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 several distinct autotrophic strains. Utilising this genetic diversity, 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's activity, thereby stabilising the carbon-fixing cycle by capping a major branching flux. For the other two mutations, we observe down-regulation of several metabolic pathways and increased expression of native genes associated with the carbon-fixing module (*rpiB*) and the energy module (*fdoGH*), as well as an increased ratio of NADH/NAD+ - the cycle's electron-donor. This study demonstrates the malleability of metabolism and its capacity to switch trophic modes using only a small number of genetic changes 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 (Olofsson and Hickler 2008; Friedlingstein et al. 2022). 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 (Leger et al. 2021; Van Peteghem et al. 2022; Schwander et al. 2016). Specifically, we suggest using microbes for CO_2 valorization and produce a diverse repertoire of products including food, fuel, bio-plastics or other commodities (Claassens et al. 2019; Gleizer et al. 2020; Li et al. 2012; Nielsen and Keasling 2016). 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 (Gleizer et al. 2020; Baumschabl et al. 2022; Gassler et al. 2020; Gleizer et al. 2019). 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 (Gleizer et al. 2019; Gassler et al. 2020; Antonovsky et al. 2016; Barenholz et al. 2017; Flamholz et al. 2020; Herz et al. 2017).

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 (Gassler et al. 2020; Gleizer et al. 2019; Kim et al. 2020; Chen et al. 2020; Keller et al. 2022; Bang et al. 2020; Satanowski et al. 2020; Yishai et al. 2018). 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 (Gleizer et al. 2019) and yeast (Gassler et al. 2020). 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), into *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 (Gleizer et al. 2019). 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 isolate the changes necessary for a given phenotype from the overall set of accumulated mutations. Here we aimed to characterise 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 unspecific low-frequency mutations, i.e. ones that are generally promoted by the physiological conditions in evolution experiments. The final outcome of this pipeline is 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 set of essential mutations. We show the phenotype included a change in the activity of a central carbon metabolism enzyme and intracellular changes of the non-native carbon fixation cycle's metabolite and co-factors pools. We conclude 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, the 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 (Gleizer et al. 2019). 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. Critically, 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. This system also requires formate dehydrogenase to oxidise formate for energy (Gleizer et al. 2019).

Although adaptive lab evolution is a powerful tool, it also presents a challenge - once a new phenotype has been achieved, it is hard to separate the essential changes required 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 successful integration 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 and potentially uncovering the mechanism behind this metabolic transition.

To isolate the essential genetic changes, 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 into 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 (Gleizer et al. 2019). 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 (Phaneuf, Gosting, and Palsson 2019), 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. 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, and different from the other *rpoB* mutations observed in the ALE experiments, we left the mutation in the genome and proceeded with this strain.

This strain, mutated in both *pgi* and *rpoB* (denoted "ancestor 2.0") was then used as the starting point for another round of evolution (Fig. 1, step B2). Since it had already contained two mutations that were likely to be relevant for the desired phenotype, we expected to observe evolved autotrophic strains after fewer generations, compared to the first round, and to find fewer mutations in them. Indeed, on two separate experiments, it took roughly one month until autotrophic growth was observed, as opposed to more than 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.

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 needed to verify 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 therefore could conclude that three mutations (*pgi**, *crp** and *rpoB**) are sufficient to facilitate the autotrophic phenotype on top of the heterotrophic ancestor background. We denote the final engineered autotrophic *E. coli* strain, the “compact autotrophic *E. coli*”.

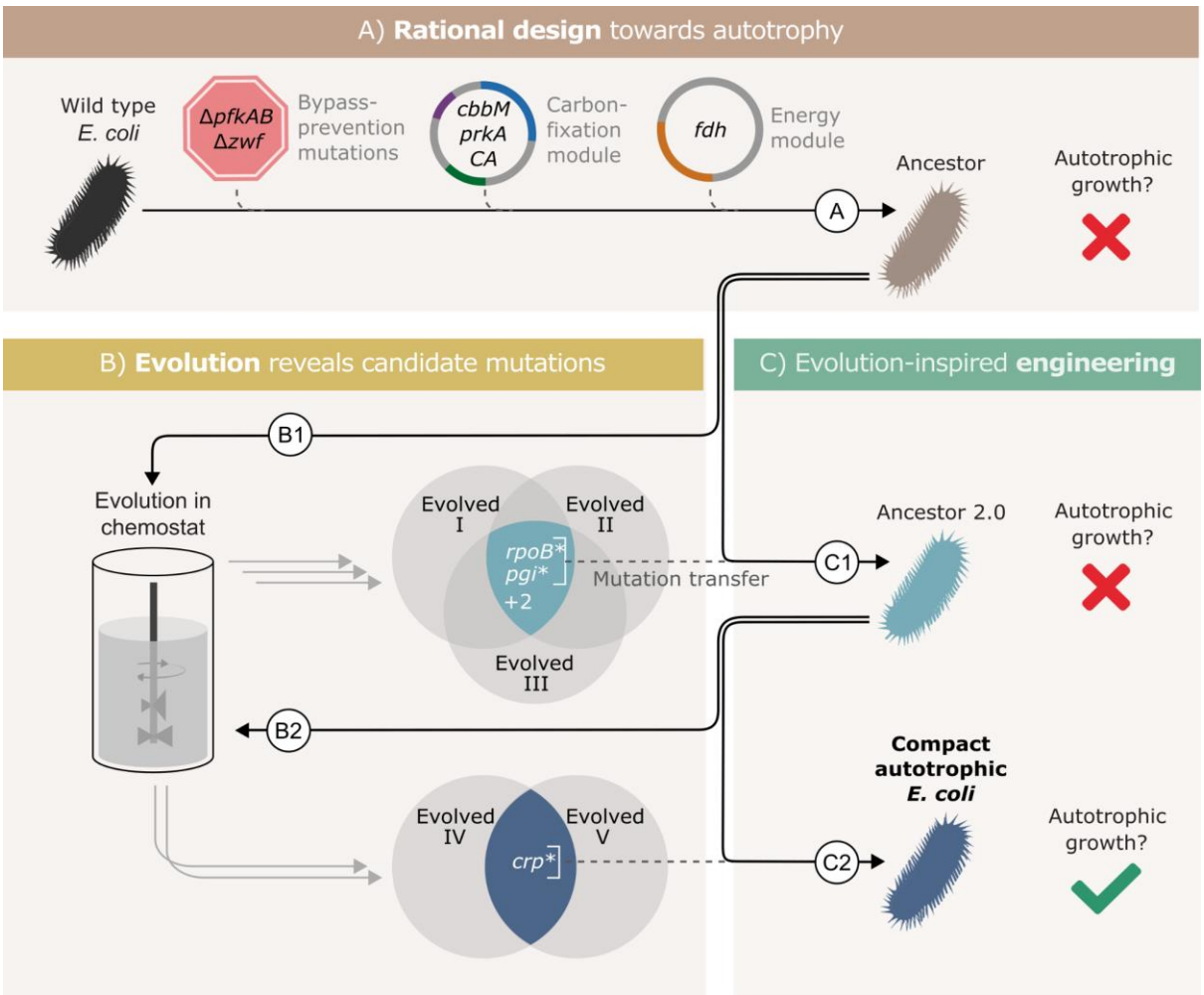

Fig. 1. Autotrophic phenotype achieved by introducing 3 mutations on top of a rationally designed ancestor. (A) We rationally designed the "wild-type" E. coli background strain (BW25113, depicted as a black bacterium) by introducing 4 enzymes: RuBisCO (cbbM), phosphoribulokinase (prkA), carbonic anhydrase (CA) and formate dehydrogenase (fdh), and 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, giving "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 into 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 new autotrophic strain with three essential mutations (*pgi*, *crp* & *rpoB*), which we denoted the *compact* strain. When comparing this compact strain to a previously characterised autotrophic strain (Gleizer et al. 2019), 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".

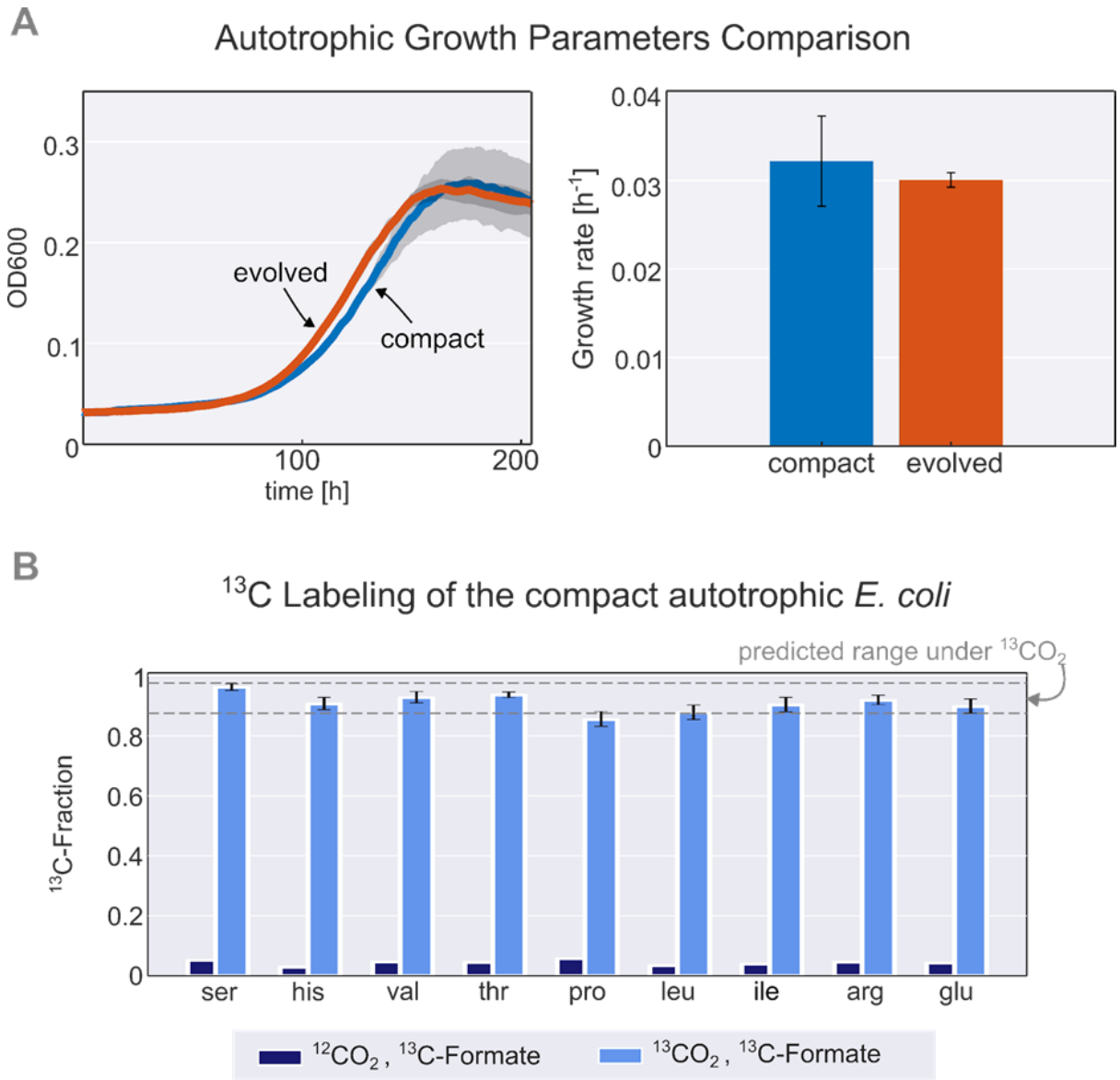

Fig. 2. Validation and characterisation of the autotrophic phenotype. (A) left: growth curve of an isolated evolved clone vs the engineered compact strain in liquid M9 minimal media supplemented with 45mM sodium formate and sparged with a gas mixture of 10% CO_2, 5% oxygen, and 85% nitrogen. right: calculated growth rate using linear regression of the engineered compact strain (light blue) and evolved strain (orange), the calculated doubling time with 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 the standard deviation of the mean (150 µl culture + 50 µl 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 labelled formate (light blue, mean of n=3). This reached the expected ^{13}C labelling fraction (see methods) of the biosynthesized amino acids. When grown in naturally labelled CO_2 and ^{13}C labelled 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 does not 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 a dependency on carboxylation by RuBisCO for growth even when consuming pentose sugars which were supplemented during the chemostat evolution (see methods, Antonovsky et al. 2016; Gleizer et al. 2019). This dependency was expected to direct the evolution towards increased usage of the synthetic rPP cycle and, eventually, to autotrophic growth.

Once autotrophic growth was achieved, and the enabling gene set was found, we wanted to test if these auxiliary knockouts (*pfkA pfkB* and *zwf*) were still required for the phenotype. We introduced 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, although its final OD was half and its growth rate was less than half of that of the rationally designed compact strain still harbouring the auxiliary knockouts (Fig. S2). This indicates that the "autotrophic enabling gene set" is sufficient to achieve the trophic mode change and that the knockouts are helpful but not essential for the final autotrophic growth. As discussed below, this demonstrates that, in some cases, metabolic rewiring could be viewed as a temporary "*metabolic scaffold*" which can be removed from the genotype once the desired phenotype is achieved.

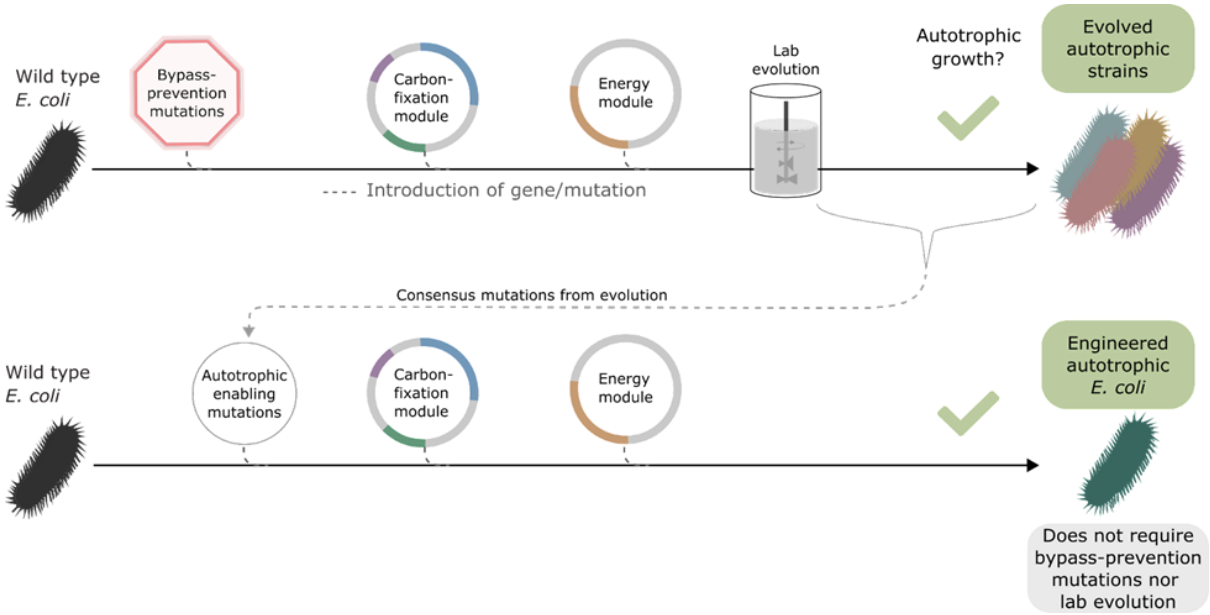

Fig. 3. Auxiliary knockouts are not required for final phenotype. We transformed a wild-type BW25113 E. coli strain (black bacteria) with the carboxylating and energy module plasmids (grey circles with coloured gene annotations), and inserted 3 auxiliary genomic knockouts (red octagon) to rewire metabolism 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.

We thus established a new engineered autotrophic *E. coli* strain containing a set of three essential and sufficient mutations, and showed that the bypass-prevention knockouts introduced before the ALE are themselves not necessary for the phenotype. Due to the small required number of genetic modifications, this strain provides an opportunity to derive guidelines for future engineering efforts that use the rPP cycle. Having this goal in mind, we focused our attention on identifying the relevant phenotypes of each of these mutations and explaining their adaptive advantage. 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 (Barenholz et al. 2017). 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 (Barenholz et al. 2017). 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.

We tested the kinetic rates of the isomerization of F6P to G6P of two purified Pgi enzymes, wild type Pgi and Pgi H386Y, using a spectroscopic coupled assay, where G6P was coupled to $NADP^+$ reduction (Widjaja et al. 1998) (Fig. 4B; see also methods). The measured k_{cat} of the wild type Pgi was ≈130 [s^{-1}] (±26) which is in line with previous computational predictions of a k_{cat} of ≈200 [s^{-1}](Davidi et al. 2016). 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 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 due to reduced consumption by Pgi. 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.

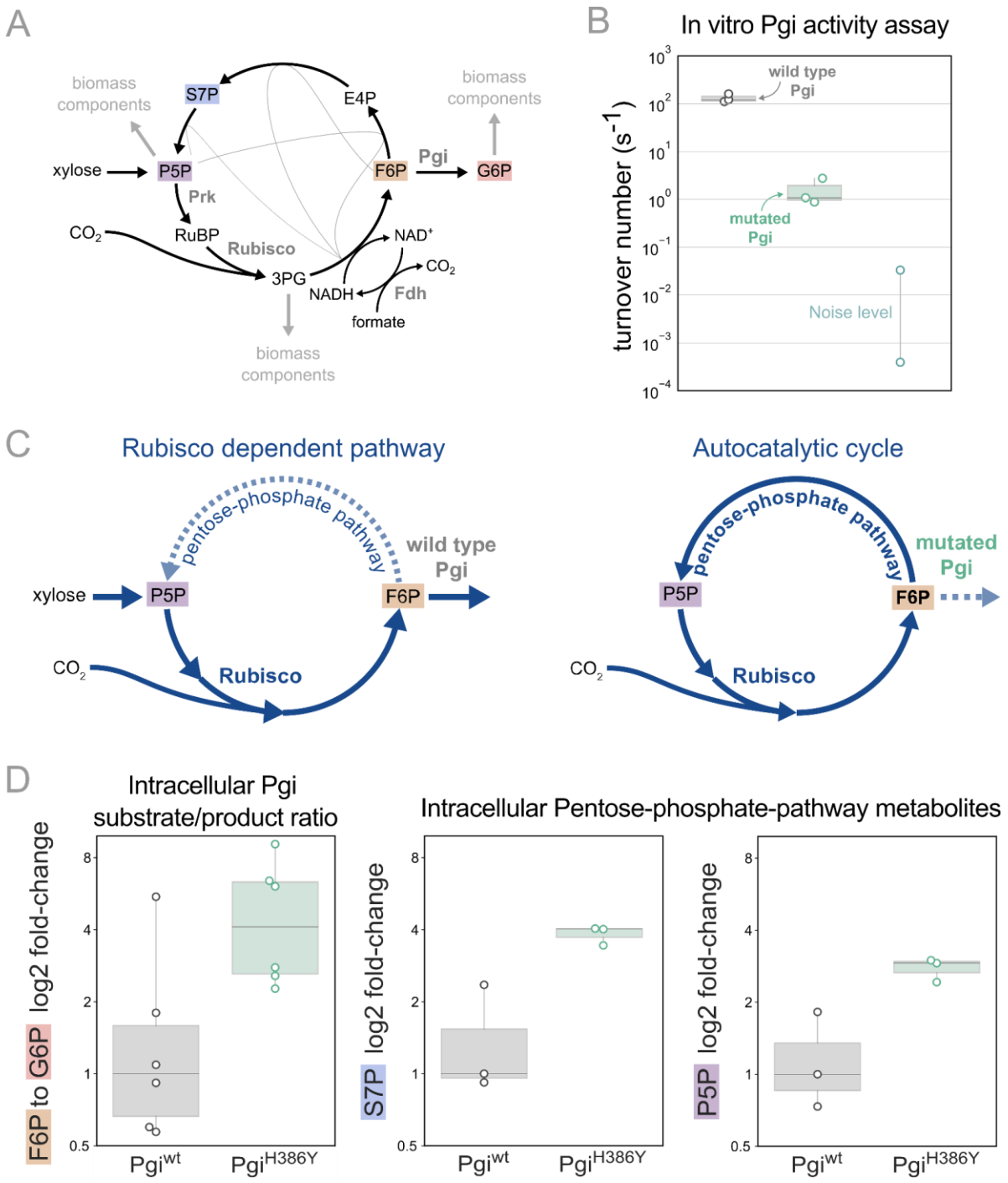

Fig. 4. **Pgi mutation is required to partition flux towards regeneration of autocatalytic cycle substrate. (A)** Metabolic scheme of rPP cycle components when growing on xylose. **(B)** In vitro spectrophotometric coupled assay determined that the rate of the isomerization of fructose-6-phosphate (F6P) 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.

Proteomic analysis reveals up-regulation of rPP cycle and formate-associated genes alongside down-regulation of catabolic genes

The other two mutations in the compact set - *crp* (H22N) and *rpoB* (A1245V), are global effectors in the cell. Crp is one of the most general transcription factors in the cell and has a role in many processes from metabolism to osmoregulation and biofilm formation (Landis, Xu, and Johnson 1999; Jackson, Simecka, and Romeo 2002; Ammar, Wang, and Rao 2018). The position H22 in Crp is part of "activation region 2" of the protein, an interaction region with RNA polymerase (Fig. S8). This position was shown to have an effect on promoter specificity (Niu et al 1996) and was previously reported to have a role in the integration of the rPP cycle in *E. coli* (Herz et al. 2017).

Similarly, mutations in *rpoB*, a subunit of RNA polymerase, have been shown to have a large effect on the cell transcriptome (Conrad et al. 2010; Utrilla et al. 2016). The observed A1245V mutation in *rpoB* is situated within the core of RNA polymerase, proximal to its interaction with the DNA (figure S8). Furthermore, this genetic variation was previously documented in the adaptation of the *E. coli* B REL1206 strain to high temperatures (Tenaillon et al. 2012) and is closely positioned to another previously reported *rpoB* mutation (K1242Q) that appeared in adaptation to loss of Pgi activity (Charusanti et al. 2010).

Depending on the transcribed gene target, the effect of the two mutations might be additive, antagonistic, or synergistic. Since each one of these mutations individually (in combination with the *pgi* mutation) is not sufficient to achieve autotrophic growth, it is reasonable to assume that only the target genes whose levels of expression change significantly in the double-mutant are the ones relevant for the autotrophic phenotype. Given the accumulated knowledge described above, these mutations are likely to have a broad effect on the cell. We chose to further elucidate their effect using a proteomics approach. As such, we conducted two experiments in an attempt to find changes in expression that are relevant for the adaptation to the new trophic mode. The first experiment compared the proteome of an evolved clone to its ancestor. The second experiment similarly compared an engineered BW25113 (BW) strain, containing the two regulatory mutations from the compact set (i.e., *crp* H22N and *rpoB* A1245V), to a wild type BW strain.

Since our aim was to observe only direct effects induced by the genetic changes, rather than by the environment, we measured the proteomes of the strains under identical conditions and dilution rates. Ideally, the comparison should be done in conditions that require autotrophic growth (since some effects might depend on the lack of an organic carbon source), however, all strains except the evolved clone cannot grow in such conditions. Therefore, we decided on growing the cells in the conditions present at the beginning of the lab evolution experiment, namely a xylose-limited chemostat. The dilution rate was set to $\approx 0.03\ h^{-1}$ leading to a doubling time of ≈24h (methods).

In order to identify the genes with a large and significant fold change (fold change >4, p-value <0.05) we generated a volcano plot for each such comparison (Fig. S6). We then proceeded by comparing the output of the first experiment (evolved divided by ancestor, Fig. S6A) to that of the double mutant divided by the base BW strain (Fig. S6B), since both mutations are essential for the phenotype and are known to affect expression. We focused on genes with common trends of expression level in both sets (Fig. 5).

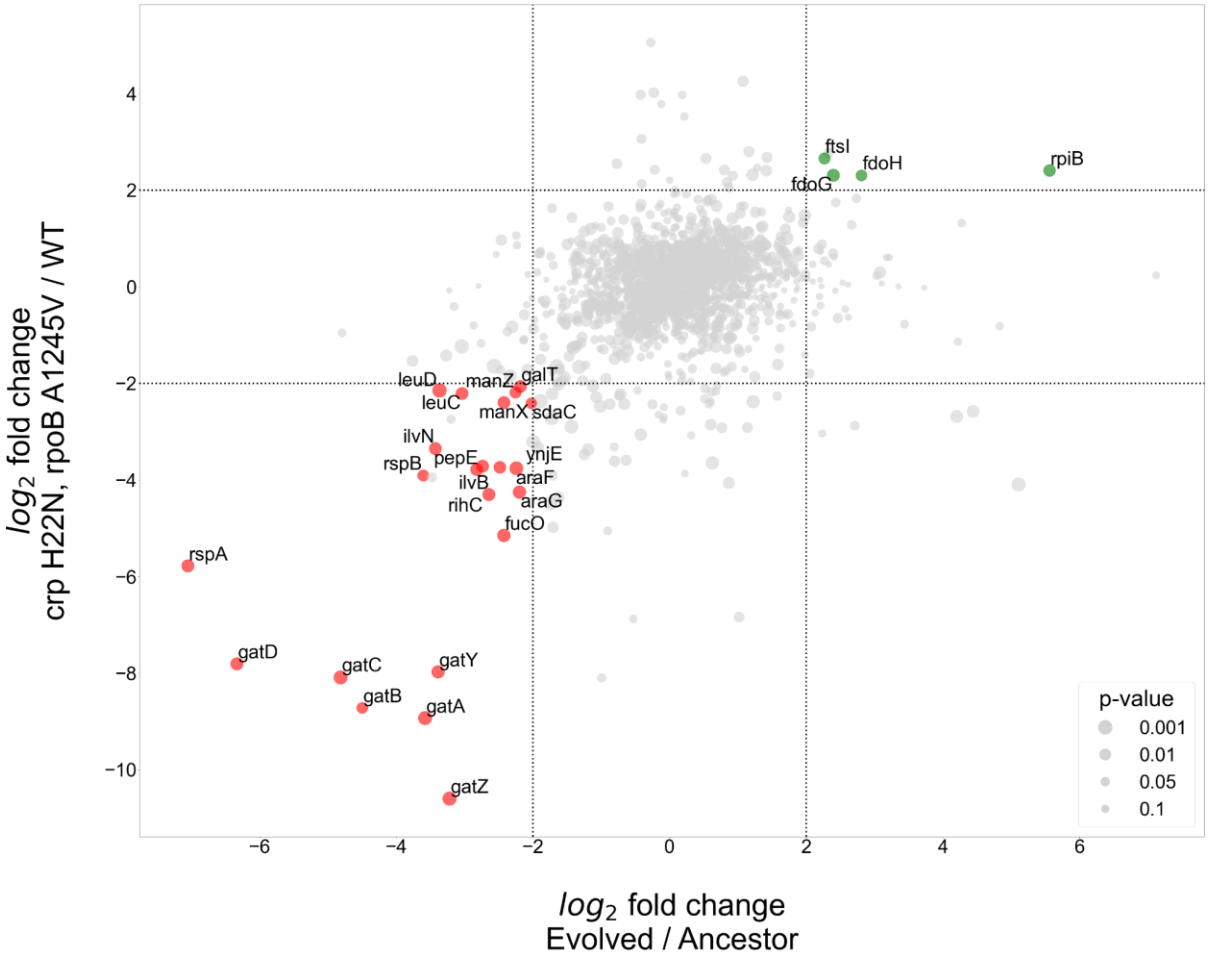

Fig. 5. Proteomics analyses of fold change in evolved/ancestor expression levels versus double mutant/WT. Comparison of the log2 fold change in the evolved autotrophic strain divided by the ancestor expression versus BW rpoB A1245V & crp H22N, divided by the WT. The size of dots represents significance, only genes with >4 fold-change who showed significance (<0.05 p-value) were annotated.

The proteomics analysis yielded few genes that had an upregulation of >4-fold in both the evolved versus ancestor and the double mutant versus BW base strain experiments. Two of these up-regulated genes, *fdoG* and *fdoH*, are subunits of the native formate dehydrogenase O enzyme complex (FDH-O) and are therefore related to the energy module. These genes were upregulated in both the evolved strain and in the strain with the *crp* H22N and *rpoB A1245V* double mutation (≈5-fold increase in both cases). The FDH-O enzyme complex takes part in the respiratory chain formed by formate dehydrogenase and dimethyl sulfoxide (DMSO) reductase. Here, the transfer of electrons from formate to DMSO (via a quinone electron carrier) is coupled to the generation of a proton-motive force across the cytoplasmic membrane for energy formation.

Another observed upregulated protein was ribose-phosphate-isomerase B (*rpiB*), which was up-regulated in the evolved strain (>40-fold increase) and in the double mutant ≈5-fold increase (Fig. 5). *rpiB* is an inducible isoenzyme of the constitutively expressed ribose-phosphate-isomerase A, a core enzyme in the rPP cycle. This might serve as evidence for a higher flux requirement through the pentose-phosphate-pathway, supported by expressing a second isoenzyme. Other explanations may include some form of regulation, perhaps allosteric, inhibiting RpiA activity, making the strain dependent on RpiB activity, or non-specific enzymatic activity of RpiB which is required by the cell. The fourth up-regulated gene - *ftsI*, a peptidoglycan DD-transpeptidase - does not have a clear explanation at this time.

At the other end of the expression spectrum, several pathways were down-regulated in both the evolved and in the BW double mutant strain (Fig. 5). Down-regulation of the starvation response, metabolic and PTS component genes seems to be at the heart of the adaptation to autotrophic growth, with several operons experiencing >4-fold decrease in expression. Most notably the *gat*, *ara* and *man* operons responsible for galactitol, arabinose and mannose degradation, respectively. Again, this is evident in both the evolved and the double mutant strains (Fig. 5). Other highly down-regulated metabolic pathways include amino acid biosynthesis, specifically valine and leucine, moreover we observed a down regulation in *rspA & rspB*, which encode for starvation sensing proteins. We suggest these results mean that many pathways usually up-regulated in *E. coli* during starvation may be redundant or even counter-productive in an autotrophic context, and decreasing their expression could result in increased fitness.

We additionally tested whether separating between the two regulatory mutations (*crp* and *rpoB*) might provide a clearer picture, e.g. if different up- or down-regulated genes are targeted by only one of these mutations and not the other. Therefore, we used single-mutant BW strains and compared them to the same BW wild-type as before. Essentially, the results did not change our conclusions and although in most cases the effect of the two mutations was cumulative, we did find at least one case where they were antagonistic (Fig. S7). For example, *fdoG* and *fdoH* showed a ≈4-fold *decrease* compared to wild type when the *crp* mutation was introduced, but (as stated earlier) had a ≈5-fold abundance *increase* in the double mutant, suggesting an epistatic effect between the *rpoB* and *crp* mutations (Fig. S6B & Fig. S7B). All the comparisons based on the single and double mutant strains can be found in the supplementary.

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

Along with the effect on expression, we suspected the two transcription-associated mutations (*crp* & *rpoB)* to also have a metabolic and physiological effect. Since Crp and RpoB are both essential to the autotrophic phenotype and are known to indirectly interact in the cell (Savery, Rhodius, and Busby 1996; Niu et al. 1996; Frendorf et al. 2019), we address them as one unit. The proteomics analysis, where the strain with both mutations seem to have a mutual effect that is in some cases different than in the individual mutants, supported this decision. Mutations in *crp* and *rpoB* are known to affect growth at different conditions such as types of media or carbon sources (Utrilla et al. 2016; Conrad et al. 2010; Cheng et al. 2014; Frendorf et al. 2019). 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 *fdh* exhibited >25% increase in the final OD compared to the wild type expressing *fdh* (Fig. 6A&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. 6A).

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. 6C).

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 but we hypothesise that it is related to the increased expression of Fdo observed in the proteomic analysis. We also note that as the proteomics results suggest - the mutations in global transcriptional machinery like Crp & RpoB could have many other effects on the autotrophic phenotype *(Haverkorn van Rijsewijk et al. 2011)* which await future exploration.

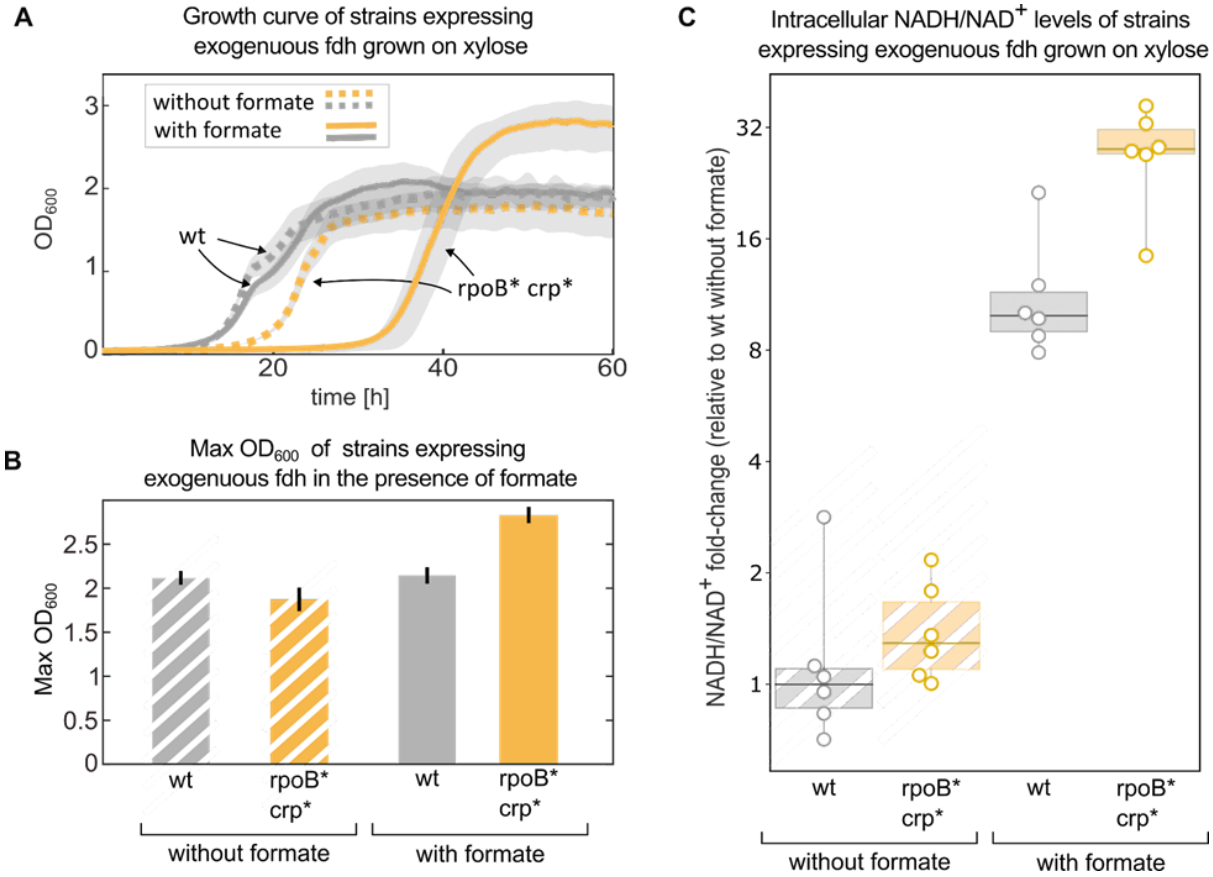

Fig. 6. mutations in *rpoB* and *crp* increase yield and intracellular NADH/NAD⁺ levels in fdh-expressing *E. coli* in the presence of formate. Growth experiment of BW25113 wild-type (grey) compared to a *crp* H22N *rpoB* A1245V mutant (orange). Both strains express *fdh*. **(A)** strains were grown 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 genomic 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, proteomic 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 three 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; (ii) facilitating the integration of the energy module via decoupling of formate harvesting into energy and redox, thereby increasing the $NADH/NAD^+$ ratio, which could have both thermodynamic and kinetic benefits (iii) and freeing protein resources by decreasing unneeded expression.

The mutation in *pgi* affects the activity of its gene product phosphoglucoisomerase - a metabolic enzyme that catalyses a reaction that branches out of the non-native rPP cycle. The enzyme converts F6P (a cycle intermediate) to G6P (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 (Gleizer et al. 2019; Gassler et al. 2020; Antonovsky et al. 2016; Barenholz et al. 2017; Flamholz et al. 2020; Herz et al. 2017). 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 RuBiCO-dependent strain resulted in increased concentration of rPP cycle intermediates. This exemplifies the key role branch-points play in autocatalytic cycles (Gleizer et al. 2019; Gassler et al. 2020; Antonovsky et al. 2016; Barenholz et al. 2017; Flamholz et al. 2020; Herz et al. 2017), making them a leading target for rational attempts to integrate non-native autocatalytic cycles into metabolism.

The remaining two mutations act as global regulators, which presents a challenge in distinguishing their specific adaptive impacts from their broader effects on cellular expression patterns. Using proteomic analyses, we have managed to identify certain genes that had a major change in expression when one of the regulating mutations was introduced, and when both were introduced together.

One main observation was the down-regulation of several metabolic pathways including catabolism of alternative carbon sources and amino acid biosynthesis. We suggest this is an adaptive role in improving cell resource allocation, and a means to modulate resource allocation, possibly freeing up proteome resources from alternative carbon metabolism proteins. This includes proteins such as PTS system components or ATP dependent transporters, that tend to incur high energetic costs.

Other reasons these catabolic and transport pathways may result in reduced fitness could include competition over metabolites with the rPP cycle, or by causing other metabolic blunders such as dead end products with potential toxicity effects under the new flux regime. Another observed effect the *crp* mutation shows is the up-regulation of *rpiB*, a typically unexpressed isoenzyme to one of the reactions in the rPP cycle. This expression increase could be beneficial or even essential in a scenario where *rpiA*, the main enzyme catalysing this reaction, cannot support the flux required by the rPP cycle, for example due to kinetic capacity limitations or allosteric inhibition. Further work is needed to investigate if up-regulation of *rpiB* is indeed essential for the autotrophic phenotype.

We also suggest these mutations play a role in improving the utilisation of formate, which is supported by the up-regulation of *fdoG* & *fdoH*. We found that a strain introduced with both the *crp* and *rpoB* mutations demonstrated a pronounced up-regulation of over 4-fold and also a concrete fitness advantage in growth conditions which include formate. Metabolomics analysis showed this strain also has a significant increase

in the NADH/NAD^+ ratio. NADH is the essential cofactor for the electron transfer reaction of the non-native rPP cycle (catalysed by the D-glyceraldehyde-3P dehydrogenase GAPDH) and a product of the non-native formate-based energy module. The importance of this up-regulation is also supported by natural chemolithoautotrophic metabolism, which uses two distinct ways to harvest energy and reducing power from the electron-donor. For example, in the model chemolithoautotroph *Cupriavidus necator*, the oxidation of formate is catalysed by two distinct types of FDH: a soluble, NAD^+-linked enzyme (S-FDH; EC 1.2.1.2) and a membrane-bound enzyme coupled directly to the respiratory chain (Friedrich, Bowien, and Friedrich 1979; Oh and Bowien 1998).

We suspect utilising formate dehydrogenase O (FDH-O) in the periplasm is enabling the cell to also utilise formate directly on the membrane to create a proton motive force for energy production without the need to exhaust the intracellular NADH pool, thus allowing for its accumulation. Increased levels of NADH creates favourable thermodynamics for the rPP cycle flux and could also offer improved kinetics. This result is in line with previous models, predicting that high cofactor levels are required for a stable metabolic state of the rPP cycle (Janasch et al. 2019)**.** We suggest this observation offers an avenue for future rational design attempts.

We continue by hypothesising 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.

Another reported observation is 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. Using metabolic growth-coupling as a temporary “metabolic scaffold” that can be removed, 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 (Aigner et al. 2017). 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. Aidelberg, Guy, Benjamin D. Towbin, Daphna Rothschild, Erez Dekel, Anat Bren, and Uri Alon. 2014. "Hierarchy of Non-Glucose Sugars in Escherichia Coli." *BMC Systems Biology* 8 (December): 133.

2. Aigner, H., R. H. Wilson, A. Bracher, L. Calisse, J. Y. Bhat, F. U. Hartl, and M. Hayer-Hartl. 2017. "Plant RuBisCo Assembly in with Five Chloroplast Chaperones Including BSD2." *Science* 358 (6368): 1272–78.

3. Ammar, Ehab M., Xiaoyi Wang, and Christopher V. Rao. 2018. "Regulation of Metabolism in Escherichia Coli during Growth on Mixtures of the Non-Glucose Sugars: Arabinose, Lactose, and Xylose." *Scientific Reports* 8 (1): 609.

4. Antonovsky, Niv, Shmuel Gleizer, Elad Noor, Yehudit Zohar, Elad Herz, Uri Barenholz, Lior Zelcbuch, et al. 2016. "Sugar Synthesis from CO2 in Escherichia Coli." *Cell* 166 (1): 115–25.

5. 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 (February): 2006.0008.

6. Bang, Junho, Chang Hun Hwang, Jung Ho Ahn, Jong An Lee, and Sang Yup Lee. 2020. "Escherichia Coli Is Engineered to Grow on CO2 and Formic Acid." *Nature Microbiology*. https://doi.org/10.1038/s41564-020-00793-9.

7. Barenholz, Uri, Dan Davidi, Ed Reznik, Yinon Bar-On, Niv Antonovsky, Elad Noor, and Ron Milo. 2017. "Design Principles of Autocatalytic Cycles Constrain Enzyme Kinetics and Force Low Substrate Saturation at Flux Branch Points." *eLife* 6 (February). https://doi.org/10.7554/eLife.20667.

8. 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.

9. Baumschabl, Michael, Özge Ata, Bernd M. Mitic, Lisa Lutz, Thomas Gassler, Christina Troyer, Stephan Hann, and Diethard Mattanovich. 2022. "Conversion of CO2 into Organic Acids by Engineered Autotrophic Yeast." *Proceedings of the National Academy of Sciences of the United States of America* 119 (47): e2211827119.

10. Charusanti, Pep, Tom M. Conrad, Eric M. Knight, Karthik Venkataraman, Nicole L. Fong, Bin Xie, Yuan Gao, and Bernhard Ø. Palsson. 2010. "Genetic Basis of Growth Adaptation of Escherichia Coli after Deletion of Pgi, a Major Metabolic Gene." *PLoS Genetics* 6 (11): e1001186.

11. Chen, Frederic Y-H, Hsin-Wei Jung, Chao-Yin Tsuei, and James C. Liao. 2020. "Converting Escherichia Coli to a Synthetic Methylotroph Growing Solely on Methanol." *Cell* 182 (4): 933–46.e14.

12. Cheng, Kian-Kai, Baek-Seok Lee, Takeshi Masuda, Takuro Ito, Kazutaka Ikeda, Akiyoshi Hirayama, Lingli Deng, et al. 2014. "Global Metabolic Network Reorganization by Adaptive Mutations Allows Fast Growth of Escherichia Coli on Glycerol." *Nature Communications* 5: 3233.

13. 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.

14. Claassens, Nico J., Charles A. R. Cotton, Dennis Kopljar, and Arren Bar-Even. 2019. "Making Quantitative Sense of Electromicrobial Production." *Nature Catalysis*. https://doi.org/10.1038/s41929-019-0272-0.

15. Conrad, Tom M., Michael Frazier, Andrew R. Joyce, Byung-Kwan Cho, Eric M. Knight, Nathan E. Lewis, Robert Landick, and Bernhard Ø. Palsson. 2010. "RNA Polymerase Mutants Found through Adaptive Evolution Reprogram Escherichia Coli for Optimal Growth in Minimal Media." *Proceedings of the National Academy of Sciences of the*

United States of America 107 (47): 20500–505.

16. Davidi, Dan, Elad Noor, Wolfram Liebermeister, Arren Bar-Even, Avi Flamholz, Katja Tummler, Uri Barenholz, Miki Goldenfeld, Tomer Shlomi, and Ron Milo. 2016. "Global Characterization of in Vivo Enzyme Catalytic Rates and Their Correspondence to in Vitro Kcat Measurements." *Proceedings of the National Academy of Sciences of the United States of America* 113 (12): 3401–6.

17. 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.

18. "Engineering Cellular Metabolism." 2016. *Cell* 164 (6): 1185–97.

19. Flamholz, Avi I., Eli Dugan, Cecilia Blikstad, Shmuel Gleizer, Roee Ben-Nissan, Shira Amram, Niv Antonovsky, et al. 2020. "Functional Reconstitution of a Bacterial CO2 Concentrating Mechanism in Escherichia Coli." *eLife* 9 (October). https://doi.org/10.7554/eLife.59882.

20. Frendorf, Pernille Ott, Ida Lauritsen, Agnieszka Sekowska, Antoine Danchin, and Morten H. H. Nørholm. 2019. "Mutations in the Global Transcription Factor CRP/CAP: Insights from Experimental Evolution and Deep Sequencing." *Computational and Structural Biotechnology Journal* 17 (June): 730–36.

21. Friedlingstein, Pierre, Michael O'Sullivan, Matthew W. Jones, Robbie M. Andrew, Luke Gregor, Judith Hauck, Corinne Le Quéré, et al. 2022. "Global Carbon Budget 2022." *Earth System Science Data* 14 (11): 4811–4900.

22. Friedrich, C. G., B. Bowien, and B. Friedrich. 1979. "Formate and Oxalate Metabolism in Alcaligenes Eutrophus." *Journal of General Microbiology* 115 (1): 185–92.

23. Gassler, Thomas, Michael Sauer, Brigitte Gasser, Michael Egermeier, Christina Troyer, Tim Causon, Stephan Hann, Diethard Mattanovich, and Matthias G. Steiger. 2020. "The Industrial Yeast Pichia Pastoris Is Converted from a Heterotroph into an Autotroph

Capable of Growth on CO2." *Nature Biotechnology* 38 (2): 210–16.

24. Gleizer, Shmuel, Yinon M. Bar-On, Roee Ben-Nissan, and Ron Milo. 2020. "Engineering Microbes to Produce Fuel, Commodities, and Food from CO2." *Cell Reports Physical Science* 1 (10): 100223.

25. Gleizer, Shmuel, Roee Ben-Nissan, Yinon M. Bar-On, Niv Antonovsky, Elad Noor, Yehudit Zohar, Ghil Jona, et al. 2019. "Conversion of Escherichia Coli to Generate All Biomass Carbon from CO2." *Cell* 179 (6): 1255–63.e12.

26. 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.

27. Haverkorn van Rijsewijk, Bart R. B., Annik Nanchen, Sophie Nallet, Roelco J. Kleijn, and Uwe Sauer. 2011. "Large-Scale 13C-Flux Analysis Reveals Distinct Transcriptional Control of Respiratory and Fermentative Metabolism in Escherichia Coli." *Molecular Systems Biology* 7 (March): 477.

28. Herz, Elad, Niv Antonovsky, Yinon Bar-On, Dan Davidi, Shmuel Gleizer, Noam Prywes, Lianet Noda-Garcia, et al. 2017. "The Genetic Basis for the Adaptation of E. Coli to Sugar Synthesis from CO 2." *Nature Communications* 8 (1): 1–10.

29. Jackson, Debra W., Jerry W. Simecka, and Tony Romeo. 2002. "Catabolite Repression of Escherichia Coli Biofilm Formation." *Journal of Bacteriology* 184 (12): 3406–10.

30. Janasch, Markus, Johannes Asplund-Samuelsson, Ralf Steuer, and Elton P. Hudson. 2019. "Kinetic Modeling of the Calvin Cycle Identifies Flux Control and Stable Metabolomes in Synechocystis Carbon Fixation." *Journal of Experimental Botany* 70 (3): 973–83.

31. Keller, Philipp, Michael A. Reiter, Patrick Kiefer, Thomas Gassler, Lucas Hemmerle, Philipp Christen, Elad Noor, and Julia A. Vorholt. 2022. "Generation of an Escherichia Coli Strain Growing on Methanol via the Ribulose Monophosphate Cycle." *Nature Communications* 13 (1): 5243.

32. Kim, Seohyoung, Steffen N. Lindner, Selçuk Aslan, Oren Yishai, Sebastian Wenk, Karin Schann, and Arren Bar-Even. 2020. "Growth of E. Coli on Formate and Methanol via the Reductive Glycine Pathway." *Nature Chemical Biology* 16 (5): 538–45.

33. Landis, L., J. Xu, and R. C. Johnson. 1999. "The cAMP Receptor Protein CRP Can Function as an Osmoregulator of Transcription in Escherichia Coli." *Genes & Development* 13 (23): 3081–91.

34. Leger, Dorian, Silvio Matassa, Elad Noor, Alon Shepon, Ron Milo, and Arren Bar-Even. 2021. "Photovoltaic-Driven Microbial Protein Production Can Use Land and Sunlight More Efficiently than Conventional Crops." *Proceedings of the National Academy of Sciences of the United States of America* 118 (26). https://doi.org/10.1073/pnas.2015025118.

35. Li, Han, Paul H. Opgenorth, David G. Wernick, Steve Rogers, Tung-Yun Wu, Wendy Higashide, Peter Malati, Yi-Xin Huo, Kwang Myung Cho, and James C. Liao. 2012. "Integrated Electromicrobial Conversion of CO2 to Higher Alcohols." *Science* 335 (6076): 1596.

36. Niu, W., Y. Kim, G. Tau, T. Heyduk, and R. H. Ebright. 1996. "Transcription Activation at Class II CAP-Dependent Promoters: Two Interactions between CAP and RNA Polymerase." *Cell* 87 (6): 1123–34.

37. Oh, Jeong-Il, and Botho Bowien. 1998. "Structural Analysis of the Fds Operon Encoding the NAD+-Linked Formate Dehydrogenase of Ralstonia Eutropha *." *The Journal of Biological Chemistry* 273 (41): 26349–60.

38. Olofsson, Jörgen, and Thomas Hickler. 2008. "Effects of Human Land-Use on the Global Carbon Cycle during the Last 6,000 Years." *Vegetation History and Archaeobotany* 17 (5): 605–15.

39. Phaneuf, P. V., D. Gosting, and B. O. Palsson. 2019. "ALEdb 1.0: A Database of Mutations from Adaptive Laboratory Evolution Experimentation." *Nucleic Acids*. https://academic.oup.com/nar/article-abstract/47/D1/D1164/5144149.

40. Popov, V. O., and V. S. Lamzin. 1994. “NAD(+)-Dependent Formate Dehydrogenase.” *Biochemical Journal* 301 (Pt 3) (Pt 3): 625–43.

41. Satanowski, Ari, Beau Dronsella, Elad Noor, Bastian Vögeli, Hai He, Philipp Wichmann, Tobias J. Erb, Steffen N. Lindner, and Arren Bar-Even. 2020. “Awakening a Latent Carbon Fixation Cycle in Escherichia Coli.” *Nature Communications* 11 (1): 5812.

42. Savery, N., V. Rhodius, and S. Busby. 1996. “Protein-Protein Interactions during Transcription Activation: The Case of the Escherichia Coli Cyclic AMP Receptor Protein.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 351 (1339): 543–50.

43. Schwander, Thomas, Lennart Schada von Borzyskowski, Simon Burgener, Niña Socorro Cortina, and Tobias J. Erb. 2016. “A Synthetic Pathway for the Fixation of Carbon Dioxide in Vitro.” Science 354 (6314): 900–904.

44. Shi, Wei, Yanan Jiang, Yibin Deng, Zigang Dong, and Bin Liu. 2020. “Visualization of Two Architectures in Class-II CAP-Dependent Transcription Activation.” PLoS Biology 18 (4): e3000706.

45. Tenaillon, Olivier, Alejandra Rodríguez-Verdugo, Rebecca L. Gaut, Pamela McDonald, Albert F. Bennett, Anthony D. Long, and Brandon S. Gaut. 2012. “The Molecular Diversity of Adaptive Convergence.” Science 335 (6067): 457–61.

46. 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.

47. Utrilla, Jose, Edward J. O’Brien, Ke Chen, Douglas McCloskey, Jacky Cheung, Harris Wang, Dagoberto Armenta-Medina, Adam M. Feist, and Bernhard O. Palsson. 2016. “Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-Scale Mechanism of Adaptive Evolution.” *Cell Systems* 2 (4): 260–71.

48. Van Peteghem, L., M. Sakarika, S. Matassa, I. Pikaar, R. Ganigué, and K. Rabaey. 2022.

"Towards New Carbon-Neutral Food Systems: Combining Carbon Capture and Utilization with Microbial Protein Production." *Bioresource Technology* 349 (April): 126853.

49. Widjaja, Arief, Masahiro Shiroshima, Takashi Oka, Masahiro Yasuda, Hiroyasu Ogino, Kazutaka Miyatake, Hiroshi Nakajima, and Haruo Ishikawa. 1998. "The Kinetics and Mechanism of a Reaction Catalyzed by Bacillus Stearothermophilus Phosphoglucose Isomerase." *Journal of Fermentation and Bioengineering* 86 (3): 324–31.

50. Yishai, Oren, Madeleine Bouzon, Volker Döring, and Arren Bar-Even. 2018. "In Vivo Assimilation of One-Carbon via a Synthetic Reductive Glycine Pathway in Escherichia Coli." *ACS Synthetic Biology* 7 (9): 2023–28.

51. 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.

Acknowledgements: We thank Alon Barshap, Alon Savidor, 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, Yafit Sugas, Yinon Bar-on, Yotam David, Yuval Kushmaro and Yuval Rosenberg for their support of this project. We acknowledge the De Botton Protein Profiling institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science for their services and contribution to 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 S8

Data S1 to S17

Materials and Methods

Strains

We generated an engineered ancestor strain for chemostat evolution based on the *Escherichia coli* BW25113 strain (Grenier et al. 2014). We used P1 transduction (Thomason, Costantino, and Court 2007) to transfer knockout alleles from the KEIO strain collection (Thomason, Costantino, and Court 2007; Baba et al. 2006) 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 KmR selection marker was removed by using the FLP recombinase encoded by the pCP20 temperature-sensitive plasmid (Cherepanov and Wackernagel 1995). 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). The transfer of mutations from evolved clones for the generation of ancestor 2.0, and other mutated strains was also accomplished by P1 transduction as reported in Herz et al. 2017. P1 lysate was prepared from Keio collection strains that had a Kanamycin (Km) resistance marker located near the desired allele (as specified in Supplementary Data S1). The strains carrying the desired mutated allele were transduced with P1 phages containing the resistance marker and then plated on Km plates. Donor strains were selected based on the presence of the desired mutated allele and the Km resistance marker in their colonies. A P1 lysate from each donor strain was used to transfer the desired mutations to a recipient strain. Because the transferred DNA fragment to the recipient was relatively large (approximately 100 kb), including the selection marker, it often carried adjacent genes. Genotyping was performed using Sanger sequencing. For further genomic modification, pCP20, a temperature-sensitive plasmid encoding the FLP recombinase from the Clone and Strain Repository unit at WIS, was used following standard procedures. This allowed for the removal of Km resistance markers, enabling iterative rounds of genome editing.

Plasmids

To create the pFDH plasmid, an *E. coli* codon optimised DNA sequence of formate dehydrogenase from the methytholotrophic bacterium *Pseudomonas sp. 101* (Popov and Lamzin 1994) was cloned with an N-terminal His-tag into a pZ plasmid (Expressys, Germany) under a constitutive promoter and with a strong ribosome binding site (rbs B of (Zelcbuch et al. 2013)). The plasmid has a pMB1 origin of replication and therefore is present in high copy number. We replaced the KmR selection marker on the plasmid with a StrepR 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 (Gleizer et al. 2019). Details regarding the pCBB plasmid are reported in (Antonovsky et al. 2016).

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 CO2 (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% CO2 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 labelling experiment

A culture of cells that were growing on naturally labelled sodium formate in an elevated CO2 (10%, naturally labelled) incubator (New Brunswick S41i CO2 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" labelling 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 CO2 (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" labelling 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}CO2 (Cambridge Isotope Laboratories, USA) + 90% air or 10% ^{12}CO2 + 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 labelling 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 labelling experiment was conducted in triplicates (n=3), with ^{12}CO2 the triplicates were pooled and measured as three technical replicates. The predicted range of labelling under ^{13}CO2 was calculated using the following equation: $f_n = 1 - 0.5^n$, where f is the ^{13}C fraction and n is the number of doublings, as the remaining unlabelled biomass is decreased 2-fold at each doubling. The ^{13}C fraction was calculated for 3 and 5 doubling to create the predicted range.

Growth comparison between engineered autotrophic strains

The growth was conducted in a DASBox mini fermentation system (Eppendorf, Germany). The starting volume of each bioreactor was 150 ml M9 media supplemented with 60 mM Formate. The minimal media also included trace elements and vitamin B1 was omitted. The atmosphere was elevated 10% CO_2 5% O_2 and 85% N_2. The culture was grown to saturation and then diluted in a ratio of 1:30, to obtain 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 (Antonovsky et al. 2016). Briefly, for protein bound amino acids ≈3 mL of culture at OD600 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) H2O + 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 $^{13}C6$-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 fi 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 (Herz et al. 2017; Antonovsky et al. 2016). The prepared libraries were sequenced by a Miseq machine (Illumina). Analysis of the sequencing data was performed as previously described in (Herz et al. 2017; Antonovsky et al. 2016; Gleizer et al. 2019) with the breseq software (Barrick et al. 2014). Cultivation Conditions for Metabolome Sampling

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 CO2 (10%) shaker-incubator (New Brunswick S41i CO2 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 phosphates, extracts were mixed with a ^{13}C-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 (Donati et al. 2021). 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 ^{12}C and ^{13}C peak heights was used to quantify metabolites. $^{12}C/^{13}C$ 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 ^{12}C peak heights were used to quantify metabolites. Retention time of hexose-phosphates were determined with authentic standards of glucose-6-P

Chemostat proteomics experiment

The growth was conducted in a DASBox mini fermentation system (Eppendorf, Germany). The starting volume of each bioreactor was 150 ml M9 media supplemented with either 30 mM and 10mM D-xylose for the evolved and ancestor samples or only 10mM D-xylose for BW including compact set mutations. The minimal media also included trace elements and vitamin B1 was omitted. In experiments including the evolved strain and the ancestor strain, the atmosphere had elevated CO_2 (10%), and the chemostat dilution rate was set to 0.03 h^{-1} corresponding to a doubling time of 24h.

Proteomics analysis

The cell pellets were lysed and subjected to in-solution tryptic digestion using the S-Trap method (by Protifi). The resulting peptides were analyzed using nanoflow liquid chromatography (nanoAcquity) coupled to high resolution, high mass accuracy mass spectrometry (Thermo Q Excative HFX). Each sample was analyzed on the instrument separately in discovery mode in a random order. Raw data was processed with MaxQuant v1.6.6.0. The data was searched with the Andromeda search engine against the E. coli BW25113 (K12) protein database as downloaded from Uniprot, appended with the mutant protein sequences and common lab protein contaminants. Search parameters included the following modifications: Fixed modification- cysteine carbamidomethylation. Variable modifications- methionine oxidation, and asparagine and glutamine deamidation. The quantitative comparisons were calculated using Perseus v1.6.0.7. Decoy hits were filtered out and only proteins that were detected in at least 2 replicates of at least one experimental group were kept.

Data, Materials, and Software Availability

All data and code are stored in a publicly available GitLab repository (https://gitlab.com/milo-lab-public/compact-autocoli).

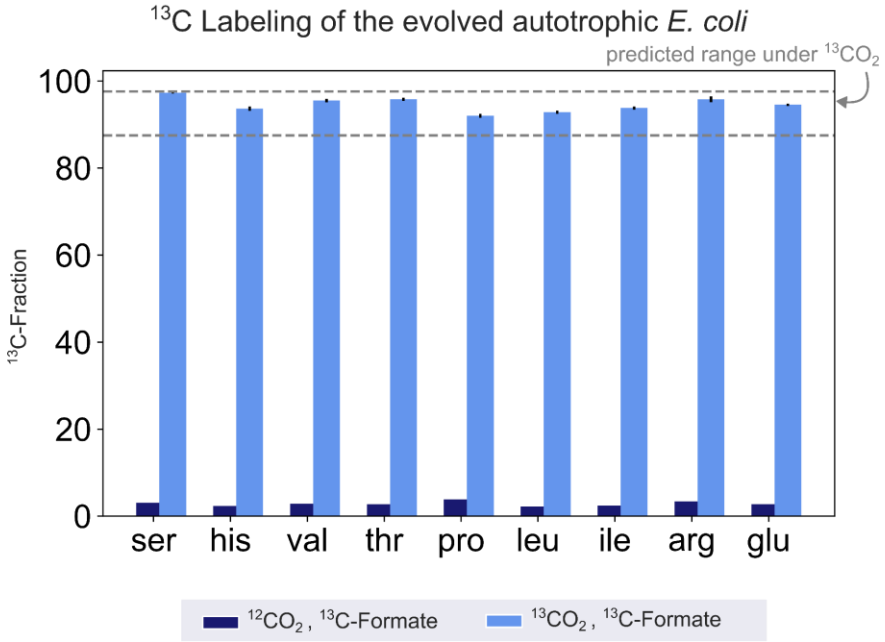

Fig. S1. The fractional contribution of $^{13}CO_2$ to various protein-bound amino acids evolved autotrophic strain after ≈4 doublings on $^{13}CO_2$ and ^{13}C labelled formate (light blue, mean of n=3) reached the expected ^{13}C labelling fraction of the biosynthesized amino acids (see methods). When grown in naturally labelled CO_2 and ^{13}C labelled 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.

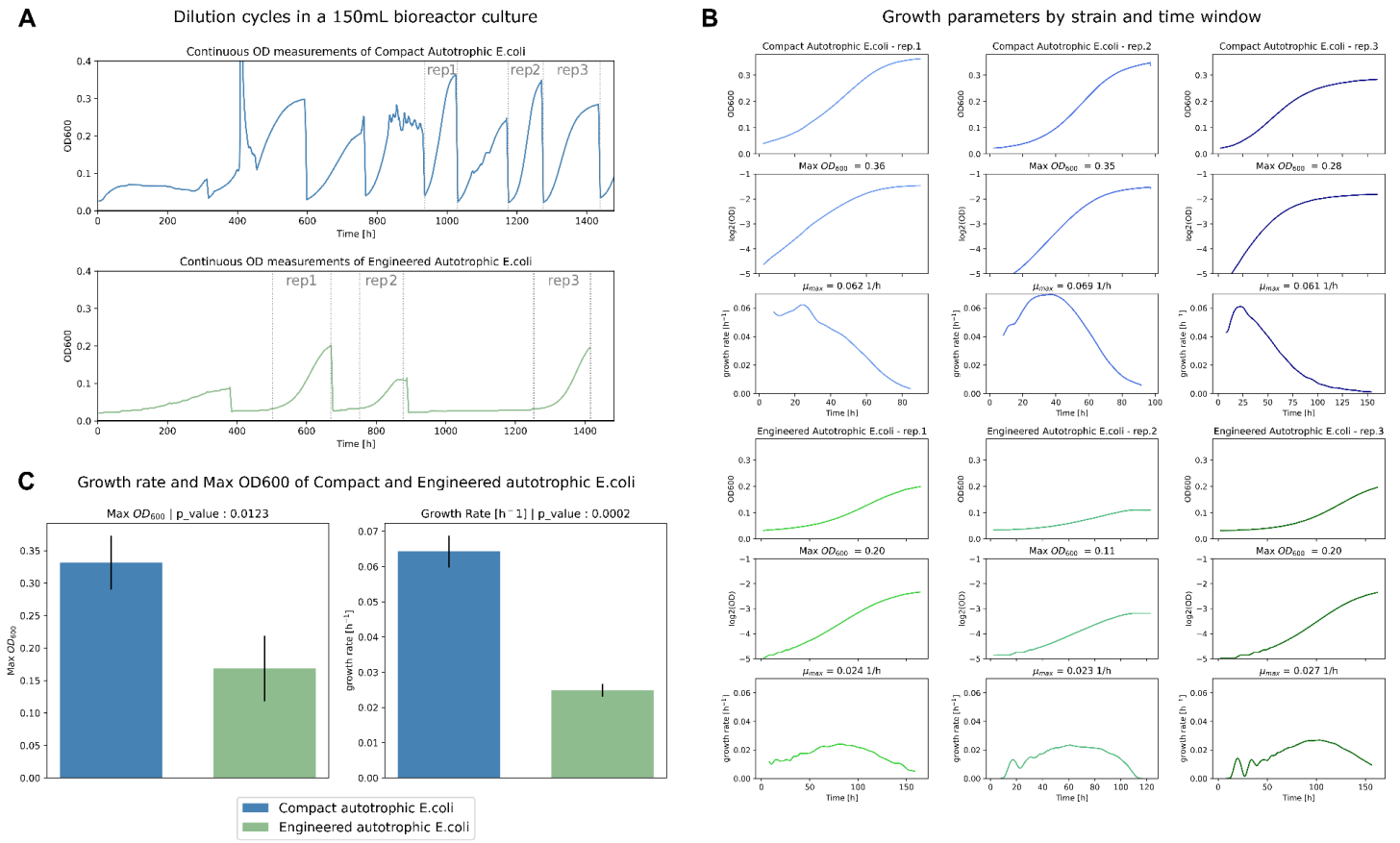

Fig. S2. Growth of Compact versus Engineered autotrophic E. coli (A) Growth curve of continuous cultures in liquid M9 minimal media supplemented with 60mM sodium formate and sparged with a gas mixture of 10% CO_2, 5% oxygen, and 85% nitrogen. Top: Compact autotrophic E. coli (green). Bottom: Engineered autotrophic strain (blue). The experiment was conducted in a DASBox mini fermentation system (Eppendorf, Germany) and the cultures were diluted in a 1:30 ratio once reaching stationarity. Dotted lines (grey) demarcate the manually selected time-windows used for the different technical replicates. (B) Growth parameters of the different technical replicates (as marked in panel A) that were used to calculate the maximal OD600. Growth rates were calculated by linear regression of the log-transformed OD600 values (data and source code can be found at https://gitlab.com/milo-lab-public/compact-autocoli). Top: replicates of the Compact autotrophic strain (green). Bottom: replicates of the Engineered autotrophic strain (blue). (C) Maximal OD600 and calculated growth rate of the Compact autotrophic E. coli (green) and the Engineered autotrophic E. coli (blue). Bar heights represent averages (n=3) of the maximal ODs and growth rates of each replicate. Error bars represent standard errors. P-values are based on Student's t-test with equal variance.

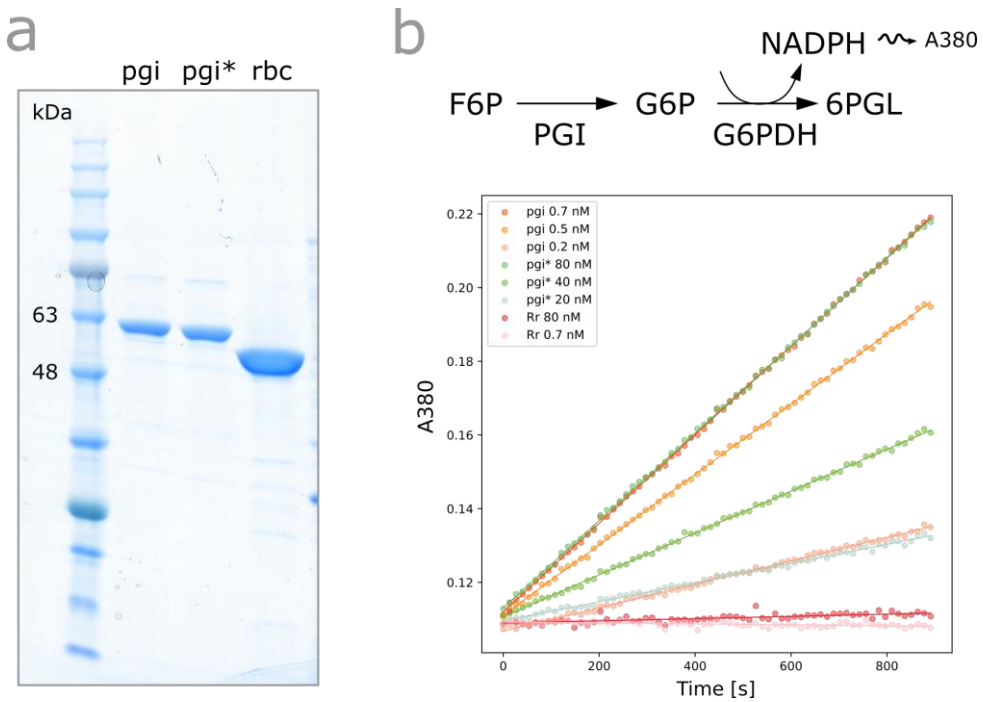

Fig. S3. In vitro characterization of Pgi^{H386Y}. (a) SDS-PAGE of purified Pgi wild-type, Pgi H386Y and Rubisco after expression in E. coli. **(b)** Spectrophotometric coupled assay determining the rate of the isomerization of fructose-6-phosphate (F6P) to glucose-6-phosphate (G6P) for purified Pgiwild-type and PgiH386Y. Rubisco was used as a negative control setting the noise level.

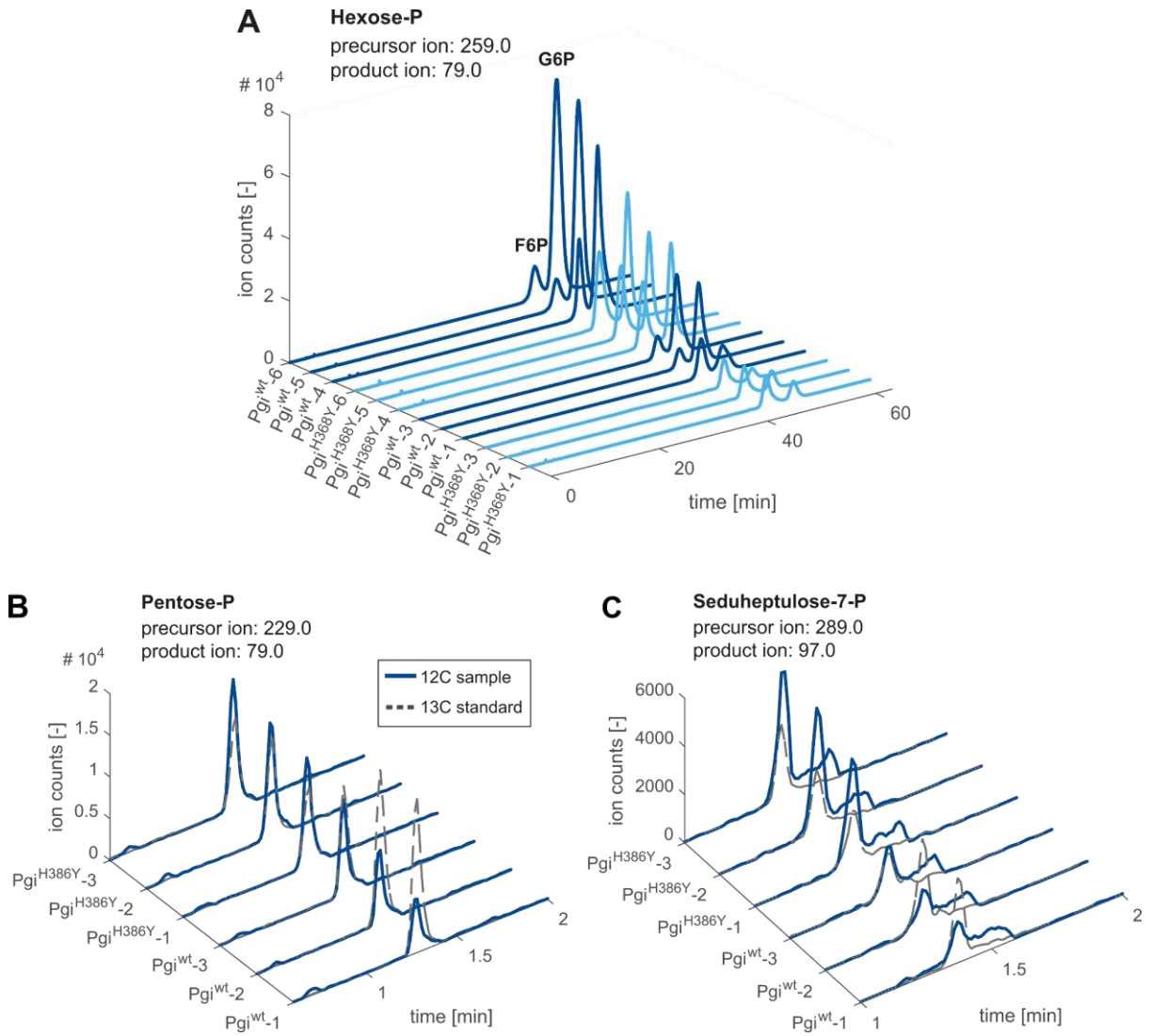

Fig. S4. LC-MS/MS chromatogram of intracellular metabolites involved in the autocatalytic cycle in *E. coli* ancestor with Pgi^{H386Y} mutation and ancestor with a wild-type Pgi.

(A) Chromatogram of fructose-6-phosphate (F6P) and glucose-6-phosphate (G6P). Precursor ion (259.0) and product ion (79.0) were used for ^{12}C 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 ^{12}C 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 ^{12}C chromatograms (continuous line) and ^{13}C chromatograms (dashed line) with precursor ion and product ion used for ^{12}C detection as indicated. Samples were measured using an Acquity BEH Amide column and the ratio of ^{12}C (sample) and ^{13}C (internal standard) peak heights was used to quantify metabolites. **(C)** sedoheptulose-7P (S7P) chromatogram of the same samples and conditions as indicated in B.

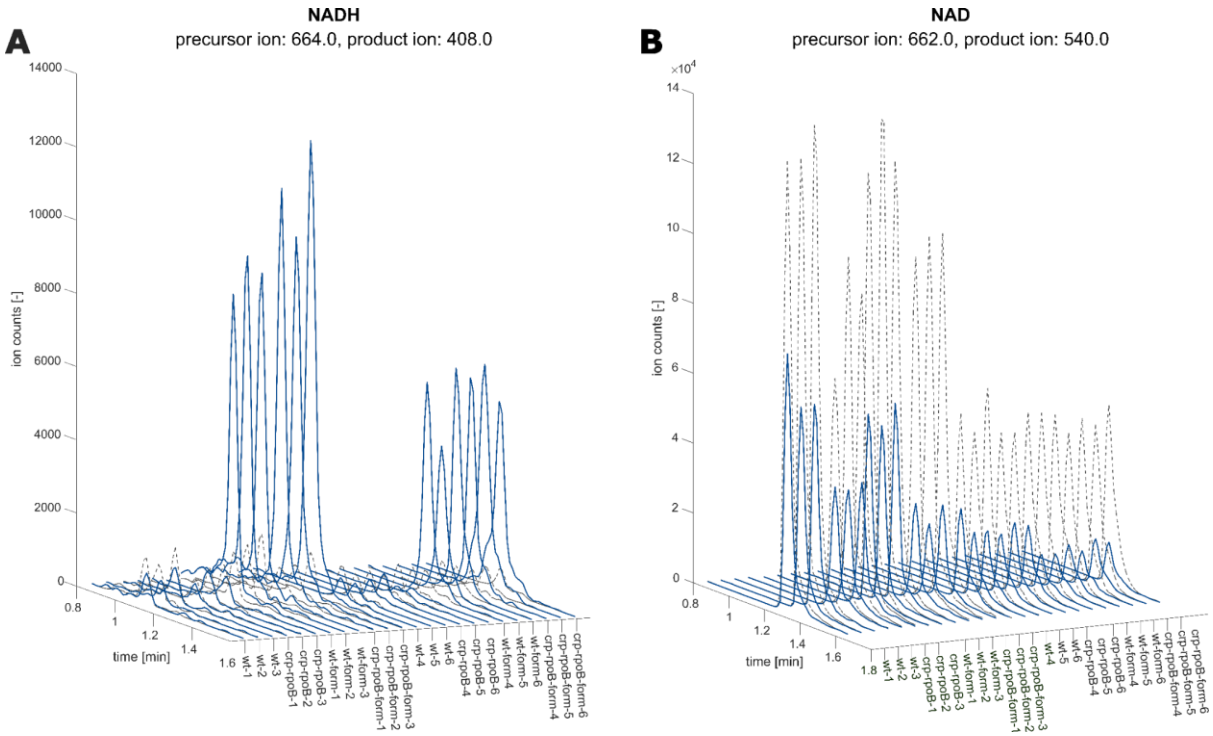

Fig. S5. LC-MS/MS chromatograms of NADH and NAD in fdh-expressing *E. coli* in the presence and absence of formate. **(A)** NADH chromatogram of *E. coli* BW25113 wild-type compared to a crpH22N rpoBA1245V mutant. Experiments were performed in two batches of n=3 cultures. Chromatograms are sorted accordingly. Overlay of ^{12}C chromatograms (continues line) and ^{13}C chromatograms (dashed line). Precursor ion and product ion used for ^{12}C 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 ^{12}C (sample) and ^{13}C (internal standard) peak heights was used to quantify metabolites.

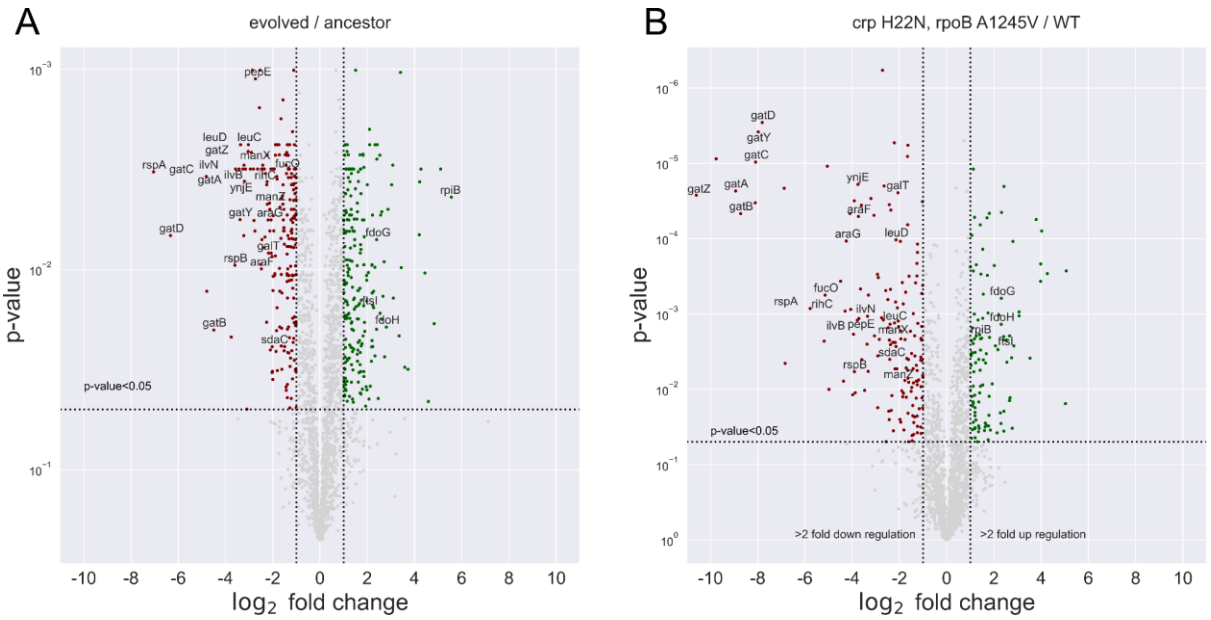

Fig. S6. proteomics analysis of evolved strain and BW double-mutant The differential protein expression between different strains. Red dots represent downregulated genes, Green - up-regulated. Only the genes that were had >4 fold change and <0.05 p-value in both sets were annotated. **(A)** Evolved / ancestor. **(B)** *rpoB* A1245V, *crp* H22N / WT.

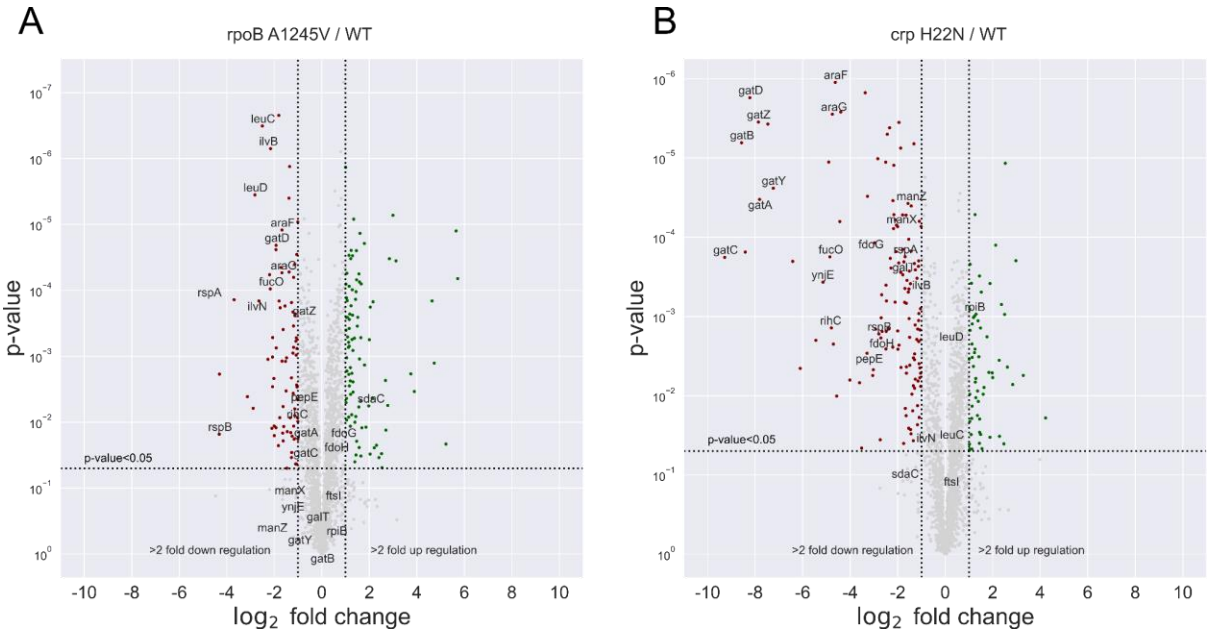

Fig. S7. proteomics analysis of BW mutants The differential protein expression between different strains. Red dots represent downregulated genes, Green - up-regulated. Only the genes that were had >4 fold change and <0.05 p-value in both of figure S6 A&B sets were annotated **(A)** *rpoB* A1245V / WT. **(B)** *crp* H22N / WT.

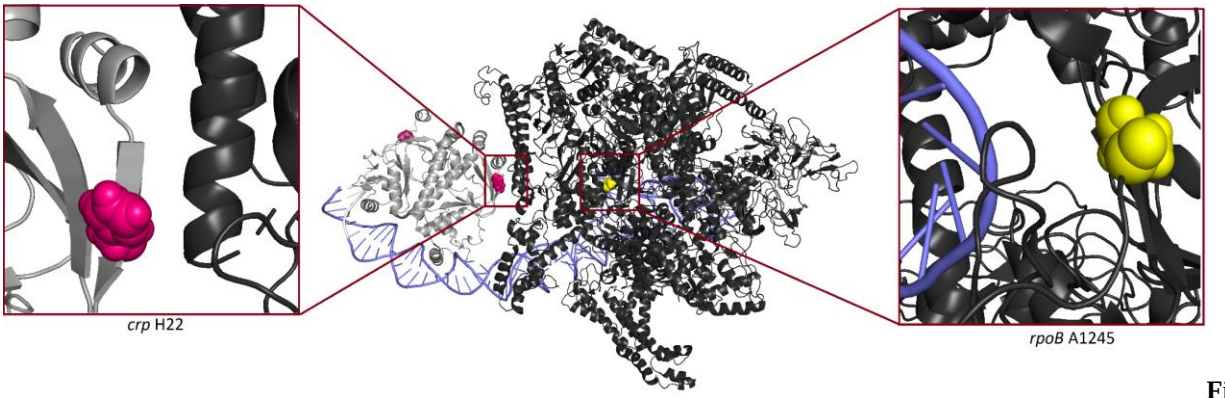

Fig. S8. 3D structure of wild type RNA polymerase and cAMP receptor protein In light grey Crp, in black RNA polymerase (the holoenzyme which includes RpoB), in purple DNA, mutated positions are shown in pink (Crp H22) and in yellow (RpoB A1245). Crp H22 is located in activation region 2 of Crp. The protein structure was taken from (Shi et al. 2020).

Data S1. (GitLab)
iterative evolution mutations

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

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

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

Data S5. (GitLab)
pgi kinetic assays data

Data S6. (GitLab)
G6PvsF6PResults

Data S7. (GitLab)
PPP metabolites

Data S8. (GitLab)
Growth experiment WT vs DM FDH variants CO2 40mMFor DATA

Data S9. (GitLab)
Growth experiment WT vs DM FDH variants CO2 40mMFor MAP

Data S10. (GitLab)
Growth experiment WT vs DM FDH variants CO2 NoFormate DATA

Data S11. (GitLab)
Growth experiment WT vs DM FDH variants CO2 NoFormate MAP

Data S12. (GitLab)
rpoB crp metabolomics

Data S13. (GitLab)
proteomics Evo Anc

Data S14. (GitLab)
proteomics WT variants

Data S15. (GitLab)
Compact AutoColi dasgib log

Data S16. (GitLab)
Scaffoldless Autocoli dasgib log

Data S17. (GitLab)
CRP and RNA polymerase with mutations