCTTTACCAATGCCGTTACGCCGATCGCGAACGCTC
GGACCGATCGCACTCCGGTTCTGTTCCTGACCGGTTCTGGTGCTCTTCGTGATGACGAAACCAACAC
CCTGCAGGCCGGTATTGATCAGGTGGCCATGGCGGCCCCGATCACGAAATGGGCTCATCGTGTTAT
GGCAACTGAACACATCCCGCGTCTGGTTATGCAGGCCATTCGTGCCGCTCTGAGCGCCCCACGTGG
CCCGGTGCTGCTGGATCTGCCATGGGACATCCTGATGAACCAAATCGATGAAGATTCCGTTATCATC
CCAGACCTGGTGCTGTCTGCTCACGGTGCCCATCCAGACCCGGCTGACCTGGACCAGGCTCTGGC
ACTGCTGCGTAAAGCCGAACGCCCAGTTATCGTACTGGGCTCCGAGGCGTCCCGCACCGCACGCAA
GACCGCACTGAGCGCATTCGTAGCGGCGACCGGTGTACCGGTTTTCGCTGACTATGAAGGCCTGTC
CATGCTGAGCGGCCTGCCGGACGCTATGCGTGGCGGCCTGGTGCAGAACCTGTACTCCTTTGCAAA
AGCTGATGCAGCTCCGGACCTGGTACTGATGCTGGGTGCTCGTTTCGGTCTGAACACCGGTCATGG
TTCCGGTCAACTGATCCCGCATTCTGCTCAGGTGATCCAGGTGGATCCAGACGCGTGTGAACTGGG
TCGCCTGCAAGGCATCGCGCTGGGTATCGTGGCTGATGTAGGTGGCACCATTGAAGCGCTGGCTCA
GGCGACCGCACAGGACGCCGCGTGGCCGGACCGCGGCGACTGGTGCGCCAAGGTAACTGACCTG
GCCCAGGAGCGTTACGCTTCCATCGCGGCTAAATCCAGCTCTGAACATGCGCTGCACCCGTTCCAC
GCTTCTCAGGTTATCGCGAAACACGTGGACGCAGGCGTGACCGTCGTTGCGGATGGTGGCCTGACT
TATCTGTGGCTGTCCGAAGTTATGTCTCGTGTCAAACCAGGCGGCTTCCTGTGCCACGGCTATCTGA
ACAGCATGGGTGTAGGCTTCGGTACTGCCCTGGGTGCGCAGGTTGCGGATCTGGAGGCAGGTCGT
CGTACCATCCTGGTGACCGGCGACGGCTCTGTTGGTTATTCCATTGGCGAATTCGACACCCTGGTAC
GCAAACAGCTGCCGCTGATTGTAATTATCATGAACAACCAGTCTTGGGGCTGGACCCTGCACTTTCA
GCAGCTGGCCGTTGGTCCTAACCGTGTCACCGGCACCCGCCTGGAAAATGGTTCCTATCACGGCGT
TGCTGCGGCATTCGGTGCTGATGGTTACCACGTCGACTCTGTCGAGAGCTTCAGCGCCGCTCTGGC
TCAGGCACTGGCACACAACCGCCCGGCATGCATCAACGTTGCTGTGGCCCTGGACCCGATCCCGC
CGGAGGAACTGATCCTGATTGGCATGGACCCGTTTGCGGGCTCCACGGAGAATCTGTATTTCCAATC
CGGCGCG

lmACDH -
ATGAGCCTGGAAGATAAGGACCTGCGTAGCATCCAAGAAGTGCGTAACCTGATCGAATCTGCGAATA
AGGCCCAAAAAGAACTGGCGGCGATGTCACAGCAACAGATTGATACCATCGTGAAAGCGATTGCCG
ACGCAGGCTATGGTGCGCGTGAAAAACTGGCTAAGATGGCGCACGAAGAAACGGGCTTTGGTATTT
GGCAGGATAAAGTTATCAAGAACGTCTTCGCCTCGAAGCATGTCTACAACTACATCAAGGATATGAA
GACCATCGGTATGCTGAAAGAAGACAACGAAAAGAAAGTTATGGAAGTCGCAGTGCCGCTGGGCGT
GGTTGCTGGTCTGATTCCGTCAACCAATCCGACCTCGACGGTGATCTACAAAACGCTGATTTCAATC
AAGGCGGGCAACAGTATCGTGTTTAGCCCGCACCCGAATGCCCTGAAAGCAATTCTGGAAACCGTC
CGCATTATCTCAGAAGCGGCCGAAAAAGCAGGCTGCCCGAAGGGTGCTATTTCGTGTATGACCGTT

CCGACGATCCAAGGCACCGATCAGCTGATGAAACATAAGGACACCGCTGTCATTCTGGCAACGGGC
GGTTCTGCGATGGTGAAAGCAGCTTATAGCTCTGGCACCCCGGCAATTGGTGTGGGTCCGGGCAAC
GGTCCGGCCTTTATTGAACGTAGTGCGAATATCCCGCGTGCGGTTAAACACATCCTGGATTCCAAGA
CCTTCGACAACGGTACGATTTGCGCCAGCGAACAGTCTGTCGTGGTTGAACGTGTCAATAAAGAAGC
TGTGATCGCGGAATTTCGCAAGCAAGGCGCACACTTCCTGAGTGATGCTGAAGCGGTGCAGCTGGG
CAAATTCATTCTGCGTCCGAACGGTAGCATGAATCCGGCGATTGTGGGCAAAAGCGTGCAACATATC
GCAAACCTGGCAGGTCTGACCGTGCCGGCCGATGCACGTGTTCTGATTGCGGAAGAAACGAAAGTT
GGCGCCAAGATCCCGTATAGTCGCGAAAAACTGGCCCCGATTCTGGCATTTTACACCGCTGAAACGT
GGCAGGAAGCATGCGAACTGAGCATGGATATTCTGTACCATGAAGGCGCTGGTCACACCCTGATTAT
CCATAGCGAAGACAAAGAAATTATCCGTGAATTTGCACTGAAAAAGCCGGTTTCTCGCCTGCTGGTC
AACACGCCGGGCGCGCTGGGCGGCATTGGTGCCACCACGAATCTGGTTCCGGCACTGACGCTGGG
CTGTGGTGCTGTCGGCGGTAGTTCCTCATCGGATAATATCGGTCCGGAAAACCTGTTTAATATTCGT
CGCATCGCCACCGGCGTGCTGGAACTGGAAGACATTCGCGAAGGCGGTAGC

yDHAK -
ATGTCCGCTAAATCGTTTGAAGTCACAGATCCAGTCAATTCAAGTCTCAAAGGGTTTGCCCTTGCTAA
CCCCTCCATTACGCTGGTCCCTGAAGAAAAAATTCTCTTCAGAAAGACCGATTCCGACAAGATCGCA
TTAATTTCTGGTGGTGGTAGTGGACATGAACCTACACACGCCGGTTTCATTGGTAAGGGTATGTTGA
GTGGCGCCGTGGTTGGCGAAATTTTTGCATCCCCTTCAACAAAACAGATTTTAAATGCAATCCGTTTA
GTCAATGAAAATGCGTCTGGCGTTTTATTGATTGTGAAGAACTACACAGGTGATGTTTTGCATTTTGG
TCTGTCCGCTGAGAGAGCAAGAGCCTTGGGTATTAACTGCCGCGTTGCTGTCATAGGTGATGATGTT
GCAGTTGGCAGAGAAAAGGGTGGTATGGTTGGTAGAAGAGCATTGGCAGGTACCGTTTTGGTTCAT
AAGATTGTAGGTGCCTTCGCAGAAGAATATTCTAGTAAGTATGGCTTAGACGGTACAGCTAAAGTGG
CTAAAATTATCAACGACAATTTGGTGACCATTGGATCTTCTTTAGACCATTGTAAAGTTCCTGGCAGG
AAATTCGAAAGTGAATTAAACGAAAAACAAATGGAATTGGGTATGGGTATTCATAACGAACCTGGTGT
GAAAGTTTTAGACCCTATTCCTTCTACCGAAGACTTGATCTCCAAGTATATGCTACCAAAACTATTGG
ATCCAAACGATAAGGATAGAGCTTTTGTAAAGTTTGATGAAGATGATGAAGTTGTCTTGTTAGTTAAC
AATCTCGGCGGTGTTTCTAATTTTGTTATTAGTTCTATCACTTCCAAAACTACGGATTTCTTAAAGGAA
AATTACAACATAACCCCGGTTCAAACAATTGCTGGCACATTGATGACCTCCTTCAATGGTAATGGGTT
CAGTATCACATTACTAAACGCCACTAAGGCTACAAAGGCTTTGCAATCTGATTTTGAGGAGATCAAAT
CAGTACTAGACTTGTTGAACGCATTTACGAACGCACCGGGCTGGCCAATTGCAGATTTTGAAAAGAC
TTCTGCCCCATCTGTTAACGATGACTTGTTACATAATGAAGTAACAGCAAAGGCCGTCGGTACCTATG
ACTTTGACAAGTTTGCTGAGTGGATGAAGAGTGGTGCTGAACAAGTTATCAAGAGCGAACCGCACAT
TACGGAACTAGACAATCAAGTTGGTGATGGTGATTGTGGTTACACTTTAGTGGCAGGAGTTAAAGGC
ATCACCGAAAACCTTGACAAGCTGTCGAAGGACTCATTATCTCAGGCGGTTGCCCAAATTTCAGATTT
CATTGAAGGCTCAATGGGAGGTACTTCTGGTGGTTTATATTCTATTCTTTTGTCGGGTTTTTCACACG
GATTAATTCAGGTTTGTAAATCAAAGGATGAACCCGTCACTAAGGAAATTGTGGCTAAGTCACTCGGA
ATTGCATTGGATACTTTATACAAATATACAAAGGCAAGGAAGGGATCATCCACCATGATTGATGCTTT
AGAACCATTCGTTAAAGAATTTACTGCATCTAAGGATTTCAATAAGGCGGTAAAAGCTGCAGAGGAAG
GTGCTAAATCCACTGCTACATTCGAGGCCAAATTTGGCAGAGCTTCGTATGTCGGCGATTCATCTCA
AGTAGAAGATCCTGGTGCAGTAGGCCTATGTGAGTTTTTGAAGGGGGTTCAAAGCGCCTTG
cmFDH-
ATGAAAATCGTCCTGGTTCTGTATGATGCGGGCAAGCACGCGGCGGATGAAGAAAAACTGTACGGC
TGCACCGAAAATAAGCTGGGCATTGCCAACTGGCTGAAAGATCAGGGCCATGAACTGATCACCACG
AGTGATAAGGAAGGTGAAACCTCCGAACTGGACAAACATATTCCGGATGCTGACATTATCATTACCA
CGCCGTTTCACCCGGCGTATATCACGAAGGAACGCCTGGATAAAGCCAAGAATCTGAAATCAGTGGT
TGTCGCAGGCGTTGGTTCGGACCATATCGATCTGGACTACATTAACCAGACCGGCAAAAAGATTAGC
GTCCTGGAAGTGACGGGTAGCAATGTGGTTTCTGTGGCTGAACACGTCGTGATGACCATGCTGGTT
CTGGTCCGTAACTTTGTCCCGGCCCATGAACAAATCATTAATCACGATTGGGAAGTGGCGGCCATCG
CTAAGGATGCGTATGACATTGAAGGCAAAACCATCGCCACGATTGGCGCAGGTCGTATCGGTTACC
GCGTTCTGGAACGTCTGCTGCCGTTCAACCCGAAAGAACTGCTGTATTACGATTATCAGGCCCTGCC
GAAAGAAGCAGAAGAAAAAGTGGGCGCGCGTCGCGTTGAAAATATCGAAGAACTGGTGGCCCAAGC
AGACATTGTGACCGTTAACGCTCCGCTGCATGCGGGCACGAAAGGTCTGATCAACAAGGAACTGCT
GAGTAAGTTCAAGAAGGGTGCATGGCTGGTTAACACCGCCCGCGGTGCAATCTGCGTTGCAGAAGA
TGTCGCAGCTGCGCTGGAATCAGGCCAGCTGCGTGGTTATGGCGGTGACGTCTGGTTCCCGCAACC
GGCACCGAAAGATCATCCGTGGCGTGACATGCGCAACAAATATGGCGCTGGTAATGCGATGACCCC

GCACTACAGCGGCACCACGCTGGATGCTCAGACCCGCTATGCGGAAGGCACGAAAAACATTCTGGA
AAGCTTTTTCACCGGTAAATTCGATTACCGTCCGCAAGACATCATTCTGCTGAATGGCGAATATGTGA
CGAAAGCGTACGGTAAACACGATAAAAAGGGCGGTTCCCTCGAG

1. Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* 37:6984–90.

2. Shetty RP, Endy D, Knight TF (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2:5.

3. MIT (2003) Registry of Standard Biological Parts [database on the internet]. Available at: www.partsregistry.org [Accessed August 20, 2010].

4. Ginkgo Bioworks Standard Assembly of Biobricks. Available at: http://ginkgobioworks.com/support/BioBrick_Assembly_Manual.pdf [Accessed May 30, 2013].

5. Baba T et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006.0008.

6. Thomason LC, Costantino N, Court DL (2007) in *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*.

7. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97:6640–6645.

8. Mosbacher TG, Mueller M, Schulz GE (2005) Structure and mechanism of the ThDP-dependent benzaldehyde lyase from Pseudomonas fluorescens. *FEBS J* 272:6067–76.

9. Brandt GS et al. (2009) Snapshot of a reaction intermediate: analysis of benzoylformate decarboxylase in complex with a benzoylphosphonate inhibitor. *Biochemistry* 48:3247–57.

10. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using Rosetta3. *PLoS One* 6:e19230.

11. Eiben CB et al. (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat Biotechnol* 30:190–2.

12. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* 82:488–92.

13. Otwinowski Z, Minor W (1997) in *Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A*, eds Carter, Jr. CW, Sweet RM (Academic Press, New York, NY), pp 307–326.

14. Winn MD et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–42.

15. McCoy AJ et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.

16. Skubák P, Murshudov GN, Pannu NS (2004) Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr D Biol Crystallogr* 60:2196–201.

17. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.

18. Buescher JM, Moco S, Sauer U, Zamboni N (2010) Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal Chem* 82:4403–12.

19. Kleijn RJ et al. (2010) Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis. *J Biol Chem* 285:1587–96.

20. Millard P, Letisse F, Sokol S, Portais J-C (2012) IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics* 28:1294–6.

21. Maniatis, T., Fritsch, E.F., Sambrook J (1982) *Molecular cloning: a laboratory manual.* (CSHL Press).

22. Bar-Even A, Noor E, Flamholz A, Milo R (2012) Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochim Biophys Acta* 1827:1039–1047.