Title: Autotrophic growth of *E. coli* is achieved by a small number of genetic changes

Authors: Roee Ben-Nissan[1†], Eliya Milshtein[1†], Vanessa Pahl[2], Benoit de Pins[1], Ghil Jona[3], Dikla Levi[3], Hadas Yung[1], Noga Nir[1], Dolev Ezra[1], Shmuel Gleizer[1], Hannes Link[2], Elad Noor[1], Ron Milo[1]

Affiliations:

[1]Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel.

[2]Interfaculty Institute for Microbiology and Infection Medicine Tübingen, University of Tübingen, Tübingen, Germany.

[3]Department of Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel.

*Corresponding author. Email: ron.milo@weizmann.ac.il

†These authors contributed equally to this work

Abstract: Synthetic autotrophy is a promising avenue to sustainable bioproduction from CO_2. Here, we use iterative laboratory evolution to generate genetically diverse autotrophic strains. We identify that just three mutations are sufficient for *E. coli* to grow autotrophically, when introduced alongside non-native energy (formate dehydrogenase) and carbon-fixing (Rubisco, phosphoribulokinase, carbonic anhydrase) modules. The mutated genes are involved in glycolysis (*pgi*), central-carbon regulation (*crp*), and RNA transcription (*rpoB*). The *pgi* mutation reduces the enzyme activity, thereby stabilising the carbon-fixing cycle by capping a major branching flux. The other two mutations increase the ratio of $NADH/NAD^+$ - the cycle's electron-donor. This study demonstrates the malleability of metabolism and evolution's capacity to switch trophic modes on laboratory time-scales and could facilitate transforming other heterotrophic organisms into autotrophs.

Introduction

Anthropogenic interference and intensive burning of fossil fuels releases large amounts of sequestered carbon in the form of CO_2, a main driver of climate change (*1, 2*). Direct air carbon capture, by any means, is urgently needed to help reduce emissions. Here, we focus on biology based carbon fixation as a potential avenue for CO_2 capture which has been shown to be a potentially efficient production platform(*3–5*). Specifically, use microbes for CO_2 valorization and produce a diverse repertoire of products including food, fuel, bio-plastics or other commodities(*6–9*). Natural autotrophs already metabolise CO_2 efficiently, yet most are challenging to cultivate and manipulate genetically compared to heterotrophic model organisms like *E. coli*. Here we explore an emerging alternative approach – converting heterotrophic model organisms into tractable autotrophs for bio-production(*7, 10–12*). In addition to promising engineering applications, studying the transition from heterotrophic to autotrophic metabolism can uncover fundamental principles about the structure and regulation of carbon fixation metabolism(*11–16*) .

Recently, with the increasing interest in platforms for sustainable production, synthetic biologists have been able to successfully manipulate different organisms to enable major metabolic transitions (*11, 12, 17–22*). These efforts resulted in the successful introduction of non-native C1 utilisation pathways, and even included successful transitions from heterotrophic into completely autotrophic metabolisms in bacteria (*12*) and yeast (*11*). These previous studies introduced the reductive pentose phosphate (rPP) cycle, the most prominent carbon fixation pathway in nature also known as the Calvin-Benson cycle, in *Escherichia coli* (*E. coli*) and *Komagataella phaffii* (previously *Pichia Pastoris*). Indeed, we showed that *E. coli* was able to utilise the rPP cycle for the synthesis of all biomass carbon - thus converting it from a heterotroph to an autotroph(*12*); However, this transition required continuous culture in selective conditions for several months (adaptive lab evolution) and resulted in many uninterpretable mutations, exposing a large knowledge gap.

Using adaptive lab evolution presents a challenge as it is hard to separate the essential changes allowing a phenotype from the overall set of accumulated mutations. Here we aimed at characterising the landscape of mutations required for autotrophic growth. Studying mutations that arise in multiple lineages can help us better understand the biological basis for the heterotroph-to-autotroph transition. To achieve this goal, we created a pipeline that enabled us to distinguish mutations essential for autotrophy from low-frequency mutations that promoted the adaptive evolution experiments and determine a compact essential set of changes permitting autotrophy in *E. coli*. We then proceeded with further analyses to shed light on the mechanism behind this compact essential set of mutations, showing the phenotype included a change in the activity of a central carbon metabolism enzyme and in intracellular changes of the non-native carbon fixation cycle's metabolite and co-factors pools. We continue by speculating about the role of these metabolic changes.

Results

A compact set of mutations that enable autotrophic growth was identified using rational design, iterative lab evolution and genetic engineering

Adaptive lab evolution is an effective tool that enables integration of non-native metabolic pathways. In order to harness the power of adaptive lab evolution, a selection system must be implemented, to direct the evolution towards the desired function. In our case, selection for the activity of the non-native rPP cycle. To ensure that some level of carbon fixation by rubisco becomes essential for growth, we knocked out phosphofructokinase (*pfkA* and *pfkB*), and 6-phosphate-1-dehydrogenase (*zwf*) which creates a stoichiometric imbalance and growth arrest when growing on five carbon sugars such as xylose. This imbalance could be rescued by a metabolic bypass - the phosphorylation of ribulose-5-P by Prk and the

subsequent carboxylation by rubisco, thereby coupling growth to the heterologously expressed rPP enzymes. This setup allowed us to evolve *E. coli*, a natural heterotroph, into an autotroph, using a non-native rPP cycle to utilise CO_2 as a sole carbon-source and Formate-dehydrogenase to oxidise formate for energy (*12*). However, using adaptive lab evolution also presents a challenge - once a new phenotype has been achieved, it is hard to separate the essential changes allowing for its appearance from the overall set of accumulated mutations. In many cases, mutations occur on the background of highly adaptive ones, leading to a spread of neutral "hitchhiker" mutations in the population. In other cases, mutations that appeared at the beginning of the evolution are no longer needed for the final phenotype. Our goal was to find a set of mutations that, if introduced into a heterotrophic wild type *E. coli*, would result in the usage of the non-native carbon fixation and energy modules. This would enable autotrophic growth based solely on genetic engineering without the need for further evolutionary adaptation. Isolating the essential genetic changes would facilitate studying the mechanism behind this metabolic transition. To do so, we used a workflow that included three main stages as shown in Fig. 1. A) rational design - introduction of required heterologous genes and metabolic knockouts to enforce Rubisco-dependent growth; B) adaptive lab evolution - revealing mutation candidates by growing the designed strain in autotrophic-selecting conditions over many generations until an autotrophic phenotype is obtained; C) evolution-inspired genetic engineering - introducing the most promising mutations revealed in stage B into the designed strain and testing it for autotrophic growth. Stages B and C were repeated until no further evolution was required, i.e. the designed strain in stage C had an autotrophic phenotype. To find the most promising mutations in stage B, we applied at least one of two criteria: mutations in genes that occur repeatedly in different autotrophic lineages (but are uncommon in other adaptive lab evolution experiments) or mutations that when reverted, result in a loss of phenotype. We discovered these "consensus" mutations by sequencing isolated autotrophic clones, obtained from repeated adaptive lab evolution experiments - similar to the conditions we used in Gleizer et al. 2019 (Fig. 1; steps B1, B2). In practice, we inoculated the strain denoted as Ancestor to a chemostat with limited xylose and excess formate. After ≈3 months (≈60 chemostat generations) we isolated from the chemostat a clone able to grow under autotrophic conditions (Methods), and compared its mutation set to those of two previously evolved autotrophic clones(*12*). The intersection of the three clones contained 4 genes that were mutated in all autotrophic isolated clones: the central carbon metabolism enzyme phosphoglucoisomerase (*pgi*), the beta subunit of the RNA polymerase (*rpoB*), poly A polymerase (*pcnB*) and the MalT DNA-binding transcriptional activator (*malT*) (Fig. 1, step B1). *rpoB*, *pcnB* and *malT* are genes that are commonly mutated in adaptive lab evolution experiments(*23*), suggesting that they have little relevance for autotrophic growth. Therefore, we chose to focus on the *pgi* mutation. We introduced the H386Y *pgi* mutation into the rationally designed ancestor (Fig. 1, step C1), and the engineered strain was tested for growth in autotrophic conditions (methods). No growth was observed in those conditions, which meant that additional mutations were necessary in order to achieve autotrophy. Genomic sequencing revealed that during the genetic manipulation for introducing the *pgi* mutation, another mutation in *rpoB* (A1245V) appeared in the genome as well. In parallel, reverting a variety of RNA polymerase mutations in other evolved strains back to their wild-type allele, showed that mutations in RNA polymerase are in fact essential to the phenotype. Therefore, despite the fact that it was unintentional, we decided to leave the mutation in the genome and move on with this strain. We then used this strain for another round of evolution (Fig. 1, step B2). Since we started the evolution experiment with a strain mutated in both *pgi* & *rpoB* (denoted "ancestor 2.0") we expected that the desired phenotype appears in fewer generations, and therefore fewer mutations. Indeed, on two separate experiments, the autotrophic phenotype appeared at a considerably shorter period of time when harbouring both mutations. In both cases, it took roughly one month until autotrophic growth was observed, compared to >90 days in previous attempts. Even though these strains (Fig. 1, "Evolved 4" and "Evolved 5") both had multiple mutations, only one mutation was shared between these independent evolution experiments - a non-synonymous mutation in the cAMP receptor protein, *crp* (H22N). When we introduced this mutation to the "ancestor 2.0" background in the evolution-inspired genetic engineering stage (Fig. 1, step C2), the strain was immediately able to grow under autotrophic conditions. Therefore, three mutations (*pgi**, *crp** and *rpoB**) were sufficient to facilitate the autotrophic phenotype. Since the RNA polymerase mutation was introduced into this strain (*rpoB* A1245V) as a byproduct of genetic engineering and never observed in the adaptive

evolution experiments, we tested its essentiality to the autotrophic phenotype. We found that a reversion of *rpoB* A1245V back to the wild-type allele causes a loss of the autotrophic phenotype. We denote the final engineered autotrophic *E. coli* strain, with only 3 mutations on top of the heterotrophic ancestor strain, the "compact autotrophic *E. coli*".

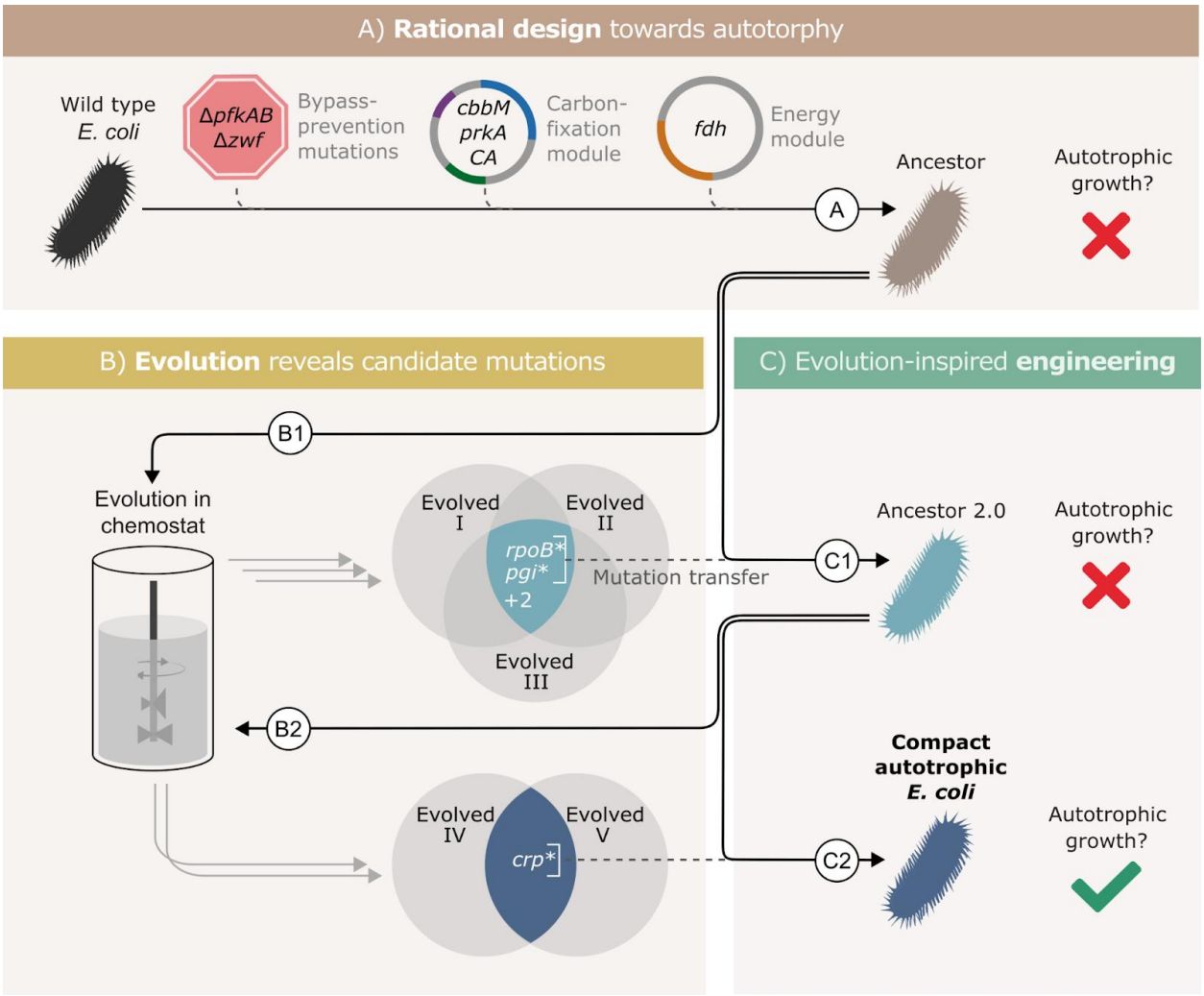

Fig. 1. Autotrophic phenotype achieved by introducing 3 mutations on top of a rationally designed ancestor. (A) We rationally designed the "wild-type" *E. coli* background strain (BW25113, depicted as a black bacterium) by introducing 4 enzymes: RuBisCO (*cbbM*), phosphoribulokinase (*prkA*), carbonic anhydrase (*CA*) andformate dehydrogenase (*fdh*), and by 3 genomic knockouts: glucose-6-phosphate dehydrogenase (*zwf*) and phosphofructokinase A&B (*pfkAB*). We denote the resulting strain as "ancestor" (brown bacterium). **(B)** We tested the resulting strain for autotrophic growth and, since it didn't grow, we used it for adaptive lab evolution in xylose-limited chemostats - i.e., conditions that select for higher carbon fixation flux. Altogether, we were able to isolate three evolved clones (two distinct strains from one chemostat experiment, Evolved I & II, and another strain from a second experiment, Evolved III, step B1). **(C)** Out of four consensus mutations that we identified, two - *rpoB* and *pgi* - were incorporated into the ancestor to get "ancestor 2.0" (light blue bacterium, step C1). Once more, we tested for autotrophic growth and were unsuccessful. Therefore, we initiated another round of adaptive lab evolution experiments using ancestor 2.0 growing in two xylose limited chemostats and isolated two evolved clones (one from each chemostat, step B2). The two clones shared a single consensus mutation, in *crp*. We thus created a new strain from ancestor 2.0 by introducing the *crp* mutation to it (blue bacterium, step C2). This strain could grow autotrophically and thus we achieved a compact autotrophic strain.

We verified the genotype using whole genome sequencing (methods). The results included the rationally designed knockouts (Δ*pfkA*, Δ*pfkB*, Δ*zwf*), heterologous plasmids (energy and carbon-fixing modules) and the three introduced mutations (*pgi**, *crp** and *rpoB**). During the genetic engineering process, two additional mutations occurred unintentionally: in the genes uhpT and yejG. uhpT is a hexose transporter and is unlikely to have any effect on the autotrophic phenotype, especially because the mutation was an early stop codon (Q7*) and likely a loss of function. As the function of yejG is unknown, we wanted to ensure that it is not essential to the autotrophic phenotype. Therefore, we reverted the mutation back to its wild-type allele and found that indeed the cells were still autotrophic.

Thus, we constructed a compact strain, which had three essential mutations (*pgi*, *crp* & *rpoB*). When comparing this compact strain to a previously characterised autotrophic strain(*12*), it was nearly identical in terms of growth rate, lag time and yield (Fig. 2A). Using ^{13}C labelling, we verified that all the biomass carbon stems from CO_2(Fig. 2B). We termed the three introduced mutations (*pgi*, *crp* & *rpoB*) together with the carbon fixation machinery (*cbbM*, *prkA* & *CA*) and the energy module (*fdh*) the "autotrophic enabling gene set".

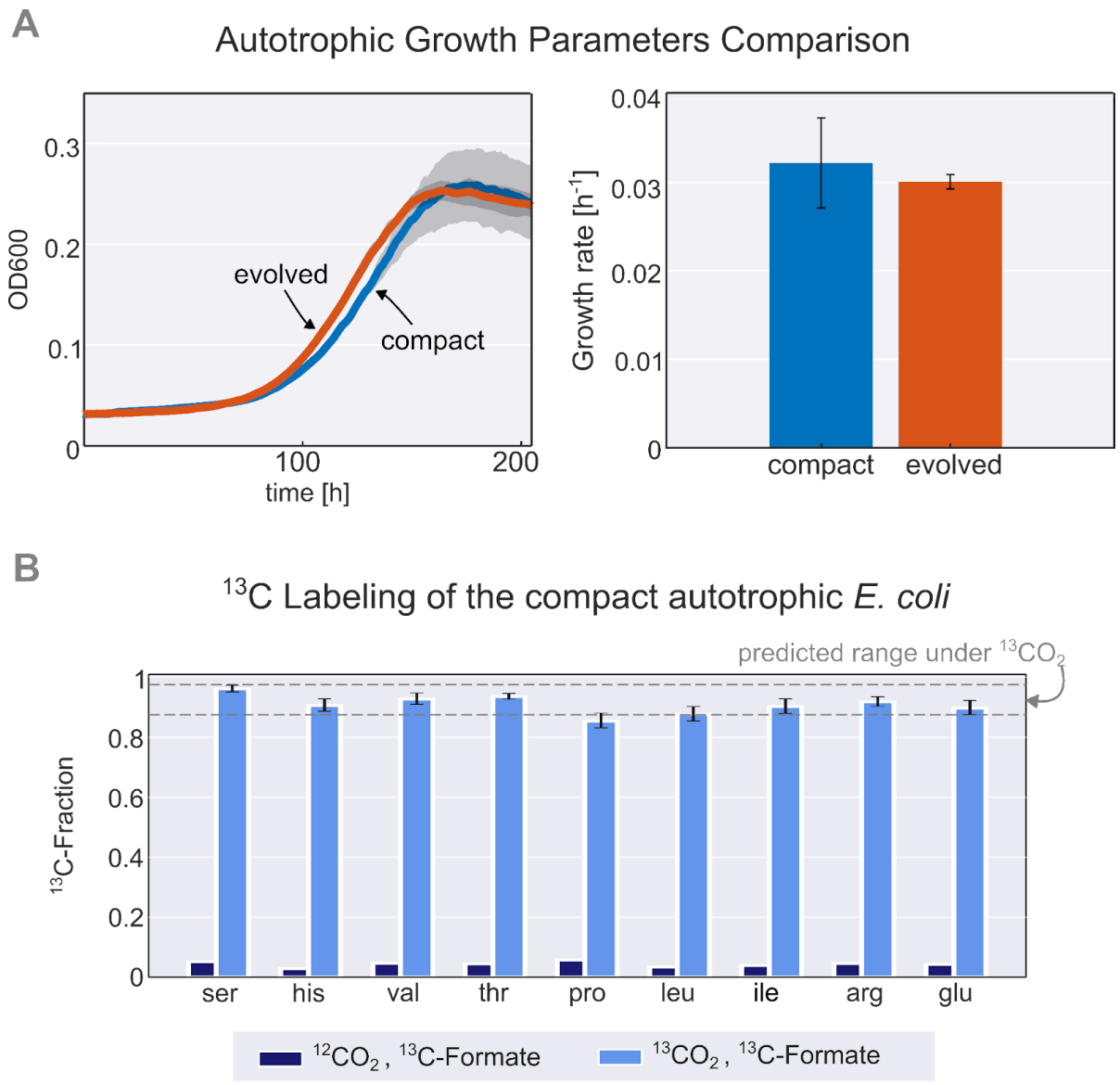

Fig. 2. Validation and characterisation of the autotrophic phenotype. (A) left: growth curve of an isolated evolved clone vs the engineered compact strain in liquid M9 minimal media supplemented with 45mM sodium formate and sparged with a gas mixture of 10% CO_2, 5% oxygen, and 85% nitrogen. right: calculated growth rate using linear regression of the engineered compact strain (light blue) and evolved strain (orange), the calculated doubling time at the given conditions is about 24h ($\mu \approx 0.03h^{-1}$). Growth was carried out in triplicates (n=3) in a Spark plate reader with gas control, dark grey area corresponds to standard deviation (150ul culture + 50ul mineral oil, to prevent evaporation). **(B)** The fractional contribution of $^{13}CO_2$ to various protein-bound amino acids of the compact autotrophic strain after ≈4 doublings on $^{13}CO_2$ and ^{13}C labeled formate (light blue, mean of n=3) reached the expected ^{13}C labeling fraction of the biosynthesized amino acids. When grown in naturally labeled CO_2 and ^{13}C labeled formate (dark blue, n=1, technical triplicate) the ^{13}C content dropped to close to natural abundance. Experiments with $^{13}CO_2$ as the substrate were carried out in air-tight (i.e., sealed) growth vessels.

Introduction of autotrophy-enabling gene set into wild-type background doesn't require bypass-prevention for growth

As described above, in order to select for autotrophic growth, the native *E. coli* was metabolically re-wired by 3 auxiliary genomic knockouts (Δ*pfkA*, Δ*pfkB*, Δ*zwf*). These knockouts created dependency on carboxylation of CO_2 by Rubisco for growth even when consuming pentose sugars, which were supplemented during the chemostat evolution (methods) (*12, 13*). This should direct the evolution towards increased usage of the synthetic rPP cycle. Once autotrophic growth was achieved, and the enabling gene set was found, we wanted to test if these knockouts (*pfkA pfkB* and *zwf*) are still required for the phenotype, now that the evolutionary selection process was completed. We tested this by introducing the set of autotrophy-enabling genes into a wild-type (BW25113) *E. coli* background (methods, Fig. 3). After doing so, we discovered an unintentional genomic mutation, in the ribosomal 16S subunit (rrsA) that was introduced during the genomic manipulation process. The engineered *E. coli* strain was able to grow autotrophically, albeit reaching 2-fold less than the final OD of the rationally designed compact strain still harbouring the auxiliary knockouts. This indicates that the "autotrophic enabling gene set" is sufficient to achieve the trophic mode change and that the knockouts are not essential for the final autotrophic growth. As discussed below, this indicates that in some cases metabolic rewiring could be used as a temporary "*metabolic scaffold*" and could be removed from the genotype once the desired phenotype is achieved. We proceeded by conducting various comparative assays, including metabolomics and physiology experiments to test the effect of the three mutations on the cell.

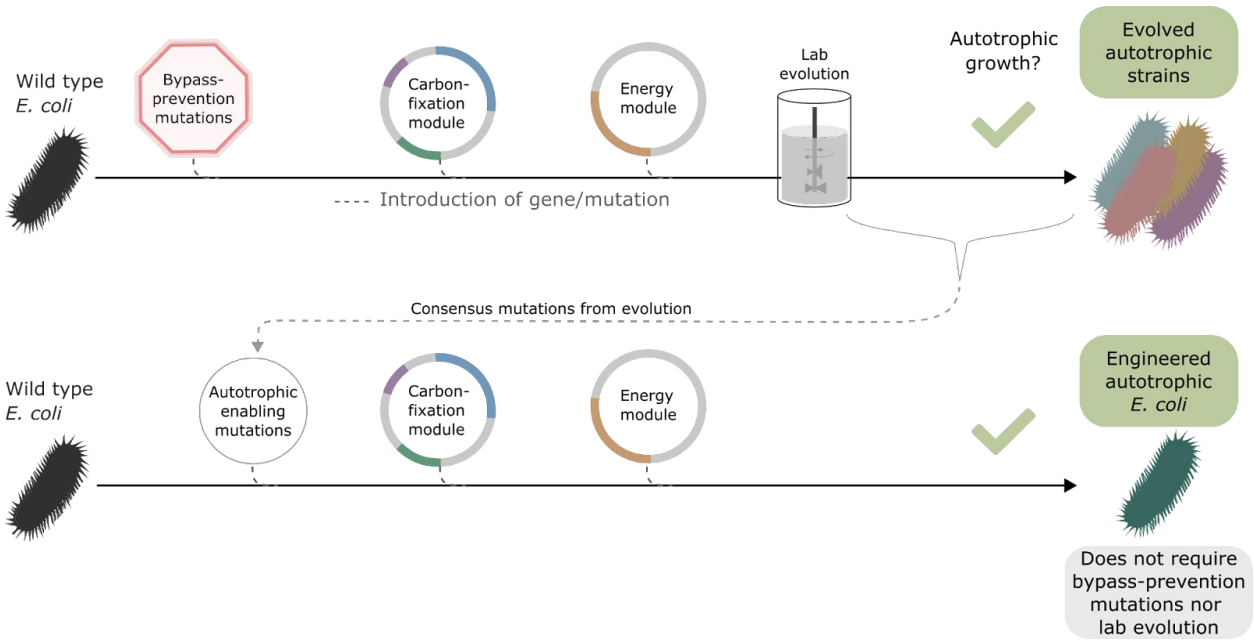

Fig. 3. Auxiliary knockouts are not required for final phenotype. We transformed a wild-type BW25113 *E. coli* strain (black bacteria) with the carboxylating and energy module plasmids (grey circles with coloured gene annotations), and inserted 3 auxiliary genomic knockouts (red octagon) to rewire metabolism toward carboxylating dependency. This strain was used for iterative evolution experiments in order to generate diverse autotrophic strains and reveal mutations candidates for rational design, i.e. consensus mutations. The identified autotrophic enabling mutations (thin grey circle) were then introduced in a wild type strain expressing the heterologous plasmids but without the auxiliary knockouts, the final strain was able to grow in autotrophic conditions. Dashed lines represent gene/mutation introduction.

The compact autotrophic *E. coli* contains only three genomic mutations and provides an opportunity to derive design principles that might serve as guidelines for future engineering efforts that use the rPP cycle. Therefore, we designed experiments that aimed to identify the relevant phenotypes of each of these mutations. Because the mutations were essential for the autotrophic growth, we could not use autotrophic conditions to compare them to a strain with wild-type alleles. Therefore, we chose the most suitable combinations of strains and conditions that could isolate the effect of the pertinent mutation.

Modulation of a metabolic branch-point activity increased the concentration of rPP metabolites

Metabolic pathways that regenerate and synthesise more of their own metabolites are referred to as autocatalytic cycles. Within the autotrophic *E. coli* the rPP cycle is autocatalytic. Due to the inherent positive feedback mechanism, autocatalytic cycles tend to be unstable, and therefore the fluxes of entry and exit points (bifurcation/branch points) need to be balanced (*14*). Such tuning is needed since any disruption of the balance between cycling and branching fluxes could result in the depletion or alternatively toxic accumulation of intermediate metabolites and cycle arrest. We previously predicted that mutations in branch points will be needed to stabilise the steady state flux within the rPP cycle (*14*). In line with this prediction we find that the mutation in Phosphoglucoisomerase (Pgi), a key branch point in the rPP cycle, follows this design principle. Pgi consumes fructose-6-phosphate (F6P), one of the cycle intermediates, converting it into glucose-6-phosphate (G6P), a precursor for cell membrane biosynthesis. Thus, Pgi regulates flux out of the autocatalytic cycle (Fig. 4A). In the different adaptive lab evolution experiments, we found three distinct mutations in the *pgi* gene. The first, H386Y, is a non-synonymous mutation occurring in one of the catalytic residues in the active site, and is part of the autotrophic enabling gene set. The second was a complete knockout, a 22KB chromosomal deletion, including also 16 other genes. The

third was an early stop codon E72*. These observations led us to suspect the H386Y mutation in *pgi* decreases or even completely eliminates the activity of the enzyme. To test this, the kinetic rates of the isomerization of F6P to G6P of two purified Pgi enzymes, wild type Pgi and Pgi H386Y, were determined using a spectroscopic coupled assay, where G6P was coupled to $NADP^+$ reduction (*24*) (Fig. 4B; see also methods). The measured k_{cat} of the wild type Pgi was ≈130 [s^{-1}] (±26) which is inline with previous computational predictions of a k_{cat} of ≈200 [s^{-1}] (*25*). The mutated Pgi showed weak activity with a k_{cat} of ≈1.5 [s^{-1}](±1). Therefore the *pgi* mutation is in line with the original expectation that branching point regulation is necessary, and the nature of the observed mutations in *pgi* supports the notion that their role is to reduce the efflux from the non-native autocatalytic cycle. Pgi is part of the glycolysis/gluconeogenesis, where significant flux is required in wild type *E. coli*. Accordingly, wild type Pgi catalyses a reaction that is causing a strong efflux of F6P from the cycle, diminishing the regeneration of ribulose-1,5-bisphosphate (RuBP) which is usually not needed in the native metabolism but is essential for the rPP autocatalytic activity. In the mutated version of Pgi the efflux capacity is diminished, which can stabilise the cycle (Fig. 4C). Following the same logic we can expect an increase in the F6P substrate metabolite pool. Therefore, we measured intracellular sugar-phosphates in the *pgi* mutant strain and a strain with a wild-type *pgi* allele by liquid chromatography-tandem mass spectrometry (LC-MS/MS). For this comparison, we used the rubisco-dependent ancestor strain as the genetic background. We found that the ratio of F6P to G6P was about 3 times higher in the *pgi* mutant strain relative to the wild-type (Fig. 4D). Furthermore, the *pgi* mutant had higher levels of metabolites within the rPP cycle (Fig. 4D), confirming the stabilising function of the mutation. These results lead us to conclude that the recurring mutations in the *pgi* gene are mediating the integration of the non-native carbon fixation genes into a stable autocatalytic cycle. This is achieved by reducing the ratio of the efflux from F6P to G6P relative to the flux in the cycle, thereby redirecting it towards the rPP cycle and the regeneration of RuBP.

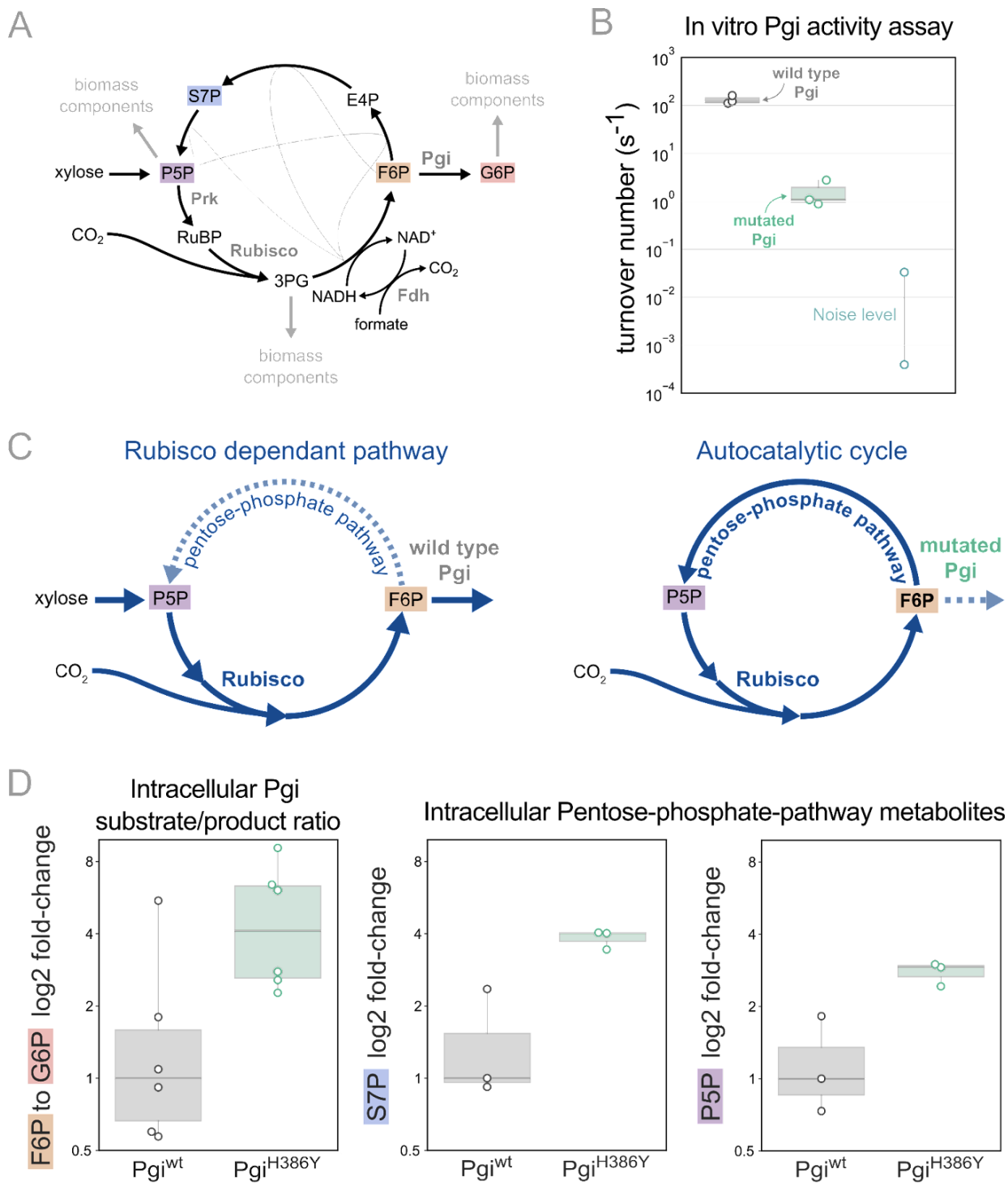

Fig. 4. **Pgi mutation is required to partition flux towards regeneration of autocatalytic cycle substrate. (A)** Metabolic scheme of rPP cycle components when growing on xylose. **(B)** In vitro spectrophotometric coupled assay determined that the rate of the isomerization of fructose-6-phosphate (F6P) to glucose-6-phosphate (G6P) is ≈100-fold lower for purified Pgi^{H386Y} (≈1.5 [s^{-1}]) compared to $Pgi^{wild\text{-}type}$ (≈130 [s^{-1}]), n=3. Noise level was determined by negative control samples containing a non-Pgi enzyme. **(C)** Left: The wild type *pgi*, competes for its substrate F6P with the autocatalytic cycle, resulting in low F6P pool and low regeneration rate of ribulose-5-phosphate (Ru5P), which means that the rubisco-dependent pathway requires constant xylose supply. Right: The mutated Pgi (green) has reduced activity, which increases the ratio between the flux in the cycle and efflux which is needed for a stable regenerative flux towards Ru5P (via the pentose-phosphate-pathway), thus enabling an autotrophic cycle. **(D)** Left: Measured

relative intracellular ratio of F6P and G6P of the ancestor background harbouring a Pgi^{H386Y} mutation compared to the ancestor with a wild-type Pgi, both grown in rubisco dependent conditions, n=6 cultures p-value < 0.05. Right: Relative intracellular abundance of pentose-phosphate-pathway metabolites - sedoheptulose-7P (S7P) and total pool of Pentose-phosphates (Ribulose-5P, Ribose-5P, Ribose-1P, and Xylulose-5P - denoted P5P) in a Pgi^{H386Y} strain versus the $Pgi^{wild\text{-}type}$ strain, both growing in a rubisco-dependent manner, n = 3 cultures.

Mutations in regulatory genes lead to increased availability of the carbon fixation cycle electron donor NADH

Along with the branch-point mutation in *pgi*, the compact autotrophic *E. coli* has two transcription-associated mutations, one in the gene *crp*, a global metabolism regulator and one in *rpoB*, a component of the transcriptional machinery. Since Crp and RpoB are known to physically interact in the cell (*26–28*), we address them as one unit, as it is hard to decouple the effect of one from the other. Mutations in *crp* and *rpoB* are known to affect growth at different conditions such as types of media or carbon sources (*28–31*). Therefore we tested whether this was also the case for the conditions under which the cells evolved in the chemostat. We performed a growth experiment of the double mutant (*crp* H22N, *rpoB* A1245V) expressing *fdh* in a minimal medium containing xylose and formate. Under such conditions, the strain harbouring both mutations and the *fdh* exhibited >25% increase in the final OD compared to the wild type expressing *fdh* (Fig. 5A&B). At the same time they also showed a longer (>10 hours) lag time. Notably, when grown solely on xylose, in the absence of formate, the increase in yield was not observed, and the extended lag was less pronounced (Fig. 5A).

Since NADH is the electron donor for carbon fixation, and a product of formate consumption via Fdh, we used metabolomics to measure the ratio of $NADH/NAD^+$ in these strains in similar conditions (Methods). All of the measured ratios were normalised to the wild-type (with *fdh*) growing in the absence of formate. The addition of 30mM formate, which can serve as an electron donor via oxidation to CO_2 by Fdh, increased $NADH/NAD^+$ ratio by >20-fold in the double mutant, significantly higher than the ≈10-fold increase in the wild-type strain (p-value <0.05). No significant difference was observed between the strains in the absence of formate (Fig. 5C). NADH could affect growth in several ways, as discussed below. The specific reason for the increase in final OD or lag time remains inconclusive at this point. We also note that mutations in global transcriptional regulators like Crp have many effects on the autotrophic *E. coli*(*32*) which await future exploration.

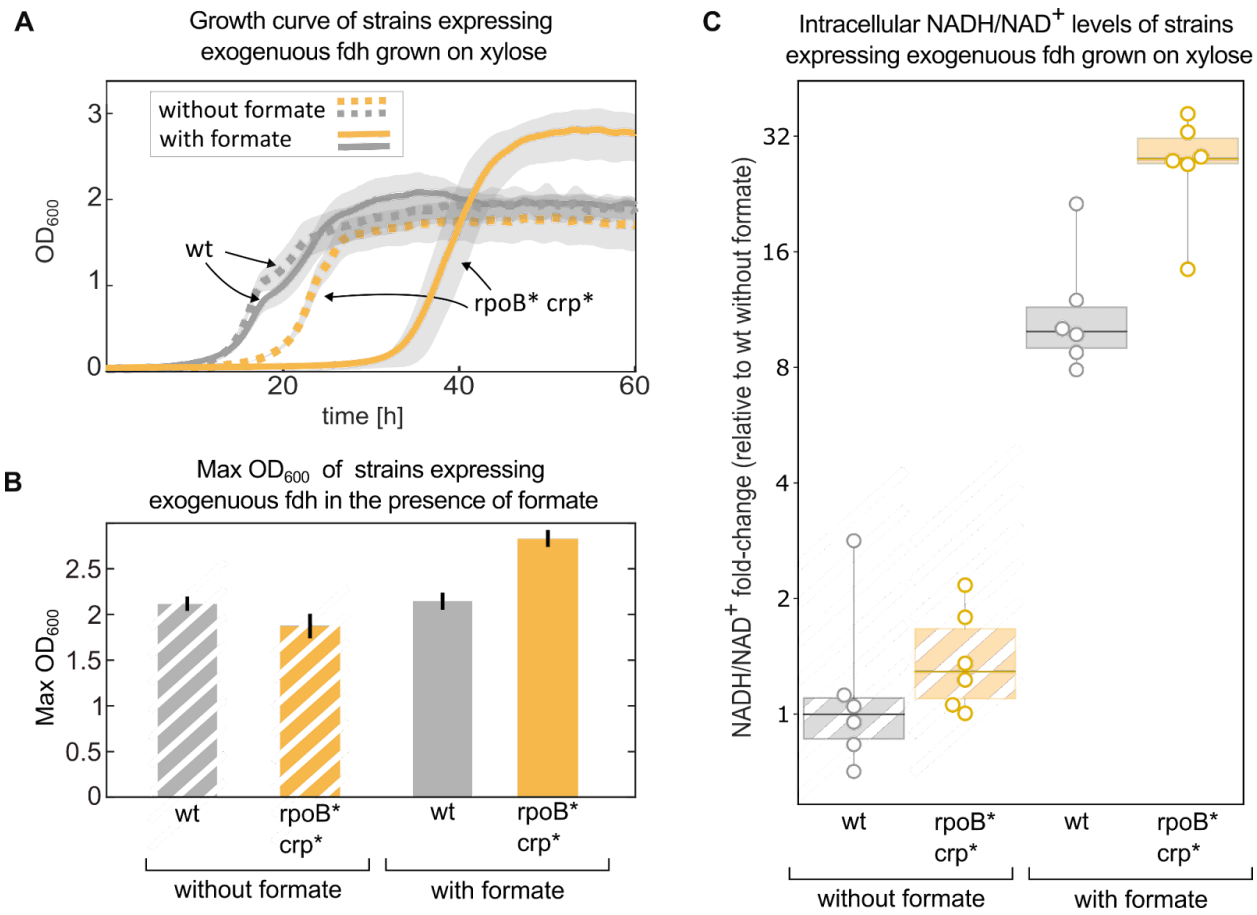

Fig. 5. mutations in *rpoB* and *crp* increase yield and intracellular NADH/NAD⁺ levels in fdh-expressing *E. coli* in the presence of formate. Growth experiment of BW25113 wild-type (grey) compared to a *crp* H22N *rpoB* A1245V mutant (orange). Both strains express *fdh*. **(A)** strains were grown in 1g/L xylose and 40mM formate (solid line) or in the absence of formate (dashed line). The experiment was executed in a 96 well plate in 10% CO_2 atmosphere, n=6 repeats. Lines represent the mean, light-grey background represents the standard deviation. **(B)** Maximal OD_{600} of the wild-type (grey) and mutant (orange) strains grown in 1g/L xylose and 40mM formate or in the absence of formate. Bar heights represent averages (n=6) of the median of the top 10 OD measurements of each replicate. Error bars represent standard error. **(C)** Intracellular $NADH/NAD^+$ ratio of the wild-type (grey) was compared to the mutant (orange). The strains were grown in 2g/L xylose and 30mM formate, or in the absence of formate. The y-values are $NADH/NAD^+$ ratios as fold-changes relative to wild-type without formate. Boxes represent 25-75 percentile ranges and dark lines represent median values. All data points are depicted without removing outliers.

Discussion

In this study, we demonstrated how iterative evolution and engineering were able to narrow down the genetic mutations required for autotrophy in *E. coli* from tens of mutations - to only three mutated genes. The three mutated genes - *pgi*, *crp* and *rpoB* - serve as a compact set, which when introduced into a wild-type background and supplemented with the carbon fixing and energy modules, achieve an autotrophic phenotype. By metabolic and physiological characterisation we were able to distil some of the effects of these mutations and understand their role in facilitating the autotrophic phenotype. We suggest that the two main effects achieved by the autotrophy-enabling mutations are: (i) integrating the non-native carbon fixation cycle by enhancing flux within the cycle over outgoing flux; and (ii) increasing the $NADH/NAD^+$ ratio, which could have both thermodynamic and kinetic benefits.

The mutation in *pgi* affects the activity of a gene that encodes a phosphoglucoisomerase - a metabolic enzyme that catalyses a reaction that branches out of the non-native rPP cycle. The enzyme converts F6P to G6P, with the former being a cycle intermediate, and the latter a precursor required for membrane biosynthesis. In a previous study, this mutation was found to be essential for the integration of the rPP autocatalytic cycle and was part of a minimal set for hemi-autotrophic growth(*16*). As our results suggest, the mutated enzyme shows a significant reduction in activity, which decreases the efflux from the rPP cycle to biosynthesis. We hypothesised that its role is to generate a new, higher, steady-state concentration for the rPP cycle metabolites, thus stabilising the cycle. This was supported by our observation that introducing a *pgi* mutation into a metabolically rewired Rubisco-dependent strain resulted in increased concentration of rPP cycle intermediates. This exemplifies the key role branch-points play in autocatalytic cycles(*12–14, 16*), making them a leading target for rational attempts to integrate non-native autocatalytic cycles into metabolism.

The other two mutations, *crp* and *rpoB*, encode for proteins that interact with one another. We therefore addressed them simultaneously, finding that these mutations lead to a significant increase in NADH/NAD$^+$ ratio - essential cofactors for the electron transfer reaction of the non-native rPP cycle (catalysed by the D-glyceraldehyde-3P dehydrogenase GAPDH). Under glycolytic conditions, the net reaction flows in the direction of D-glyceraldehyde-3P oxidation and, therefore, needs to be reversed for the rPP cycle. Increased levels of NADH, the reaction product in glycolytic conditions, facilitates such reversal, by tilting the thermodynamics in the opposite direction. Moreover, this increased ratio could offer a solution not only to a thermodynamic requirement, but also to a kinetic one. Since the amount of flux is a function of both the enzyme and substrate concentration - if we assume the enzyme concentration has not changed, increasing the substrate pool, as the ratio suggests, would result in increased flux. This result is in line with previous models, predicting that high cofactor levels are required for a stable metabolic state of the rPP cycle(*33*) and offers an avenue for future rational designs. Since both *crp* and *rpoB* affect gene expression globally, the role they play in the autotrophic phenotype likely extends beyond controlling the NADH level(*32*). In the future, this could be explored with various methods, such as proteomics, transcriptomics, ChIP-seq, or other assays for transcription factor activity. We speculate that such regulatory mutations are to be expected in adaptive laboratory evolution experiments. Especially during selection for a major shift (novelty) in the phenotype. A major shift will often necessitate a modulation of the activity of many cellular components. Lab evolution is limited by population size and number of generations, which means it can only explore a small part of the genotypic space. Therefore, sets of three or more mutations, where each mutation on its own does not confer an advantage, have a negligible probability to occur. On the other hand, a single mutation in a global regulator could concurrently perturb the activity of hundreds of components and might be the only viable solution to capture several required changes simultaneously.

We also show here how metabolic rewiring can serve as a "*metabolic scaffold*" to direct lab evolution. Since scaffold mutations are put in place for the purposes of selection, they can be removed once the goal is achieved. A common limitation of lab evolution is its tendency to follow evolutionary trajectories which increase fitness by deactivating or losing the heterologously expressed genes, thereby arriving at "dead-end" local optimum. A "metabolic scaffold" as utilised here can block these dead-end trajectories by coupling a target activity (here rubisco carboxylation) to growth, thereby increasing the likelihood of trajectories that lead to the desired phenotype. Once the phenotype is achieved, removal of the metabolic scaffold should be possible since growth is by definition dependent on the non-native genes. This could serve as a promising strategy for achieving minimally perturbed genotypes in future metabolic engineering attempts.

The approach presented here, of harnessing iterative lab evolution to find a compact set of essential mutations, strikingly shows the extreme malleability of metabolism and evolution capacity to switch trophic modes in laboratory time-scales with very few mutations. It can also serve as a promising pipeline for other genetic engineering attempts. Having a compact and well-defined genotype in a model organism allowed us to reveal metabolic adaptation steps needed for autotrophic growth, and could be used for future

exploration of the design principles of the rPP cycle operation and its components(*34*). The harnessing and rationalisation of the lab evolution process and metabolic adaptation steps elucidated here could facilitate the introduction of the rPP cycle and other carbon fixation cycles into other organisms of interest. Next steps could include integrating metabolic bio-production modules and alternative energy sources. Now that synthetic autotrophy in *E. coli* can be achieved with only a few defined steps, it expands the playground for more labs to join the quest for sustainable bio-production from CO_2.

References and Notes

1. J. Olofsson, T. Hickler, Effects of human land-use on the global carbon cycle during the last 6,000 years. *Veg. Hist. Archaeobot.* **17**, 605–615 (2008).

2. P. Friedlingstein, M. O'Sullivan, M. W. Jones, R. M. Andrew, L. Gregor, J. Hauck, C. Le Quéré, I. T. Luijkx, A. Olsen, G. P. Peters, W. Peters, J. Pongratz, C. Schwingshackl, S. Sitch, J. G. Canadell, P. Ciais, R. B. Jackson, S. R. Alin, R. Alkama, A. Arneth, V. K. Arora, N. R. Bates, M. Becker, N. Bellouin, H. C. Bittig, L. Bopp, F. Chevallier, L. P. Chini, M. Cronin, W. Evans, S. Falk, R. A. Feely, T. Gasser, M. Gehlen, T. Gkritzalis, L. Gloege, G. Grassi, N. Gruber, Ö. Gürses, I. Harris, M. Hefner, R. A. Houghton, G. C. Hurtt, Y. Iida, T. Ilyina, A. K. Jain, A. Jersild, K. Kadono, E. Kato, D. Kennedy, K. Klein Goldewijk, J. Knauer, J. I. Korsbakken, P. Landschützer, N. Lefèvre, K. Lindsay, J. Liu, Z. Liu, G. Marland, N. Mayot, M. J. McGrath, N. Metzl, N. M. Monacci, D. R. Munro, S.-I. Nakaoka, Y. Niwa, K. O'Brien, T. Ono, P. I. Palmer, N. Pan, D. Pierrot, K. Pocock, B. Poulter, L. Resplandy, E. Robertson, C. Rödenbeck, C. Rodriguez, T. M. Rosan, J. Schwinger, R. Séférian, J. D. Shutler, I. Skjelvan, T. Steinhoff, Q. Sun, A. J. Sutton, C. Sweeney, S. Takao, T. Tanhua, P. P. Tans, X. Tian, H. Tian, B. Tilbrook, H. Tsujino, F. Tubiello, G. R. van der Werf, A. P. Walker, R. Wanninkhof, C. Whitehead, A. Willstrand Wranne, R. Wright, W. Yuan, C. Yue, X. Yue, S. Zaehle, J. Zeng, B. Zheng, Global carbon budget 2022. *Earth Syst. Sci. Data*. **14**, 4811–4900 (2022).

3. D. Leger, S. Matassa, E. Noor, A. Shepon, R. Milo, A. Bar-Even, Photovoltaic-driven microbial protein production can use land and sunlight more efficiently than conventional crops. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2015025118.

4. Towards new carbon–neutral food systems: Combining carbon capture and utilization with microbial protein production. *Bioresour. Technol.* **349**, 126853 (2022).

5. T. Schwander, L. Schada von Borzyskowski, S. Burgener, N. S. Cortina, T. J. Erb, A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*. **354**, 900–904 (2016).

6. N. J. Claassens, C. A. R. Cotton, D. Kopljar, A. Bar-Even, Making quantitative sense of electromicrobial production. *Nature Catalysis*. **2** (2019), pp. 437–447.

7. S. Gleizer, Y. M. Bar-On, R. Ben-Nissan, R. Milo, Engineering Microbes to Produce Fuel, Commodities, and Food from CO2. *Cell Reports Physical Science*. **1**, 100223 (2020).

8. H. Li, P. H. Opgenorth, D. G. Wernick, S. Rogers, T.-Y. Wu, W. Higashide, P. Malati, Y.-X. Huo, K. M. Cho, J. C. Liao, Integrated electromicrobial conversion of CO2 to higher alcohols. *Science*. **335**, 1596 (2012).

9. J. Nielsen, J. D. Keasling, Engineering Cellular Metabolism. *Cell*. **164** (2016), pp. 1185–1197.

10. M. Baumschabl, Ö. Ata, B. M. Mitic, L. Lutz, T. Gassler, C. Troyer, S. Hann, D. Mattanovich, Conversion of CO2 into organic acids by engineered autotrophic yeast. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2211827119 (2022).

11. T. Gassler, M. Sauer, B. Gasser, M. Egermeier, C. Troyer, T. Causon, S. Hann, D. Mattanovich, M. G. Steiger, The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2. *Nat. Biotechnol.* **38**, 210–216 (2020).

12. S. Gleizer, R. Ben-Nissan, Y. M. Bar-On, N. Antonovsky, E. Noor, Y. Zohar, G. Jona, E. Krieger, M. Shamshoum, A. Bar-Even, R. Milo, Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell*. **179**, 1255–1263.e12 (2019).

13. N. Antonovsky, S. Gleizer, E. Noor, Y. Zohar, E. Herz, U. Barenholz, L. Zelcbuch, S. Amram, A. Wides, N. Tepper, D. Davidi, Y. Bar-On, T. Bareia, D. G. Wernick, I. Shani, S. Malitsky, G. Jona, A. Bar-Even, R. Milo, Sugar Synthesis from CO2 in Escherichia coli. *Cell*. **166**, 115–125 (2016).

14. U. Barenholz, D. Davidi, E. Reznik, Y. Bar-On, N. Antonovsky, E. Noor, R. Milo, Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points. *Elife*. **6** (2017), doi:10.7554/eLife.20667.

15. A. I. Flamholz, E. Dugan, C. Blikstad, S. Gleizer, R. Ben-Nissan, S. Amram, N. Antonovsky, S. Ravishankar, E. Noor, A. Bar-Even, R. Milo, D. F. Savage, Functional reconstitution of a bacterial CO2 concentrating mechanism in Escherichia coli. *Elife*. **9** (2020), doi:10.7554/eLife.59882.

16. E. Herz, N. Antonovsky, Y. Bar-On, D. Davidi, S. Gleizer, N. Prywes, L. Noda-Garcia, K. L. Frisch, Y. Zohar, D. G. Wernick, Others, The genetic basis for the adaptation of E. coli to sugar synthesis from CO 2. *Nat. Commun.* **8**, 1–10 (2017).

17. S. Kim, S. N. Lindner, S. Aslan, O. Yishai, S. Wenk, K. Schann, A. Bar-Even, Growth of E. coli on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* **16**, 538–545 (2020).

18. F. Y.-H. Chen, H.-W. Jung, C.-Y. Tsuei, J. C. Liao, Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol. *Cell*. **182**, 933–946.e14 (2020).

19. P. Keller, M. A. Reiter, P. Kiefer, T. Gassler, L. Hemmerle, P. Christen, E. Noor, J. A. Vorholt, Generation of an Escherichia coli strain growing on methanol via the ribulose monophosphate cycle. *Nat. Commun.* **13**, 5243 (2022).

20. J. Bang, C. H. Hwang, J. H. Ahn, J. A. Lee, S. Y. Lee, Escherichia coli is engineered to grow on CO2 and formic acid. *Nature Microbiology*. **5** (2020), pp. 1459–1463.

21. A. Satanowski, B. Dronsella, E. Noor, B. Vögeli, H. He, P. Wichmann, T. J. Erb, S. N. Lindner, A. Bar-Even, Awakening a latent carbon fixation cycle in Escherichia coli. *Nat. Commun.* **11**, 5812 (2020).

22. O. Yishai, M. Bouzon, V. Döring, A. Bar-Even, In Vivo Assimilation of One-Carbon via a Synthetic Reductive Glycine Pathway in Escherichia coli. *ACS Synth. Biol.* **7**, 2023–2028 (2018).

23. P. V. Phaneuf, D. Gosting, B. O. Palsson, A. M. Feist, ALEdb 1.0: a database of mutations from adaptive laboratory evolution experimentation. *Nucleic Acids Res.* **47**, D1164–D1171 (2019).

24. A. Widjaja, M. Shiroshima, T. Oka, M. Yasuda, H. Ogino, K. Miyatake, H. Nakajima, H. Ishikawa, The kinetics and mechanism of a reaction catalyzed by Bacillus stearothermophilus phosphoglucose isomerase. *J. Ferment. Bioeng.* **86**, 324–331 (1998).

25. D. Davidi, E. Noor, W. Liebermeister, A. Bar-Even, A. Flamholz, K. Tummler, U. Barenholz, M. Goldenfeld, T. Shlomi, R. Milo, Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3401–3406 (2016).

26. N. Savery, V. Rhodius, S. Busby, Protein-protein interactions during transcription activation: the case of the Escherichia coli cyclic AMP receptor protein. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **351**, 543–550 (1996).

27. W. Niu, Y. Kim, G. Tau, T. Heyduk, R. H. Ebright, Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell*. **87**, 1123–1134 (1996).

28. P. O. Frendorf, I. Lauritsen, A. Sekowska, A. Danchin, M. H. H. Nørholm, Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing. *Comput. Struct. Biotechnol. J.* **17**, 730–736 (2019).

29. J. Utrilla, E. J. O'Brien, K. Chen, D. McCloskey, J. Cheung, H. Wang, D. Armenta-Medina, A. M. Feist, B. O. Palsson, Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution. *Cell Syst*. **2**, 260–271 (2016).

30. T. M. Conrad, M. Frazier, A. R. Joyce, B.-K. Cho, E. M. Knight, N. E. Lewis, R. Landick, B. Ø. Palsson, RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20500–20505 (2010).

31. K.-K. Cheng, B.-S. Lee, T. Masuda, T. Ito, K. Ikeda, A. Hirayama, L. Deng, J. Dong, K. Shimizu, T. Soga, M. Tomita, B. O. Palsson, M. Robert, Global metabolic network reorganization by adaptive mutations allows fast growth of Escherichia coli on glycerol. *Nat. Commun.* **5**, 3233 (2014).

32. B. R. B. Haverkorn van Rijsewijk, A. Nanchen, S. Nallet, R. J. Kleijn, U. Sauer, Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli. *Mol. Syst. Biol.* **7**, 477 (2011).

33. M. Janasch, J. Asplund-Samuelsson, R. Steuer, E. P. Hudson, Kinetic modeling of the Calvin cycle identifies flux control and stable metabolomes in Synechocystis carbon fixation. *J. Exp. Bot.* **70**, 973–983 (2019).

34. H. Aigner, R. H. Wilson, A. Bracher, L. Calisse, J. Y. Bhat, F. U. Hartl, M. Hayer-Hartl, Plant RuBisCo assembly in with five chloroplast chaperones including BSD2. *Science*. **358**, 1272–1278 (2017).

35. Grenier, Frédéric, Dominick Matteau, Vincent Baby, and Sébastien Rodrigue. 2014. "Complete Genome Sequence of Escherichia Coli BW25113." Genome Announcements 2 (5). https://doi.org/10.1128/genomeA.01038-14.

36. Thomason, Lynn C., Nina Costantino, and Donald L. Court. 2007. "E. Coli Genome Manipulation by P1 Transduction." Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.] Chapter 1 (July): Unit 1.17.

37. Baba, Tomoya, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A. Datsenko, Masaru Tomita, Barry L. Wanner, and Hirotada Mori. 2006. "Construction of Escherichia Coli K-12 In-frame, Single-gene Knockout Mutants: The Keio Collection." Molecular Systems Biology 2 (1): 2006.0008.

38. Cherepanov, P. P., and W. Wackernagel. 1995. "Gene Disruption in Escherichia Coli: TcR and KmR Cassettes with the Option of Flp-Catalyzed Excision of the Antibiotic-Resistance Determinant." Gene 158 (1): 9–14.

39. Popov, V. O., and V. S. Lamzin. 1994. "NAD(+)-Dependent Formate Dehydrogenase." Biochemical Journal 301 (Pt 3) (August): 625–43.

40. Zelcbuch, Lior, Niv Antonovsky, Arren Bar-Even, Ayelet Levin-Karp, Uri Barenholz, Michal Dayagi, Wolfram Liebermeister, et al. 2013. "Spanning High-Dimensional Expression Space Using Ribosome-Binding Site Combinatorics." Nucleic Acids Research 41 (9): e98.

41. Barrick, Jeffrey E., Geoffrey Colburn, Daniel E. Deatherage, Charles C. Traverse, Matthew D. Strand, Jordan J. Borges, David B. Knoester, Aaron Reba, and Austin G. Meyer. 2014. "Identifying Structural Variation in Haploid Microbial Genomes from Short-Read Resequencing Data Using Breseq." BMC Genomics 15 (November): 1039.

42. Donati, Stefano, Michelle Kuntz, Vanessa Pahl, Niklas Farke, Dominik Beuter, Timo Glatter, José Vicente Gomes-Filho, Lennart Randau, Chun-Ying Wang, and Hannes Link. 2021. "Multi-Omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms That Buffer Decreases of Enzymes in E. Coli Metabolism." Cell Systems 12 (1): 56–67.e6.

Acknowledgments: We thank Alon Barshap, Aliza Fedorenko, Avi Flamholz, Avihu Yona, Daria Fedorova, Emanuel Avrahami, Ifat, Goldstein, Ilana Rogachev, Ido Cohen, Lior Greenspoon, Lior Shachar, Margarita Gortikov, Merav Hagag, Niv Antonovsky, Ofir Shechter, Ron Sender, Samuel Lovat, Tasneem Bareia, Tali Wiesel, Yinon Bar-on, Yuval Kushmaro, Yuval Rosenberg and Yafit Sugas for their support of this project. This research was generously supported by the Mary and Tom Beck Canadian Center for Alternative Energy Research, the Schwartz-Reisman Collaborative Science Program, the Ullmann Family Foundation and the Yotam Project. Prof. Ron Milo is the Head of the Mary and Tom Beck Canadian Center for Alternative Energy Research and the incumbent of the Charles and Louise Gartner Professorial Chair. R.B-N. is a Weizmann SAERI fellow. E.M. is a fellow of the Ariane de Rothschild Women Doctoral Program. HL and VP acknowledge funding from the Cluster of Excellence EXC 2124 from the Deutsche Forschungsgemeinschaft.

Funding:

Mary and Tom Beck Canadian Center for Alternative Energy Research (RM)

Schwartz-Reisman Collaborative Science Program (RM)

Ullmann Family Foundation and the Yotam Project (RM)

Sustainability and Energy Weizmann Doctoral Fellowship (RBN)

Ariane de Rothschild Women Doctoral Program (EM)

Cluster of Excellence EXC 2124 from the Deutsche Forschungsgemeinschaft (HL, VP)

Author contributions:

Conceptualization: RBN, EM, EN and RM

Formal analysis: RBN, EM, VP, BdP, HL, EN and RM

Data curation: RBN, EM, VP, BdP, EN

Methodology: RBN, EM, VP, BdP, GJ, SG, HL, EN and RM

Investigation: RBN, EM, VP, BdP, GJ, DL, HY, NN, DE and SG

Visualization: RBN, EM, VP, BdP, EN and RM

Funding acquisition: RBN, EM, VP, HL, EN and RM

Project administration: RBN, EM, EN and RM

Resources: HL and RM

Software: RBN, EM, VP, BdP, EN

Supervision: HL, EN and RM

Validation: RBN, EM, VP, BdP, HL, EN and RM

Writing – original draft: RBN, EM, EN and RM

Competing interests:

We declare the following provisional patent related to the manuscript, "An Engineered Autotrophic E. coli Strain for CO2 Conversion to Organic Materials."

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S4

References (12,13,16,35–42)

Data S1 to S10

Materials and Methods

Strains

We generated an engineered ancestor strain for chemostat evolution based on the Escherichia coli BW25113 strain(*35*). We used P1 transduction(*36)* to transfer knockout alleles from the KEIO strain collection *(37)* to our engineered strain, and to knock out the genes phosphofructokinase (*pfkA* and *pfkB*), and 6-phosphate-1-dehydrogenase (*zwf*). Following the transduction of each knockout allele the Km^R selection marker was removed by using the FLP recombinase encoded by the pCP20 temperature-sensitive plasmid(*38*). Loss of the selection marker and the temperature sensitive plasmid were validated by replica-plating the screened colonies and PCR analysis of the relevant loci. The cells were then transformed with the pCBB plasmid(13) (accession number KX077536) and a pFDH plasmid with a constitutive promoter controlling the expression of the *fdh* gene that resulted from Gleizer et al. 2019(*12*).

Plasmids

To create the pFDH plasmid, an *E. coli* codon optimised DNA sequence of formate dehydrogenase from the methytholotrophic bacterium Pseudomonas sp. 101(*39*) was cloned with an N-terminal His-tag into a pZ plasmid (Expressys, Germany) under a constitutive promoter with and a strong ribosome binding site (rbs B of *(40*). The plasmid has a pMB1 origin of replication and therefore is present in high copy number. We replaced the Km^R selection marker on the plasmid with a $Strep^R$ marker. An 8 bp deletion appeared in the promoter region of the first evolved clone. This plasmid was isolated and was the plasmid used for all consequential evolution experiments and autotrophic strains *(12*). Details regarding the pCBB plasmid are reported in *(13)*.

Growth tests

The growth test experiments were conducted in 96 well-plates. The final volume of each well was 200 µL (50 µL of mineral oil and 150 µL culture). The media consisted of M9 media supplemented with varying concentrations of the relevant carbon source, and trace elements (without addition of vitamin B1). Bacterial cells were seeded from a culture tube. Growth temperature was set to 37°C, and either aerated with ambient air or air with elevated CO_2 (10%) either with ambient or reduced oxygen (5%). OD600 measurements were taken every 10-30 minutes using a Tecan Spark plate reader.

Chemostat evolution experiments

The evolutionary experiment was conducted in a Bioflo 110 chemostat (New Brunswick Scientific, USA) at a working volume of 0.7L and a dilution rate of $0.02h^{-1}$ (equivalent to a doubling time of ≈33 hours) at 37°C. The chemostat was fed with media containing 4 g/L sodium formate and 0.5 g/L D-xylose as sole carbon sources. This amount of xylose in the feed makes xylose the limiting nutrient for cell growth in the chemostat. Gradually the concentration of xylose in the feed was reduced until a phenotype was observed. Chloramphenicol (30 mg/L) and streptomycin (100 mg/L) were added to the feed media. Aeration of the chemostat was done through a DASGIP MX4/4 stand-alone gas-mixing module (Eppendorf, Germany) with a composition of 10% CO_2 and 5% oxygen and 85% air at a flow rate of 40 sL/hr. To monitor the chemostat, a weekly sampling protocol was performed. Samples were taken for media analysis and phenotyping (inoculation of the bacteria on minimal media containing formate and lacking D-xylose). The optical density of each extracted sample was measured using a spectrophotometer (Ultrospec 10 Cell density meter, Amersham Biosciences) and a standard 10 mm polystyrene cuvette (Sarstedt, Germany).

^{13}C Isotopic labeling experiment

A culture of cells that were growing on naturally labeled sodium formate in an elevated CO_2 (10%, naturally labeled) incubator (New Brunswick S41i CO_2 incubator shaker, Eppendorf, Germany) were diluted 10-fold into fresh M9 media with 30 mM ^{13}C-formate sodium salt (Sigma Aldrich) to a total volume of 10 ml of culture. In the "open" labeling setup, growth was carried out in 125 ml glass shake flasks (n=3), which allow free exchange of gases between the headspace of the growth vessel and the gas mixture of the incubator. The flasks were placed inside an elevated CO_2 (10%) shaker-incubator (New Brunswick) with 37°C. After ≈4 doublings, the cells were harvested for subsequent analysis of protein bound amino acids and intracellular metabolites. In the "closed" labeling setup, growth was carried out in 250 mL glass shake flasks with a transparent extension which allows measurement of the optical density of the culture without opening it. After ≈3 doublings, the cells were diluted 8-fold into flasks covered with air-tight rubber septa (SubaSeal, Sigma Aldrich). Then, the headspace of the flask was flushed with a gas mixture containing 10% $^{13}CO_2$ (Cambridge Isotope Laboratories, USA) + 90% air or 10% $^{12}CO_2$ + 90% air generated by a DASGIP MX4/4 stand-alone gas-mixing module (Eppendorf, Germany). The flasks were placed in a 37°C shaker incubator. After ≈4 doublings, the cells were harvested for subsequent analysis of protein bound amino acids. Glass flasks used in the labeling experiments were pretreated by heating in a 460°C furnace for 5 hours to evaporate any excess carbon sources that could remain in the vessels from previous utilizations. Each labeling experiment was conducted in triplicates (n=3), with $^{12}CO_2$ the triplicates were pooled and measured as three technical replicates.

Sample preparation for liquid chromatography coupled to mass spectrometry and mass analysis of biomass components

After harvesting the biomass culture samples were prepared and analyzed as described in (*13*). Briefly, for protein bound amino acids ≈3 mL of culture at OD_{600} turbidity of ≈0.1-0.15 were pelleted by centrifugation for 5 minutes at 8,000 g. The pellet was suspended in 1 mL of 6N HCl and incubated for 16 hours at 110°C. The acid was subsequently evaporated with a nitrogen stream, resulting in a dry hydrolysate. Dry hydrolysates were resuspended in 0.6 mL of miliQ water, centrifuged for 5 minutes at 14,000 g and the supernatant was used for injection into the LCMS. Hydrolyzed amino acids were separated using ultra performance liquid chromatography (UPLC, Acquity - Waters, USA) on a C-8 column (Zorbax Eclipse XBD - Agilent, USA) at a flow rate of 0.6 mL/min and eluted off the column using a hydrophobicity gradient. Buffers used were: A) H_2O + 0.1% formic acid and B) acetonitrile + 0.1% formic acid with the following gradient: 100% of A (0-3 min), 100% A to 100% B (3-9 min), 100% B (9-13 min), 100% B to 100% A (13-14 min), 100% A (14-20 min). The UPLC was coupled online to a triple quadrupole mass spectrometer (TQS - Waters, USA). Data was acquired using MassLynx v4.1 (Waters, USA). We selected amino acids which have peaks at distinct retention time and m/z values for all isotopologues and also showed correct ^{13}C labeling fractions in control samples that contained protein hydrolyzates of WT cells grown with known ratios of $^{13}C_6$-glucose to ^{12}C-glucose.

The ^{13}C fraction of each metabolite was determined as the weighted average of the fractions of all the isotopologues of the metabolite, as depicted in the equation below:

$$\%^{13}C = \frac{\sum_{i=0}^{n} f_i \cdot i}{n}$$

where n is the number of carbons in the compound (e.g., for the amino acid serine n=3) and f_i is the relative fraction of the i-th isotopologue.

Whole-genome sequencing

DNA extraction (DNeasy blood & tissue kit, QIAGEN) and library preparation procedures were carried out as previously described in*(16)*. The prepared libraries were sequenced by a Miseq machine (Illumina). Analysis of the sequencing data was performed as previously described in *(12,13,16)* with the breseq software *(41)*.

Cultivation Conditions for Metabolome Sampling

For the G6P and F6P, P5P, and S7P measurements: Single colonies were transferred into liquid 3 mL M9 media containing 0.2 g/L xylose, 2g/L formate, Chloramphenicol (30 mg/L) and streptomycin (100 mg/L) from fresh plates with the same medium. The M9 pre-cultures were adjusted to a starting OD600 of 0.05 into 10mL culture in 250 mL glass shake flasks. The flasks were placed inside an elevated CO_2 (10%) shaker-incubator (New Brunswick S41i CO_2 incubator shaker, Eppendorf, Germany) at 37°C.

NADH/NADH+ measurements: Single colonies were transferred into 10 mL M9* xylose streptomycin from fresh M9* xylose streptomycin plates, and cultivated for 24h while shaking at 37°C. M9 pre-cultures were adjusted to a starting OD600 of 0.1 into 12-well plates, with 2 mL of medium in each well. Strains were cultivated in triplicates with or without formate (30 mM), added at the beginning of the culture. Optical density at 600nm was monitored every 10 min using a plate reader (Tecan, Spark). Plates were then rapidly transferred to a thermostatically controlled hood at 37°C and kept shaking during the sampling procedure.

Metabolomics Measurements

Cultivations were performed as described above. Culture aliquots were vacuum-filtered on a 0.45 µm pore size filter (HVLP02500, Merck Millipore). Filters were immediately transferred into a 40:40:20 (v-%) acetonitrile/methanol/water extraction solution at -20°C. Filters were incubated in the extraction solution for at least 30 minutes. Subsequently, metabolite extracts were centrifuged for 15 minutes at 13,000 rpm at -9°C and the supernatant was stored at -80°C until analysis. For measurements of NADH, NAD, Sedoheptulose-7P and the pool of pentose-P, extracts were mixed with a 13C-labelled internal standard in a 1:1 ratio. LC-MS/MS analysis was performed with an Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) as described previously*(42)*. An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30°C, and the injection volume was 3 µL. LC solvents in channel A were either water with 10 mM ammonium formate and 0.1% formic acid (v/v) (for acidic conditions, NADH/NAD), or water with 10 mM ammonium carbonate and 0.2% ammonium hydroxide (for basic conditions, Sedoheptulose-7P and pool of Pentoses). LC solvents in channel B were either acetonitrile with 0.1% formic acid (v/v) (for acidic conditions) or acetonitrile without additive (for basic conditions). LC columns were an Acquity BEH Amide (30 x 2.1 mm, 1.7 µm) for acidic conditions, and an iHILIC-Fusion(P) (50 x 2.1 mm, 5 µm) for basic conditions. The gradient for basic and acidic conditions was: 0 min 90% B; 1.3 min 40 % B; 1.5 min 40 % B; 1.7 min 90 % B; 2 min 90 % B. The ratio of 12C and 13C peak heights was used to quantify metabolites. 12C/13C ratios were normalized to OD at the time point of sampling. For the measurements of intracellular glucose-6-P and fructose-6-P, metabolic extracts were 10x concentrated using a vacuum evaporation (Eppendorf Concentrator plus) and resuspended in 40:40:20 (v-%) acetonitrile/methanol/water extraction solution. Temperature of the column oven was 30°C, and the injection volume was 5 µL. LC solvent in channel A was water with 10 mM ammonium formate and 0.1% formic acid (v/v), and in channel B was acetonitrile with 0.1% formic acid (v/v). The LC column was a Shodex HILICpak VG-50 2D (2.0 x 150 mm). The gradient was: 0 min 90% B; 58 min 70 % B; 3 min 90 % B. The 12C peak heights were used to quantify metabolites. Retention time of hexose-phosphates were determined with authentic standards of glucose-6-P

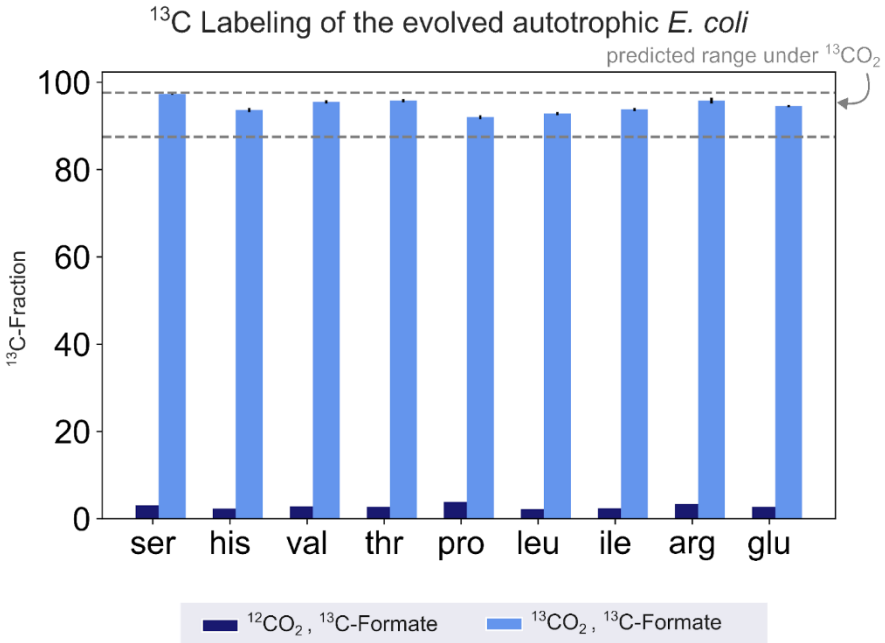

Fig. S1. The fractional contribution of $^{13}CO_2$ to various protein-bound amino acids evolved autotrophic strain after ≈4 doublings on $^{13}CO_2$ and ^{13}C labeled formate (light blue, mean of n=3) reached the expected ^{13}C labeling fraction of the biosynthesized amino acids. When grown in naturally labeled CO_2 and ^{13}C labeled formate (dark blue, n=1, technical triplicate) the ^{13}C content dropped to close to natural abundance. Experiments with $^{13}CO_2$ as the substrate were carried out in air-tight (i.e., sealed) growth vessels.

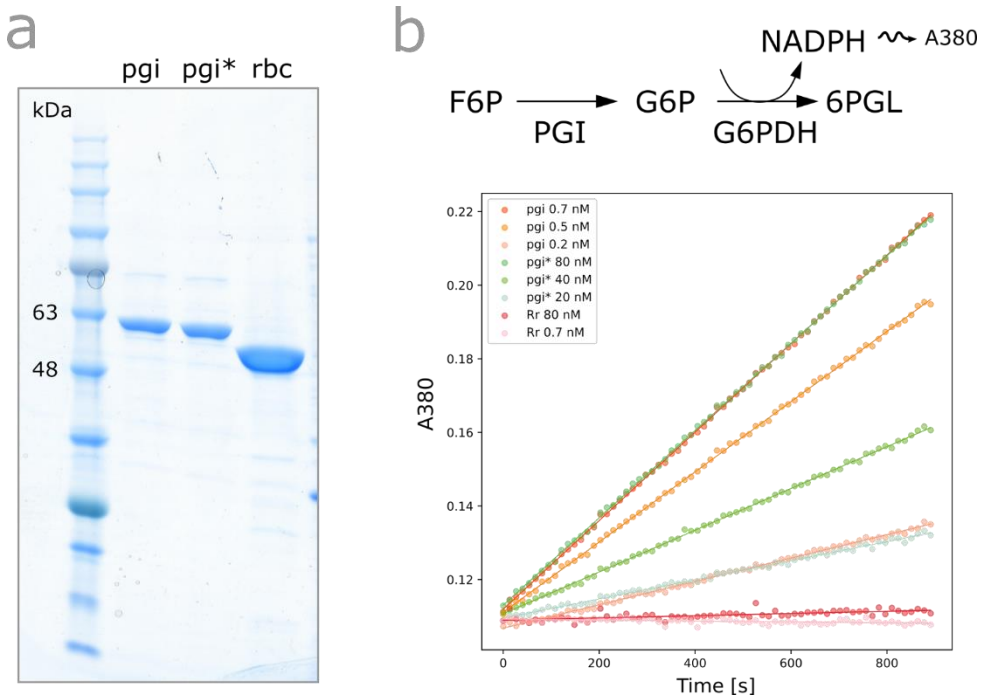

Fig. S2. In vitro characterization of Pgi^{H386Y}. (a) SDS-PAGE of purified Pgiwild-type, PgiH386Y and Rubisco after expression in E. coli. **(b)** Spectrophotometric coupled assay determining the rate of the isomerization of fructose-6-phosphate (F6P) to glucose-6-phosphate (G6P) for purified Pgiwild-type and PgiH386Y. Rubisco was used as a negative control setting the noise level.

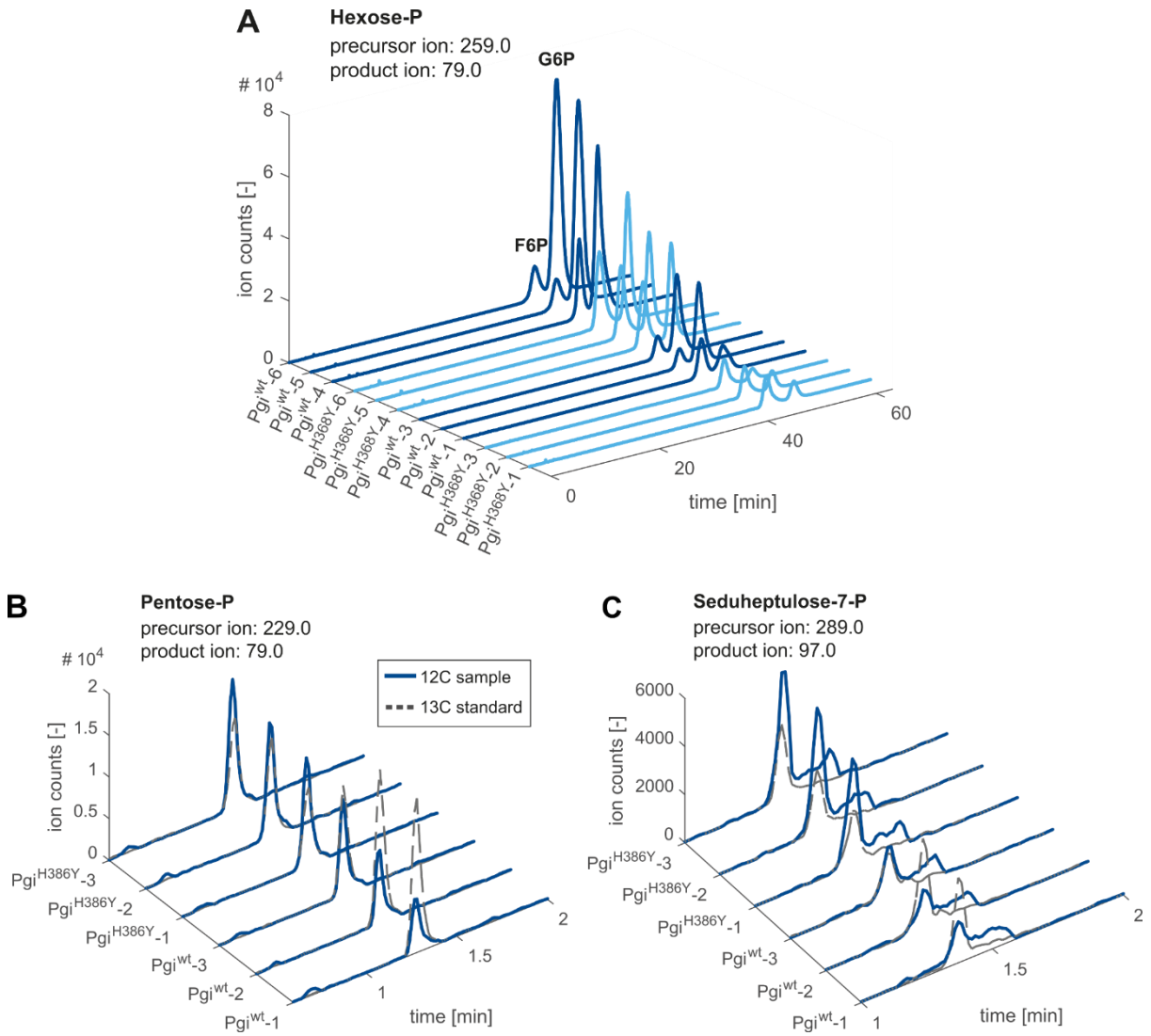

Fig. S3. LC-MS/MS chromatogram of intracellular metabolites involved in the autocatalytic cycle in *E. coli* ancestor with Pgi^{H386Y} mutation and ancestor with a wild-type Pgi.

(A) Chromatogram of fructose-6-phosphate (F6P) and glucose-6-phosphate (G6P). Precursor ion (259.0) and product ion (79.0) were used for 12C detection. Strains (n=6) are grown in rubisco dependent conditions.. Samples were measured using a Shodex HILICpak VG-50 2D. The retention time of hexose-phosphates were determined with authentic standards of G6P and F6P and the 12C peak heights were used to quantify metabolites. Experiments were performed in two batches of n=3 cultures. First batch samples (replicates 1-3) were used for measurements in A,B and C. **(B)** Chromatogram of total pool of Pentose-phosphates (Ribulose-5P, Ribose-5P, Ribose-1P, and Xylulose-5P - denoted P5P). Strains were grown in a rubisco-dependent manner (n=3) cultures. Overlay of 12C chromatograms (continuous line) and 13C chromatograms (dashed line) with precursor ion and product ion used for 12C detection as indicated. Samples were measured using an Acquity BEH Amide column and the ratio of 12C (sample) and 13C (internal standard) peak heights was used to quantify metabolites. **(C)** Sedoheptulose-7P (S7P) chromatogram of the same samples and conditions as indicated in B.

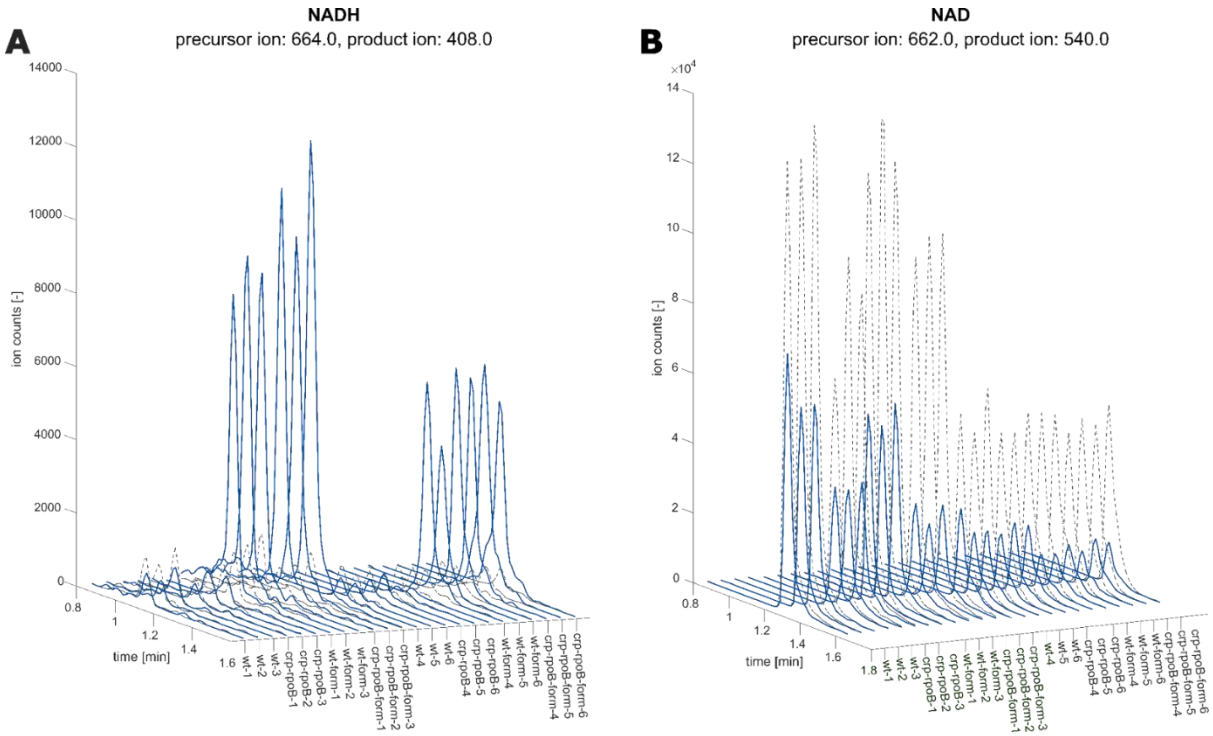

Fig. S4. LC-MS/MS chromatograms of NADH and NAD in fdh-expressing *E. coli* in the presence and absence of formate. **(A)** NADH chromatogram of *E. coli* BW25113 wild-type compared to a crpH22N rpoBA1245V mutant. Experiments were performed in two batches of n=3 cultures. Chromatograms are sorted accordingly. Overlay of 12C chromatograms (continues line) and 13C chromatograms (dashed line). Precursor ion and product ion used for 12C detection are indicated. Strains are grown with 30 mM formate (blue) or without the addition of formate (gray). **(B)** NAD chromatogram, same samples as in A. The ratio of 12C (sample) and 13C (internal standard) peak heights was used to quantify metabolites.

Data S1. (separate file)
iterative evolution mutations

Data S2. (separate file)
Compact and evolved labeling run curated peaks summary

Data S3. (separate file)
compact vs evolved growth experiment DATA 45mM 5 pctO2

Data S4. (separate file)
compact vs evolved growth experiment MAP 45mM 5 pctO2

Data S5. (separate file)
pgi_kinetic_assays_data

Data S6. (separate file)
G6PvsF6PResults

Data S7. (separate file)
PPP metabolites

Data S8. (separate file)
Growth experiment 60hr WT vs DM FDH variants CO2 40mMFor DATA and map

Data S9. (separate file)
Growth experiment WT vs DM FDH variants No formate CO2 MAP DATA and map

Data S10. (separate file)
rpoB and crp mutant metabolomics