



HAL
open science

Genetic and biocatalytic basis of formate dependent growth of *Escherichia coli* strains evolved in continuous culture

Valérie Delmas, Nadia Perchat, Oriane Monet, Marion Fouré, Ekatarina Darii, David Roche, Ivan Dubois, Emilie Pateau, Alain Perret, Volker Döring, et al.

► To cite this version:

Valérie Delmas, Nadia Perchat, Oriane Monet, Marion Fouré, Ekatarina Darii, et al.. Genetic and biocatalytic basis of formate dependent growth of *Escherichia coli* strains evolved in continuous culture. *Metabolic Engineering*, 2022, 72, pp.200-214. 10.1016/j.ymben.2022.03.010 . hal-03638763

HAL Id: hal-03638763

<https://hal.science/hal-03638763>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Genetic and biocatalytic basis of formate dependent growth of *Escherichia coli* strains evolved in continuous culture.

Valérie A. Delmas, Nadia Perchat, Oriane Monet, Marion Fouré, Ekatarina Darii, David Roche, Ivan Dubois, Emilie Pateau, Alain Perret, Volker Döring, Madeleine Bouzon*

Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry-Courcouronnes, France

*** Corresponding author**

E-mail address: mbouzon@genoscope.cns.fr

Postal address: CEA Genoscope, 2 rue Gaston Crémieux, 91057 Evry-Courcouronnes, France

Keywords

Escherichia coli, formate assimilation, one carbon metabolism, evolution, continuous culture, mutation analysis, lipoamide dehydrogenase, methylene-H₄F dehydrogenase/cyclohydrolase

Highlights

- *E. coli* was adapted to grow on formate as sole carbon source in continuous culture
- ¹³C-tracing proved that formate is assimilated *via* the reductive glycine pathway
- Point mutations in the genes *folD* and *lpd* encoding the two dehydrogenases of the pathway are essential for growth on formate
- Activity analysis of mutated FolD and Lpd revealed altered inhibition patterns

1

2 **Abstract**

3 The reductive glycine pathway was described as the most energetically favorable
4 synthetic route of aerobic formate assimilation. Here we report the successful
5 implementation of formatotrophy in *Escherichia coli* by means of a stepwise adaptive
6 evolution strategy. Medium swap and turbidostat regimes of continuous culture were
7 applied to force the channeling of carbon flux through the synthetic pathway to
8 pyruvate establishing growth on formate and CO₂ as sole carbon sources. Labeling
9 with ¹³C-formate proved the assimilation of the C1 substrate via the pathway
10 metabolites. Genetic analysis of intermediate isolates revealed a mutational path
11 followed throughout the adaptation process. Mutations were detected affecting the
12 copy number (gene *ftfL*) or the coding sequence (genes *folD* and *lpd*) of genes which
13 specify enzymes implicated in the three steps forming glycine from formate and CO₂,
14 the central metabolite of the synthetic pathway. The mutation R196S present in
15 methylene-tetrahydrofolate dehydrogenase/cyclohydrolase (FolD) abolishes the
16 inhibition of cyclohydrolase activity by the substrate formyl-tetrahydrofolate. The
17 mutation R273H in lipoamide dehydrogenase (Lpd) alters substrate affinities as well
18 as kinetics at physiological substrate concentrations likely favoring a reactional shift
19 towards lipoamide reduction. In addition, genetic reconstructions proved the
20 necessity of all three mutations for formate assimilation by the adapted cells. The
21 largely unpredictable nature of these changes demonstrates the usefulness of the
22 evolutionary approach enabling the selection of adaptive mutations crucial for
23 pathway engineering of biotechnological model organisms.

24

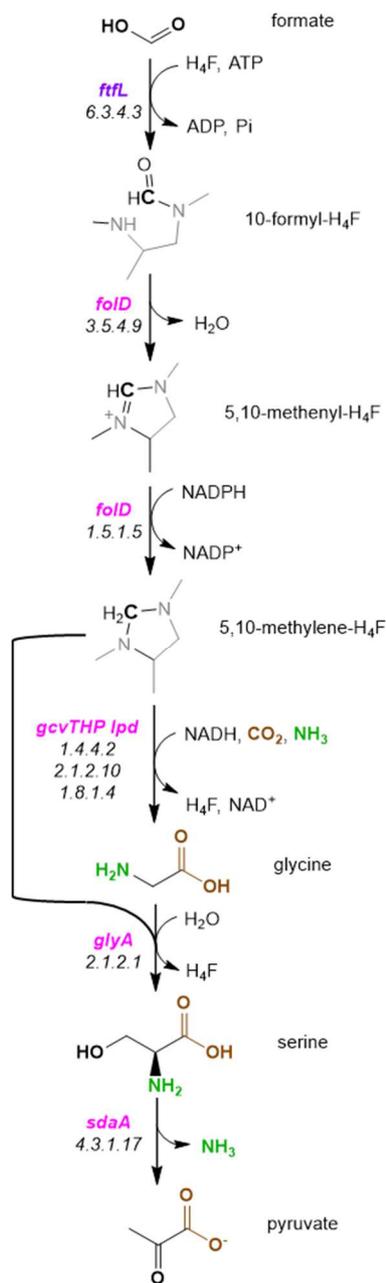
1 **1. Introduction**

2 The utilization of carbon dioxide (CO₂) and of its reduced derivatives formate and
3 methanol as feedstocks for microbial strains holds potential for the biotechnological
4 production of chemicals and fuels (Alissandratos and Easton, 2015; Chistoserdova, 2018;
5 Cotton et al., 2020; Jiang et al., 2021; Yishai et al., 2016). Production processes reducing
6 CO₂ to formate using electrolytic systems (Jouny et al., 2018) or to methanol via
7 hydrogenation (Bowker, 2019) are actively developed and optimized with the prospect of
8 relying on energy supplied from renewable sources.

9 Autotrophic and methylotrophic organisms can grow on one carbon (C1) compounds as
10 sole carbon or as carbon and energy source. However, compared with biotechnological
11 platform organisms like *Escherichia coli*, these organisms generally present limits in terms
12 of growth and product range they can efficiently produce. As a consequence, a number of
13 synthetic biology approaches are currently undertaken to endow *E. coli* cells with hemi-
14 autotrophic (Antonovsky et al., 2016; Gleizer et al., 2019), methylotrophic (Chen et al.,
15 2015; Keller et al., 2020; Meyer et al., 2018; Muller et al., 2015; Siegel et al., 2015), or
16 formatotrophic growth (Kim et al., 2020; Yishai et al., 2018) through the expression of
17 suitable assimilatory pathways. Generally, enzymatic activities catalyzing key steps of the
18 respective natural route are introduced in the cells, which, together with native metabolic
19 reactions, constitute the new assimilation route broadening the trophic mode of the host
20 organism. In most cases, the expression of heterologous genes following a merely rational
21 pathway design does not directly lead to the targeted growth phenotype but requires
22 mutations affecting the activity of pathway enzymes, regulatory modules or side activities.
23 Since the prediction of mutations optimizing fluxes within an artificial metabolic network is
24 often beyond current knowledge, an evolutionary approach is commonly applied which
25 implicates serial dilutions of cell populations or chemostats under selective growth
26 conditions.

1 The implementation of the Calvin-Benson-Bassham cycle in *E. coli* enabling assimilation of
2 CO₂ as sole carbon source for biomass sugars required the selection of mutations
3 modulating the activity of biosynthetic enzymes mobilizing the carbon pool of the cycle and
4 of global metabolic regulators (Antonovsky et al., 2016; Herz et al., 2017). Likewise,
5 methanol consumption became essential for growth of rationally designed derivatives of *E.*
6 *coli* harboring a heterologous methanol dehydrogenase and the ribulose-monophosphate
7 cycle enzymes through serial dilutions in an adaptive medium containing methanol and
8 gluconate as sole complex carbon source (Meyer et al., 2018). Again, non-synonymous
9 mutations in genes coding for metabolic regulators and enzymes susceptible to lower the
10 pool of pathway intermediates were selected. More recently, a synthetic methylotrophic *E.*
11 *coli* was obtained by coupling a design, build and test strategy with laboratory evolution
12 (Chen et al., 2020).

13 In a recently published work, the linear reductive glycine pathway (Figure 1), which
14 assimilates formate and CO₂ into central carbon metabolism, has been implemented in *E.*
15 *coli* (Bar-Even et al., 2013; Kim et al., 2020). All enzymatic steps of the pathway were
16 installed in a modular manner by assembling the required genes into three distinct
17 synthetic operons and by inserting the latter into the *E. coli* chromosome at various
18 insertion sites. The final genetic construct exhibited a very low growth rate when fed on
19 formate as sole carbon and energy source in air enriched with 10 % CO₂. Subsequent
20 serial subculturing improved growth and selected for intergenic cis-acting mutations that
21 increased the level of expression of key enzymatic activities.



1

2

3 **Figure 1.** The reductive glycine pathway for assimilation of formate into biomass (adapted from
 4 Bar-Even et al., 2013). Genes encoding the enzymes catalyzing the pathway reactions are
 5 indicated, with *E. coli* genes depicted in pink, the heterologous gene *fftL* coding for formate H₄F-
 6 ligase depicted in purple. The respective enzymatic activities are indicated by their EC numbers.
 7 Structures in grey correspond to a subpart of tetrahydrofolate (H₄F).

8

9 In our earlier study (Döring et al., 2018) we reported the implementation of key steps of the
 10 reductive glycine pathway in *E. coli* auxotrophs selected for formate assimilation resorting

1 to long-term cultivation protocols (Supplementary Figure S1). In a first stage, formate was
2 established as precursor of the essential tetrahydrofolate (H₄F)-bound activated C1-
3 compounds, *i.e.* formyl-, methenyl-, methylene- and methyl-H₄F in an *E. coli* strain with a
4 defective C1-metabolism. Based on the genetic background of a formate-complemented
5 isolate, a glycine auxotrophic strain was constructed and submitted to the next selection
6 stage, aiming to enforce the condensation of formate-derived methylene-H₄F with CO₂ and
7 ammonia for synthesizing the amino acid glycine. Formate-dependent growth of a glycine-
8 requiring bacterial population was selected after a short period of cultivation under
9 selective conditions. With the exception of the gene *ftfL* from *Clostridium kluyveri*, which
10 codes for a formate-H₄F ligase (a missing enzymatic activity in *E. coli*), no heterologous
11 activity was expressed in the selection strains, thus formate assimilation relied on recruited
12 chromosomally encoded enzymes.

13 In the present study, we report the adaptation of *E. coli* to formate and CO₂ usage as sole
14 carbon sources through assimilation via the reductive glycine pathway by applying our
15 strategy of directed evolution in continuous culture. The bacterial population depending on
16 formate for C1 and glycine synthesis described earlier (Döring et al., 2018) was adapted to
17 faster growth in a turbidostat. One robustly growing isolate was genetically modified to be
18 used as starting point of an adaptive evolution toward growth on formate and CO₂ as
19 carbon source and acetate as energy source. We identified adaptive mutations fixed in the
20 genomes of all intermediate and final formate depending isolates. Genes coding for
21 enzymes of the reductive glycine pathway were found to harbor non-synonymous
22 mutations. To understand the functional role of these and other mutations, which
23 accumulated throughout the stepwise evolution process, we characterized the activity of
24 mutant pathway enzymes, quantified gene expression and conducted reverse-genetic
25 analysis. Our findings point to the fine-tuning of metabolic fluxes by variant enzymes and

1 to the flexibility of the central carbon metabolism of bacteria like *E. coli* capable of adapting
2 to new trophic modes upon only a few mutations.

3

4 **2. Materials and methods**

5 2.1. Chemicals, media and cultivation conditions

6 Chemicals were generally purchased from Sigma-Aldrich. (6-R,S)-5,10-methylene-5,6,7,8-
7 H₄F calcium salt and (6-R,S)-5,10-methenyl-H₄F chloride were from Schircks Laboratories
8 (Switzerland). Lipoamide was from abcr GmbH (Germany). Dihydrolipoamide was
9 synthesized by borohydride reduction of lipoamide as previously described (Reed et al.,
10 1958); purity (> 95%) was checked by nuclear magnetic resonance and infusion mass-
11 spectrometry analysis.

12 Bacteria were routinely grown aerobically in lysogeny broth (LB rich medium) or in minimal
13 saline defined medium (MS) supplemented with D-glucose (2 g/L) as a carbon source.
14 Growth media were solidified with 15 g/L agar for the preparation of Petri dishes. Bacterial
15 cultures were incubated at 37°C either in a standard incubator or in an incubator with a
16 controlled humidified atmosphere containing a mix of 95% air - 5 % CO₂. The nutrients
17 required for the growth of glycine auxotrophs or of C1 pathway-defective bacteria were
18 added to the culture medium at the following final concentrations: 2 mM glycine, 0.3 mM L-
19 methionine, 0.3 mM thymidine, 0.3 mM inosine, 6 μM DL-pantothenate. When required,
20 antibiotics were added at the following concentrations: 100 mg/L ampicillin; 100 mg/M
21 spectinomycin; 180 mg/L erythromycin, 25 mg/L chloramphenicol and 30 mg/L kanamycin.

22

23 2.2. Strain construction

24 The strains used or constructed in this study were all derivatives of the wild type *E. coli*
25 K12 strain MG1655. Their relevant genotypes and filiations are listed in Supplementary
26 Table S1. The desired genetic contexts were obtained by serial P1 transductions of gene

1 knockouts substituted by antibiotic resistance cassettes according to the method of Miller
2 (Miller, 1972). Mutated genes of interest were mobilized in the desired recipient cells by
3 co-transduction with closely linked kanamycin markers originating from the Keio *E. coli*
4 knock-out collection (Baba et al., 2006). Care was taken that the transduced DNA
5 fragment did not introduce or suppress any mutations other than the one which was
6 targeted by considering genetic distances between mutated loci. In case of close vicinity
7 on the genome (up to 1.5 minute), sequence conformity of the transductants was verified
8 by PCR amplification and sequencing. Resistance cassettes were removed by flippase
9 reaction after transformation with the plasmid pCP20.

10

11 2.3. Bacterial growth assays

12 A Microbiology Reader Bioscreen C apparatus (Thermo Fisher Scientific) was used for
13 growth curve experiments. It consists of a thermostatic incubator and a culture growth
14 monitoring device (OD reader). Overnight bacterial cultures were washed once in MS
15 medium and diluted 100-fold in the growth medium; 200 μ l aliquots of the cell suspensions
16 were distributed into honeycomb 100-wells plates. Each experiment was performed in
17 triplicate. The plates were incubated at 37°C under continuous agitation. Bacterial growth
18 was followed by recording optical densities at 600 nm every 15 minutes for 24 hours if not
19 otherwise stated.

20

21 2.4. Continuous culture

22 Evolution experiments in continuous culture were carried out using automaton-driven GM3
23 fluidic self-cleaning cultivation devices (Mutzel and Marlière, 2000). A population of
24 growing bacteria from strain G4369 (glycine auxotroph) was adapted to utilization of
25 formate and CO₂ as precursors of C1-moieties and glycine under medium-swap regime
26 (Döring et al., 2018). Growth rate of the formate-complemented population was further

1 improved by selecting the fastest variants in the population under a turbidostat regime. At
2 regular 10 minute intervals the optical density of the culture was measured and compared
3 to a fixed threshold (OD_{600 nm} value of 0.4). When the measured OD_{600 nm} exceeded the
4 threshold, a pulse of fresh nutrient medium (1.6 mL) was injected into the culture, the
5 volume of which was kept constant (16 mL). The monitoring of the turbidity ensured that
6 the biomass in the vessel remained constant and that the bacteria grew at their maximal
7 growth rate. A continuous gas flow of controlled composition (95% air + 5% CO₂) through
8 the culture vessel ensured constant aeration and growth in suspension by counteracting
9 cell sedimentation.

10 A population of growing bacteria from strain G4985 was adapted to use formate and CO₂
11 as sole carbon sources and acetate as energy source under medium-swap regime. This
12 regime enables gradual adaptation of a bacterial population to grow in an initially non-
13 permissive stressing medium (Marlière et al., 2011). The growing culture can be diluted by
14 either permissive or stressing medium. The choice between the two dilution media
15 depends on the turbidity of the culture with respect to a set OD threshold (OD_{600nm} value of
16 0.4). When the OD_{600nm} exceeds the threshold, a pulse of stressing medium is injected,
17 otherwise a pulse of permissive medium. Dilutions are triggered every 10 min with a set
18 volume of medium, thus imposing a generation time on the cell population. The culture
19 conditions fixed for the culture of G4985 bacteria were as follows: permissive medium: MS
20 40 mM NH₄Cl 15 mM glycine 15 mM acetate 45 mM formate; stressing medium: MS 40
21 mM NH₄Cl 15 mM acetate 45 mM formate; aeration gas: air + 5% CO₂; fixed generation
22 time: 8 hours.

23

24 2.5. Whole genome sequencing and mutation analysis

1 Pair-end libraries (2x150 bp) were prepared with one μg of genomic DNA from the
2 ancestor strain G3745 and the evolved strains (G3806, G4400, G4463, G5222 and
3 G5225) and sequenced using a MiSeq sequencer (Illumina).
4 High-throughput sequencing data were analyzed using the PALOMA bioinformatic pipeline
5 implemented in the MicroScope platform (Vallenet et al., 2019)
6 (<https://mage.genoscope.cns.fr/microscope/home/>). In a first step, reads were mapped
7 onto the *E. coli* MG1655 reference (NC_000913.3) using the SSAHA2 package (v.2.5.1)
8 (Ning, 2001). Only unique matches having an alignment score equal to at least half of their
9 length were retained as seeds for full Smith-Waterman realignment (Smith and Waterman,
10 1981) with a region extended on both sides by five nucleotides of the reference genome.
11 All computed alignments then were screened for discrepancies between read and
12 reference sequences and a score based on coverage, allele frequency, quality of bases,
13 and strand bias was computed for each detected event to assess its relevance. The
14 mutations (single nucleotide variations and short insertions or deletions) with a score
15 superior to 0.8 with at least five supporting reads were retained.

16

17 2.6. Carbon labeling analysis

18 Bacteria of strains G5222 and G5225 were grown on MS medium NH_4Cl 40 mM
19 supplemented with either 45 mM sodium formate- ^{12}C 15 mM sodium acetate- $^{13}\text{C}_2$ (Sigma
20 Aldrich 282014) or 45 mM sodium formate- ^{13}C (Sigma Aldrich 279412) 15 mM sodium
21 acetate- ^{12}C in the presence of 95% air + 5% CO_2 at 37°C for 48h.

22 Metabolomes were prepared using a method adapted from
23 ([http://www.polyomics.gla.ac.uk/assets/downloads/MSMetabolomicsPrepCells-
24 Aug2013.pdf](http://www.polyomics.gla.ac.uk/assets/downloads/MSMetabolomicsPrepCells-Aug2013.pdf)). Briefly, the bacteria were cooled on ice and centrifuged (3200 rcf /10 min –
25 4°C) and the pellets suspended in 200 μL of the mix water/methanol/chloroform (1:3:1),
26 then frozen (in a dry ice/ethanol bath, -80°C)/thawed (room temperature) twice. The

1 lysates were transferred to microtubes, agitated at high speed for 1 hour at 4 °C
2 (Thermomixer, Eppendorf) then centrifuged (15000 rcf/3min at 4°C). The supernatants
3 were transferred to microtubes and evaporated in vacuum (Speedvac). The lyophilisates
4 were conserved at -80°C.

5 LCMS/MSⁿ analyses were performed on a Dionex TCC-3000RS chromatographic system
6 (Thermo Fisher Scientific, Courtaboeuf, France) coupled to an Orbitrap Elite mass
7 spectrometer (Thermo Fisher Scientific, Courtaboeuf, France) equipped with a HESI
8 source. For HPLC separation, a SeQuant ZICpHILIC column 5 µm, 4.6 × 150 mm (Merck,
9 Darmstadt, Germany) thermostated at 40°C was used. Mobile phase flow rate was set at
10 0.5 ml/min, and the injection volume was 10 µL. Aqueous solution of 10 mM (NH₄)₂CO₃
11 (pH of 9.9) was used as phase A and acetonitrile was used as phase B. The following
12 gradient conditions were applied for elution: 2 min equilibration step at 80% of phase B; 20
13 min linear gradient from 80 to 40% of phase B; 8 min isocratic elution at 40% of B, return
14 to 80% of phase B in 5 min and a reconditioning step of 15 min.

15 Mass spectra and fragmentation spectra were recorded in the ESI positive and negative
16 ionization modes. Ion spray (IS) was set at -4.2 kV for negative mode and 4.5 kV for
17 positive mode. Capillary temperature was set at 250°C. Sheath gas, auxiliary gas and
18 sweep gas flow rates were set at 60, 44 and 0 arbitrary units, respectively. The mass
19 resolving power of the detector was 60000 for MS and 120000 for MSⁿ experiments. Raw
20 data were analyzed using the Qual-browser module of Xcalibur version 2.2 (Thermo Fisher
21 Scientific, Courtaboeuf, France). Prior to analysis, valine, glutamate and GSH standards
22 (Sigma-Aldrich) were analyzed under the same conditions in order to determine typical
23 retention times, and, in the case of GSH, fragmentation conditions.

24

25 2.7. Quantitative PCR analysis of *ftfL* gene expression

1 The relative expression levels of the *C. kluyveri* *fftL* gene and the native gene *panB* in
2 bacteria of the evolved strain G3806 and of its ancestor strain G3745 were evaluated by
3 RT-qPCR analysis. The bacteria were cultivated until the OD_{600nm} reached 0.5-0.7. The
4 appropriate volume of a RNA stabilization solution was then added, the samples
5 centrifuged for 10 min at 4500 rpm and the supernatants discarded; the cell pellets were
6 frozen at -80°C. Total RNA was extracted using the RNeasy kit (QIAGEN). Contaminating
7 genomic DNA was eliminated by treatment with Turbo DNase (Ambion, Life
8 Technologies), and the absence of DNA from purified total RNA was verified by PCR.
9 Total RNA was quantified using a Qubit fluorimeter (Life Technologies). cDNA was
10 produced with the SuperScript Vilo cDNA synthesis kit (Invitrogen) and purified using the
11 Ampure kit (Beckman-Coulter). A Stratagene Full Velocity kit (Integrated Genomics) and
12 Mx3005P real-time PCR system (Agilent Technologies) were used for qPCR. The primer
13 pairs used for qPCR were designed to amplify an approximately 200 bp region from the
14 housekeeping gene *panB*: 5'-TTAGAAGCTGCTGGGGCACA (forward) and 5'-
15 CCGTTTCGGCGAGGAAATT (reverse); and from the heterologous gene *fftL*: 5'-
16 TAACCGCCCATGACCTAAAG (forward) and 5'- TTCTGCACCAAGATCTGCAC (reverse).
17 Primer concentrations were optimized and dissociation curves were generated for each of
18 the primer pairs to verify the amplification of a single PCR product. Amplifications were
19 performed in duplicate using cDNA equivalent to 20 ng of total RNA in four sequential
20 three-fold dilutions. The amplification cycling conditions were: initial denaturation at 95°C
21 for 3 min, then 40 cycles of denaturation at 95°C for 30 s; annealing at 60°C for 30 s; and
22 extension at 72°C for 30 s. Data were analyzed using MxPro QPCR software (Agilent).
23 The entire experiment, from exponentially growing bacteria to RNA extraction, cDNA
24 synthesis and qPCR, was performed twice. The comparative Ct (Cycle threshold) method
25 was used to compare the abundance of the mRNAs transcripts of *fftL* gene in the wild type
26 and evolved strains (G3745 and G3806, respectively) as well as in reconstructed strain

1 G4552. The differences (ΔC_t) between C_t for the internal control amplicon *panB* and C_t for
2 the target amplicon *fftL* were determined.

3

4 2.8. Protein expression and purification

5 The wild type genes *fold* (Uniprot P24186) and *lpd* (Uniprot P0A9P0) as well as the
6 mutated genes *fold* R191S and *lpd* R273H were amplified by PCR and ligated into
7 pET22b(+) vector (Novagen) to generate in each case a protein with an N-terminal 6x-
8 histidine tag. The resulting plasmids were introduced into *E. coli* BL21 (DE3) by chemical
9 transformation. Standard methods were used for cell culture, induction of gene expression,
10 cell extract preparation and nickel affinity purification of His-tagged recombinant proteins.

11

12 2.9. FoLD activity assays

13 The kinetic parameters of dehydrogenase and cyclohydrolase activities of the bifunctional
14 enzymes FoLD were determined from duplicate experiments by non-linear analysis of initial
15 rates using SigmaPlot version 9.0 (Systat Software). Spectrophotometric assays were
16 performed at 25°C in a Safas UV mc2 double beam spectrophotometer.

17 *Forward FoLD cyclohydrolase assay.* 5,10-methenyl-H₄F, stable at acidic pH, was
18 dissolved in a buffer consisting of 100 mM potassium phosphate pH 7.3 and 36 mM 2-
19 mercaptoethanol for 20 minutes with constant stirring. Then, the solution was titrated to pH
20 2.5 with HCl 12N and incubated for 15 minutes before use. The concentration of 5,10-
21 methenyl-H₄F was determined spectrophotometrically ($\epsilon_{356} = 24\,900\text{ M}^{-1}\text{ cm}^{-1}$).

22 Cyclohydrolase activity was determined by following the disappearance of 5,10-methenyl-
23 H₄F at 355 nm in a continuous spectrophotometric assay. The reactions were carried out
24 in 100 μ l of an activity buffer consisting of 25 mM MOPS pH 7.3, 2 mM potassium
25 phosphate, 36 mM 2-mercaptoethanol in the presence of [0-45 μ M] 5,10-methenyl-H₄F.
26 The reactions were initiated by the addition of 40 ng of the enzyme. Measured initial rates

1 were corrected for the non-enzymatic rate of 5,10-methenyl-H₄F spontaneous conversion
2 to 10-formyl-H₄F under the same conditions.

3 *Reverse FOLD cyclohydrolase assay.* 10-Formyl-H₄F was prepared by dissolving 5,10-
4 methenyl-H₄F in N₂-sparged buffer consisting of 25 mM alanine (pH 8.5) and 36 mM 2-
5 mercaptoethanol as described previously (Pawelek and MacKenzie, 1998). The reactions
6 were performed in 100 µl buffer consisting of 25 mM MOPS pH 7.3, 2 mM potassium
7 phosphate, 36 mM 2-mercaptoethanol and [0-450 µM] 10-formyl-H₄F in the presence or
8 absence of 1 mM 2',5'-ADP. The assays were initiated by the addition of 170 ng of the
9 enzyme and the production of 5,10-methenyl-H₄F was continuously measured
10 spectrophotometrically at 355 nm.

11 *Forward FOLD dehydrogenase assay.* 5,10-methylene-H₄F, slightly soluble in water and
12 stable at pH>9, was dissolved in a buffer consisting of 100 mM Tris pH 8, 0.2N NaOH and
13 36 mM 2-mercaptoethanol. The solution was titrated to pH 12. 5,10-methylene-H₄F
14 dehydrogenase activity was determined by measuring the production of NADPH using a
15 spectrofluorometric assay with excitation and emission wavelengths set at 340 nm and
16 445 nm, respectively (Safas Xenius XC). Assays were performed in 350 µl reactional
17 mixture consisting of 25 mM MOPS pH 7.3, 2 mM potassium phosphate, 36 mM 2-
18 mercaptoethanol, 5,10-methylene-H₄F in the presence of NADP⁺ and were initiated by the
19 addition of 40 ng of the enzyme. The kinetic constants for 5,10-methylene-H₄F and NADP⁺
20 were evaluated by varying the concentration of one substrate while keeping the other
21 substrate at saturating concentration (80 µM and 2 mM respectively). Previously, a
22 standard range of NADPH was produced in order to relate percentage of EMI and NADPH
23 concentration.

24 *Reverse FOLD dehydrogenase assay.* 5,10-methylene-H₄F dehydrogenase reductive
25 activity was assayed by measuring the disappearance of 5,10-methenyl-H₄F at 350 nm in
26 a spectrophotometrically end-point assay according to Pawelek and MacKenzie, 1998.

1 Assays were performed in 100 μ l of an activity buffer consisting of 25 mM MOPS pH 7.3, 2
2 mM potassium phosphate, 36 mM 2-mercaptoethanol, 1.5 mM NADPH and [0-300 μ M]
3 5,10-methenyl-H₄F. The reactions were initiated by the addition of 2.5 μ g enzyme and
4 were stopped with 0.36 N HCl after 10 sec. The competing conversion of 5,10-methenyl-
5 H₄F to formylH₄folate by cyclohydrolase activity impeded the determination of reliable
6 kinetic parameters for the reverse dehydrogenase activity.

7 *Overall forward reaction assay and substrate channeling.* Substrate channeling is defined
8 as the rate of 10-formyl-H₄F production divided by the rate of [5,10-methenyl-H₄F + 10-
9 formyl-H₄F] production over a linear time course (Pawelek and MacKenzie, 1998). 5,10-
10 methylene-H₄F was prepared in a buffer consisting of 100 mM Tris pH8, 0.2 N NaOH and
11 36 mM 2-mercaptoethanol. Duplicated assays were performed in 200 μ l buffer consisting
12 of 25 mM MOPS pH 7.3, 2 mM potassium phosphate, 36 mM 2-mercaptoethanol, 200 μ M
13 5,10-methylene-H₄F and 1.5 mM NADP⁺. The reactions were initiated by the addition of 15
14 ng of the enzyme. The rate of production of 5,10-methenyl-H₄F was determined at 355 nm
15 in a continuous spectrophotometric assay. The rate of production of [5,10-methenyl-H₄F +
16 10-formyl-H₄F] was measured at 355 nm after acidification of the reaction mixtures over
17 the same time course. The rate of production of 10-formyl-H₄F was calculated by
18 subtracting the rate of production of 5,10-methenyl-H₄F from the rate of production of
19 [5,10-methenyl-H₄F + 10-formyl-H₄F].

20 *Overall reverse reaction assay.* 10-Formyl-H₄F was prepared by dissolving 5,10-methenyl-
21 H₄F in a buffer consisting of 100 mM potassium phosphate pH 7.3 and 36 mM 2-
22 mercaptoethanol with constant stirring during one hour. Three independent assays were
23 performed in 100 μ l reactional mix of 25 mM MOPS pH 7.3, 2 mM potassium phosphate,
24 36 mM 2-mercaptoethanol, 1.5 mM NADPH and 300 μ M 10-formyl-H₄F. The reactions
25 were initiated by the addition of 80 ng of enzyme and stopped with 0.36N HCl at fixed time

1 points (0.5, 1.5, 3mn). The rate of disappearance of [5,10-methenyl-H₄F + 10-formyl-H₄F]
2 was measured at 351 nm in end-point assays and used to establish the global reverse k_{cat} .

3

4 2.10. Lpd activity assays

5 *Lpd characterization.* Spectrophotometric assays were performed at 25°C in a Safas UV
6 mc2 double beam spectrophotometer. The concentration of purified Lpd enzyme was
7 determined spectrophotometrically using an extinction coefficient of 34.0 mM⁻¹.cm⁻¹ at 280
8 nm (computed using Expasy Paramtool <https://web.expasy.org/protparam/>). The
9 concentration of FAD was determined using an extinction coefficient of 15.4 mM⁻¹.cm⁻¹ at
10 446 nm. The absorbances (A) at 446 nm and 280 nm were measured and the ratio
11 A_{446nm}/A_{280nm} calculated to determine the fraction of active FAD-containing catalysts within
12 each batch of purified enzyme and to normalize the results between the different enzyme
13 forms. FAD ratios varied from 8% to 11 % of total protein. Assays of Lpd-catalyzed
14 oxidative and reductive reactions were conducted in 100 mM Na phosphate, 100 mM KCl,
15 and 8 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) pH 7.6. The substrates
16 lipoamide and dihydrolipoamide were solubilized in DMSO; final DMSO concentrations in
17 reaction mixes were less than 10% throughout all assays. NAD⁺ and NADH solutions were
18 prepared in water.

19 *Lpd-catalyzed lipoamide reduction assay.* The influence of the NADH/NAD ratio on the
20 reductive reaction was determined with 300 μM NADH, 5 mM lipoamide and NAD⁺ at
21 concentrations varying from 0 to 15 mM. NADH kinetic parameters were determined by
22 varying its concentration (0, 10, 20, 30, 40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 250
23 and 500 μM) in the presence of a saturating concentration of lipoamide (3 mM) and of
24 NAD⁺ 1 mM. Lipoamide kinetic parameters were determined in the presence of 0.195 mM
25 NADH and 6 mM NAD⁺ and varying concentrations of lipoamide (0, 0.2, 0.5, 1, 3, 5 and 6
26 mM) at 340 nm in a continuous spectrophotometric assay. Reactions were initiated with

1 the appropriate amount of enzyme and monitored by recording the linear disappearance of
2 NADH at 340 nm. Kinetic constants were determined by non-linear analysis of initial rates
3 from duplicate experiments using SigmaPlot 9.0.

4 *Lpd-catalyzed dihydrolipoamide oxidation assay.* Kinetic parameters for NAD⁺ were
5 determined by varying its concentration (0, 0.5, 1, 2, 4, 6, 10, 20, and 40 mM) in the
6 presence of a saturating concentration of dihydrolipoamide (4 mM). Kinetic parameters for
7 dihydrolipoamide were determined by varying its concentration (0, 0.5, 1, 2, 4, 6, 10, 20,
8 and 40 mM) in the presence of a saturating concentration of NAD⁺ (4 mM). Final
9 concentration of the enzyme varied from 0.03 to 3 μM. The reactions were monitored by
10 recording the linear accumulation of NADH at 340 nm. Kinetic constants were determined
11 by nonlinear analysis of initial rates from duplicate experiments using SigmaPlot 9.0
12 (Systat Software, Inc.).

13

14 **3. Results**

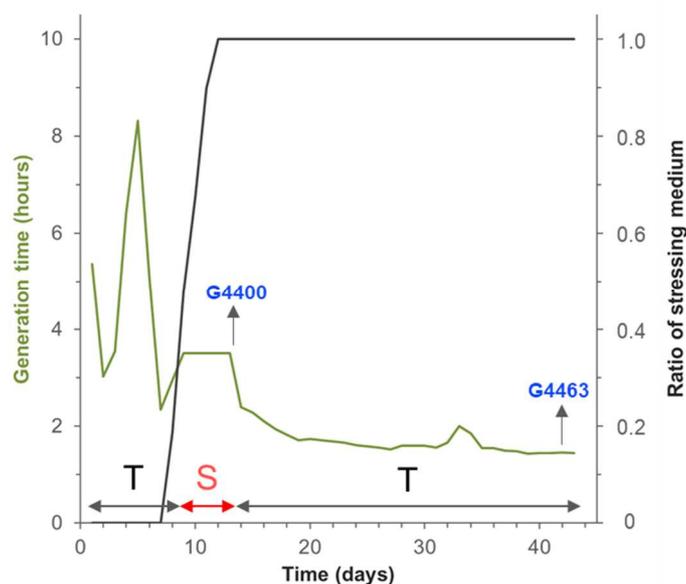
15

16 **3.1. Adaptation of *E. coli* to efficient formate assimilation**

17 As previously described (Döring et al., 2018), an adaptive evolution strategy was applied
18 to sequentially optimize the early enzymatic steps of the reductive glycine pathway in
19 appropriate selection strains of *E. coli* (Table 1 and Supplementary Figure S1 for strain
20 lineage; Supplementary Table S1 for exhaustive strain description).

21 In the first stage, the *E. coli* C1-auxotroph G3745 ($\Delta glyA \Delta(tdh-kbl) \Delta ltaE \Delta gcvR ffl^+ C.$
22 *kluuyveri \Delta gcvTHP*) was gradually adapted to use formate as sole precursor of the H₄F-
23 bound C1 moieties by applying the continuous cultivation medium swap protocol. Under
24 this cultivation regime, a bacterial suspension of strain G3745 was alternatively fed with a
25 permissive medium (containing formate and the C1 metabolites inosine and methionine) or
26 a stressing medium (containing formate but no C1 metabolite), the choice depending on

1 the bacterial growth dynamic, until the population evolved to grow in 100 % stressing
2 medium.
3 Based on the genetic background of an isolate from this culture evolved to formate-
4 complementation of C1 requirements (strain G3806), the glycine auxotroph G4369 ($\Delta glyA$
5 $\Delta(tdh-kbl)$ $\Delta ltaE$ $\Delta aceBAK$ $ftfL^+$ $pdeH$ $\Delta 3bp$ $folD$ R191S) was submitted to the next
6 selection stage aiming to establish formate as precursor of the H₄F-bound C1 moieties and
7 the amino acid glycine. Again, a medium swap regime of continuous cultivation was
8 applied, which alimented the cell population with either a permissive medium (MS glucose
9 glycine formate) or a stressing medium (MS glucose formate), for weaning cells off glycine
10 (Figure 2). In this latter experiment, the aeration atmosphere of the culture consisted in a
11 mix 95% air - 5% CO₂, as CO₂ is a substrate of the glycine cleavage system, which
12 catalyzes the reductive synthesis of glycine (Figure 1). Eighty generations under medium
13 swap regime were necessary for the glycine auxotroph G4369 to evolve towards the
14 assimilation of formate and CO₂ into glycine at a fixed generation time of 3.5 hours (Figure
15 2).
16



17

1 **Figure 2:** Adaptation of glycine-requiring G4369 bacteria ($\Delta glyA \Delta (tdh-kbl) \Delta ltaE \Delta aceBAK fflL+$
2 $pdeH \Delta 3bp folD R191S$) to utilization of formate and CO₂ as C1 and glycine precursors by directed
3 evolution in continuous culture. The glycine-auxotrophic bacteria were initially adapted to grow in
4 permissive medium (MS glucose 100 mM NH₄Cl 0.5 mM glycine 30 mM sodium formate) under
5 turbidostat regime (**T**, black). From day 8, medium swap regime (**S**, red) was applied to the culture
6 with a steady generation time set to 3.5 h. Maximum ratio of stressing medium (MS glucose 100
7 mM NH₄Cl 30 mM sodium formate) was reached at day 13 after 80 generations under medium
8 swap regime. The growth of the formate-complemented bacterial population was further
9 accelerated under turbidostat regime (**T**, black) in stressing medium. The aeration gas of the
10 culture consisted in a mix of 95% air – 5% CO₂ from day 1 to day 43. Green line: evolution of the
11 generation time of the culture. Black line: daily ratio of stressing medium dilution pulses. Grey
12 arrows show the time-point of culture samples from which the indicated strains were isolated.

13

14 Once stable growth was attained in the stressing medium, the formate-complemented
15 glycine-requiring cell population was further cultivated under turbidostat feeding mode to
16 select for higher growth rate (Materials and methods, section 2.4). After 29 days of
17 continuous growth (425 generations), the generation time of the bacterial population had
18 dropped from 2h20 to 1h45 (Figure 2).

19 The mutations fixed in the strain G4463 isolated from a sample of the adapting bacterial
20 population taken at this time point were identified by whole-genome sequencing and
21 mutational analysis (refer to Materials and methods, section 2.5). They are listed in Table
22 1 together with the mutations previously found for the formate-complemented C1
23 auxotrophic intermediate strain G3806 and the formate-complemented glycine auxotrophic
24 intermediate strain G4400 obtained at the end of the medium swap evolution (Döring et al.,
25 2018) (Supplementary Table S2 for an exhaustive list). The mutations present in the
26 ancestor G3806 (formate-complemented C1 auxotroph) of the isolate G4463, thus
27 selected during the first evolution step, were conserved. Few additional mutations were
28 fixed during the adaptation step towards efficient assimilation of formate into glycine. A
29 mutation in the ribosomal gene *rrfF* was probably selected during the medium swap-driven
30 adaptation, as the isolate G4400 sampled by the time formate fulfilled C1 and glycine

1 requirements of the adapting glycine-requiring bacterial population, already harbors this
2 mutation (Table 1, Supplementary Figure S1). The gene *lpd*, which codes for the lipoamide
3 dehydrogenase (Lpd), was mutated in the formate-complemented glycine auxotroph
4 G4463 (*lpd* R273H) sampled after growth acceleration under turbidostat regime. The Lpd
5 enzyme functions as the dehydrogenase subunit of the glycine cleavage system - which is
6 expected to catalyze the reductive synthesis of glycine - and is a component of both
7 pyruvate and 2-ketoglutarate dehydrogenases.

Table 1: Genetic divergences in the genomes of the evolved formate-dependent strains with respect to the genome of their corresponding ancestor strain.

selection	ancestor	cultivation regime	isolate	point mutations and short in/dels			mutated genes	rearrangements	
				total	intragenic	intergenic		non synonymous mutations and short in/dels	amplifications
formate as C1 source	G3745 <i>ΔglyA ΔltaE Δ(tdh-kbl) ΔgcvTHP ftfL⁺</i>	medium swap	G3806	4	3	1	<i>folD R191S purR D307V pdeH Δ3bp</i>	-	-
			G4369 <i>ΔglyA ΔltaE Δ(tdh-kbl) ftfL⁺ ΔaceBAK folD R191S purR D307V pdeH Δ3bp</i>	medium swap	G4400	3	1	2	<i>rrfF</i>
formate as glycine source		turbidostat	G4463	4	3	1	<i>rrfF lpd R273H purF Q243K</i>	<i>fetA-ybck^{***}</i> 55.5 kb	-
formate as carbon source	G4985 <i>ΔltaE Δ(tdh-kbl) ftfL⁺ ΔaceBAK folD R191S purR D307V pdeH Δ3bp rrfF lpd R273H purF Q243K</i>	medium swap	G5222	27	20	7	<i>lrp nadR nadK rho rpoD gltA argS btuR helD hyfC menA metF mrdA trkG ybiW yfaL yfiQ yjjM ypfH yqjH</i>	<i>fetA-ybck^{***}</i> 55.5 kb	e14 prophage 15 kb <i>yggS-yghO⁺</i> 34.4 kb
			G5225	27	21	6	<i>lrp nadR nadK rho rpoD iscR gltA btuR helD hyfC menA metF mrdA nrfB trkG ybiW yfaL yfiQ yjjM ypfH yqjH</i>	<i>fetA-ybck^{***}</i> 55.5 kb	e14 prophage 15 kb <i>yggS-yghO⁺</i> 34.4 kb

*duplicated region contains gene *folD*; **duplicated region contains gene *ftfL*; ***amplified region (x4) contains gene *folD*; *deleted region contains *mutY*.

1 We conducted reverse genetic experiments to substantiate the contribution of the
 2 mutations to the trophic phenotypes, focusing on genes coding for enzymes of the
 3 reductive glycine pathway or related to C1 metabolism. Notably, we targeted the 3 base
 4 pair deletion ($\Delta 3\text{bp}$) which localizes in the gene *pdeH* 339 base pair upstream of the
 5 heterologous gene *ftfL* of *C. kluyveri* encoding formate-H₄F ligase activity inserted in the
 6 chromosome, the methylene-H₄F dehydrogenase cyclohydrolase-encoding gene *folD*
 7 (R191S), the lipoamide dehydrogenase-encoding gene *lpd* (R273H), as well as the
 8 transcriptional repressor-encoding gene *purR* (D307V) (Table 1).

9 We used phage P1-mediated transduction either to substitute the wild type genes for the
 10 mutated genes in the evolved isolates or the mutated genes for the wild type genes in the
 11 unevolved ancestor strains (Materials and methods, section 2.2). The gene *pdeH* co-
 12 transduced with the neighboring gene *ftfL* flanked by a kanamycin resistance cassette.
 13 The genes *folD*, *purR* and *lpd*, which are not directly selectable, were mobilized by co-
 14 transduction with a closely linked kanamycin marker originating from the Keio *E. coli*
 15 knock-out collection ($\Delta allE::kan$, $\Delta ydhL::kan$, $\Delta yachH::kan$, respectively) (Baba et al., 2006).
 16 The various genetic constructs obtained either in the evolved context of C1-auxotroph
 17 G3806 or in the unevolved context of the C1-auxotrophic ancestor strain G3745 (for
 18 genotype refer to Supplementary Table S1) were tested for growth with formate as sole C1
 19 source. Table 2 compiles the results from growth-tests on semi-solid media (see
 20 Supplementary Figure 2 for all growth patterns).

21

22 **Table 2.** Impact of adaptative mutations on the growth phenotype of formate-complemented C1
 23 auxotrophic strains.

Strain	Relevant genotype	Growth on MS glucose glycine medium supplemented with the indicated nutrients*		
		0	C1	formate
G3745	ancestor strain <i>$\Delta glyA \Delta taE \Delta (tdh-kbl) \Delta gcvTHP ftfL^+$</i>	-	+	-
G3806	evolved strain <i>$\Delta glyA \Delta taE \Delta (tdh-kbl) \Delta gcvTHP ftfL^+$</i>	-	+	+

foID R191S purR D307V pdeH Δ3bp

G4234	G3806 <i>Δalle::kan</i>	-	+	+
G4235	G3806 <i>Δalle::kan foID wt</i>	-	+	+/-
G4371	G3806 <i>ΔydhL::kan</i>	-	+	+
G4346	G3806 <i>ΔydhL::kan purR wt</i>	-	+	+
G4284	G3745 <i>Δalle::kan</i>	-	+	-
G4372	G3745 <i>Δalle::kan foID R191S</i>	-	+	-
G4286	G3745 <i>ΔydhL::kan</i>	-	+	-
G4418	G3745 <i>ΔydhL::kan purR D307V</i>	-	+	-
G4578	G3745 <i>Δalle foID R191S ΔydhL::kan purR D307V</i>	-	+	-
G4552	G3745 <i>pdeH Δ3bp</i>	-	+	+

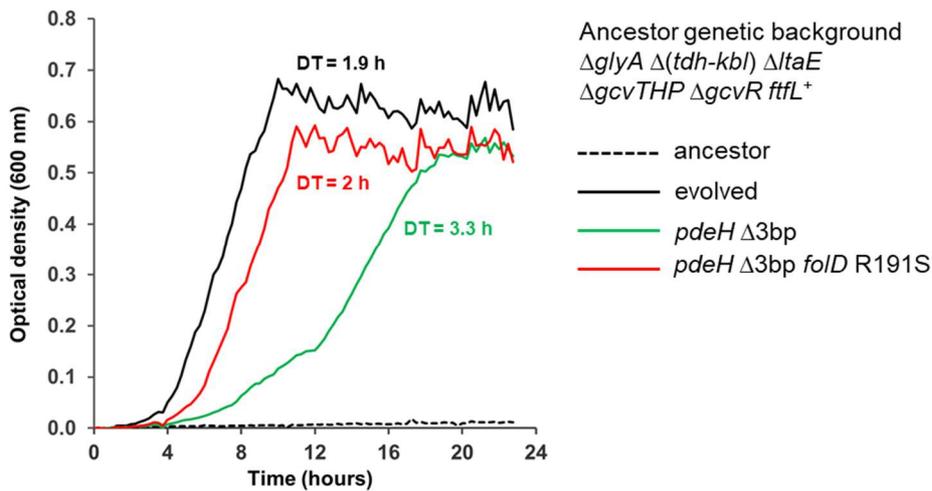
1 *Bacteria were streaked on semi-solid MS medium glucose 2 mM glycine supplemented with C1 metabolites (methionine
2 0.3 mM inosine 0.3 mM thymidine 0.3 mM DL-pantothenate 6 μM) or 10 mM formate and incubated for 48 h at 37°C. -:
3 no growth. +: robust growth. +/-: weak growth.

4

5 Neither the restitution of the wild type *purR* gene in strain G3806 (formate-complemented
6 C1 auxotroph, derived strain G4346), nor the transfer of the *purR* allele D307V in the
7 genetic context of the ancestor G3745 (C1 auxotroph, derived strain G4418) did affect
8 growth of any of the resulting strains (Supplementary Fig. 2A). The *purR* D307V mutation
9 might be neutral or affect the purine nucleotide biosynthesis with a limited impact on
10 growth that remained undetectable in our tests. When the mutated FoID variant R191S
11 was reverted to wild type in the evolved strain G3806, growth of the resulting strain G4235
12 slowed down but was not abolished (Supplementary Fig.2A). This mutation was not
13 sufficient to promote growth by itself in the unevolved ancestor (strain G4372,
14 Supplementary Fig.2B). Despite repeated attempts, we did not succeed in replacing the
15 mutated version of *pdeH* by its wild type counterpart in the chromosome of the evolved
16 strain G3806. Yet, the mutated *pdeH* gene was successfully introduced into the unevolved
17 ancestor C1-auxotroph G3745. The resulting strain G4552 sustainably grew when formate
18 was provided as sole C1 source (Table 2, Supplementary Fig.2B). When the two mutated
19 genes *pdeH* Δ3bp and *foID* R191S were present in the ancestor genetic background the

1 resulting strain G4827 grew almost as well as the evolved strain G3806 when fed with
2 formate as sole C1 source (Figure 3), showing that the 3 base pair deletion in *pdeH* has a
3 major impact, the mutation in *folD* a minor impact on the growth phenotype of strain
4 G3806.

5



6
7

8 **Figure 3:** Growth of ancestor strain G3745 and transduction derivatives harboring the mutations
9 *folD* R191S and *pdeH* $\Delta 3bp$ in liquid MS glucose medium supplemented with 2 mM glycine and 10
10 mM formate at 37 °C in a Bioscreen C growth reader. Dotted black line: ancestor strain G3745;
11 black line: evolved strain G3806; green line: Strain G4826, derivative of G3745 harboring *pdeH*
12 $\Delta 3bp$ mutation; red line: strain G4827, derivative of G3745 harboring *pdeH* $\Delta 3bp$ and *folD* R191S
13 mutations. Doubling time (DT) is indicated in the figure.

14

15 To determine the contribution of the mutated *lpd* allele to the fast growth phenotype
16 obtained under turbidostat regime, we restored the wild type *lpd* gene in the evolved
17 formate-complemented glycine auxotroph G4463 by P1 phage-mediated co-transduction
18 with the neighboring Keio marker $\Delta yacH::kan$. The growth of the recombinant derivatives
19 was then tested on semi solid MS glucose medium plus formate and CO₂ as precursors for
20 C1 and glycine synthesis. The results presented in Supplementary Figure S3 are compiled
21 in Table 3.

22

1 **Table 3.** Impact of the mutated *lpd* R273H gene on the growth phenotype of formate-
 2 complemented glycine auxotrophic strains.

Strain	Relevant genotype	Growth on MS glucose medium supplemented with the indicated nutrients*		
		0	glycine	formate 5% CO ₂
G4369	ancestor strain <i>ΔglyA ΔltaE Δ(tdh-kbl) ftfL⁺ ΔaceBAK folD R191S purR D307V pdeH Δ3bp</i>	-	+	-
G4400	evolved strain <i>ΔglyA ΔltaE Δ(tdh-kbl) ftfL⁺ ΔaceBAK folD R191S purR D307V pdeH Δ3bp rrfF</i>	-	+	+/-
G4463	evolved strain <i>ΔglyA ΔltaE Δ(tdh-kbl) ftfL⁺ ΔaceBAK folD R191S purR D307V pdeH Δ3bp rrfF lpd R273H</i>	-	+	+
G5037	G4463 <i>Δyach::kan lpd wt</i>	-	+	+/-
G5039	G4463 <i>Δyach::kan</i>	-	+	+

3 *Bacteria were streaked on semi-solid MS medium glucose 100 mM NH₄Cl supplemented with 2 mM glycine or 30 mM
 4 formate and incubated for 48 h at 37°C in air + 5% CO₂. -: no growth. +: robust growth. +/-: weak growth.

5

6 The wild type *lpd*-bearing strain G5037 formed colonies on formate-supplemented MS
 7 medium glucose plates as slowly as did strain G4400 (formate-complemented glycine
 8 auxotroph). By contrast, the control transductant G5039 obtained from G4463 by
 9 transferring only the marker *Δyach::kan*, grew in the presence of formate and CO₂ as
 10 glycine precursors almost as well as its mother strain, demonstrating that the growth
 11 acceleration observed in the turbidostat was directly related to the mutated Lpd enzyme.

12 To investigate if the mutations affecting the enzymes of the reductive glycine pathway
 13 were necessary and sufficient for formate-complemented growth of a glycine auxotroph,
 14 we introduced the mutated alleles of the genes *pdeH* (Δ3bp, next to the gene *ftfL*), *folD*
 15 (R191S) and *lpd* (R273H) into the unevolved glycine auxotroph G3715 (Supplementary
 16 Table S1). As shown in Supplementary Figure S4, among the various genetic constructs
 17 obtained tested for growth on formate and CO₂ as C1 and glycine precursors, only the
 18 strain G5137, which contains the three mutations, was able to grow.

19

3.2. Functional consequences of the mutations *pdeH* Δ 3bp, *foiD* R191S and *lpd* R273H entailing efficient formate assimilation.

3.2.1. *The short deletion upstream of formate-H₄F ligase-encoding gene cis-regulates its level of transcription.*

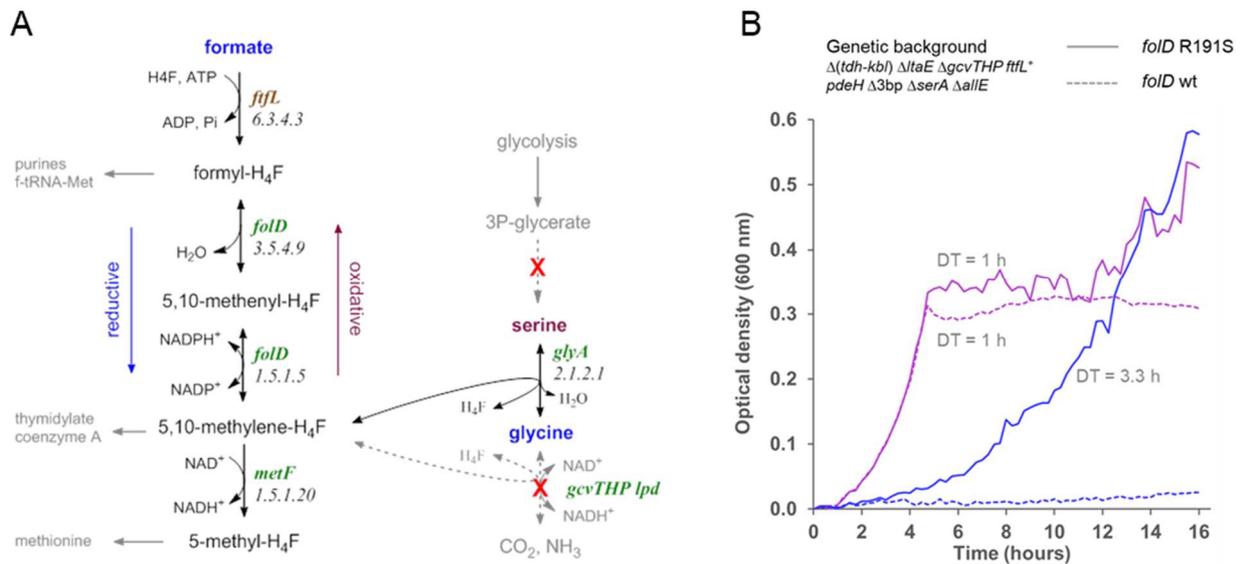
The activity of the GMP phosphodiesterase PdeH could not be directly related to the growth phenotype of strain G3806 (formate-complemented C1 auxotroph). Most probably, the Δ 3bp deletion in the *pdeH* gene, which is located upstream of the chromosomal insertion site of the heterologous gene *fflL*, is situated within a cis-regulating element which modulates the level of expression of the formate-H₄F ligase. Comparative RT-qPCR analysis of mRNA transcripts confirmed the impact on transcription of the deletion (Table 4), revealing the 60 times enhancement of gene *fflL* transcription level in the evolved strain G3806 and in the reconstructed strain G4552 in comparison to the wild type ancestor G3745 (C1 auxotroph). Most likely, this transcriptional boost augments the availability of formyl-H₄F in the cells, thus facilitating formate assimilation.

Table 4. Comparison of the transcriptional level of gene *fflL* *C. kluuyveri* in different genetic contexts determined by RT-qPCR. The relative levels of expression of *fflL* gene were evaluated by the calculation of Δ Ct using the housekeeping gene *panB* as endogenous control in bacteria from non-evolved strain G3745, evolved strain G3806 and reconstructed strain G4552. Inter-strain fold differences in expression of gene *fflL* were determined by the calculation of $\Delta\Delta$ Ct using Δ Ct of the non-evolved G3745 bacteria as reference.

strain	Δ Ct Ct <i>panB</i> <i>E. coli</i> – Ct <i>fflL</i> <i>C. kluuyveri</i>	$\Delta\Delta$ Ct	amplification factor $2^{\Delta\Delta$ Ct}
G3745	-1.60 +/- 0.06		
G3806	4.18 +/- 0.10	5.78	55
G4552	4.44 +/- 0.10	6.04	66

3.2.2. *Mutation in FoiD cyclohydrolase/methylene-H₄F dehydrogenase lifts inhibition of reductive reaction.*

1 In wild type *E. coli* strain MG1655, which does not express any formate-H₄F ligase activity,
 2 the H₄F-bound C1 moieties originate from the cleavage of the amino acids serine and
 3 glycine and the subsequent transfer of a hydroxymethyl group onto H₄F forming the
 4 metabolic intermediate methylene-H₄F. The latter is then interconverted to formyl-H₄F or
 5 methyl-H₄F by reversible redox reactions catalyzed, respectively, by *folD*-encoded
 6 bifunctional dehydrogenase-cyclohydrolase and *metF*-encoded methylene-H₄F reductase
 7 (Figure 4A). The selective constraints applied to the C1-auxotrophic cell population of
 8 strain G3745 during the evolution process towards formate assimilation into C1
 9 compounds (Döring et al, 2018) required the reductive activity of the FoLD enzyme for the
 10 production of methylene-H₄F from formate. Therefore, the *folD* R191S mutation is likely to
 11 have been selected because it enhances the non-physiological reverse function of the
 12 enzyme.



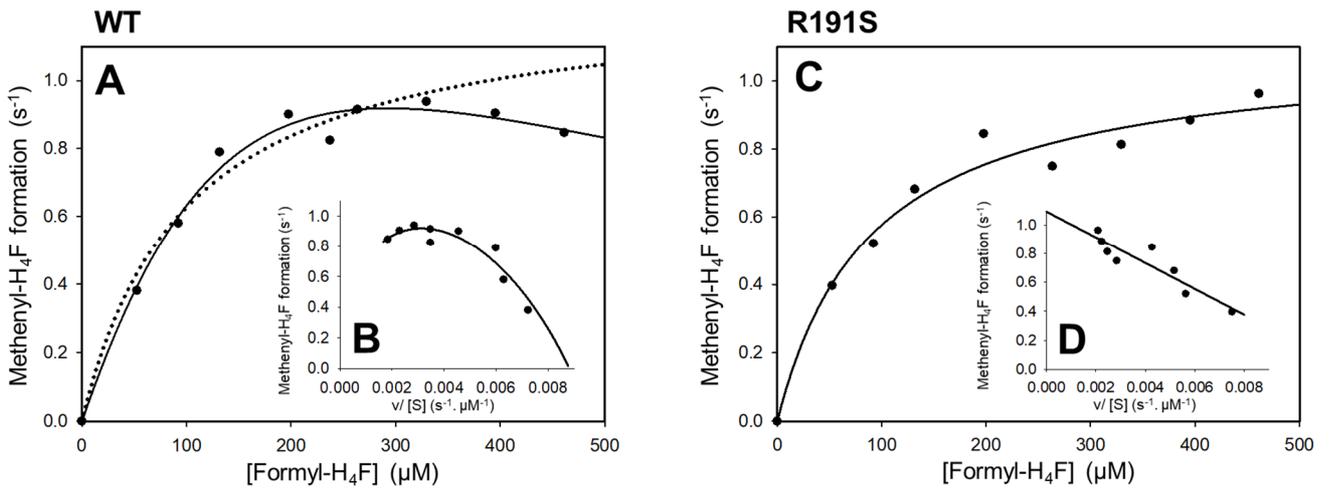
13
 14
 15 **Figure 4. A:** Serine, glycine and one-carbon metabolism in the serine-auxotrophic selection strains
 16 G4829 ($\Delta(tdh-kbl) \Delta ltaE \Delta gcvTHP ffl^+ pdeH \Delta 3bp \Delta serA \Delta allE$) and its *folD* R191S derivative
 17 G4830 ($\Delta(tdh-kbl) \Delta ltaE \Delta gcvTHP ffl^+ pdeH \Delta 3bp \Delta serA \Delta allE folD$ R191S). The deletion of gene
 18 *serA* and operon *gcvTHP* (indicated by red crosses) inactivates serine biosynthesis from 3-
 19 phosphoglycerate and H₄F-bound C1 synthesis from glycine, respectively. When serine is provided
 20 in the culture medium (indicated in purple), glycine and methylene-H₄F are formed through GlyA-
 21 mediated cleavage of serine; FoLD-catalyzed methylene-H₄F oxidation produces formyl-H₄F, which

1 is indispensable for purine synthesis. Conversely, when glycine and formate are provided
2 (indicated in blue), serine synthesis from glycine and methylene-H₄F depends on formate
3 activation to formyl-H₄F and subsequent FOLD-catalyzed reduction of formyl-H₄F. **B:** Growth of the
4 strains G4829 and G4830 on MS glucose medium supplemented with either serine (purple lines) or
5 glycine and formate (blue lines). Dotted lines: strain G4829 (*fold* wt). Solid lines: strain G4830 (*fold*
6 R191S). Doubling time (DT) is indicated in the figure.

7
8 Consequently, we elaborated a genetic context for testing *in vivo* the catalytic activity of
9 FOLD in the oxidative and the reductive direction, and compared both wild type and
10 mutated R191S forms of FOLD for their impact on growth under nutritional conditions where
11 cell proliferation strictly depends on FOLD activity. To this end, the serine auxotrophic strain
12 G4829 ($\Delta(tdh-kb)$ $\Delta ltaE$ $\Delta gcvTHP$ $ftfL^+$ $pdeH$ $\Delta 3bp$ $\Delta serA$ $\Delta allE::kan$) and its counterpart
13 G4830 carrying the *fold* R191S allele were constructed. As shown in Figure 4A, adding
14 either serine to the culture medium or glycine and formate should complement the serine
15 auxotrophy of these strains. In the case serine is added, methylene-H₄F and glycine are
16 formed from serine cleavage catalyzed by serine hydroxymethyltransferase GlyA;
17 methylene-H₄F is then oxidized to formyl-H₄F by FOLD. In the case glycine and formate are
18 provided, formate is first activated by formate-H₄F ligase to formyl-H₄F, the latter is then
19 reduced to methylene-H₄F by FOLD; serine is subsequently obtained by the condensation
20 of glycine and the C1 moiety of methylene-H₄F. As shown in Figure 4B, the two strains
21 grew similarly well when fed with serine. However, whereas strain G4830 (harboring *fold*
22 R191S) grew in the presence of glycine and formate, strain G4829 (harboring wild type
23 *fold*) did not, suggesting insufficient methylene-H₄F synthesis by wild type FOLD in this
24 strain.

25 To understand the mechanism underlying the higher activity of the FOLD mutant in the
26 reverse reactions (from formyl-H₄F to methylene-H₄F), we measured the cyclohydrolase
27 activity (formation of methenyl-H₄F from formyl-H₄F) of the FOLD wild type and the R191S
28 variant. In the presence of excess formyl-H₄F, the activity of wild type FOLD decreased,

1 pointing to substrate inhibition of the enzyme, which is unambiguously illustrated in Figure
 2 5B by the nonlinearity of the Eadie-Hofstee plot (activity rate v versus $v/[S]$) drawn from the
 3 data. The Michaelis-Menten model being inappropriate for fitting these data (Fig. 5A), we
 4 based our analysis on the standard kinetic model for substrate inhibition, in which the
 5 enzyme has catalytic activity when one substrate molecule is bound, but loses activity
 6 upon binding of the second substrate molecule.



7
 8 **Figure 5.** Comparison of the steady-state kinetics of methenyl-H₄F formation from formyl-H₄F by
 9 wild type FolD and R191S variant. **A:** Plot of reaction rate v versus $[S]$ for the wild type enzyme.
 10 Dotted line: hyperbolic curve fitted using the Michaelis-Menten equation, not taking into account
 11 the inhibited rates at high substrate concentrations:solid line, substrate inhibition curve fitted with
 12 eq. 1 (see text below). **B:** Eadie-Hofstee representation (v versus $v/[S]$) of the values presented in
 13 A. **C:** Plot of v versus $[S]$ for FolD R191S, with a line drawn using the Michaelis-Menten equation.
 14 **D:** Eadie-Hofstee representation (v versus $v/[S]$) of the values presented in C. Values correspond
 15 to the average of two replicates.

16
 17 Accordingly, values of v at various substrate concentrations were fit using nonlinear
 18 regression, based on equation 1:

$$v_i = \frac{V_{\max} [S]}{K_m + [S] + \frac{[S]^2}{K_i}} \quad (\text{eq. 1})$$

20 where K_i is the inhibition constant that defines the binding of the second, inhibiting
 21 substrate molecule. Kinetic parameters obtained from eq. 1 gave a theoretical k_{cat} value of
 22 3.1 s^{-1} (Table 5). However, data in Figure 5A show that the apparent V_{max} is 0.9 s^{-1} . This

1 reaction rate was obtained for the optimal formyl-H₄F concentration (292 μM) given by
 2 equation 2:

$$3 \quad S^{opt} = \sqrt{K_m K_I} \quad (\text{eq. 2})$$

4 The R191S mutation altered the kinetic regulation of the enzyme that no longer showed
 5 substrate inhibition. Kinetics of FoID R191S were hyperbolic and the corresponding Eadie-
 6 Hofstee plot was linear, as illustrated in Figures 5C and 5D and Table 5. The maximal
 7 activity, by contrast, was similar (0.9 vs 1.1 s⁻¹ for wild type and FoID R191S, respectively).

8
 9 **Table 5.** Kinetic parameters of wild type FoID and FoID R191S variant for reverse cyclohydrolase
 10 activity.

		Michaelian parameters		Non-Michaelian parameters		
Enzyme	Substrate	K_m (μM)	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat} (s ⁻¹)	K_i (μM)
FoID wt	formyl-H ₄ F			361 ± 229	3.1 ± 1.6	237 ± 177
FoID R191S	formyl-H ₄ F	91 ± 19	1.1 ± 0.1			

11 Experiments were conducted in the presence of 1 mM of 2',5'-ADP, an analogue and activator of
 12 the nicotinamide-dinucleotide cofactor (Pawelek and McKenzie, 1998). Values correspond to the
 13 average of two replicates.

14
 15 We also measured the reverse, reductive dehydrogenase reaction (from methenyl-H₄F to
 16 methylene-H₄F), although the simultaneous cyclohydrolase activity forming formyl-H₄F
 17 from methenyl-H₄F complicated the assay. Therefore, measured rates were probably
 18 underestimated (Pawelek and MacKenzie, 1998). Data in Supplementary Table S3
 19 suggest that the two enzymes have a comparable activity. The k_{cat} value for the overall
 20 reverse reaction (formyl-H₄F to methylene-H₄F, see Materials and methods, section 2.9 for
 21 details) was determined. For both enzyme variants, these values were found to be close to
 22 the k_{cat} values for the reverse cyclohydrolase reaction (Supplementary Table S3), meaning
 23 that almost all of the methenyl-H₄F formed by the cyclohydrolase is channeled through the

1 dehydrogenase. It also signifies that the cyclohydrolase activity is the rate-limiting step of
2 the overall reverse reductive reaction.
3 To rule out additional effects of the R191S mutation, we determined the catalytic
4 parameters of the forward reactions (from methylene-H₄F to formyl-H₄F). We found no
5 significant difference for the dehydrogenase (oxidative activity) nor the cyclohydrolase
6 activity parameters between the two enzyme variants in this direction (Supplementary
7 Tables S4 and S5). We also observed no shift in cofactor preference (NADP⁺ to NAD⁺)
8 due to the mutation (Supplementary Table S4). We therefore conclude that the FoID
9 variant enables a more efficient formate assimilation through the reductive glycine pathway
10 due to loss of inhibition of the reverse cyclohydrolase activity by its substrate formyl-H₄F,
11 in accordance with the concomitant overexpression of the formyl-H₄F ligase in the selected
12 cells (Table 4).

13

14 3.2.3. *The mutation R273H in Lpd modifies substrate affinities and reaction rate.*

15 The selection for enhanced growth on formate and CO₂ as C1 and glycine precursors was
16 linked to a mutation (R273H) in the second dehydrogenase of the reductive glycine
17 pathway, the lipoamide dehydrogenase (Lpd), which is part of the glycine cleavage
18 complex. Genomic reconstruction showed this mutation to confer a selective advantage to
19 the adapted strain. In an attempt to interpret this effect, the wild type and variant R273H
20 Lpd were kinetically characterized for both the oxidative and reductive reactions (Table 6).
21 For the oxidative reaction, the k_{cat} and the K_{m} for dihydrolipoamide of the mutated Lpd
22 R273H both decreased by a factor of ten in comparison with the wild type enzyme,
23 resulting in an unaltered catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) for dihydrolipoamide ($1.8 \cdot 10^6$ vs 1.2
24 $\cdot 10^6$ s⁻¹ · M⁻¹ for the wild type and R273H enzyme, respectively) (Table 6). Concomitantly,
25 the K_{m} of the mutant for the cofactor NAD⁺ increased by almost a factor of 10 (3.9 vs 0.44
26 mM).

1

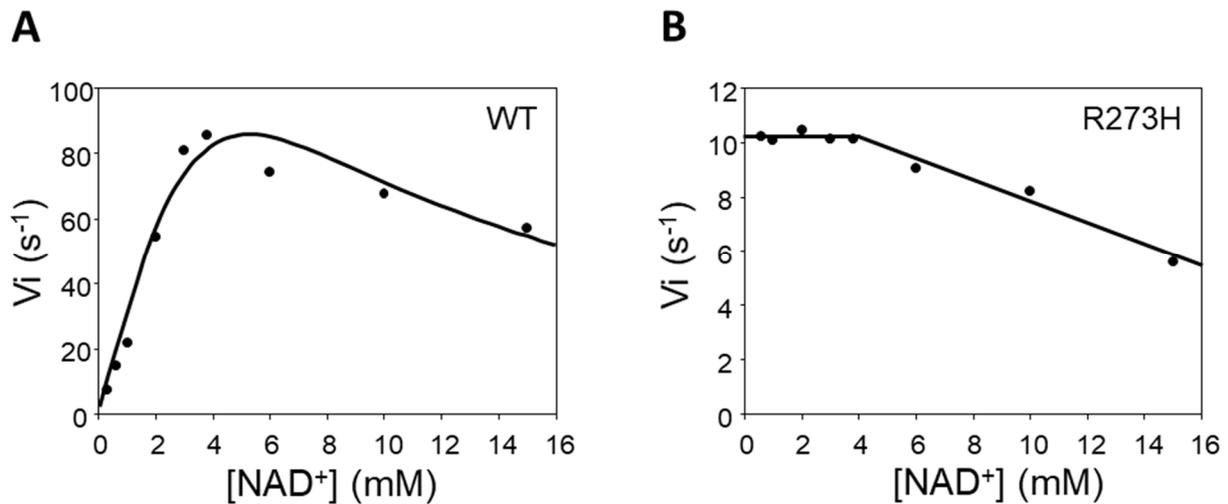
2 **Table 6.** Kinetic parameters of wild type *E. coli* lipoamide dehydrogenase and variant R273H.

Enzyme	Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}.M^{-1}$)
Lpd wt	dihydrolipoamide	0.197 ± 0.028	363 ± 65	$1.8 \cdot 10^6$
	NAD ⁺	0.437 ± 0.060		$8.3 \cdot 10^5$
Lpd R273H	dihydrolipoamide	0.026 ± 0.004	30.5 ± 1.9	$1.2 \cdot 10^6$
	NAD ⁺	3.888 ± 0.424		$7.8 \cdot 10^3$
Lpd wt	lipoamide	0.540 ± 0.090	51.4 ± 2.7	$9.5 \cdot 10^4$
	NADH	0.101 ± 0.009		$5.1 \cdot 10^5$
Lpd R273H	lipoamide	0.095 ± 0.004	3.5 ± 0.3	$3.7 \cdot 10^4$
	NADH	0.034 ± 0.011		$1.0 \cdot 10^5$

3

4 Regarding the reductive reaction, early work showed that *E. coli* Lpd is strongly inhibited
5 by NADH (Koike et al., 1960; Williams, 1965). NAD⁺ partially overcomes this inhibition and
6 it was shown that the initial velocity of the reaction relies on the concentration of added
7 NAD⁺, with an apparent binding constant of 400 μ M (Wilkinson and Williams, 1981).
8 Therefore, reductive activity measurements for the wild type enzyme were conducted in
9 the presence of 1 mM NAD⁺ as previously reported (Wilkinson and Williams, 1981). The
10 measured kinetic parameters for the wild type enzyme (Table 6) are thus apparent, since
11 the simultaneous presence of oxidized and reduced nucleotides can lead to the formation
12 of dead-end enzyme-nucleotide complexes. NADH reduces the two-electron (EH₂) to the
13 inactive four-electron (EH₄) state of the enzyme, and NAD⁺ inhibits by binding to the
14 oxidized enzyme (E_{ox}) (Schmincke-Ott and Bisswanger, 1981; Wilkinson and Williams,
15 1981).

16 It turns out that the main effect of the R273H mutation is to render the enzyme insensitive
17 to NADH as the dihydrolipoamide formation rate by the R273H variant is independent of
18 the presence of NAD⁺ (Figure 6).



1

2

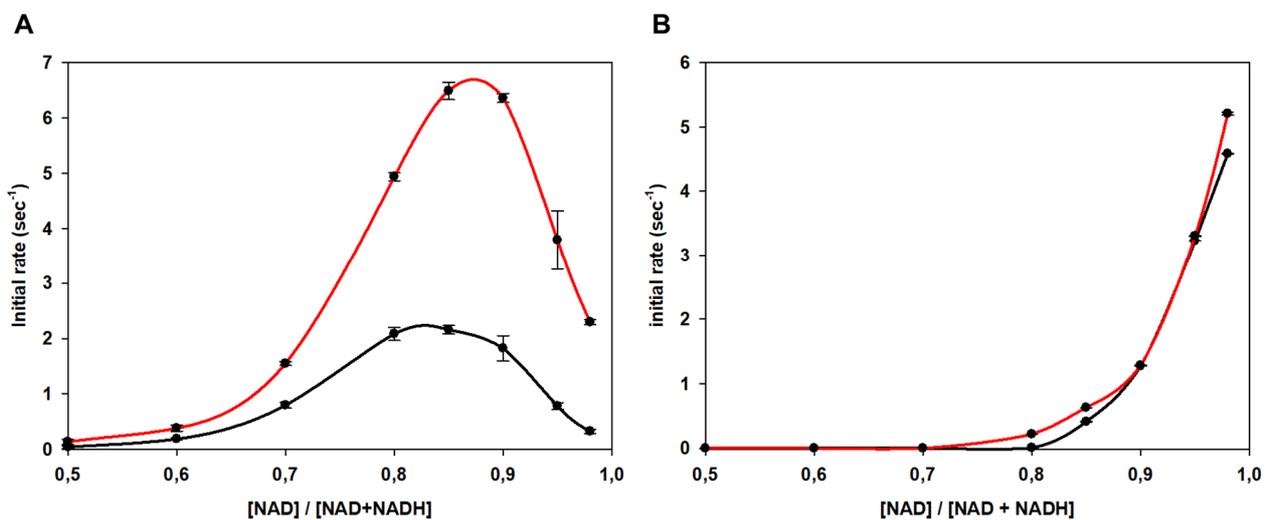
3 **Figure 6.** Comparison of the inhibition of lipoamide dehydrogenase activity by NADH of wild type
 4 Lpd and variant R273H. Initial activity rates for lipoamide reduction of wild type Lpd (A) and mutant
 5 Lpd R273H (B) were recorded as a function of NAD⁺ concentration in the presence of saturating
 6 concentrations of lipoamide (5 mM) and NADH (0.3 mM). Plotted values correspond to the average
 7 of two replicates.

8

9 Therefore, kinetic characterization of Lpd R273H was performed in the absence of added
 10 NAD⁺. The kinetic data indicated that, similar to what has been observed for the oxidative
 11 reaction, the mutation did not significantly affect the catalytic efficiency for lipoamide (9.5×10^4 vs 3.7×10^4 s⁻¹·M⁻¹ for the wild type and mutant enzyme, respectively) but decreased
 12 the k_{cat} of the mutant by more than 10 fold (3 vs 51 s⁻¹) and, to a slightly lesser extent, the
 13 K_m for lipoamide (95 vs 540 μM) (Table 6). The mutation also decreased the K_m value for
 14 NADH of the enzyme variant from 101 to 34 μM. This decrease in the K_m for NADH along
 15 with the increase in that for NAD⁺ could suggest that the mutation favors the enzyme
 16 activity in the reductive direction. Taking together these shifts in K_m with the observed
 17 insensitivity to NADH inhibition, the Lpd R273H mutant is susceptible, as part of the
 18 glycine cleavage complex, to assure an increased glycine synthesis thus endowing the cell
 19 with enhanced capabilities to use formate and CO₂ as carbon sources through the
 20 reductive glycine pathway. In order to further evaluate the possible physiological

1 advantage provided by the R273H mutation, the modulation of Lpd activity by the NAD⁺
 2 mole fraction $\text{NAD}^+ / (\text{NAD}^+ + \text{NADH})$ was determined. As has been shown for the *E. coli*
 3 pyruvate dehydrogenase complex, which is also strongly inhibited by NADH (Bremmer,
 4 1969), the rate of reaction is actually modulated by the degree of reduction of the pyridine
 5 nucleotide pool rather than by the NADH concentration (Shen and Atkinson, 1970). The
 6 experiments were performed by setting constant the total nucleotide pool at 2683 μM ,
 7 value measured for the *E. coli* cytoplasm (Bennett et al., 2009), and varying the relative
 8 amounts of the oxidized and reduced forms. As shown in Figure 7A, the Lpd R273H
 9 variant was several folds more efficient in the reductive reaction for a NAD⁺ mole fraction
 10 between 0.8 and 0.97, which is believed to correspond to the physiological range in
 11 aerobiosis (Hansen and Henning, 1966; Bennett et al., 2009; Shen and Atkinson, 1970).
 12 By contrast, the experiments showed no difference between the two enzymes in the
 13 oxidative reaction (Figure 7B).

14



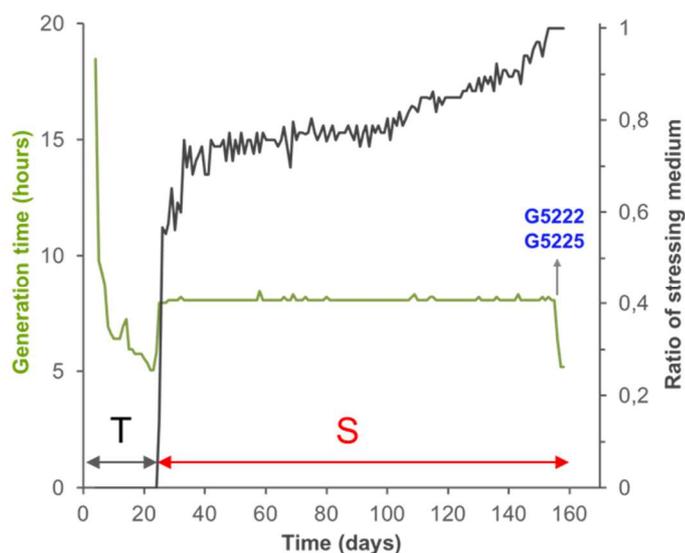
15

16 **Figure 7.** Activity of wild type Lpd and mutant Lpd R273H as a function of the NAD⁺ mole fraction
 17 of the nicotinamide cofactor pool. The relative concentrations of NAD⁺ to NADH were varied, by
 18 keeping the total cofactor amount constant at 2683 μM . **A:** Initial reaction rates of the reduction of
 19 lipoamide (1 mM). **B:** Initial reaction rates of the oxidation of dihydrolipoamide (1 mM). Each assay
 20 was repeated thrice; \pm standard deviation is indicated. Black line: wild type Lpd. Red line: mutant
 21 Lpd R273H.

1

2 3.3. Adaptation to growth on formate and CO₂ as sole carbon sources

3 Based on the results of the genetic studies outlined above, the formate-complemented
4 glycine-auxotroph G4463 was used as starting point for completing the installation of a
5 functional reductive glycine pathway in the *E. coli* lineage towards synthetic formatotrophy
6 (Supplementary Table S1 for strain genotypes and Supplementary Figure S1 for strain
7 lineage). To this aim, the *glyA* wild type gene encoding serine hydroxymethyltransferase
8 was first restored on the chromosome of G4463 strain by P1 phage-mediated
9 transduction. The resulting strain G4985 ($\Delta(tdh-kbl)$ $\Delta ltaE$ $\Delta aceBAK$ *ftfL+* *pdeH* $\Delta 3bp$ *folD*
10 *R191S* *lpd* *R273H*) was able to grow, albeit slowly, when fed with both formate and glycine
11 as carbon sources, and provided with acetate as an energy source (because the glyoxylic
12 by-pass was not functional in the metabolic context of strain G4985, acetate did not
13 contribute to a net gain of carbon atoms into biomass aside probable assimilation into fatty
14 acids). This strongly suggested that the two last steps of the reductive glycine pathway
15 were operational, *i.e.* the GlyA-catalyzed condensation of glycine with a C1-moiety to
16 serine and subsequent serine deamination to pyruvate. No growth was observed when
17 formate was provided as sole carbon source in the presence of acetate and a CO₂-
18 enriched aeration gas, pointing to insufficient glycine synthesis from formate for sustaining
19 carbon flux. To select for a higher carbon flux from formate to pyruvate through the
20 pathway, we again resorted to continuous culture protocols using automaton-driven GM3
21 devices (Figure 8).



1
2 **Figure 8.** Adaptation of G4985 bacteria ($\Delta(tdh-kbl) \Delta ltaE \Delta aceBAK ftfL+ pdeH \Delta 3bp folD$ R191S
3 *lpd* R273H) to formate utilization as sole carbon source (acetate as energy source) by directed
4 evolution in continuous culture. The bacteria were initially adapted to grow in permissive medium
5 (MS 15 mM glycine 45 mM formate 15 mM acetate) under turbidostat regime (**T**, black). From day
6 25, medium swap regime was applied to the culture with a steady generation time set to 8 hours
7 (**S**, red). Stressing medium consisted in MS 45 mM formate 15 mM acetate. The aeration gas of the
8 culture consisted in a mix of 95 % air – 5% CO₂. Maximum ratio of stressing medium was attained
9 after 460 generations. Green line: evolution of the generation time of the culture. Black line: daily
10 ratio of stressing medium dilution pulses. The grey arrow shows the time-point of the culture
11 sample from which the indicated strains were isolated.

12
13 First, the growth rate of a G4985 bacterial culture was improved under turbidostat regime
14 on a permissive medium consisting of MS medium supplemented with glycine and formate
15 as carbon sources and acetate as an energy source (refer to Materials and methods,
16 section 2.4). The aeration gas of the culture consisted of an air 95%-CO₂ 5% mix. Once
17 the culture steadily grew (generation time around 5h), a medium-swap regime was applied
18 that alternatively diluted the growing bacterial population with the same permissive
19 medium or the stressing medium consisting of MS plus formate as carbon source and
20 acetate as an energy source. The generation time of the culture was set to 8 hours
21 (Materials and methods, section 2.4). This weaning regime selected for growth in the

1 presence of ever lower concentrations of glycine throughout successive generations until
2 the bacteria no longer required the supply of this nutrient. 460 generations were necessary
3 to obtain stable growth in the stressing medium, *i.e.* growth with formate and CO₂ as
4 carbon source (acetate as energy source) (Figure 8). Clones were isolated on plates under
5 the same feeding conditions. Growth of the evolved isolates G5222 and G5225 strictly
6 depended on the joint availability of the three compounds, formate, acetate and CO₂.

7 The mutations fixed during the adaptation to formate utilization as carbon source (and
8 acetate as energy source) in the isolates G5222 and G5225 were identified by whole-
9 genome sequencing and mutational analysis (Materials and methods, section 2.5). They
10 are listed in Table 1 together with the mutations found earlier for formate-complemented
11 C1-auxotroph G3806 and glycine-auxotrophs G4400 and G4463 (Table 1 and
12 Supplementary Table S2 for an exhaustive list). Single nucleotide polymorphisms as well
13 as short insertions and deletions were detected by mapping the sequence reads onto the
14 genome of the ancestor strain G4369 (glycine auxotroph) used as reference. Twenty-
15 seven single nucleotide polymorphisms were fixed in the evolved strains, 20 non-
16 synonymous intragenic and 7 intergenic mutations in strain G5222, 21 non-synonymous
17 intragenic and 6 intergenic in strain G5225 (Supplementary Table S2). None of these
18 mutations affected a gene encoding an enzyme of the reductive glycine pathway. Several
19 transcriptional regulators were mutated (Lrp, Rho, NadR). The analysis of the read
20 coverage of the genome revealed a chromosomal deletion of 34.4 kb. The deleted
21 fragment contains the *mutY* gene coding for an A:G mismatch glycosylase involved in
22 DNA repair. A defective *mutY* allele confers a mutator phenotype to *E. coli*, increasing the
23 rate of G:C to T:A transversions (Nghiem et al., 1988). Accordingly, 27 out of the 29
24 mutations detected were transversions.

25

3.4. Genetic studies and metabolic labeling experiments confirm formatotrophic growth via the reductive glycine pathway

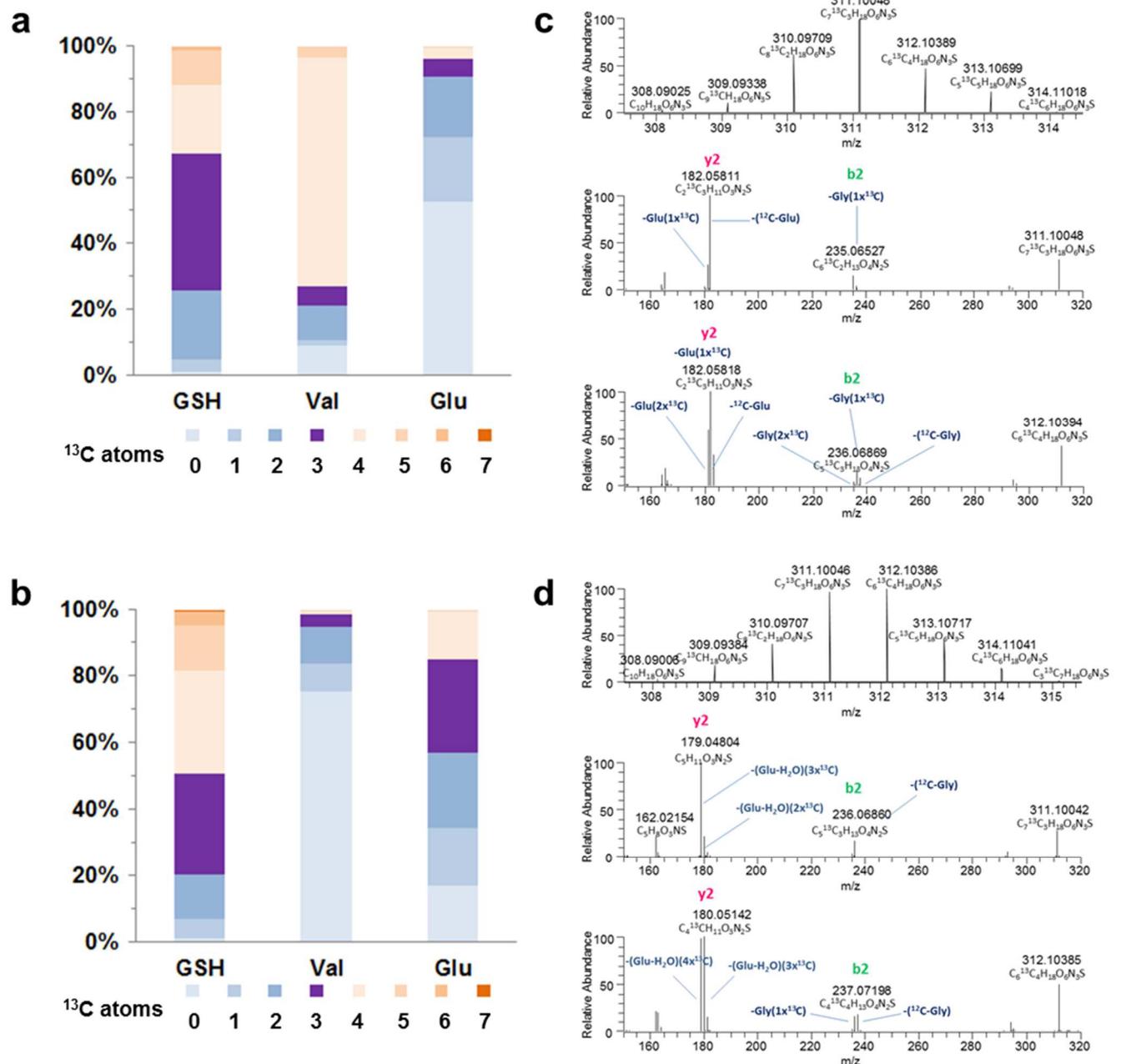
To confirm that formate indeed sustained central carbon metabolism for biomass synthesis through the reductive glycine pathway, we tested the essentiality of the key enzymes of the pathway for growth in the conditions of selection (MS formate acetate + 5% CO₂). The heterologous gene *ftfL* of formate H₄F-ligase, *gcvTHP* operon encoding three subunits of the glycine cleavage system, *glyA* encoding serine hydroxymethyltransferase as well as the *sdaA* gene of serine deaminase were replaced one by one by a selectable deletion marker on the chromosome of G5222 and G5225 strains (Supplementary Table S1). As shown in Supplementary Figure S5, none of the resulting derivative strains was able to grow in the selective conditions applied, contrary to the undeleted mother strains. These results strongly suggested that all targeted enzymatic steps were essential for formate metabolic assimilation as sole carbon source. By contrast, the deletion of the gene *serA* (encoding 3-phosphoglycerate dehydrogenase, the first enzyme of canonical serine biosynthesis route from the glycolysis intermediate 3-phosphoglycerate) from the chromosome of the strains G5222 and G5225 did not affect formate-dependent growth of the derived strains (Supplementary Figure S5). This strongly supported that, in the evolved strains G5222 and G5225 grown on formate, serine and glycine did not derive from 3-phosphoglycerate, but most probably were synthesized from formate and CO₂ according to the reductive glycine pathway.

Furthermore, we questioned the essentiality of the mutated genes *pdeH* (Δ 3bp), *folD* (R191S) and *lpd* (R273H) for formate assimilation as a carbon source by restoring one by one the wild type counterpart into the evolved strains G5222 and G5225 by P1 phage-mediated transduction. As shown in Supplementary Figure S6, the resulting strains had lost the ability to grow when fed with formate as sole carbon source. Thus, the mutations,

1 which were fixed during the first two adaptive evolution steps, remained necessary to the
2 formatotrophic growth phenotype.

3 To demonstrate that growth on formate proceeds via glycine and serine, two intermediates
4 of the reductive glycine pathway, and to shed light on carbon fluxes, we carried out carbon
5 labeling experiments by feeding bacteria of strains G5222 and G5225 either on ^{12}C -
6 formate plus $^{13}\text{C}_2$ -acetate or on ^{13}C -formate plus ^{12}C -acetate in the presence of 95% air -
7 5% CO_2 . ^{13}C -labeling patterns of the abundant tripeptide glutathione (γ -glutamyl-cysteinyl-
8 glycine) as well as of the amino acids valine and glutamate present in the extracted cell
9 metabolomes were determined using liquid chromatography coupled to mass spectrometry
10 (refer to Materials and methods, section 2.6).

11 When the ^{13}C -atoms originated from ^{13}C -formate, the analysis of the mass spectrum of
12 glutathione indicated that a major fraction of glutathione molecules contained at least three
13 ^{13}C -labels (Figure 9a). Mass-spectrometry fragmentation analysis (Döring et al., 2018) of
14 the molecules carrying three ^{13}C -labels (major ion in the isotopologue cluster at m/z 311)
15 clearly assigned one ^{13}C -label to glycine, two ^{13}C -labels to the cysteinyl moiety and no
16 label to the glutamyl moiety (Figure 9c). The second most abundant ion at m/z 312, which
17 carries four ^{13}C -labels, contained one additional ^{13}C -label in the glutamyl moiety of the
18 tripeptide. This labeling pattern is in agreement with the synthesis of glycine and serine
19 (the biosynthetic precursor of cysteine) through the reductive glycine pathway, for which
20 ^{13}C -labeling was expected at C_2 of glycine and at C_2 and C_3 of serine (Supplementary
21 Figure S7). A major fraction of valine molecules harbored four ^{13}C -labels (Figure 9a).



1
2 **Figure 9.** Labeling patterns of glutathione and of the amino acids valine and glutamate confirm the
3 assimilation of formate via the reductive glycine pathway. Isotopic distribution of glutathione, valine
4 and glutamate in strains G5222 grown in 95% air - 5% CO₂ on MS medium supplemented with
5 ¹³C-formate and ¹²C-acetate (a) or ¹²C-formate and ¹³C-acetate (b). LC-MS/MS² fragmentation
6 pattern of glutathione in the strain G5222 grown in 95% air - 5% CO₂ on MS medium
7 supplemented with ¹³C-formate and ¹²C-acetate (c) or ¹²C-formate and ¹³C-acetate (d).
8
9 As valine originates from the condensation of two pyruvate molecules with the loss of one
10 carboxyl carbon, we could infer that pyruvate is labeled in carbon C₂ and C₃ as predicted
11 for its synthesis via the reductive glycine pathway (Supplementary Figure S7).

1 Interestingly, glutamate, which derives from α -ketoglutarate, was partially labeled as well,
2 with 48 % of the molecules carrying at least one ^{13}C -label (Figure 9a). This is a clear
3 indication of an active anaplerotic carbon flux, assuring the tricarboxylic acid cycle to
4 operate and to provide essential biomass precursors.

5 When the ^{13}C -atoms originated from ^{13}C -acetate (Figure 9b,d), fragmentation analysis of
6 the protonated glutathione molecules carrying three or four ^{13}C -labels (major ions in the
7 isotopologue cluster at m/z 311 and m/z 312) assigned the ^{13}C -labels mostly to the
8 glutamyl moiety of the tripeptide. Neither glycy- nor cysteinyl-residues of glutathione
9 contained ^{13}C atoms in the ion at m/z 311. These residues remained also mostly unlabeled
10 in the ion at m/z 312 (Figure 9d). These results indicated that ^{13}C carbon atoms of acetate
11 were not channeled towards glycine and serine synthesis. A more complex picture of
12 metabolic labeling with ^{13}C -acetate was obtained for valine, which is directly produced
13 from pyruvate. Most valine molecules (75%) were unlabeled (Figure 9b), in agreement with
14 ^{12}C -formate sustaining pyruvate synthesis through the reductive glycine pathway
15 (Supplementary Figure S7). However, a small but notable fraction of valine contained one
16 or more ^{13}C -labels, most likely reflecting carbon flux from tricarboxylic acid cycle to
17 pyruvate through the malic enzyme reaction. As stated above, this flux does not signify
18 any net gain of biomass from acetate. The fact that glutamate was mostly multi-labeled
19 (Figure 9b) indicates an active carbon flux through the tricarboxylic acid cycle. However, a
20 significant amount of glutamate molecules (17%) was unlabeled. Again, this revealed
21 metabolic flow from the upper part of the central carbon metabolism towards tricarboxylic
22 acid cycle through anaplerotic reactions and acetyl-CoA production from pyruvate,
23 although acetate was added in the culture medium.

24

25

26

4. Discussion

In the present study, we evolved *E. coli* strains to grow on formate and CO₂ as carbon source applying different continuous culture regimes, which enabled the selection of new growth phenotypes and of faster growth towards optimized metabolic carbon flux. Carbon assimilation in the selected cells proceeds via the reductive glycine pathway (Figure 1), which has been proposed as the most favorable aerobic route for synthetic formatotrophy in terms of biomass yield and flexibility, enabling also methanol assimilation (Bar-Even et al., 2013; Cotton et al., 2020). The linearity of the reductive glycine pathway, which passes through essential metabolic intermediates, *i.e.* the folate cofactors, glycine and serine, enabled a modular selection strategy for the installation of formate assimilation, linking cellular growth of purposely constructed *E. coli* auxotrophs to a stepwise increasing part of the carbon biomass produced from formate (Döring et al., 2018; Yishai et al., 2018). With the exception of formate-H₄F ligase necessary to charge H₄F with the oxidized C1 feedstock to form 5,10-formyl-H₄F, all enzymes of the synthetic pathway belong to the canonical metabolic repertoire of *E. coli* cells.

Synthesizing pyruvate from two formate and one CO₂ molecules in *E. coli* via the reductive glycine route comes with the major challenge of redirecting cellular dehydrogenase activities, usually operating in the oxidative sense during heterotrophic growth on multicarbon reduced compounds, towards reduction. This involves the bifunctional 5,10-methylene-H₄F dehydrogenase/5,10-methenyl-H₄F cyclohydrolase FoID and the lipoamide dehydrogenase Lpd as part of the glycine cleavage enzymatic complex, which together transfer the six electrons necessary for pyruvate formation from either NADPH or NADH. Consequently, mutations affecting catalytic activity and expression of these enzymes were found in isolates obtained at different steps of the continuous culture adaptations.

1 The mutation R191S in F_oID appeared during the first adaptation stage to formate
2 utilization as C₁ precursor (strain G3806). The substitution of serine for arginine at position
3 191 had very limited effect on both the affinity of NADPH and the reaction rate in the
4 oxidative and reductive sense, and did not measurably improve the coupling between
5 dehydrogenase and cyclohydrolase activities of the bifunctional enzyme. In an early study,
6 F_oID of *E. coli* was described to be inhibited by formyl-H₄F (Dev and Harvey, 1978). We
7 confirmed this substrate inhibition for the cyclohydrolase activity of the wildtype enzyme
8 operating in the reductive sense. By contrast, high formyl-H₄F concentrations had no effect
9 on the cyclohydrolase activity of the R191S variant. The loss of substrate inhibition
10 provides an explanation for the emergence of this mutation in the population, which had
11 evolved to grow on formate as precursor of C₁-compounds, and the sustained presence of
12 this mutation over the entire evolution towards formatotrophy. The presumably high formyl-
13 H₄F pool in strain G3806, a consequence of the 60 times higher transcriptional rate of
14 gene *ftfL* measured in this strain, favors its growth on formate. For a full benefit of this rise
15 in formyl-H₄F concentration, the elimination of substrate inhibition in variant R191S
16 allowing for an efficient reduction of formyl-H₄F to methylene-H₄F is a prerequisite, as
17 demonstrated by the lowered growth on glycine and formate of the G3806 derived strain
18 G4235 carrying the *fold* wild type gene (Fig. 2A). Interestingly, Yishai et al. (2018)
19 observed that overexpression of *E. coli* wild type gene *fold* placed under a strong
20 constitutive promoter from a medium-copy number plasmid did not promote growth of a
21 glycine auxotroph in the presence of formate under selective conditions. In this study,
22 glycine complementation was obtained when the two genes *mdhA* and *fch* of the
23 facultative methylotroph *Methylobacterium extorquens*, which encode the independent
24 enzymes methylene-H₄F dehydrogenase and cyclohydrolase, were co-expressed. These
25 enzymes, as part of the C₁-assimilating serine cycle, function in the reductive sense in *M.*
26 *extorquens*, as opposed to *E. coli* F_oID (Crowther et al., 2008).

1 The activity of F_oD seemed to remain a bottleneck for formatotrophic growth, as illustrated
2 by the chromosomal amplifications of the gene locus during evolution (Table 1). While the
3 R191S mutation was sufficient for growth on formate as C1 source, a more stringent
4 formate dependence required the doubling (formate as C1 and glycine precursor) and
5 even the quadruplicating (accelerated growth, formate as biomass precursor) of the
6 mutated locus. It must be noted that the amplification of the *ftfL* locus followed a different
7 path: the duplication of the locus, found upon growth on formate as C1 and glycine
8 precursor, disappeared during later stages of the evolution, suggesting that other
9 mutations lowered the need for *ftfL* overexpression.

10 The selection for enhanced growth on formate and CO₂ as C1 and glycine precursors was
11 linked to a mutation in the second dehydrogenase of the pathway, the lipoamide
12 dehydrogenase, which is part of the GCV complex. The Lpd variant R273H showed loss of
13 activity inhibition by NADH and was more active in the reductive sense at
14 NAD/(NADH+NAD) ratios in the physiological range. The lack of sensitivity of Lpd R273H
15 towards NADH evokes the behavior of the pig heart enzyme (Matthews and Williams,
16 1976). The Lpd redox potentials from *E. coli* and pig heart have been determined and their
17 impact on catalytic properties hypothesized (Wilkinson and Williams, 1981; Wilkinson and
18 Williams, 1979). It was suggested that the larger gap between the E_{ox}/EH₂ and EH₂/EH₄
19 redox potentials of the pig enzyme (-66 mV) compared to the value determined for the *E.*
20 *coli* enzyme (-53 mV), accounts for the easier reduction of this latter enzyme to the
21 inactive 4-electron state. This suggests that the R273H mutation might alter the enzyme's
22 redox properties by augmenting the thermodynamic barrier against EH₂ reduction by
23 NADH. While the residue R273 is not expected to interact with FAD (Chandrasekhar et al.,
24 2013), the redox potential of the flavin cofactor may be influenced by residues located
25 outside its direct vicinity. To verify this hypothesis, a measurement of the corresponding
26 redox potentials of the R273H variant would have to be performed. In an earlier study, Kim

1 and co-workers constructed an *E. coli* strain unable to grow under anaerobic conditions
2 due to a high NADH/NAD⁺ ratio in the cells. Growth was rescued by point mutations in the
3 gene *lpd* coding for two enzyme variants (H322Y / E354K) with strongly reduced sensitivity
4 for NADH (Kim et al., 2008).

5 The adaptive mutations affecting the enzymes catalyzing the reactions of the reductive
6 glycine pathway were indispensable for growth of the evolved bacteria on formate as sole
7 carbon source. Additional mutations, which were fixed during the ultimate evolution step
8 implementing formate utilization as a carbon source, impacted several transcriptional
9 regulators such as the leucine responsive regulatory protein Lrp, which regulates about
10 10% of all genes (Tani et al., 2002), the global regulator of transcription termination and
11 gene expression Rho (Richardson, 1990), as well as NadR, a repressor of transport and
12 *de novo* synthesis of the nicotinamide cofactors (Gerasimova and Gelfand, 2005). The
13 RNA polymerase subunit RpoD was also mutated. While the activities of these proteins
14 are not directly linked to the formate assimilatory pathway, their mutant forms might trigger
15 pleiotropic effects on cell growth via an altered expression of numerous genes. The
16 transcription regulator Rho, the repressor NadR as well as various subunits of the RNA
17 polymerase were repeatedly found to be mutated in the course of long-term laboratory
18 evolution experiments and are key contributors for the acquisition of novel complex
19 phenotypes (Conrad et al., 2009; Wyttock et al., 2018).

20 Kim *et al.*, 2020 recently reported the implementation of the reductive glycine pathway in
21 *E. coli*, also following a stepwise approach of formate assimilation. However, they resorted
22 to a rational design and build strategy. Stable growth on formate as carbon and acetate as
23 energy source required that all enzymes involved in the pathway were expressed from
24 genes inserted in the chromosome and placed under strong synthetic promoters or a
25 modified RBS for enhanced translation. Using the *M. extorquens* genes *mdhA* and *fch*
26 coding for methylene-H4F dehydrogenase and cyclohydrolase, they obtained slow but

1 sustained growth without secondary mutations modulating enzyme activity or expression.
2 By contrast, our evolutionary approach based on successive selections was successful in
3 establishing the desired growth phenotypes through a limited number of mutations in key
4 enzymes of the targeted pathway (unless a mutator phenotype was acquired), pointing to
5 the finely tunable flexibility of enzymatic catalysts and the remarkable adaptability of *E. coli*
6 metabolism. When Kim and coworkers went on and installed formatotrophy in their strains
7 by expressing a heterologous formate dehydrogenase providing energy from formate, they
8 applied a directed evolution protocol to lower the doubling time of their strain from around
9 70 hours to less than 10, again showing the power of this method for the realization of
10 ambitious projects of synthetic biology.

Author contributions

V.D., M.B. conceived and supervised the project. A.P. designed and supervised the biochemical experiments. V.A.D. constructed all genetic contexts, performed growth assays, prepared metabolomes. N.P., O.M., M.F. performed FoID and Lpd enzymatic analysis. D.R. performed bioinformatic mutational analysis. E.D. performed LC/MS experiments and analyzed fragmentation data. I.D. conducted continuous cultures in GM3 automatons. E.P. performed RT-qPCR experiments. V.D., M.B., A.P. analyzed data. V.D., M.B., A.P. wrote the manuscript. All authors commented on and approved the final manuscript.

Declaration of competing interests

The authors declare no competing financial interest.

Acknowledgements

This research was supported by the Commissariat à l'Énergie Atomique (CEA), France and by the European Union's Horizon 2020 research and innovation program under grant agreement No 763911. The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the MicroScope annotation platform. The authors thank Dr. Arren Bar-Even for fruitful discussions, Aude Perdereau and Genoscope's sequencing team for genome sequencing, Julien Patrouix for technical support for gas-mixing GM3 devices, Nathalie Vega-Czarny

for technical assistance with continuous cultures, Peggy Sirvain for protein purifications and Carine Vergne-Vaxelaire and Aurélie Fossey-Jouenne for dihydrolipoamide synthesis.

Appendix A. Supplementary data

The following is the Supplementary data to this article:

Supplementary Figures S1 to S7

Supplementary Tables S1 to S5

References

- Alissandratos, A., Easton, C. J., 2015. Biocatalysis for the application of CO₂ as a chemical feedstock. *Beilstein Journal of Organic Chemistry*. 11, 2370-87.
- Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N., Davidi, D., Bar-On, Y., Bareia, T., Wernick, David G., Shani, I., Malitsky, S., Jona, G., Bar-Even, A., Milo, R., 2016. Sugar Synthesis from CO₂ in *Escherichia coli*. *Cell*. 166, 115-125.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., Mori, H., 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*. 2, 2006 0008.
- Bar-Even, A., Noor, E., Flamholz, A., Milo, R., 2013. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochim Biophys Acta*. 1827, 1039-47.
- Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., Rabinowitz, J. D., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol*. 5, 593-9.
- Bowker, M., 2019. Methanol Synthesis from CO₂ Hydrogenation. *ChemCatChem* 11, 4238-4246.
- Bremmer, J., 1969. Pyruvate Dehydrogenase, Substrate Specificity and Product Inhibition. *European Journal of Biochemistry*. 8, 535-540.
- Chandrasekhar, K., Wang, J., Arjunan, P., Sax, M., Park, Y.-H., Nemeria, N. S., Kumaran, S., Song, J., Jordan, F., Furey, W., 2013. Insight to the Interaction of the Dihydrolipoamide Acetyltransferase (E2) Core with the Peripheral Components in the *Escherichia coli* Pyruvate Dehydrogenase Complex via Multifaceted Structural Approaches. *Journal of Biological Chemistry*. 288, 15402-15417.

- Chen, F. Y., Jung, H. W., Tsuei, C. Y., Liao, J. C., 2020. Converting *Escherichia coli* to a Synthetic Methyloph Growing Solely on Methanol. *Cell*. 182, 933-946 e14.
- Chen, Z., Geng, F., Zeng, A. P., 2015. Protein design and engineering of a *de novo* pathway for microbial production of 1,3-propanediol from glucose. *Biotechnology Journal*. 10, 284-9.
- Chistoserdova, L., 2018. Applications of methylophs: can single carbon be harnessed for biotechnology? *Current Opinion in Biotechnology*. 50, 189-194.
- Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y., Palsson, B. O., 2009. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biology*. 10, R118.
- Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S., Bar-Even, A., 2020. Renewable methanol and formate as microbial feedstocks. *Current Opinion in Biotechnology*. 62, 168-180.
- Crowther, G. J., Kosaly, G., Lidstrom, M. E., 2008. Formate as the main branch point for methylophic metabolism in *Methylobacterium extorquens* AM1. *J Bacteriol*. 190, 5057-62.
- Dev, I. K., Harvey, R. J., 1978. A complex of N5,N10-methylenetetrahydrofolate dehydrogenase and N5,N10-methenyltetrahydrofolate cyclohydrolase in *Escherichia coli*. Purification, subunit structure, and allosteric inhibition by N10-formyltetrahydrofolate. *Journal of Biological Chemistry*. 253, 4245-4253.
- Döring, V., Darii, E., Yishai, O., Bar-Even, A., Bouzon, M., 2018. Implementation of a Reductive Route of One-Carbon Assimilation in *Escherichia coli* through Directed Evolution. *ACS Synthetic Biology*. 7, 2029-2036.
- Gerasimova, A. V., Gelfand, M. S., 2005. Evolution of the NadR regulon in *Enterobacteriaceae*. *Journal of Bioinformatics and Computational Biology*. 03, 1007-1019.
- Gleizer, S., Ben-Nissan, R., Bar-On, Y. M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., Milo, R., 2019. Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO₂. *Cell*. 179, 1255-1263 e12.
- Hansen, R.G., Henning, U., 1966. Regulation of pyruvate dehydrogenase activity in *Escherichia coli* K12. *Biochim Biophys Acta*. 122, 355-358.
- Herz, E., Antonovsky, N., Bar-On, Y., Davidi, D., Gleizer, S., Prywes, N., Noda-Garcia, L., Lyn Frisch, K., Zohar, Y., Wernick, D. G., Savidor, A., Barenholz, U., Milo, R., 2017. The genetic basis for the adaptation of *E. coli* to sugar synthesis from CO₂. *Nature Communications*. 8, 1705.
- Jiang, W., Hernandez Villamor, D., Peng, H., Chen, J., Liu, L., Haritos, V., Ledesma-Amaro, R., 2021. Metabolic engineering strategies to enable microbial utilization of C1 feedstocks. *Nat Chem Biol*. 17, 845-855.
- Jouny, M., Luc, W., Jiao, F., 2018. General Techno-Economic Analysis of CO₂ Electrolysis Systems. *Ind. Eng. Chem. Res*. 57, 2165-2177.
- Keller, P., Noor, E., Meyer, F., Reiter, M. A., Anastassov, S., Kiefer, P., Vorholt, J. A., 2020. Methanol-dependent *Escherichia coli* strains with a complete ribulose monophosphate cycle. *Nature Communications*. 11, 5403.

- Kim, Y., Ingram, L. O., Shanmugam, K. T., 2008. Dihydrolipoamide Dehydrogenase Mutation Alters the NADH Sensitivity of Pyruvate Dehydrogenase Complex of *Escherichia coli* K-12. *J. Bacteriol.* 190, 3851-3858.
- Kim, S., Lindner, S. N., Aslan, S., Yishai, O., Wenk, S., Schann, K., Bar-Even, A., 2020. Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nat Chem Biol.*
- Koike, M., Shah, P. C., Reed, L. J., 1960. alpha-Keto acid dehydrogenation complexes. III. Purification and properties of dihydrolipoic dehydrogenase of *Escherichia coli*. *J Biol Chem.* 235, 1939-43.
- Marlière, P., Patrouix, J., Döring, V., Herdewijn, P., Tricot, S., Cruveiller, S., Bouzon, M, Mutzel, R., 2011. Chemical Evolution of a Bacterium's Genome. *Angew. Chem. Int. Ed.* 50, 7109–7114
- Matthews, R. G., Williams, C. H., Jr., 1976. Measurement of the oxidation-reduction potentials for two-electron and four-electron reduction of lipoamide dehydrogenase from pig heart. *Journal of Biological Chemistry.* 251, 3956-3964.
- Meyer, F., Keller, P., Hartl, J., Groninger, O. G., Kiefer, P., Vorholt, J. A., 2018. Methanol-essential growth of *Escherichia coli*. *Nature Communications.* 9, 1508.
- Miller, J. H., 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, N.Y.
- Muller, J. E. N., Meyer, F., Litsanov, B., Kiefer, P., Potthoff, E., Heux, S., Quax, W. J., Wendisch, V. F., Brautaset, T., Portais, J. C., Vorholt, J. A., 2015. Engineering *Escherichia coli* for methanol conversion. *Metabolic Engineering.* 28, 190-201.
- Mutzel, R., Marlière, P., Method and device for selecting accelerated proliferation of living cells in suspension. 2000. Patent WO2000034433 A1
- Nghiem, Y., Cabrera, M., Cupples, C. G., Miller, J. H., 1988. The *mutY* gene: a mutator locus in *Escherichia coli* that generates G.C--T.A transversions. *Proceedings of the National Academy of Sciences.* 85, 2709-2713.
- Ning, Z., 2001. SSAHA: A Fast Search Method for Large DNA Databases. *Genome Res.* 11, 1725–1729.
- Pawelek, P. D., MacKenzie, R. E., 1998a. Methenyltetrahydrofolate cyclohydrolase is rate limiting for the enzymatic conversion of 10-formyltetrahydrofolate to 5,10-methylenetetrahydrofolate in bifunctional dehydrogenase-cyclohydrolase enzymes. *Biochemistry.* 37, 1109-15.
- Reed, L. J., Koike, M., Levitch, M. E., Leach, F. R., 1958. Studies on the Nature and Reactions of Protein-Bound Lipoic Acid. *Journal of Biological Chemistry.* 232, 143-158.
- Richardson, J. P., 1990. Rho-dependent transcription termination. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression.* 1048, 127-138.
- Schmincke-Ott, E., Bisswanger, H., 1981. Dihydrolipoamide Dehydrogenase Component of the Pyruvate Dehydrogenase Complex from *Escherichia coli* K12. *European Journal of Biochemistry.* 114, 413-420.
- Shen, L. C., Atkinson, D. E., 1970. Regulation of Pyruvate Dehydrogenase from *Escherichia coli*: Interactions of adenylate energy charge and other regulatory parameters. *Journal of Biological Chemistry.* 245, 5974-5978.
- 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., Baker, D., 2015. Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci U S A.* 112, 3704-9.
- Smith, T. F., and Waterman, M. S., 1981. Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197.
- Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O., Matthews, R. G., 2002. Adaptation to famine: A family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci U S A.* 99, 13471-13476.
- Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., Médigue, C., 2019. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* 48, D579-D589.
- Wilkinson, K. D., Williams, C. H., 1981. NADH inhibition and NAD activation of *Escherichia coli* lipoamide dehydrogenase catalyzing the NADH-lipoamide reaction. *Journal of Biological Chemistry.* 256, 2307-2314.
- Wilkinson, K. D., Williams, C. H., Jr., 1979. Evidence for multiple electronic forms of two-electron-reduced lipoamide dehydrogenase from *Escherichia coli*. *Journal of Biological Chemistry.* 254, 852-862.
- Williams, C. H., 1965. Studies on Lipoyl Dehydrogenase from *Escherichia coli*. *Journal of Biological Chemistry.* 240, 4793-4800.
- Wytock, T. P., Fiebig, A., Willett, J. W., Herrou, J., Fergin, A., Motter, A. E., Crosson, S., 2018. Experimental evolution of diverse *Escherichia coli* metabolic mutants identifies genetic loci for convergent adaptation of growth rate. *PLoS genetics.* 14, e1007284.
- Yishai, O., Bouzon, M., Döring, V., Bar-Even, A., 2018. In Vivo Assimilation of One-Carbon via a Synthetic Reductive Glycine Pathway in *Escherichia coli*. *ACS Synthetic Biology.* 7, 2023-2028.
- Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H., Bar-Even, A., 2016. The formate bio-economy. *Current Opinion in Chemical Biology.* 35, 1-9.