# Protein allocation and utilization in the versatile chemolithoautotroph *Cupriavidus necator*

Michael Jahn[1], Nick Crang[1], Markus Janasch[1], Andreas Hober[1], Björn Forsström[1], Kyle Kimler[1,2], Alexander Mattausch[1,3], Qi Chen[1], Johannes Asplund-Samuelsson[1], Elton P. Hudson[1]

## Affiliation

[1]School of Engineering Sciences in Chemistry, Biotechnology and Health, Science for Life Laboratory, KTH – Royal Institute of Technology, Stockholm, Sweden.

[2]current affiliation: The Broad Institute and Boston Children's Hospital, Boston, MA, United States of America

[3]current affiliation: Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Heidelberg, Germany

## Corresponding Author and Lead Contact

Elton P. Hudson, paul.hudson@scilifelab.se

## Keywords

*Cupriavidus necator*; *Ralstonia eutropha*; cellular economy; resource allocation; resource balance; genome scale model; proteomics; mass spectrometry; substrate limitation

# Highlights

- A large fraction of the *C. necator* proteome is related to environmental readiness
- Highly utilized enzymes are more abundant and less variable
- Autotrophy related enzymes are largely underutilized
- Re-assimilation of $CO_2$ *via* CBB cycle does not provide a fitness benefit on sugars

## Summary

Bacteria must balance the different needs for substrate assimilation, growth functions, and resilience in order to thrive in their environment. Of all cellular macromolecules, the bacterial proteome is by far the most important resource and its size is limited. Here, we investigated how the highly versatile 'knallgas' bacterium *Cupriavidus necator* reallocates protein resources when grown on different limiting substrates and with different growth rates. We determined protein quantity by mass spectrometry and estimated enzyme utilization by resource balance analysis modeling. We found that *C. necator* invests a large fraction of its proteome in functions that are hardly utilized. Of the enzymes that are utilized, many are present in excess abundance. One prominent example is the strong expression of CBB cycle genes such as Rubisco during growth on fructose. Modeling and mutant competition experiments suggest that $CO_2$-reassimilation through Rubisco does not provide a fitness benefit for heterotrophic growth, but is rather an investment in readiness for autotrophy.

## Introduction

*Cupriavidus necator* (formerly *Ralstonia eutropha*) is a model aerobic lithoautotroph and formatotroph, and is notable for production of the storage polymer polyhydroxybutyrate (PHB) [Yishai et al., 2016, Brigham, 2019]. *Cupriavidus necator* H16 (hereafter abbreviated *C. necator*) is a soil-dwelling bacterium with a large genome (~6,600 genes) distributed on two chromosomes and one megaplasmid [Pohlmann et al, 2006]. It features a wide arsenal of metabolic pathways for xenobiotics degradation, hydrogen and formate oxidation, carbon fixation *via* the Calvin-Bensson-Bassham (CBB) cycle, and utilization of nitrate/nitrite as alternative electron acceptors (de-nitrification) [Cramm, 2008]. Several operons for substrate assimilation are present in multiple copies, often on different chromosomes (e.g. *cbb* operon, hydrogenases, formate dehydrogenases). A detailed reconstruction of its metabolic network suggested that it can metabolize 229 compounds [Park et al., 2011]. Interestingly, *C. necator* prefers organic acids as growth substrate over sugars. The only sugar that supports growth is fructose, which is metabolized *via* the Entner-Doudoroff (ED) pathway [Alagesan et al., 2018]. Although the metabolic versatility of *C. necator* is interesting from a biotechnological point of view, we wondered if it does not come at a considerable cost for the cell. In particular, whether the expression of various substrate assimilation pathways is efficiently

regulated under different conditions, and if gene expression is optimal to maximize growth or rather another trait such as environmental readiness. The 'cellular economy' concept entails that an organism has a limited pool of (enzyme) resources and must re-allocate resources to different functions in order to meet the current environmental needs [Molenaar et al., 2009, Scott et al., 2014, Hui et al., 2015]. A prime example is the switch from energy-efficient, high-enzyme-cost respiration to energy-inefficient, but low-enzyme-cost fermentation during overflow metabolism [Basan et al., 2016, Sanchez et al., 2017]. The protein economy has been studied experimentally and with dedicated metabolic models in heterotrophic microorganisms like *E. coli* [Scott et al., 2014, O'Brien et al., 2016] and *S. cerevisiae* [Metzl-Raz et al, 2017, Sanchez et al., 2017]. More recently, resource allocation was studied in photoautotrophic bacteria (*Synechocystis* sp.) [Jahn et al., 2018, Zavrel et al., 2019]. There, a large investment in the $CO_2$-fixation (2-7% protein mass is Rubisco) and photosynthesis machinery (20-40% protein mass are antennae and photosystems) may reduce proteome space for ribosomes, resulting in lower growth rates than heterotrophs.

Previous studies of *C. necator* grown in different trophic conditions have shown that gene expression is regulated in a condition-dependent manner [Schwartz et al., 2009, Kohlmann et al., 2011, Kohlmann et al., 2014]. For example, CBB cycle genes are strongly expressed during autotrophic growth but were also upregulated on fructose [Shimizu et al., 2015], prompting the question of whether such expression is of benefit or evolutionary advantage. To date, protein allocation and utilization has not been investigated. We wondered, if and how *C. necator* would reallocate protein resources when confronted with different types and strengths of substrate limitation. We also wondered to what extent a versatile soil bacterium would express unutilized or underutilized proteins. And finally, how protein allocation and enzyme utilization would scale with growth rate. To address these questions, we designed a multivariate set of growth experiments. *C. necator* was cultivated in bioreactors at steady state conditions using four limiting substrates and five different growth rates. We quantified the cellular proteome using LC-MS/MS and trained a genome-scale resource allocation model with our data [Bulovic et al., 2019, Goelzer et al., 2015]. We found that *C. necator* allocates its resources in response to the imposed environmental challenges, but invests more than 40% of its protein mass in genes that are either unlikely to be utilized or have no known function. Enzyme utilization in the central carbon metabolism was markedly different between pathways, with enzymes in the proximity

of substrate assimilation (upper glycolysis, CBB cycle) showing higher variability, higher absolute abundance, and higher utilization than enzymes involved in supply of biomass precursors (tricarboxylic acid cycle (TCA), pyruvate metabolism). $CO_2$-assimilation enzymes were also expressed in heterotrophic growth regimes but were found to provide no fitness benefit.

# Results

### *C. necator* expresses most of its annotated genes

In order to access cellular states that were optimally acclimated to a nutrient limitation, we cultivated *C. necator* in chemostat bioreactors. We selected four limiting growth substrates as interesting entry points to metabolism (Figure 1 A). Fructose was chosen as it is the only known sugar (apart from sugar alcohols) that *C. necator* utilizes [Orita et al, 2012]. It is taken up *via* a specific ABC transporter and metabolized in the ED pathway. Succinate was chosen as an entry point to the TCA cycle. Formate was chosen because *C. necator* has the special ability to utilize it as both energy and carbon source. Formatotrophic growth closely resembles lithoautotrophic growth regarding the utilized enzymes [Cramm, 2008]. Formate ($COOH^-$) is first oxidized by formate dehydrogenases (FDH) to $CO_2$ with simultaneous reduction of $NAD^+$ to NADH. The $CO_2$ is then fixed *via* the CBB cycle. Finally, growth on fructose with limiting ammonium was chosen as we expected a dedicated response to N-limitation by adjustment of gene expression and flux ratios between different pathways. For each limitation, four independent bioreactor cultivations were performed with dilution rate (equalling growth rate μ) increasing step-wise from 0.05 to 0.1, 0.15, 0.2, and 0.25 $h^{-1}$ (Figure S1 A) and subsequent sampling for proteomics. The substrate limitation in chemostats was verified by determining the residual carbon concentration in culture supernatants using HPLC (Figure S1 B). For ammonium limitation, a high concentration of residual fructose was determined, as expected when nitrogen is limiting. All other conditions showed no or very low concentration of residual substrate.

We analyzed the proteome of *C. necator* for all conditions of the chemostat cultivations (four substrate limitations, five growth rates, four biological replicates). We employed a label-free quantification strategy with a feature propagation approach, allowing us to significantly increase the coverage of protein quantification [Weisser et al., 2017]. More than 4,000 proteins were quantified in each individual sample (Figure S2 A). Altogether,

5,357 proteins out of 6,614 annotated genes were quantified in at least one condition (81.0%), and 4,260 proteins were quantified with at least two peptides (Figure S2 B). The proteomics data can be accessed through an interactive web application at https://m-jahn.shinyapps.io/ShinyProt. Based on the distribution of protein abundance 99% of the proteome by mass was quantified. An analysis of sample similarity based on expression revealed that low growth rates are more similar to each other, and that growth on formate is most unlike the other conditions (Figure S2 C). Gene expression in terms of proteome mass fraction was unequally distributed over the genome (Figure 1 B): 78.7% of protein mass was encoded by chromosome 1, 16.4% encoded by chromosome 2, and 5.4% by pHG1. Chromosome 2 and pHG1 thus encode predominantly specialized functions, as predicted by *in silico* analyses [Pohlmann et al., 2006, Fricke et al., 2009]. On chromosome 2, highly expressed genes were the *cbb* operon (CBB Cycle, pentose phosphate pathway (PPP), Figure S2 D), glycolysis related genes (*pgi, zwf*), and the methionine synthase *metE*. On pHG1, highly expressed were the second copy of the *cbb* operon as well as *hox*/*hyp* operons (soluble and membrane bound hydrogenases, up to 3% of proteome by mass). The majority of pHG1 encoded protein mass is therefore related to autotrophic growth. Note that the two copies of the *cbb* operon are 99% identical on amino acid sequence level and can not be distinguished well by LC-MS/MS (abundance of ambiguous peptides was allocated to both copies). Promoter activity studies have shown that expression levels from both operons were similar [Gruber et al., 2017]. As we also cultivated *C. necator* on formate, we were interested in the expression of formate dehydrogenase (FDH) genes (Figure S2 E). *C. necator* is equipped with two types of FDH, soluble S-FDH (operons *fds* and *fdw* on chromosome 1 and 2, respectively) and membrane-bound M-FDH (*fdo* and *fdh* operons, the latter present in two copies on chromosome 1 and 2, respectively). In contrast to *cbb* genes, which were expressed under both fructose and formate growth, expression of FDHs was induced only during growth on formate, and the soluble dehydrogenase (*fds*) was the predominant form.

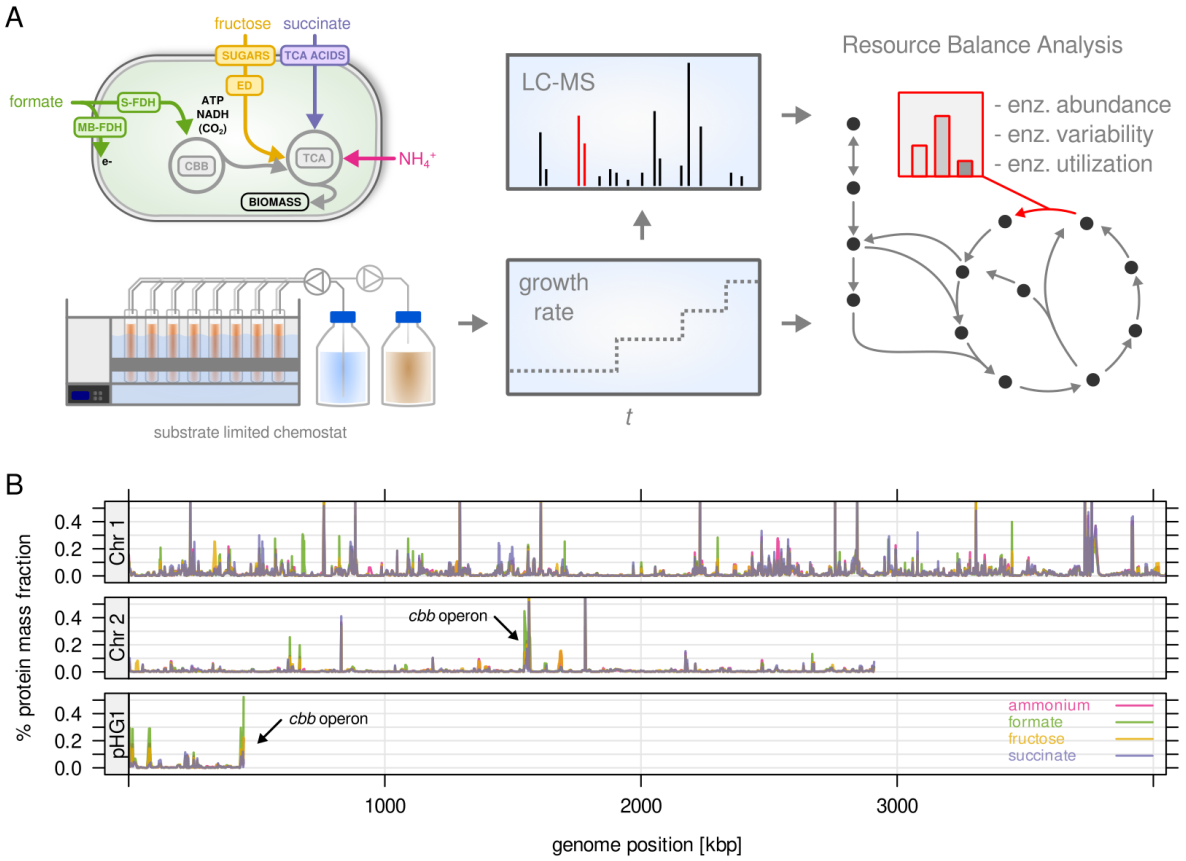


***Figure 1***. C. necator *expresses most of its annotated genes.* ***A)*** *Four different limitations were chosen covering different entry points to central metabolism. Cells were cultivated in chemostat bioreactors and dilution rate (equals growth rate) was stepwise increased from 0.05 to 0.25* $h^{-1}$*. The proteome was analyzed by LC-MS/MS. Enzyme abundance was used to constrain a resource balance analysis (RBA) model, and enzyme utilization was investigated for the different limitations.* ***B)*** *Protein mass fraction (%) of all proteins (5,357) mapped to their respective genes on chromosome 1, 2, and megaplasmid pHG1 (mean of four substrate limitations,* $\mu$*=0.25* $h^{-1}$*). Density is mean protein mass fraction for a sliding window of five genes. The genes of the* cbb *operon (arrows) are the most expressed regions on chromosome 2 and pHG1.*

### A large fraction of the *C. necator* proteome is related to environmental readiness

We next explored how the proteins of *C. necator* are utilized during the different growth modes. We created a resource balance analysis (RBA) model [Bulovic et al, 2019] based on a previous genome-scale metabolic reconstruction of *C. necator* (1,360 reactions) [Park et

al., 2011]. The RBA model predicts optimal flux distributions as in flux balance analysis (FBA), but also takes kinetic parameters and enzyme abundance into account (Methods). DNA replication, transcription, translation, and protein folding were included as lumped reactions (macromolecular machines) with protein subunit composition and rate estimates taken from literature (Methods, Table S1). Each enzyme or macromolecular machine imparts a protein cost, with the total protein pool being limited. RBA models can predict trade-offs between high- and low-enzyme cost pathways, increase of ribosome abundance with growth rate, and upper boundaries on growth in substrate-replete conditions [Goelzer et al. 2015, Sanchez et al., 2017, Salvy et al., 2020]. The *C. necator* RBA model was constrained using a set of parameters obtained from proteomics data, the UniProt database, and literature (Methods, Figure S3, Table S1). A critical parameter for RBA is the enzyme efficiency $k_{app}$ of each reaction, which links the reaction rate to the abundance of its catalyzing enzyme. These were obtained by estimating the metabolic flux boundaries per reaction (using flux sampling), and then dividing maximal flux by unit enzyme allocated to the reaction [Goelzer et al., 2015, Davidi & Milo, 2017, Bulovic et al., 2019].

We used the constrained resource allocation model to analyze the non-utilized and the under-utilized fraction of the *C. necator* proteome. The non-utilized proteome fraction consists of enzymes that do not carry flux in any of the tested conditions. To quantify this fraction, we performed a series of RBA model simulations corresponding to the experimental conditions of the chemostats. The model predicted optimal flux distribution and enzyme abundance to maximize growth rate for each of the four different substrate limitations. The model was generally able to reproduce experimentally determined protein allocation using fitted (optimal) $k_{app}$ values (Figure S4 A). However, these simulations may predict one out of many possible solutions to the protein allocation problem. In order to estimate the total number of usable reactions independent from the optimal set of $k_{app}$, we performed 200 simulations per substrate limitation where $k_{app}$ was randomly sampled from the $k_{app}$ distribution. This converged to maximally 550 utilized reactions per condition. (Figure S4 B). In total, 587 of 1,360 reactions were utilized at least once in all simulations, 280 reactions were used in all simulations on all substrates (core reactions), and 28 reactions were used in only one particular limitation. We mapped the *C. necator* proteome quantification data onto RBA model reactions to categorize proteins as: 1) not included in the model, 2) included but non-utilized enzymes, 3) utilized enzymes, and 4) utilized machinery (Figure 2 A). The

non-modeled proteome fraction comprised on average 38% of the proteome mass (0.26 g/gDCW, 4,041 proteins), and was slightly dependent on condition. Non-utilized enzymes were low-abundant in mass (0.03 g/gDCW, 400 proteins) compared to the utilized enzyme fraction (0.27 g/gDCW, 823 proteins). Macromolecular machinery averaged 0.12 g/gDCW for 93 annotated proteins. Non-utilized enzymes were not enriched in a particular functional category, while the non-modeled protein fraction was enriched in functions for transport, transcription (factors), and post-translational modification (Figure 2 B). A large group of proteins has no annotated function. Taking non-modeled and non-utilized proteins together, 43% of the *C. necator* proteome (by mass) is unlikely to be utilized in the tested conditions, or involved in processes not covered by the RBA model. We also estimated the protein mass encoded by essential genes per utilization category (Figure 2 A, shaded area). Gene essentiality was determined by sequencing a randomly barcoded transposon library with 60,000 mutants after growth on rich medium (RB-TnSeq workflow) [Rubin et al, 2015, Wetmore et al., 2015]. Transposon insertion density of a gene was used to sort it into one of three different categories, 'essential' (496 genes), 'probably essential' (149), or 'non-essential' (4,712). On average, 47% of utilized enzymes (by mass) were encoded by essential genes, while only 19% and 3% of the non-modeled and non-utilized protein mass, respectively, was essential. We conclude that a large portion of the *C. necator* proteome is associated with nutrient scavenging and regulatory adaptation to new environments.

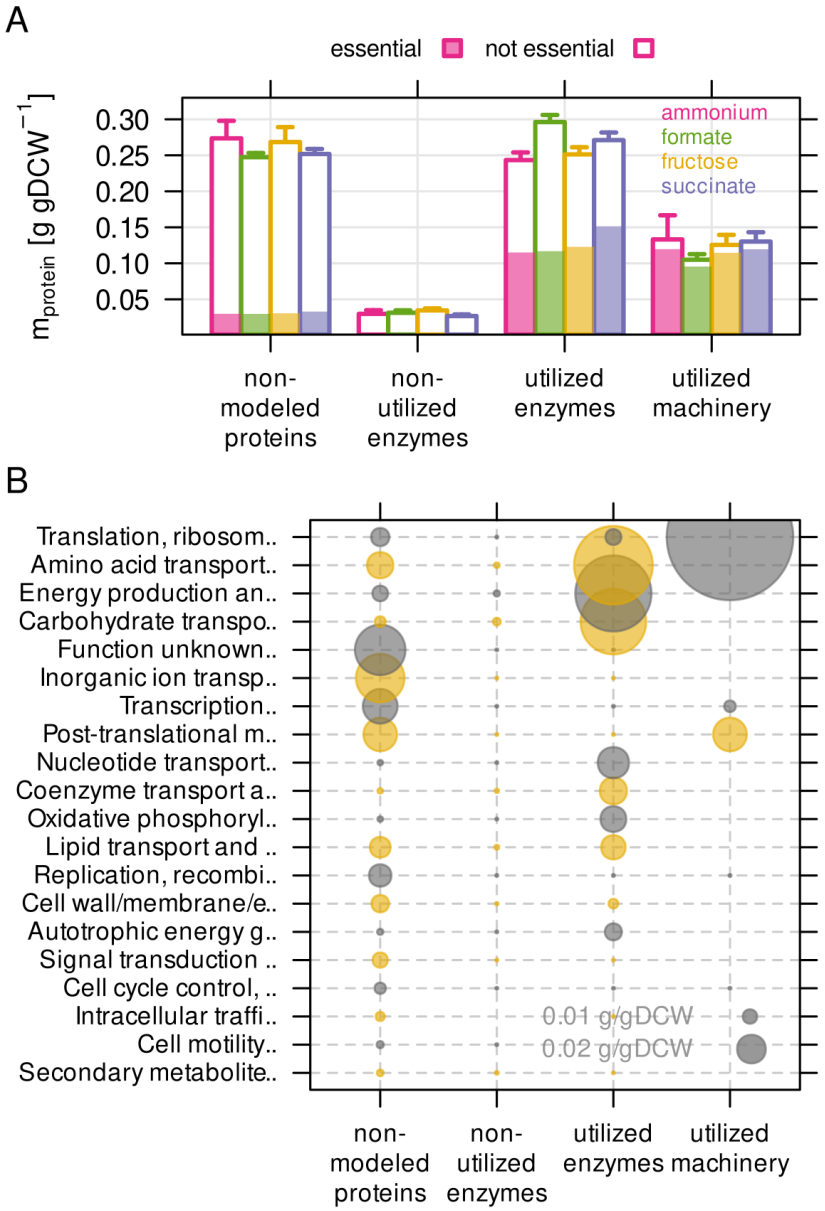


***Figure 2.*** *The non-modeled and non-utilized proteome of* C. necator *is related to environmental readiness.* ***A)*** *A series of model simulations was conducted with randomly sampled enzyme efficiency* $k_{app}$ *(n=200) to obtain the maximum number of potentially utilized reactions in each growth condition. The* C. necator *proteome (5,357 proteins) was allocated to each of four utilization categories and protein mass summed up per category. Protein mass encoded by essential genes is indicated as shaded area in bars. Bars represent mean of four biological replicates, whiskers represent standard deviation.* ***B)*** *Average protein mass by utilization category and functional group. Alternating color (grey and yellow) for bubbles*

*are used in alternating rows.*

**Highly utilized enzymes are more abundant, less variable, and often essential**

The *under*-utilized proteome fraction is a subset of the utilized fraction. Generally, metabolic flux through a reaction can be correlated to the associated enzyme abundance. The rate of a reaction $v_R$ is then the product of the enzyme efficiency $k_{app}$ and the concentration of the enzyme that catalyzes the reaction ($v_R = k_{app} \cdot [E]$) [Davidi & Milo, 2017]. Under steady-state conditions, optimal gene expression would adjust enzyme abundance proportional to the flux that it is supposed to carry (metabolic demand), keeping utilization of the enzyme constant. If enzyme abundance and flux do not change proportionally between different conditions or growth rates, utilization has changed. To estimate the degree of utilization, we compared experimental protein allocation to model predictions at different growth rates. The RBA model predicts the minimal required enzyme abundance to drive a metabolic reaction, assuming full substrate saturation of the enzyme. Although full saturation of all enzymes is not realistic [Reznik et al., 2017, Janasch et al., 2018], it is a useful assumption to determine enzyme utilization. Utilization $U_E$ is calculated by dividing the predicted minimal enzyme abundance by the experimentally determined enzyme abundance [Davidi & Milo, 2017]:

$$U_E\,[\%] \;=\; [E]_{minimal} \,/\, [E]_{measured} \cdot 100$$

We first looked at utilization of the macromolecular machines (Figure S5). Only two of these, ribosomes and chaperones, had a considerable protein mass allocated to them. The abundance of ribosomal proteins increased linearly with growth rate, as observed in other bacteria [Scott et al, 2014, Peebo et al., 2015, Jahn et al., 2018]. The RBA model simulations accurately predicted expansion of ribosomes with increasing growth rate, but failed to predict incomplete reduction of ribosomes at low growth rate (Figure S5 B). This can be explained by the evolutionary benefit that cells gain from keeping a ribosome reserve for nutrient upshifts [Mori et al., 2017]. The ribosome reserve led to a decrease in utilization at low growth rate regardless of the limiting substrate (Figure S5 C).

Next, we examined metabolic enzyme utilization by comparing experimental and simulated protein abundance. All metabolic reactions/enzymes of the RBA model that had associated proteins quantified by MS were included in the analysis (n=1,012). For each

enzyme, the average utilization in the four limiting conditions ($\mu$=0.25 $h^{-1}$) was determined, and then used to group enzymes into three categories: low (≤33%, n=738), moderate (33-66%, n=134) and high utilization (>66%, n=140). Highly utilized enzymes are therefore predominantly enzymes utilized in several of the four limiting conditions. There were significant differences between these three groups: Highly utilized enzymes were on average more abundant in terms of protein mass (g/gDCW) (Figure 3 A). We also calculated variability in enzyme abundance by determining the coefficient of variation (CV) of allocated protein mass across the four different conditions (Figure 3 B). For example, formate dehydrogenase (FDH) was strongly expressed in only one out of four conditions (growth on formate) and therefore showed high variability (CV=1.25), and low average utilization (23%). Altogether, variability was significantly lower for moderately and highly utilized enzymes. These observations support the notion that *C. necator* optimizes the cost-benefit ratio of gene expression by keeping utilization high for highly abundant enzymes. Similarly, low variation in gene expression of highly-utilized enzymes could provide a fitness benefit in conditions changing on a short time scale. Constitutive expression of such genes can buffer substrate and metabolite surges. Finally, we wondered if utilization of enzymes is also correlated to essentiality of the associated gene(s) as determined by RB-TnSeq from our transposon mutant library. Enzymes were sorted into, 'essential', 'probably essential', or 'non-essential' based on the essentiality of their associated genes (Methods, Figure 3 C). We found that enzymes with intermediate and high utilization were more likely to be encoded by an essential gene compared to lowly utilized enzymes.

A closer inspection of the central carbon metabolism of *C. necator* revealed that enzyme abundance, variability, and utilization was markedly different between major pathways. The enzymes of lower glycolysis (PGM, ENO, PYK, PDH) showed moderate to high utilization but low variability (Figure 3 D), clearly distinct from the enzymes in upper glycolysis and the CBB cycle. The average abundance of lower glycolysis enzymes was also lower. This trend continued with reactions down-stream of glycolysis/gluconeogenesis, such as the reactions of pyruvate metabolism and the TCA cycle. In contrast, reactions of the ED pathway and particularly CBB cycle were marked by much higher variability and moderate to high abundance. This supports the conclusion that gene expression regulation in *C. necator* is hierarchically organized: Enzymes close to the entry point of substrates into central metabolism are expressed 'on demand', and therefore show high variability, high absolute

abundance, and high utilization, at least in some growth regimes. Enzymes downstream of substrate assimilation show lower expression and variability, perhaps owing to their universal role in providing biomass precursors (TCA, pyruvate metabolism). A lower protein investment per catalytic activity allows for larger reserves of these enzymes. The low utilization of TCA and pyruvate metabolism may provide a benefit for robustness, by avoiding full saturation.

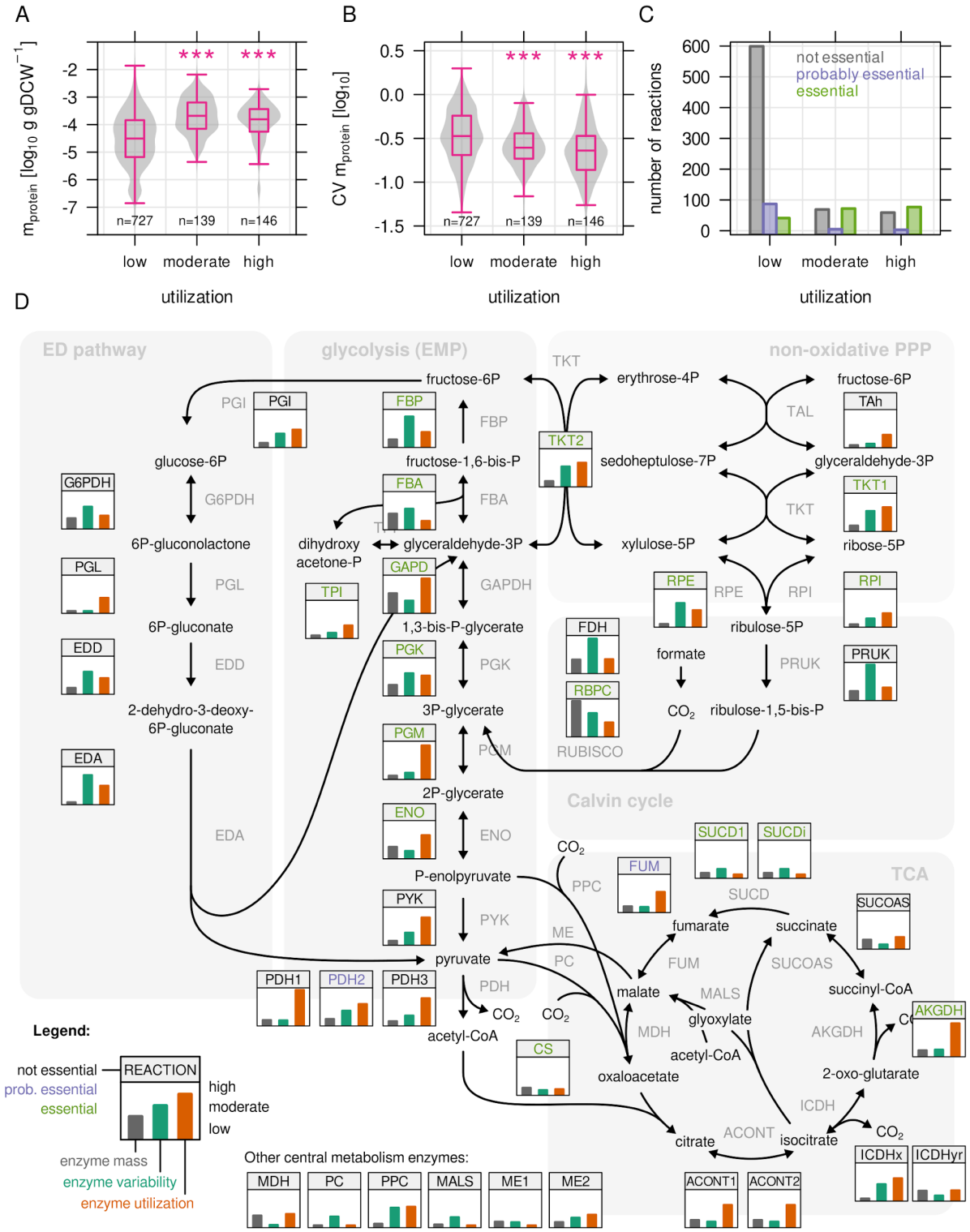


***Figure 3.*** *Highly utilized enzymes are more abundant, less variable, and often essential.* ***A)*** *Protein mass in g/gDCW allocated to enzymes with low, moderate, and high utilization. Enzymes with moderate and high utilization are significantly more abundant ($p=2.4x10^{-31}$ and*

*$7.5x10^{-18}$, respectively. Student's t-test, two-sided).* ***B)*** *Coefficient of variation (CV) as a measure of variability in enzyme abundance. Enzymes with moderate and high utilization have significantly lower variability ($p=2.0x10^{-6}$ and $2.1x10^{-9}$, respectively. Student's t-test, two-sided).* ***C)*** *Number of reactions associated with at least 1 essential gene, or at least 1 probably essential gene, or no essential gene at all, broken down by utilization.* ***D)*** *Map of* C. necator*'s central carbon metabolism. Inset figures show mean enzyme abundance, variability, and mean utilization over the four limiting conditions ($\mu=0.25\ h^{-1}$, four biological replicates). Values were rescaled from the respective minimum and maximum to a range of 0 to 1. Enzyme abbreviations are colored according to essentiality as described in C).*

**Autotrophy-related enzymes are largely underutilized**

The high average abundance and variability of the CBB cycle enzymes is particularly interesting. While phosphoribulokinase (PRUK) and Rubisco (RBPC) are specific for the purpose of $CO_2$-fixation, the other enzymes overlap with sugar phosphate metabolism (glycolysis/gluconeogenesis, pentose phosphate pathway) providing precursors that are essential for growth. We wondered if the expression of these enzymes is optimally regulated based on the metabolic demands of the four different substrate limitations. We compared the predicted (optimal) abundance with the experimentally measured abundance for important enzymes of the CBB cycle (Figure 5 A). On formate, the protein concentration of these enzymes increased with growth rate and therefore estimated flux, correlating with RBA model predictions. A positive correlation was also found for fructose-limited growth, but a negative correlation for succinate and ammonium limitation. Rubisco was highly abundant even during growth on fructose where the model did not predict flux through the CBB cycle (up to 0.02 g/gDCW or 3% of the proteome by mass). With the exception of Rubisco and PRUK, the CBB cycle enzymes are encoded by three different copies on the *C. necator* genome. Two of these are arranged in the *cbb* operons on chromosome 2 and pHG1, while the respective third copy on chromosome 1 is the evolutionarily most ancestral [Pohlmann et al., 2006, Fricke et al., 2009]. Expression of the ancestral enzymes is regulated differently than the *cbb* operons, with lower average protein abundance that is independent of substrate and growth rate (Figure S6).

When estimating the utilization of *cbb* enzymes, we found that utilization was high for formate due to the obligatory flux through the CBB cycle, but low for other conditions (Figure 5 B). It was not zero for some reactions that are required to drive lower glycolysis for

catabolism of fructose (PGK, GAPDH), or the non-oxidative PPP for the purpose of nucleotide synthesis (transketolase reactions TKT1/2). We conclude that *C. necator* keeps large amounts of underutilized CBB enzymes (0.024 to 0.04 g/gDCW, or 3.5 to 5.9% of the proteome depending on substrate) whose abundance is not warranted by the expected fluxes from glycolysis/gluconeogenesis or nucleotide biosynthesis. The underutilized enzyme mass may be in preparation for autotrophic or formatotrophic growth, even when such substrates are not in reach.

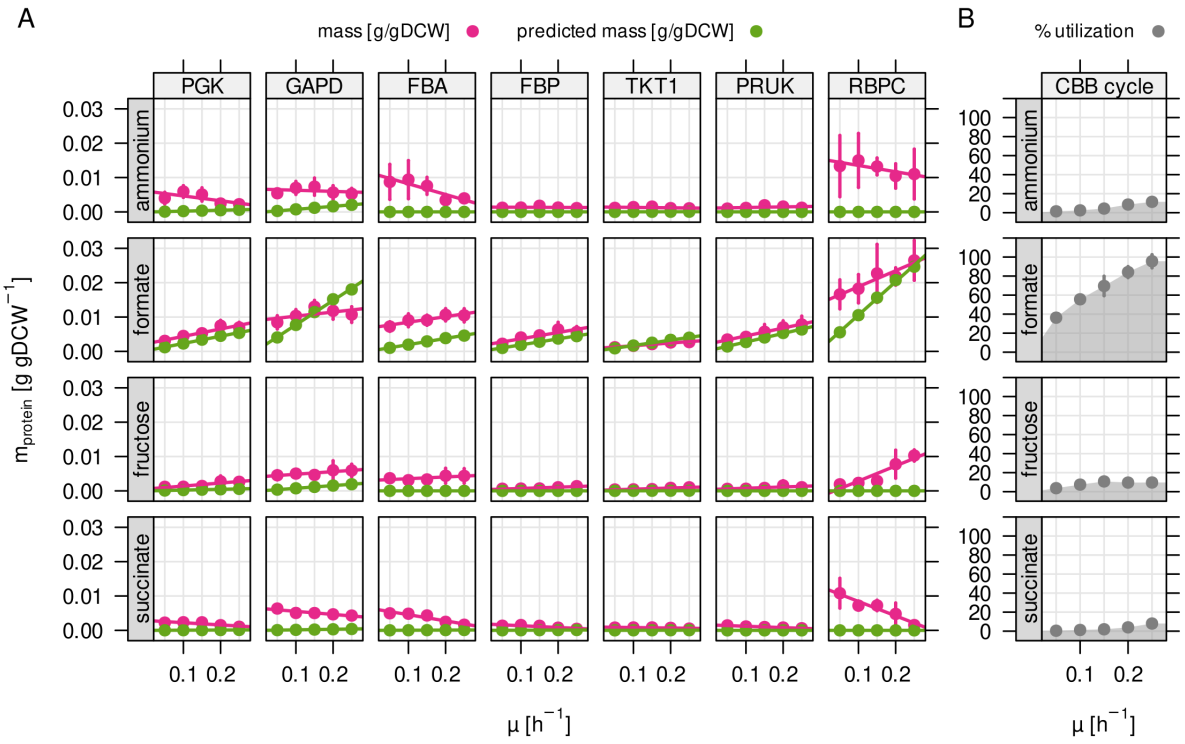


***Figure 4.*** *Autotrophy-related enzymes are largely underutilized.* ***A)*** *Experimentally determined and model-predicted protein concentration for the seven most abundant enzymes of the CBB cycle (points and error bars represent mean and standard deviation of four biological replicates, respectively). PGK, phosphoglycerate kinase; GAPD, glyceraldehyde-3-phosphate dehydrogenase; FBA, fructose bisphosphate aldolase; FBP, fructose bisphosphatase; TKT1, transketolase; PRUK, phosphoribulokinase; RBPC, ribulose bisphosphate carboxylase.* ***B)*** *Total utilization of the enzymes in A). Utilization was calculated as the sum of predicted (optimal) enzyme abundance divided by the sum of experimentally measured abundance.*

**Reassimilation of $CO_2$ does not provide a fitness benefit for *C. necator***

*C. necator* appears to keep large amounts of Rubisco (and other CBB cycle enzymes) under-utilized in heterotrophic conditions. This assumption was based on RBA model simulations where flux through the CBB cycle was absent on any substrate other than formate. However, the RBA model only finds optimal solutions to maximize growth while other objectives are possible as well. It was recently reported that *C. necator* fixes $CO_2$ *via* the CBB cycle during heterotrophic growth on fructose [Shimizu *et al.*, 2015]. The authors speculated that reassimilation of $CO_2$ by Rubisco could provide a benefit for carbon yield ($^{13}C$ labeled $CO_2$ was enriched in PHB granules). The additional protein burden of CBB enzymes could come at the cost of lower growth rate, representing a yield-growth rate trade-off. To test if reassimilation of emitted $CO_2$ improves carbon yield, we performed RBA model simulations with fructose as primary energy and carbon source and forced flux through Rubisco (Figure 5 A). We simulated five different $CO_2$ fixation rates (0 to 5 mmol $gDCW^{-1}$ $h^{-1}$) at a fructose uptake rate of 3.2 mmol $gDCW^{-1}$ $h^{-1}$. However, neither biomass yield nor growth rate was improved in any of the simulations (Figure 5 B, C). The metabolic flux was diverted from the ED pathway towards the non-oxidative PPP in order to provide ribulose-5-phosphate precursors for $CO_2$ fixation (Figure 5 D). Simultaneously, the high energy requirement for $CO_2$ fixation led to higher flux through the TCA cycle in order to generate additional NADH and ATP. Respiration and $O_2$ consumption was also predicted to increase, while no net reduction of $CO_2$ emission was found. Simulations suggested instead that the cells emit more $CO_2$ when $CO_2$ fixation is enforced, an apparent paradox caused by the lack of additional energy.

Finally, we tested experimentally if expression of CBB genes conveys a fitness benefit during growth on different carbon sources. To this end, the barcoded transposon library (pool of 60,000 mutants) was cultivated in fructose-, succinate-, and formate-limited chemostat bioreactors (dilution rate of 0.1 $h^{-1}$). The continuous feed fixes the growth rate and selects cells with higher substrate affinity or biomass yield [Wides & Milo, 2018]. The composition of the mutant pool was probed after 8 and 16 generations of growth using next generation sequencing. The fitness contribution of each gene was estimated by the degree of enrichment or depletion of mutants over time [Wetmore et al., 2015]. Surprisingly, we found that fitness of *cbb* mutants was largely unchanged, even during growth on formate where the activity of the CBB cycle is strictly essential for growth (Figure 5 E). The only

notable exception was *cbbR*, the transcriptional regulator of the *cbb* operon. Compared to other *cbb* genes encoding important enzymes of the CBB cycle, *cbbR* mutants had a strong fitness penalty on formate (Figure 5 F, fitness ≤ -6). Knockout of *cbbR* leads to a 100 fold down-regulation of *cbb* gene expression [Shimizu et al., 2015]. *CbbP,* encoding phosphoribulokinase, and *cbbS/L* encoding the two subunits of Rubisco showed no significant change in fitness on any substrate. These results show that knockout of a *cbb* gene can be fully compensated by the second copy of the *cbb* operon without compromising growth. The (functional) *cbbR* regulator is only present in one copy of the *cbb* operon (chromosome 2), which explains its essentiality. *CbbR* mutants had no significant fitness penalty on fructose or succinate, suggesting that the activity of the CBB cycle is either neutral to growth or the effect is too small to detect with our method. We reproduced these experiments with a cultivation regime that primarily selects for faster growth rate (medium pulses every 2 h) and obtained the same results (Figure S7 A, B). We conclude that (re-) fixation of $CO_2$ during heterotrophic growth is unlikely to convey a fitness benefit without additional energy (e.g. from $H_2$ oxidation). We hypothesize that the up-regulation of Rubisco on fructose is a 'byproduct' of up-regulation of other glycolysis related genes of the *cbb* operon.

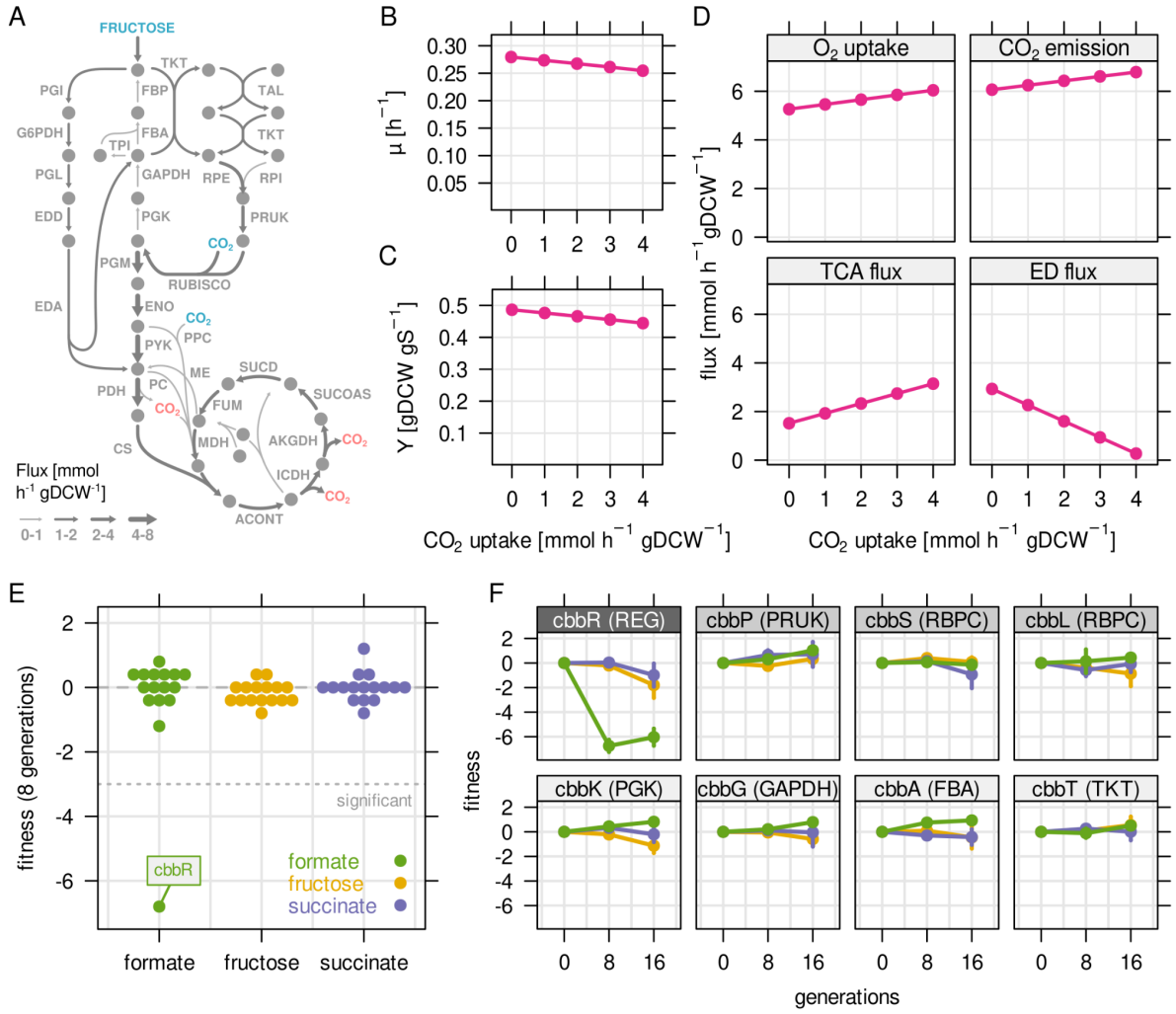


***Figure 5**. Reassimilation of $CO_2$ is unlikely to provide a fitness benefit for* C. necator*. RBA model simulations were performed for a fixed fructose uptake rate combined with five different $CO_2$ fixation rates. **A)** Example metabolic flux map for a fructose uptake rate of 3.2 mmol $gDCW^{-1}$ $h^{-1}$ and $CO_2$ fixation rate of 3 mmol $gDCW^{-1}$ $h^{-1}$. Blue - uptake of fructose and $CO_2$, red - emission of $CO_2$. **B)** Predicted growth rate μ. **C)** Biomass yield Y in gDCW g $fructose^{-1}$. **D)** Net flux through selected reactions for the same simulations as in B) and C). For the TCA cycle, flux through citrate synthase was used as a proxy. For the Entner-Doudoroff (ED) pathway, flux through 6-phosphogluconolactonase (EDD) was used as a proxy. **E)** Fitness for all* cbb *genes determined by growth competition of a barcoded transposon knockout library on three different substrates. **F)** Fitness over time for selected* cbb *genes of the pHG1 encoded operon, except* cbbR *which is located on chromosome 2. Chromosome 2 encoded* cbb *genes are not shown due to low transposon insertion frequency. Points and error bars represent mean and standard deviation of four biological replicates, respectively. Grayscale labels indicate role in CBB pathway: dark gray - transcriptional regulator, moderate gray - specific for $CO_2$ fixation, light gray - overlapping*

*role in glycolysis/CBB cycle.*

**The central metabolism of *C. necator* is highly redundant**

We have previously established that many enzymes in the central carbon metabolism of *C. necator* are encoded by strictly essential genes (Figure 3 D). However, most reactions are annotated with more than one (iso-) enzyme. We therefore expanded our gene fitness analysis to all enzymes of central carbon metabolism in order to find conditionally essential genes. The reactions of central carbon metabolism were grouped into four different pathways, CBB cycle including FDH, ED pathway, pyruvate metabolism and TCA cycle (Figure 6 A), and the fitness of all genes associated with these reactions was quantified. The majority of genes showed no significant fitness penalty (or benefit) when knocked out. Only a few genes showed a significant decrease in fitness, and the effect on fitness was substrate-specific. Four genes encoding subunits of a soluble FDH (*fdsABDG*) showed significantly reduced fitness on formate. This demonstrates that *fds* does not only encode the most expressed FDH (Figure S2), but most likely also the one that carries the majority of the flux from formate to $CO_2$ (Figure 6 B). No other annotated FDH genes had a similar fitness penalty (Table S2). Another conditionally essential gene on formate was *ppc,* encoding the PEP-carboxylase (PPC). The reaction has no other annotated (iso-) enzymes and was predicted by RBA to carry most flux on formate compared to the other carbon sources (Figure 6 B). On fructose, genes for four consecutive enzyme reactions had significantly reduced fitness, *pgl*, *edd1* and *eda* from the ED pathway, as well as *pdhA* encoding the E1 component of pyruvate dehydrogenase (Figure 6 C, Table S2). For EDD another isoenzyme is annotated (*edd2*) that could not compensate for the *edd1* knockout. For PDH, five alternative loci are annotated, all of which did not rescue *pdhA* knockout (Table S2). On succinate, only two gene knockouts have significantly reduced fitness, malic enzyme *maeA* and *pdhA*. Both associated reactions (ME and PDH) carry significant flux on succinate according to RBA simulations (Figure 6 D). Malic enzyme has one more annotated gene, *maeB*, with different cofactor specificity (NADPH instead of NADH), which could not compensate for the loss of *maeA* (Table S2). We conclude that the central carbon metabolism of *C. necator* has a very high degree of redundancy. Apart from a core set of essential genes encoded on chromosome 1, many enzyme functions can be compensated by alternative copies. The genes that were found to be conditionally essential were either

present with only one copy (*pgl*, *eda*, *ppc*), or the alternative enzymes could not compensate for their loss (*edd2*, *maeB, pdhA2,* alternative FDHs). The degree of essentiality for these genes was correlated to the flux carried by the enzyme (Figure 6 B, C, D).

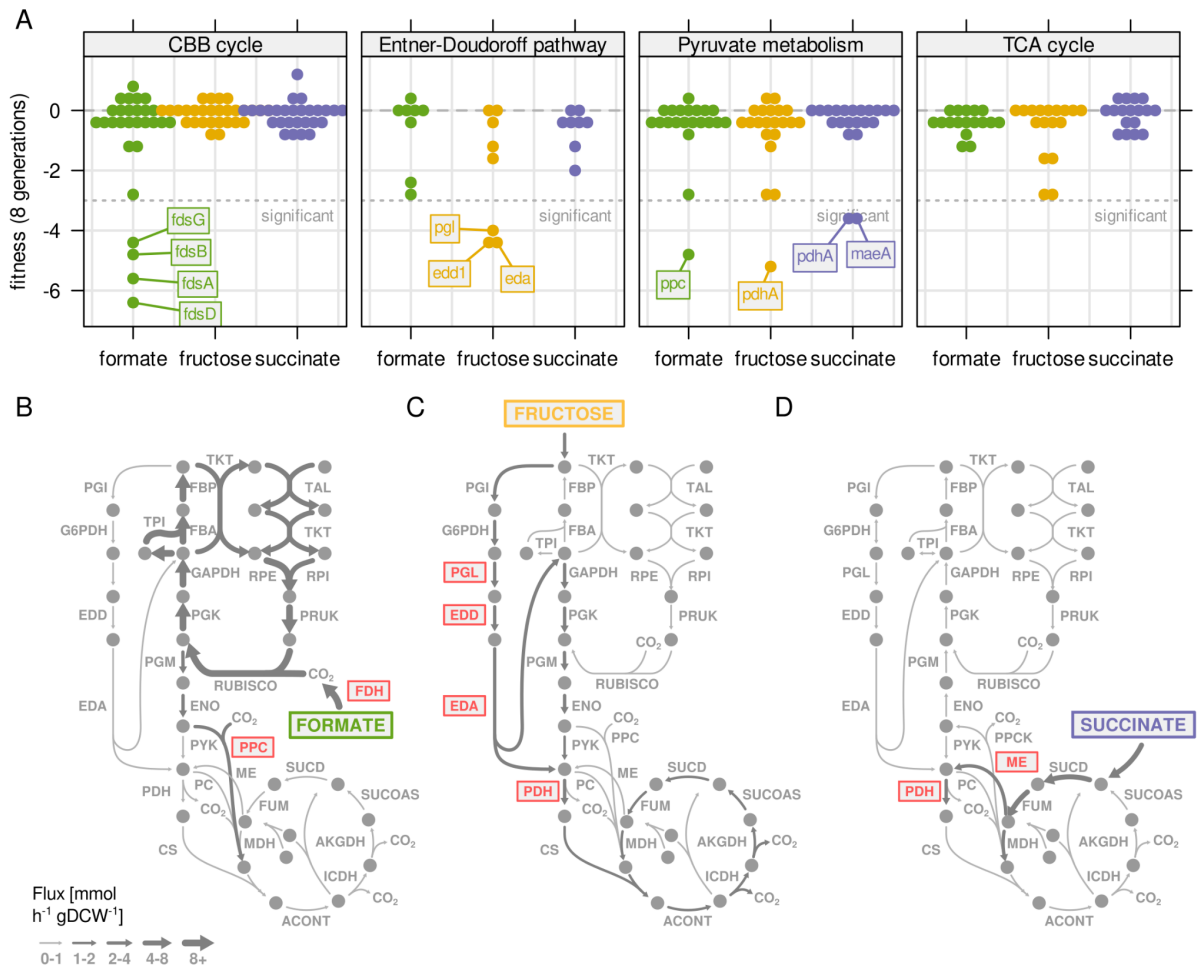


***Figure 6**. Central metabolism enzymes are highly redundant in C. necator. **A)** Fitness for all central carbon metabolism genes associated with the reactions in Figure 3 D. Fitness was determined by growth competition of a barcoded transposon knockout library on three different substrates. Genes are broken down by pathway. Dotted line - fitness ≤ -3 was regarded as significant. A summary of all reactions with significantly changed fitness is available in Table S2. **B)** Metabolic flux map for growth on formate. RBA simulation with formate uptake rate of 62 mmol gDCW$^{-1}$ h$^{-1}$. Red - reaction where annotated genes show significantly reduced fitness in growth competition from A). **C)** Same as B) for fructose with uptake rate of 3.0 mmol gDCW$^{-1}$ h$^{-1}$. **D)** Same as B) for succinate with uptake rate of 4.7 mmol gDCW$^{-1}$ h$^{-1}$.*

## Discussion

A characteristic feature of all *Burkholderiales*, to which *C. necator* belongs, is their fragmented genome organisation (2-4 replicons) [Fricke et al., 2009]. Comparative genome analysis suggested different evolutionary origins of the *C. necator* chromosomes, with chromosome 1 being more conserved among related species than chromosome 2 and pHG1 [Fricke et al., 2009]. We found that the largest fraction of protein mass (78.7%) can be attributed to chromosome 1, while chromosome 2 and the pHG1 megaplasmid only show strong expression at a few selected loci responsible for alternative lifestyles (lithoautotrophy, denitrification). Chromosome 1 also showed predominantly constitutive expression across different trophic conditions, while the few highly expressed loci on chromosome 2 and pHG1 were transcriptionally regulated. This supports the hypothesis that *C. necator* may have acquired chromosome 2 and pHG1 at a later stage of its evolutionary history and highlights the 'accessory' character of both replicons [Fricke et al., 2009].

Of the 5,357 quantified proteins only 1,223 are associated with enzymes and another 93 with central dogma machinery in the *C. necator* RBA model. Yet, utilized enzymes and machinery summed up to 57% of the protein mass, while 43% of the proteome was non-utilized, including all proteins not covered by the RBA model. Our estimate for the non-utilized protein mass in *C. necator* is higher than a previously reported estimate for *E. coli* of 26-39%, particularly regarding the non-modeled protein fraction (39% in *C. necator* compared to maximally 26% in *E. coli*) [O'Brien et al., 2016]. Another estimate for the proportion of non-utilized enzymes for *E. coli* obtained about 30% of the proteome [Davidi & Milo, 2017]. We conclude that *C. necator* not only has a larger genome compared to e.g. *E. coli*, but also expresses many genes without utilizing them in the controlled, homogeneous environments that are typical in biotechnology applications. The large non-utilized protein fraction is related to environmental readiness and likely increases fitness of *C. necator* in the variable and mixed substrate conditions typical of soil [Hewavitharana et al., 2019].

It is important to note that estimation of protein utilization is not straight-forward and prone to several sources of error. For example, many proteins in *C. necator* are not functionally annotated but could be catalytically active, eventually leading to underestimation of the utilized protein fraction. On the other hand, enzymes can have 'moonlighting' activities

so that the calculated utilization is underestimated for some enzymes and overestimated for others [Cotton et al., 2020]. Proteins involved in cell motility, cell cycling, sensing of and responding to environmental changes are generally not a part of the metabolic model, yet have vital functions for cellular fitness and are thus utilized in some way. Another challenge is to accurately assign enzyme abundance to reactions that have several annotated proteins, or a protein that is assigned to several enzymatic reactions. In these cases we divided protein abundance between different enzymes and *vice versa*.

Bearing these limitations in mind, we used the RBA model to investigate the *under*utilization of enzymes. Underutilization as used in this study serves as a proxy for the relation between maximum attainable reaction rate ($V_{max}$) and actual reaction rate, with the latter being shaped by substrate saturation, reverse flux as well as potential allosteric effectors. The estimated enzyme efficiency $k_{app}$ is influenced by these factors and can deviate from *in vitro* measured maximum turnover $k_{cat}$ [Davidi et al., 2016]. A general observation regarding utilization is the dependency on growth rate. Flux of metabolic enzymes is directly proportional to growth rate, given that all other cultivation parameters are kept constant. At low growth and low flux through metabolism, bacteria optimize fitness by reallocating protein resources from growth functions (ribosomes) to substrate assimilation (transporters) [Scott et al., 2014, Hui et al., 2015, Jahn et al., 2018]. However, this reallocation is only a gradual response and neither results in full reduction of superfluous proteome sectors, nor the shrinking of the protein pool (g protein/gDCW). The consequence is that enzyme utilization becomes low at low growth rates (O'Brien et al., 2016). *C. necator* also shows this pattern: ribosomal proteins are incompletely reduced at low growth rates, and enzymes of central metabolism generally remain highly abundant (Figure 4, Figure S5), effectively creating an underutilized enzyme reserve.

Underutilization of enzymes represents an 'efficiency sacrifice' for host fitness. Expression of excess non-metabolic proteins such as LacZ or YFP reduces bacterial growth rate [Hui et al., 2015, Jahn et al., 2018]. However, several recent experimental studies have shown that enzyme underutilization in *E. coli* central metabolism, such as in the OPP pathway and amino acid biosynthesis, provides a buffer against perturbations in environmental conditions or gene expression [Davidi & Milo, 2017, Christodoulou 2018, Sander et al., 2019]. The importance of underutilized enzymes for metabolic stability has also been shown for metabolic networks such as the CBB cycle [Barenholz et al., 2017,

Janasch et al., 2018]. We observed that highly abundant enzymes are better utilized and less variable across conditions. This is most likely a result of the evolutionary pressure on enzyme reserve costs, which increase proportionally with the abundance of enzymes.

It is of interest to compare enzyme utilization in *C. necator* to *E. coli,* a model bacterium with a different environmental niche. The central carbon metabolism pathways of *C. necator* showed differences in enzyme abundance, variability, and utilization. Abundance of enzymes for the upper EMP pathway, PPP, and CBB cycle was on average higher than for the enzymes of the ED pathway, pyruvate metabolism or TCA. This is similar to *E. coli*, where higher abundance of glycolysis enzymes was explained by high flux demand and low thermodynamic driving force [Noor et al., 2016]. But enzymes of the upper EMP pathway and PPP also showed strong transcriptional regulation (variability in gene expression), which is a marked difference to *E. coli*, where enzyme levels show low variation across multiple growth conditions [Schmidt et al., 2016], and flux is mainly regulated through allosteric interactions [Reznik et al., 2017]. Of all central carbon metabolism, the TCA cycle enzymes showed on average lowest abundance, variability and utilization. This is similar to *E. coli*, where the sole flux capacity demand suggested lower enzyme abundance than what was measured experimentally [Noor et al., 2016]. Only when reverse flux (reactions with low thermodynamic driving force) and low enzyme saturation, estimated from metabolite levels, was taken into account, was the calculated enzyme demand similar to the measured levels [Noor et al., 2016].

How was the regulatory network in *C. necator*'s central carbon metabolism shaped by its native environment? *E. coli* is adapted to regular nutrient upshifts every 2-3 hours [Mori et al, 2017]. It therefore evolved allosteric regulation to deal with quickly changing fluxes through the EMP pathway, its prime catabolic route [Reznik et al., 2017]. For *C. necator*, hexose sugars are only one out of many possible substrate classes and the flux through the (upper) EMP pathway is lower as it uses the ED pathway to catabolize sugars. A slow but more resource efficient transcriptional regulation of glycolysis could therefore provide a fitness benefit for an environment with limited and irregular substrate supply. Interestingly, only the glycolysis/PPP enzymes located on the phylogenetically young *cbb* operons are transcriptionally regulated, while the ancestral enzymes on chromosome 1 are constitutively expressed (Figure S6). These enzymes are also scattered over the chromosome and therefore not collectively regulated. The diverging regulation for glycolysis-related genes

could mark a branching point in the evolutionary history of *C. necator*. The pHG1 plasmid was likely acquired recently, based on its transmissibility and proven ability to confer hydrogenotrophic metabolism [Friedrich et al., 1981]. *Cbb* genes could either get lost or take over the function as main glycolysis enzymes from their chromosome 1 orthologs.

The two copies of the *cbb* operon in *C. necator* are of hybrid nature as CBB cycle enzymes functionally overlap with EMP glycolysis and PPP. Expression of the *cbb* operon depended on the supplied substrate and was highest for growth on formate, where CBB cycle genes are essential. However, a more complex picture emerged for *cbb* expression during other substrate limitations (increasing with μ on fructose, decreasing with μ on succinate). The *cbb* operon is transcriptionally regulated by two systems, CbbR [Bowien & Kusian, 2002] and RegA/B [Gruber et al., 2017]. RegA/B guarantees a basic level of constitutive expression, while CbbR senses the intracellular PEP concentration [Gruber et al., 2017]. PEP is an important allosteric regulator responsible for the switch between glycolytic and gluconeogenic flux in *E. coli* [Reznik et al., 2017]. In *C. necator*, growth on fructose leads to low PEP concentration, triggering *cbb* expression, while it is the other way around for succinate. This prompts the question which evolutionary benefit cells gain from *cbb* expression during heterotrophic growth? It has been proposed that reassimilation of the emitted $CO_2$ by Rubisco improves carbon yield [Shimizu et al., 2015]. Model simulations however suggested that $CO_2$-reassimilation is unlikely to provide a benefit as long as there is no additional energy source (Rubisco activity even causes a higher net $CO_2$ emission). Probing the fitness contribution of the *cbb* genes using a barcoded transposon mutant library generally confirmed this picture. Knockout of *cbbR*, which leads to a down-regulation of the entire *cbb* operon, caused a strong fitness penalty for mutants growing on formate. The same mutants showed no significant fitness change on fructose or succinate, suggesting that $CO_2$ fixation in these conditions provides no benefit. The conserved PEP-dependent transcriptional regulation of *cbb* leads to a collateral expression of Rubisco in conditions where it is not required, such as fructose. This is a remarkable example of suboptimality, where one benefit could be readiness for lithoautotrophic growth when hydrogen or formate become available. Probing the effect of *cbb* gene knockouts with the transposon library also revealed that *C. necator* can compensate the loss of any *cbb* gene by expressing the respective second copy. This finding applies to central carbon metabolism in general. Almost all enzyme functions are covered by several gene loci, so that knockout did not result in

fitness loss. Notable exceptions are the reactions of the ED pathway, PEP carboxylase (*ppc*), and malic enzyme, (*maeA*) that showed significantly reduced fitness in conditions where these reactions carry high flux.

Our results highlight the metabolic flexibility of *C. necator* and its robustness to changing environmental conditions. Its high degree of genomic redundancy makes it tolerant to gene loss, but may also lead to regulatory conflicts exemplified by *cbb* expression. A recent comparison of microbial genomes showed that the CBB cycle is accompanied by a metabolism-wide range of adaptations [Asplund-Samuelsson and Hudson, 2021]. Considering a possibly recent acquisition of the CBB cycle *via* pHG1, it is likely that *C. necator* is currently evolving to make best use of the *cbb* genes. Therefore, in addition to a readiness imperative, the ongoing adaptation process may help explain the observed suboptimal use of the CBB cycle.

## Supplemental Information

Supplemental Information includes seven figures and three tables and can be found with this article online.

## Acknowledgments

We acknowledge Julia Foyer and Arvid Gynnå for assistance in bioreactor cultivations. We like to thank Anne Goelzer for guidance with resource balance analysis modeling. This study was financially supported by the Swedish Research Council Vetenskapsrådet (Grant number 2016-06160), the Swedish Research Council Formas (Grant number 2015-939 and 2019-01491), and Novo Nordisk Fonden (Grant number NNF20OC0061469).

## Author Contributions

M. Jahn conceived the study, performed cultivation experiments, analyzed mass spectrometry and sequencing data, constructed the transposon library, performed metabolic modeling and wrote the manuscript. N.C. constructed the transposon library and performed next generation sequencing. M. Janasch performed metabolic modeling and wrote the

manuscript. A.H. and B.F. performed mass spectrometry measurements. K.K. constructed the transposon library. A.M. performed cultivations and physiological measurements. Q.C. and J.A.S. implemented fitness analysis for the transposon library. E.P.H. conceived the study and wrote the manuscript. All authors read and approved the manuscript.

## Declaration of Interests

The authors declare no competing interests.

## References


[1] Alagesan S, Minton NP, and Malys N. $^{13}$C-assisted metabolic flux analysis to investigate heterotrophic and mixotrophic metabolism in *Cupriavidus necator* H16. Metabolomics, 14, 1–10, **2018**.

[2] Asplund-Samuelsson J and Hudson EP. Wide range of metabolic adaptations to the acquisition of the Calvin cycle revealed by comparison of microbial genomes. PLOS Computational Biology, 17, e1008742, **2021**.

[3] Barenholz U, Davidi D, Reznik E, Bar-On Y, Antonovsky N, et al. Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points. eLife, 6, 1-32, **2017**.

[4] Basan M, Hui S, Okano H, Zhang Z, Shen Y, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature, 528, 99–104, **2015**.

[5] Bowien B and Kusian B. Genetics and control of $CO_2$ assimilation in the chemoautotroph *Ralstonia eutropha*. Archives of Microbiology, 178, 85–93, **2002**.

[6] Brigham C. Perspectives for the biotechnological production of biofuels from $CO_2$ and $H_2$ using *Ralstonia eutropha* and other 'knallgas' bacteria. Applied Microbiology and Biotechnology, 103, 2113–2120, **2019**.

7] Bulović A, Fischer S, Dinh M, Golib F, Liebermeister W, et al. Automated generation of bacterial resource allocation models. Metabolic Engineering, 55, 12–22, **2019**.

[8] Cotton CA, Bernhardsgrütter I, He H, Burgener S, Schulz L, et al. Underground isoleucine biosynthesis pathways in *E. coli*. eLife, 9, 1–25, **2020**.

[9] Christodoulou D, Link H, Fuhrer T, Kochanowski K, Gerosa L, et al. Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli*'s rapid response to oxidative stress. Cell Systems, 6, 569–578.e7, **2018**.

[10] Cramm R. Genomic view of energy metabolism in *Ralstonia eutropha* H16. Journal of

Molecular Microbiology and Biotechnology, 16, 38–52, **2008**.

[11] Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, et al. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to in vitro $k_{cat}$ measurements. Proceedings of the National Academy of Sciences of the United States of America, 113, 3401-3406, **2016**.

[12] Davidi D and Milo R. Lessons on enzyme kinetics from quantitative proteomics. Current Opinion in Biotechnology, 46, 81–89, **2017**.

[13] Ebrahim A, Lerman JA, Palsson BO, and Hyduke DR. COBRApy: Constraints-based reconstruction and analysis for Python. BMC Systems Biology, 7, 74, **2013**.

[14] Epshtein V and Nudler E. Cooperation between RNA polymerase molecules in transcription elongation. Science, 300, 801–805, **2003**.

[15] Fricke WF, Kusian B, and Bowien B. The genome organization of *Ralstonia eutropha* strain H16 and related species of the *Burkholderiaceae*. J Mol Microbiol Biotechnol, 16, 124–135, **2009**.

[16] Friedrich B, Hogrefe C, and Schlegel HG. Naturally occurring genetic transfer of hydrogen-oxidizing ability between strains of *Alcaligenes eutrophus*. Journal of Bacteriology, 147, 198–205, **1981**.

[17] Goelzer A, Muntel J, Chubukov V, Jules M, Prestel E, et al. Quantitative prediction of genome-wide resource allocation in bacteria. Metabolic Engineering, 32, 232–243, **2015**.

[18] Granholm V, Kim S, Navarro JC, Sjölund E, Smith RD, et al. Fast and accurate database searches with MS-GF+percolator. Journal of Proteome Research, 13, 890–897, **2014**.

[19] Gruber S, Schwab H, and Heidinger P. CbbR and RegA regulate *cbb* operon transcription in *Ralstonia eutropha* H16. Journal of Biotechnology, 257, 78–86, **2017**.

[20] Hewavitharana SS, Klarer E, Reed AJ, Leisso R, Poirier B, et al. Temporal dynamics of the soil metabolome and microbiome during simulated anaerobic soil disinfestation. Frontiers in Microbiology, 10, 2365, **2019**.

[21] Hui S, Silverman JM, Chen SS, Erickson DW, Basan M, et al. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. Molecular Systems Biology, 11, 784, **2015**.

[22] Jahn M, Vialas V, Karlsen J, Maddalo G, Edfors F, et al. Growth of cyanobacteria is constrained by the abundance of light and carbon assimilation proteins. Cell Reports, 25, 478–486.e8, **2018**.

[23] Janasch M, Asplund-Samuelsson J, Steuer R, and Hudson EP. Kinetic modeling of the Calvin cycle identifies flux control and stable metabolomes in *Synechocystis* carbon fixation. Journal of Experimental Botany, 70, 1017–1031, **2018**.

[24] Kohlmann Y, Pohlmann A, Otto A, Becher D, Cramm R, et al. Analyses of soluble and membrane proteomes of *Ralstonia eutropha* H16 reveal major changes in the protein complement in adaptation to lithoautotrophy. Journal of Proteome Research, 10, 2767–2776, **2011**.

[25] Kohlmann Y, Pohlmann A, Schwartz E, Zühlke D, Otto A, et al. Coping with anoxia: A comprehensive proteomic and transcriptomic survey of denitrification. Journal of Proteome Research, 13, 4325–4338, **2014**.

[26] Metzl-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y, et al. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. eLife, 6, 1–21, **2017**.

[27] Molenaar D, Berlo RV, Ridder DD, and Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. Molecular Systems Biology, 5, 1–10, **2009**.

[28] Mori M, Schink S, Erickson DW, Gerland U, and Hwa T. Quantifying the benefit of a proteome reserve in fluctuating environments. Nature Communications, 8, 1–8, **2017**.

[29] Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R, et al. The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. PLoS Computational Biology, 12, **2016**.

[30] O'Brien EJ, Utrilla J, and Palsson BO. Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. PLoS Computational Biology, 12, 1–22, **2016**.

[31] Orita I, Iwazawa R, Nakamura S, and Fukui T. Identification of mutation points in *Cupriavidus necator* NCIMB 11599 and genetic reconstitution of glucose utilization ability in wild strain H16 for polyhydroxyalkanoate production. Journal of Bioscience and Bioengineering, 113, 63–69, **2012**.

[32] Park J, Kim T, and Lee S. Genome-scale reconstruction and in silico analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production. BMC Systems Biology, 5, 101, **2011**.

[33] Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K, et al. Proteome reallocation in *Escherichia coli* with increasing specific growth rate. Molecular BioSystems, 11, 1184–1193, **2015**.

[34] Pohlmann A, Fricke WF, Reinecke F, Kusian B, Liesegang H, et al. Genome sequence of the bioplastic-producing "knallgas" bacterium *Ralstonia eutropha* H16. Nature

Biotechnology, 24, 1257–1262, **2006**.

[35] Reznik E, Christodoulou D, Goldford JE, Briars E, Sauer U, et al. Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity. Cell Reports, 20, 2666–2677, **2017**.

[36] Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, et al. The essential gene set of a photosynthetic organism. Proceedings of the National Academy of Sciences of the United States of America, 112, e6634–e6643, **2015**.

[37] Röst HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, et al. OpenMS: A flexible open-source software platform for mass spectrometry data analysis. Nature Methods, 13, 741–748, **2016**.

[38] Salvy P and Hatzimanikatis V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. Nature Communications, 11, **2020**.

[39] Sánchez BJ, Zhang C, Nilsson A, Lahtvee P, Kerkhoven EJ, et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Molecular Systems Biology, 13, 935, **2017**.

[40] Sander T, Farke N, Diehl C, Kuntz M, Glatter T, et al. Allosteric feedback inhibition enables robust amino acid biosynthesis in *E. coli* by enforcing enzyme overabundance. Cell Systems, 8, 66–75, **2019**.

[41] Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, et al. The quantitative and condition-dependent *Escherichia coli* proteome. Nature Biotechnology, 34, 104–110, **2016**.

[42] Schwartz E, Voigt B, Zühlke D, Pohlmann A, Lenz O, et al. A proteomic view of the facultatively chemolithoautotrophic lifestyle of *Ralstonia eutropha* H16. Proteomics, 9, 5132–5142, **2009**.

[43] Scott M, Klumpp S, Mateescu EM, and Hwa T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Molecular Systems Biology, 10, 747, **2014**.

[44] Shimizu R, Dempo Y, Nakayama Y, Nakamura S, Bamba T, et al. New insight into the role of the Calvin cycle: Reutilization of $CO_2$ emitted through sugar degradation. Scientific Reports, 5, 11617, **2015**.

[45] Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Research, 44, D447–D456, **2016**.

[46] Weisser H and Choudhary JS. Targeted feature detection for data-dependent shotgun proteomics. Journal of Proteome Research, 16, 2964–2974, **2017**.

[47] Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly barcoded transposons. mBio, 6, 1–15, **2015**.

[48] Wides A and Milo R. Understanding the dynamics and optimizing the performance of chemostat selection experiments. ArXiv.org, **2018**.

[49] Yao L, Shabestary K, Björk S, Asplund-Samuelsson J, Joensson H, et al. Pooled CRISPRi screening of the cyanobacterium *Synechocystis sp*. PCC 6803 for enhanced industrial phenotypes. Nature Communications, 11, 1–13, **2020**.

[50] Yishai O, Lindner SN, de la Cruz JG, Tenenboim H, and Bar-Even A. The formate bio-economy. Current Opinion in Chemical Biology, 35, 1–9, **2016**.

[51] Zavřel T, Faizi M, Loureiro C, Poschmann G, Stühler K, et al. Quantitative insights into the cyanobacterial cell economy. eLife, 8, 446179, **2019**.

# Methods

## Contact for reagent and resource sharing

Further information and requests for reagents may be directed to and will be fulfilled by Lead Contact Elton P. Hudson (paul.hudson@scilifelab.se).

## Method details

### Strains and cultivation

*Cupriavidus necator* H16 was obtained from the German Collection of Microorganisms and Cell Cultures, strain number DSM-428. Cells were cultivated on complete (LB) medium, or minimal medium depending on experimental setup. Minimal medium was composed of 0.78 g/L $NaH_2PO_4$, 4.18 g/L $Na_2HPO_4x2H_2O$, 1 g/L $NH_4Cl$, 0.1 g/L $K_2SO_4$, 0.1 g/L $MgCl_2x6H_2O$, 1.6 mg/L $FeCl_3x6H_2O$, 0.4 mg/L $CaCl_2$, 0.05 mg/L $CoCl_2x6H_2O$, 1.8 mg/L $Na_2MoO_4x2H_2O$, 0.13 g/L $Ni_2SO_4x6H_2O$, 0.07 mg/L $CuCl_2x2H_2O$. Depending on the experiment, 0.5 g/L D-fructose, 0.5 g/L succinate, or 1.5 g/L pH-neutralized formic acid was added as carbon source. For nitrogen limitation, the concentration of D-fructose was increased to 2 g/L and concentration of $NH_4Cl$ was reduced to 0.025 g/L. All components were added to autoclaved sodium phosphate buffer from filter-sterilized stock solutions. Batch cultures were grown in

100 mL shake flasks at 30°C and 180 RPM. Precultures of the barcoded *C. necator* transposon library were supplemented with 200 µg/mL kanamycin and 50 µg/mL gentamicin to suppress growth of untransformed *C. necator* recipient or *E. coli* donor cells.

**Chemostat bioreactors**

*C. necator* H16 (wild type) or the *C. necator* H16 transposon mutant library was cultivated in an 8-tube MC-1000-OD bioreactor (Photon System Instruments, Drasov, CZ). The system was customized to perform chemostat cultivation as described previously [Jahn et al., 2018, Yao et al., 2020]. Bioreactors (65 mL) were filled with minimal medium supplemented with the respective carbon and nitrogen source, and inoculated with an overnight preculture to a target $OD_{720nm}$ of 0.05. Bioreactors were bubbled with air at a rate of 12.5 mL/min and a temperature of 30°C. The $OD_{720nm}$ and $OD_{680nm}$ were measured every 15 min. Fresh medium was continuously added using Reglo ICC precision peristaltic pumps (Ismatec, GER). For pulsed chemostat experiments, a volume corresponding to the continuous addition of medium over a given time period was added in a single pulse every 2 h. For proteomics, 40 mL samples were taken after five retention times of continuous growth at a fixed dilution rate ($t_R$ = 1/D; for example $t_R$(D = 0.1 $h^{-1}$) = 1 / 0.1 = 10 h). For transposon library competition experiments, 15 mL samples were taken after 0, 8 and 16 generations of growth (population average). Cells were harvested by centrifugation for 10 min at 5,000 xg, 4°C, washed with 1 mL ice-cold PBS, transferred to a 1.5 mL tube, and centrifuged again for 2 min at 8,000 xg, 4°C. The supernatant was discarded and the pellet frozen at -20°C.

**Determination of biomass yield**

Substrate uptake rate $q_S$ was determined using the dilution rate D, the culture volume V, the biomass concentration $c_{bm}$ in gDCW $L^{-1}$, and the initial and residual substrate concentrations $S_i$ and $S_r$, respectively, in the following equation: $q_S = \frac{V \cdot D \cdot (S_i - S_r)}{c_{bm}}$. The biomass yield $Y_{X/S}$ for all substrates was determined by fitting a linear model to the growth rate-substrate uptake rate relationship. The biomass yield $Y_{X/S}$ for fructose, succinate, formate, and ammonium chloride was 0.45, 0.46, 0.09 and 1.6 $g_{DCW}$ $g_S^{-1}$, respectively.

**Dry cell weight determination**

Dry cell weight (DCW) was determined from $OD_{720nm}$ using a premade calibration curve. A

$log_2$ serial dilution of a *C. necator* H16 overnight culture (minimal medium with 2 g/L fructose, 1 g/L ammonium) was made with a volume of 10 mL per dilution. $OD_{720nm}$ was measured for all dilutions using a photospectrometer (SpectraMax M5, Molecular Devices), and cells were harvested by centrifugation at 5,000 xg for 10 min, washed with 1 mL mqH2O, transferred to preweighed 1.5 mL tubes and dried for 4 h at 70°C. Dried cell mass was measured on a precision scale and a relation of gDCW = 0.518 x $OD_{720nm}$ - 0.01 was determined.

**Residual substrate measurement with HPLC**

Culture supernatant was obtained after centrifugation of cell samples. A volume of 1 mL supernatant was transferred to an LC glass vial using Millex-HV PVDF syringe filter tips (Merck Millipore). The HPLC column (Aminex 300-mm HPX-87H) was equilibrated with 5 mM $H_2SO_4$ as mobile phase for 1 h, at a flow rate of 0.5 mL/min. The column was heated to 60°C. A volume of 20 µL per sample was injected to the HPLC followed by a run time of 30 min. UV-absorption was constantly detected at 210 nm wavelength. Standards with four different concentrations, 10, 50, 100 and 200 mg/L, were used for quantification of each residual substrate (succinate, formate, fructose, ammonium chloride). Calibration curves were obtained by fitting a linear equation to the concentration-absorbance relationship. Residual substrate concentration was then determined by solving the equation with the obtained absorbance measurements.

**Statistical analysis**

Bioreactor cultivations, LC-MS/MS measurement for proteomics, and library competition experiments ('BarSeq') were performed with four independent biological replicates. HPLC measurement of supernatants was performed with three biological replicates. Here, biological replicate means that samples were obtained from independently replicated bioreactor cultivations inoculated with the same preculture. The sample size of four was chosen based on the known variance from previous proteomics experiments. If not otherwise indicated in figure legends, points and error bars represent the mean and standard deviation. No removal of outliers was performed. All analyses of proteomics, modeling, and fitness data are documented in R notebooks available at https://github.com/m-jahn/R-notebook-ralstonia-proteome.

**Sample preparation for LC-MS/MS**

Frozen cell pellets were resuspended in 125 µL solubilization buffer (200 mM TEAB, 8 M Urea, protease inhibitor). 100 µL glass beads (100 µm diameter) were added to the cell suspension and cells were lysed by bead beating in a Qiagen TissueLyzer II (5 min, f = 30/s, precooled cassettes). Cell debris was removed by centrifugation at 14,000 xg, 30 min, 4°C, and supernatant was transferred to a new tube. Protein concentration was determined using the Bradford assay (Bio-Rad). For reduction and alkylation of proteins, 2.5 µL 200 mM DTT (5 mM final) and 5 µL 200 mM CAA (10 mM final) were added, respectively, and samples incubated for 60 min at RT in the dark. Samples were diluted 8x with 700 µL 200 µM TEAB. For digestion, Lys-C was added in a ratio of 1:75 w/w to protein concentration, and samples were incubated at 37°C and 600 RPM for 12 h. Trypsin was added (1:75 w/w) and samples incubated for 24 h at the same conditions. Samples were acidified with 100 µL 10% formic acid (FA) and insoluble compounds were removed by centrifugation (14,000 xg, 15 min, RT). Peptide samples were then cleaned up using a solid phase extraction (SPE) protocol in 96-well plate format (Tecan Resolvex A200) according to the manufacturer's recommendations. Briefly, the 96-well plate with SPE inserts was equilibrated with 200 µL acetonitrile (ACN) and 2x200 µL 0.6% acetic acid. A lysate volume corresponding to 40 µg protein was loaded on the plate and washed twice with 200 µL 0.6% acetic acid. Peptides were eluted from the column in 100 µL elution buffer (0.6% acetic acid, 80% ACN) and dried in a speedvac for 2 h, 37°C. Dried peptides were frozen at -80°C and dissolved in 10% FA to a final concentration of 1 µg/µL before MS measurement.

**LC-MS/MS analysis of lysates**

Lysates were analyzed using a Thermo Fisher Q Exactive HF mass spectrometer (MS) coupled to a Dionex UltiMate 3000 UHPLC system (Thermo Fisher). The UHPLC was equipped with a trap column (Acclaim PepMap 100, 75 µm x 2 cm, C18, P/N 164535, Thermo Fisher Scientific) and a 50 cm analytical column (Acclaim PepMap 100, 75 µm x 50 cm, C18, P/N ES803, Thermo Fisher Scientific). The injection volume was 2 µL out of 18 µL in which the samples were dissolved in the autosampler. Chromatography was performed using solvent A (3% ACN, 0.1% FA) and solvent B (95% ACN, 0.1% FA) as the mobile phases. The peptides were eluted from the UHPLC system over 90 min at a flow rate of 250 nL/min with the following mobile phase gradient: 2% solvent B for 4 min, 2 to 4% solvent B

for 1min, 4 to 45% solvent B for 90 min, 45 to 99% solvent B for 3 min, 99% solvent B for 10 min and 99% to 2% solvent B for 1 minute following re-equilibration of the column at 2% solvent B for 6 min. The MS was operated in a data-dependent acquisition mode with a Top 8 method. The MS was configured to perform a survey scan from 300 to 2,000 m/z with resolution of 120,000, AGC target of $1x10^6$, maximum IT of 250 ms and 8 subsequent MS/MS scans at 30,000 resolution with an isolation window of 2.0 m/z, AGC target of $2x10^5$, maximum IT of 150 ms and dynamic exclusion set to 20 s.

**Protein identification and quantification**

Thermo raw spectra files were converted to the mzML standard using Proteowizard's MSConvert tool. Peptide identification and label-free quantification were performed using OpenMS 2.4.0 in KNIME [Röst et al., 2016]. The KNIME pipeline for MS data processing was deposited on https://github.com/m-jahn/openMS-workflows (labelfree_MSGFplus_Percolator_FFI.knwf). MS/MS spectra were subjected to sequence database searching using the OpenMS implementation of MS-GF+ and Percolator [Granholm et al., 2014] with the *Cupriavidus necator* H16 reference proteome as database (NCBI assembly GCA_000009285.2, downloaded 07 January 2019). Carbamidomethylation was considered as a fixed modification on cysteine and oxidation as a variable modification on methionine. The precursor ion mass window tolerance was set to 10 ppm. The PeptideIndexer module was used to annotate peptide hits with their corresponding target or decoy status, PSMFeatureExtractor was used to annotate additional characteristics to features, PercolatorAdapter was used to estimate the false discovery rate (FDR), and IDFilter was used to keep only peptides with q-values lower than 0.01 (1% FDR). The quantification pipeline is based on the FeatureFinderIdentification workflow allowing feature propagation between different runs [Weisser et al., 2017]. MzML files were retention time corrected using MapRTTransformer, and identifications (idXML files) were combined using the IDMerger module. FeatureFinderIdentification was then used to generate featureXML files based on all identifications combined from different runs. Individual feature maps were combined to a consensus feature map using FeatureLinkerUnlabelledKD, and global intensity was normalized using ConsensusMapNormalizer (by median). Protein quantity was determined by summing up the intensities of all unique peptides per protein. Abundance of ambiguous peptides (peptides mapping to two different proteins) were shared between

proteins.

**Creation of barcoded *C. necator* transposon library**

The transposon library was prepared according to the RB-TnSeq workflow described in Wetmore et al., 2015. Briefly, *C. necator* H16 wild type was conjugated with an *E. coli* APA766 donor strain containing a barcoded transposon library. The strain is auxotrophic for DAP, the L-Lysin precursor 2,6-diamino-pimelate, to allow for counter selection. Overnight cultures of *E. coli* APA766 and *C. necator* H16 in 10 mL LB medium in shake flasks were prepared. The APA766 culture was supplemented with 0.4 mM DAP and 50 µg/mL kanamycin. 2 L of LB medium (APA766 with 0.4 mM DAP and 50 µg/mL kanamycin) in shake flasks was each inoculated with the respective pre-cultures and incubated overnight at 30°C and 180 RPM. Cells were harvested during exponential growth phase by centrifugation for 10 min, 5000 xg, RT. Supernatant was discarded, cell pellets were resuspended in residual liquid, transferred to 2 mL tubes, washed twice with 2 mL PBS, and finally resuspended in a total amount of 500 µL PBS. Cell suspensions from both strains were combined and plated on 25 cm x 25 cm large trays (Q-tray, Molecular Devices) with LB agar supplemented with 0.4 mM DAP. For conjugation, plates were incubated overnight at 30°C. Cells were then harvested from mating plates by rinsing with 200 µL PBS. The cell suspension was plated on selection plates with LB agar supplemented with 100 µg/mL kanamycin, without DAP. After colonies of sufficient size appeared, transformants were harvested by scraping all cell mass from the plate and collecting the pooled scrapings in 1.5 mL tubes. The mutant library diluted tenfold and was then immediately frozen at -80°C. For competition experiments, a 1 mL 10-fold diluted aliquot (pool of all conjugations, ~1 M CFU) was used to inoculate pre-cultures.

**Mapping of transposon mutants (TnSeq)**

A 1 mL aliquot of the diluted pooled library scrapings was used to inoculate a 50 mL LB culture (with 200 µg/mL kanamycin) and grown overnight at 30°C, 200 RPM. DNA was extracted from 1 mL of this outgrown culture using a GeneJet Genomic DNA Purification Kit (ThermoScientific) and the concentration of genomic DNA was quantified using a Qubit dsDNA HS Assay Kit (Invitrogen). A 1 µg aliquot of genomic DNA was suspended in 15 µL of 10 mM Tris buffer, placed in a microTUBE-15 AFA Beads tube (Covaris) and fragmented

into 300 bp fragments using an ME220 focused ultrasonicator (Covaris) with waveguide 500526 installed. Cycle time was increased to 60 seconds, all other settings were taken from manufacturer's recommendation for generating 350 bp fragments. Fragment end repair and adaptor ligation was performed using an NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) following the manufacturer's protocol. Size selection of NEB adaptor ligated fragments was carried out using SPRISelect magnetic beads (Beckman Coulter) following the method in the NEBNext Ultra II DNA Library Prep Kit User manual. To selectively enrich transposon-containing sequences, a 30 cycle PCR amplification was performed using the Biotin_Short_pHIMAR and NC102 primers (Table S3) using Q5 mastermix (New England Biolabs). Cycle conditions were 30 seconds 98°C followed by 30 cycles (15 seconds 98°C, 75 seconds 72°C) and a 5 minutes 72°C final extension. The biotinylated transposon-containing sequences were purified using MyOne Streptavidin T1 Dynabeads (Invitrogen) according to the manufacturer's instructions. The transposon containing DNA was then stripped from the beads by resuspending the beads in 25 µL of MilliQ water followed by incubation at 70°C for 10 minutes. The beads were separated by incubation on a magnetic stand at room temperature for 1 minute and the supernatant was recovered. Adaptors for Illumina sequencing were added via PCR amplification using Nspacer_barseq_pHIMAR (Wetmore et al., 2015) and NEBNext Index 3 Primer for Illumina (New England Biolabs). Cycle conditions were 30 seconds 98°C followed by 4 cycles (15 seconds 98°C, 75 seconds 72°C) and a 5 minutes 72°C final extension. PCR products were separated on a 1% agarose gel and gel extraction was performed on the band between 300-600 bp using a Gel Extraction Kit (ThermoScientific). The DNA concentration of the samples were quantified using a Qubit dsDNA HS Assay Kit (Invitrogen) and diluted to 2 nM. The 2 nM library was diluted, denatured and sequenced using a NextSeq 500/550 Mid Output Kit v2.5 150 Cycles, (Illumina) run on a NextSeq 550 instrument (Illumina) according to the manufacturer's instructions. Library loading concentration was 1.8 pM with a 10% phiX spike. Reads containing barcodes and genomic DNA fragments were mapped to the *C. necator* genome following the protocol from Wetmore et al., 2015. Briefly, the scripts *MapTnSeq.pl* and *DesignRandomPool.pl* from https://bitbucket.org/berkeleylab/feba/src/master/ were adapted to map reads to the reference genome, and to summarize read counts per barcode, respectively. Only barcodes mapping to the same region with at least two reads were included. The automatic pipeline

for TnSeq data analysis is available at https://github.com/m-jahn/TnSeq-pipe.

**Gene essentiality analysis**

TnSeq data from two different iterations of the transposon library were combined to obtain high insertion frequency per gene (72,443 and 57,040 mutants, respectively). Of the 129,483 transposon insertions, 23,339 mapped to intergenic regions and were excluded from essentiality analysis. Of all insertions mapping to a gene, 78.7% were localized within the central 80% of the ORF and were considered as true knockouts. Following the method from Rubin et al., 2015, a metric for essentiality was calculated, the insertion index *II*; *II* is the number of transposon insertions *n* of a gene *i* with length *k* divided by insertions per region *r* (average of 100 genes around the target position):

$$II_i = (n_i / k_i) / (n_r / k_r)$$

The *II* is bimodally distributed, one set of genes is hit by transposons at an average rate while other genes are hit with lower frequency. To determine an *II* threshold for essentiality, two gamma distributions were fitted to the assumed populations of 1) essential and 2) non-essential genes. For all possible *II*, the probability of falling into the essential and non-essential distribution was determined and a five-fold difference defined as lower and upper thresholds to count a gene as essential or non-essential, respectively. Genes with *II* between the two thresholds were flagged as ambiguous (*p* denotes the probability density function of *II* for essential and non-essential genes):

$$II_{ambiguous} = II\,[(p_{ess} < 5 \cdot p_{non-ess}) \cup (p_{non-ess} < 5 \cdot p_{ess})]$$

To estimate essentiality of enzymes/reactions instead of genes, each enzyme with at least one associated gene being essential was counted as essential, and each enzyme associated with at least one probably essential gene was counted as probably essential; all other enzymes were marked as non-essential.

**Gene fitness analysis (BarSeq)**

Frozen cell pellets from the pulsed and continuous competition experiments were

resuspended in 100 µL of 10 mM Tris and genomic DNA was extracted from 10 µL of the resuspension using a GeneJet Genomic DNA Purification Kit (ThermoScientific). Amplification of the barcodes from genomic DNA was conducted using one of the custom forward indexing primers (BarSeq_F_i7_001 - BarSeq_F_i7_036, Table S3) and the reverse phasing primer pool (BarSeq_R_P2_UMI_Univ - BarSeq_R_P2_UMI_Univ_N5). For each sample 9 µL of genomic DNA extract (≥10 ng/µL) was combined with 3 µL of a forward indexing primer (100 nM), 3 µL of the reverse phasing primer pool (100 nM) and 15 µL of Q5 Mastermix (New England Biolabs). Cycle conditions were 4 minutes at 98°C followed by 20x (30 seconds at 98°C, 30 seconds at 68°C and 30 seconds at 72°C) with a final extension of 5 minutes at 72°C. Concentrations of each sample was quantified using a Qubit dsDNA HS Assay Kit (Invitrogen). Samples were then pooled with 40 ng from up to 36 different samples being combined and run on a 1% agarose gel. Gel extraction was performed on the thick band centered around 200 bp and the concentration of the purified pooled library was quantified again via Qubit assay and diluted down to 2 nM. The 2 nM library was then diluted, denatured and sequenced using a NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (Illumina) run on a NextSeq 550 instrument (Illumina) according to the manufacturer's instructions. Library loading concentration was 1.8 pM with a 1% phiX spike. Gene fitness was calculated from read counts per barcoded mutant based on the method from Wetmore et al., 2015. Briefly, scripts from https://bitbucket.org/berkeleylab/feba/src/master/ were adapted to trim and filter reads, extract barcodes, and summarize read counts per barcode. Fitness score calculation based on the $\log_2$ fold change of read count per barcode over time was implemented as an R script. The automatic pipeline for BarSeq analysis is available at https://github.com/Asplund-Samuelsson/rebar. Altogether, fitness for 5,441 genes was quantified with an average of 6.4 insertion mutants per gene. The remaining 1,173 genes were either essential (no viable insertion mutant), probably essential (number of transposon mutants in the surrounding region too low to determine essentiality), or fitness could not be quantified with sufficient confidence (low read count). A significance threshold of $|F| \geq 3$ after at least 8 generations was chosen based on the bulk fitness distribution of mutants ($-2 \leq F \leq 2$).

**Resource Balance Analysis model**

The resource balance analysis (RBA) model for *C. necator* H16 was generated using the RBApy package [Bulovic et al., 2019]. The model and a detailed description of its generation is available at https://github.com/m-jahn/Bacterial-RBA-models/. The main input was the curated genome scale model for *C. necator* in SBML format (1,360 reactions, excluding exchange reactions), available at https://github.com/m-jahn/genome-scale-models. Amino acid sequence, subunit stoichiometry and cofactor requirements for all proteins associated with model reactions were automatically retrieved from uniprot (organism ID: 381666). Fasta files detailing the composition of the ribosome (3 rRNA and 68 proteins), chaperones (8 proteins), DNA polymerase III (8 proteins), and RNA polymerase II (9 proteins) were added manually. Rates for these macromolecular 'machines' were adopted from published values for *E. coli* (Table S1). Rates for ribosome and chaperone were taken from [Bulovic et al., 2019], rate of RNA polymerase was taken from [Epshtein et al., 2003], and rate of DNA polymerase was the average of several published values obtained from https://bionumbers.hms.harvard.edu (IDs 102052, 104938, 109251, 111770). Biomass composition of *C. necator* H16, growth- and non-growth associated maintenance were all taken from [Park et al., 2011]. A growth rate dependent flux towards PHB was added (3 mmol $gDCW^{-1}$) to obtain biomass yields corresponding to experimentally determined values. The model was calibrated by adding estimates for $k_{app}$, the apparent catalytic rate for each metabolic enzyme, following the procedure in [Bulovic et al., 2019]. For each model reaction and substrate limitation, flux boundaries were obtained from flux sampling analysis (FSA) using COBRApy [Ebrahim et al., 2013], and enzyme abundance in mmol $gDCW^{-1}$ was obtained from proteomics measurements. $k_{app}$ was determined by calculating the maximum flux per unit enzyme over all conditions. For enzymes without estimated $k_{app}$ (no flux, or no protein abundance available), the median of the $k_{app}$ distribution was used (6,296 $h^{-1}$, Figure S3 A). The average protein fraction of cell dry weight was taken from Park et al., 2011. The reported concentration of 0.68 g protein $gDCW^{-1}$ was converted to mmol amino acids $gDCW^{-1}$ by assuming an average molecular weight per amino acid of 110 g $mol^{-1}$:

$$c = \frac{0.68\ g \cdot mol \cdot 1000}{gDCW \cdot 110\ g} = 6.18\ mmol\ gDCW^{-1}$$

Proteome fraction per cellular compartment (cytoplasm, cytoplasmic membrane) was estimated based on proteomics measurements and predicted protein localization (psortb

algorithm) as input. Growth rate dependent fractions for cytoplasmic and membrane proteins were obtained by correlating growth rate and the respective mass fractions and fitting a linear model (Table S1, Figure S3 B). The same procedure was applied to estimate the non-enzymatic protein fraction per compartment. Proteins not contained in the model were categorized as non-enzymatic as they have no catalytic function in the model (Table S1, Figure S3 C).

## Data and software availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaino et al., 2016) partner repository with the dataset identifier PXD024819. Protein quantification results can be browsed and interactively analyzed using the web application available at https://m-jahn.shinyapps.io/ShinyProt. All sequencing data for TnSeq and BarSeq experiments are available at the European Nucleotide Archive with accession number PRJEB43757. The data for competition experiments performed with the transposon mutant library can be browsed and interactively analyzed using the web application available at https://m-jahn.shinyapps.io/ShinyLib/.

The openMS/KNIME workflow for MS data processing is available at https://github.com/m-jahn/openMS-workflows. The revised genome scale model of *C. necator* H16 is available at https://github.com/m-jahn/genome-scale-models. The resource balance analysis (RBA) model of *C. necator* H16 is available at https://github.com/m-jahn/Bacterial-RBA-models. The code used to process TnSeq data from raw fastq files (read trimming, filtering, mapping to genome) is available at https://github.com/m-jahn/TnSeq-pipe. The code used to process BarSeq data from raw fastq files is available at https://github.com/Asplund-Samuelsson/rebar. All analyses of proteomics, modeling, and fitness data were performed using the R programming language and are documented in R notebooks available at https://github.com/m-jahn/R-notebook-ralstonia-proteome.