Awakening a latent carbon fixation cycle in *Escherichia coli*

Ari Satanowski\(^1,5\), Beau Dronsella\(^1,5\), Elad Noor\(^2\), Bastian Vögeli\(^3\), Hai He\(^1\), Philipp Wichmann\(^1\), Tobias J. Erb\(^3,4\), Steffen N. Lindner\(^1,5\) & Arren Bar-Even\(^1\)

Carbon fixation is one of the most important biochemical processes. Most natural carbon fixation pathways are thought to have emerged from enzymes that originally performed other metabolic tasks. Can we recreate the emergence of a carbon fixation pathway in a heterotrophic host by recruiting only endogenous enzymes? In this study, we address this question by systematically analyzing possible carbon fixation pathways composed only of *Escherichia coli* native enzymes. We identify the GED (Gnd-Entner-Doudoroff) cycle as the simplest pathway that can operate with high thermodynamic driving force. This autocatalytic route is based on reductive carboxylation of ribulose 5-phosphate (Ru5P) by 6-phosphogluconate dehydrogenase (Gnd), followed by reactions of the Entner-Doudoroff pathway, gluconeogenesis, and the pentose phosphate pathway. We demonstrate the in vivo feasibility of this new-to-nature pathway by constructing *E. coli* gene deletion strains whose growth on pentose sugars depends on the GED shunt, a linear variant of the GED cycle which does not require the regeneration of Ru5P. Several metabolic adaptations, most importantly the increased production of NADPH, assist in establishing sufficiently high flux to sustain this growth. Our study exemplifies a trajectory for the emergence of carbon fixation in a heterotrophic organism and demonstrates a synthetic pathway of biotechnological interest.
The ability to assimilate inorganic carbon into biomass sets a clear distinction between autotrophic primary producers and the heterotrophs depending on them for the supply of organic carbon. Most primary production occurs via the ribulose bisphosphate (RuBP) cycle—better known as the Calvin–Benson cycle—used in bacteria, algae, and plants. Six other carbon fixation pathways are known to operate in various bacterial and archaean lineages and also contribute to primary production. Recent studies have made considerable progress in establishing carbon fixation pathways in heterotrophic organisms, with the long-term goal of achieving synthetic autotrophy, which could pave the way towards sustainable bioproduction schemes rooted in CO2 and renewable energy. Most notably, overexpression of phosphoribulokinase and Rubisco, followed by long-term evolution, enabled the industrial hosts Escherichia coli and Pichia pastoris to synthesize all biomass from CO2 via the RuBP cycle. Also, overexpression of enzymes of the 3-hydroxypropionate cycle established the activity of different modules of this carbon fixation pathway in E. coli. While such studies help us to gain a deeper understanding of the physiological changes required to adapt a heterotrophic organism to autotrophic growth, they do not, however, shed light on the origin of the carbon fixation pathways themselves.

Besides the reductive acetyl-CoA pathway and the reductive TCA cycle—both of which are believed to have originated early in the evolution of metabolism—the other carbon fixation routes are thought to have evolved by recruiting enzymes from other metabolic pathways. For example, Rubisco—the carboxylating enzyme of the RuBP cycle—probably evolved from a non-CO2-fixing ancestral enzyme, thus emerging in a non-autotrophic context. Similarly, acetyl-CoA carboxylase likely originated as a key component of fatty acid biosynthesis before being recruited into carbon fixation pathways in several prokaryotic lineages. The limited number of natural carbon fixation pathways indicates that the recruitment of endogenous enzymes to support carbon fixation is a rather exceptional event. To understand this process better we aimed to recreate it in a heterotrophic bacterium.

Here, we use a computational approach to comprehensively search for all thermodynamically feasible carbon fixation pathways that rely solely on endogenous E. coli enzymes. We identify a promising candidate route—the GED cycle—that is expected to enable carbon fixation with minimal reactions and with a high thermodynamic driving force. This synthetic route combines reductive carboxylation of ribulose 5-phosphate (Ru5P) with the Entner–Doudoroff (ED) pathway, gluconeogenesis, and the pentose phosphate pathway. We demonstrate that overexpression of key pathway enzymes together with small modifications of the endogenous metabolic network enable growth via the GED cycle, including the carboxylation step, for the biosynthesis of (almost) all biomass building blocks. Our findings indicate the feasibility of recruiting endogenous enzymes to establish a non-native carbon fixation pathway and pave the way for future establishment of synthetic autotrophy based on new-to-nature pathways.

Results
Systematic search for latent carbon fixation pathways in E. coli.
To identify possible carbon fixation pathways that can be established using only native E. coli enzymes, we used the genome-scale metabolic model of this bacterium. We assumed that all reactions are reversible and then used an algorithm to systematically uncover all possible combinations of enzymes, the net reaction of which use CO2 and cofactors (e.g., ATP, NAD(P)H) as sole substrates to produce pyruvate, a reference product commonly used to compare carbon fixation pathways (see “Methods” section and Supplementary Method 1). The pathways were then analyzed thermodynamically: for each pathway, we calculated the Max–min Driving Force (MDF), representing the smallest driving force among all pathway reactions after optimizing metabolite concentrations within a physiological range (see “Methods” section and Supplementary Method 1). We assumed an elevated CO2 concentration of 20% (200 mbar), which is easily attainable in microbial cultivation within an industrial context and further characterizes the natural habitats of E. coli, e.g., the mammalian gut. The MDF criterion enabled us to discard thermodynamically infeasible routes (having MDF < 0) and to compare the feasible pathways according to their energetic driving force, which directly affects their kinetics.

Using this approach, we identified multiple carbon fixation pathways based on endogenous E. coli enzymes (see Supplementary Table 4 for a full list). We ranked the pathways according to two key criteria that can be calculated for each of them in a straightforward manner: their MDF and the number of enzymes they require (preferring fewer enzymes, see “Methods” section and Supplementary Method 1). Pathways ranked high in terms of these criteria are expected to be simpler to establish and to operate more robustly under fluctuating physiological conditions. The pathway that was ranked highest (Fig. 1a) had the lowest number of enzymes while still supporting a high MDF (>3 kJ/mol, such that reverse enzyme flux can be minimized). This cycle is based on the reductive carboxylation of ribulose 5-phosphate (Ru5P) by 6-phosphoglucuronate 6-phosphate dehydrogenase (Gnd). The carboxylation product, 6PG, is then metabolized by the enzymes of the Entner–Doudoroff (ED) pathway—6PG dehydratase (Edd) and 2-keto-3-deoxygluconate 6-phosphate aldolase (Ea)—to produce glyceraldehyde-3-phosphate (GAP) and pyruvate (Fig. 1a). Pyruvate is subsequently converted to GAP via native gluconeogenesis, and GAP is metabolized via the pentose phosphate pathway to regenerate Ru5P, thus completing the cycle. We termed this pathway the GED (Gnd–Entner–Doudoroff) cycle, according to its key enzymes that serve to connect CO2 fixation to central metabolism.

Our computational analysis identified multiple variants of the GED cycle (Supplementary Table 4). However, as these are unnecessarily more complex than the simple GED cycle design, we decided not to consider them further. We also identified numerous pathways that do not require Gnd. These Gnd-independent pathways share a common carbon fixation strategy in which a sub-cycle converts CO2 to formate, where the release of formate is catalyzed by an oxygen-sensitive oxoacid formate lyase. Formate is then assimilated via one of several variants of the reductive glycine pathway (see Fig. 1b for an example), the activity of which was recently demonstrated in E. coli. Alternatively, formate is assimilated via a variant of the serine cycle, or, more precisely, the previously suggested serine-threonine cycle (see Fig. 1c for an example). While these formate-dependent pathways are interesting, their high oxygen sensitivity and general complexity make them less attractive. Therefore, for further investigation, we decided to focus on the GED cycle.

Properties of the GED cycle and its enzymes.
The GED cycle mirrors the structure of the canonical RuBP cycle, where phosphoribulokinase and Rubisco are replaced with Gnd, Edd, Ea, and gluconeogenic enzymes (Fig. 1). Similarly to the RuBP cycle, the GED cycle is autocatalytic and any one of its intermediates can be used as a product to be diverted towards the biosynthesis of cellular building blocks. Production of pyruvate, a key biosynthetic building block, is more ATP-efficient via the GED cycle.
The Gibbs energy of the Gnd reaction indicates that it should be in the oxidative direction required for pathway activity, with the exception of Gnd. CH2-THF corresponds to methylene-THF.

M, [CO2] oxidative decarboxylation direction (pH 7.5, ionic strength of 0.25 M, [CO2] = 200 mbar, and 1 mM concentration of the other reactants). Indeed, similar oxidative decarboxylation enzymes are known to support reductive carboxylation, for example, the malic enzyme and isocitrate dehydrogenase. While sporadic studies have reported that some Gnd variants support the reductive carboxylation of Ru5P in vitro, a comprehensive kinetic characterization of this activity in bacterial Gnd variants is lacking. More importantly, it remains unclear whether this reaction could operate under physiological conditions, where the concentrations of substrates and products are constrained; that is, substrate concentrations are not necessarily saturating and product concentrations are non-negligible.

First, we measured the kinetics of E. coli Gnd. We found Gnd to have a rather high $k_{cat}$ in the reductive carboxylation direction, approaching $6 \text{ s}^{-1}$ ($5.9 \pm 0.2 \text{ s}^{-1}$ with Ru5P as substrate, Table 1 and Supplementary Fig. 1), about twice as high as the $k_{cat}$ of most plant Rubisco variants. The affinity of Gnd towards CO2 is high enough to enable saturation under elevated CO2 concentrations: $K_M = 0.9 \pm 0.1 \text{ mM}$ (Table 1 and Supplementary Fig. 1) which is equivalent to ~3% CO2 in the headspace (at ambient pressure). Notably, these kinetic parameters are substantially better than those previously reported for a eukaryotic Gnd variant ($k_{cat} \sim 1 \text{ s}^{-1}$ and $K_M = 15 \text{ mM}^{31,32}$).

### Table 1 Kinetic parameters of E. coli Gnd in the reductive and oxidative directions.

| Reaction                      | $k_{cat}$ (s$^{-1}$) | $K_M$ (mM) |
|-------------------------------|----------------------|------------|
| Reductive carboxylation        |                      |            |
| Ribulose 5-phosphate           | 5.9 ± 0.2            | 2.8 ± 0.2  |
| NADPH                         | 5.2 ± 0.2            | 1.0 ± 0.1  |
| CO2                           | 4.7 ± 0.2            | 0.9 ± 0.1  |
| Oxidative carboxylation        |                      |            |
| 6-Phosphogluconate             | 49 ± 1               | 0.035 ± 0.002 |
| NADP$^+$                       | 46 ± 1               | 0.012 ± 0.003 |

Values indicate mean ± standard error. Fitted Michaelis-Menten curves are shown in Supplementary Fig. 1. See the “Methods” section for a detailed description of the kinetic characterization.
To check how prevalent the potential of carbon fixation via the GED cycle is, we performed a phylogenetic analysis to identify bacteria that harbor its key enzymes (see “Methods” section). We found that the enzymes of the GED cycle are ubiquitous in α-proteobacteria (962 species), gamma-proteobacteria (1684), and actinobacteria (312) (Fig. 1d). The species of these phyla might therefore be prime candidates to search for the carbon fixation activity of the GED cycle. However, in other bacterial lineages, the combined occurrence of the GED cycle enzymes is quite rare, with only ten other species harboring all key enzymes.

Selection for the activity of the GED shunt within a Δrpe context. Engineering *E. coli* for autotrophic growth via the GED cycle would be a challenging task requiring considerable metabolic adaptation of the host; for example, establishing a delicate balance between the metabolic fluxes within the cycle and those diverging out of the cycle (as found in previous efforts to establish the RuBP cycle in *E. coli*). Hence, to check the feasibility of the cycle, we focused on establishing growth via the GED shunt, representing a segment of the full cycle which consists of reductive carboxylation by Gnd and the subsequent ED pathway (blue reactions in Fig. 2; for a similar approach see He et al.37 and Meyer et al.38). As we show below, growth via this linear shunt requires the activity of most enzymes of the GED cycle but relies on a pentose substrate rather than the regeneration of Ru5P.

First, we generated an *E. coli* strain deleted in the gene encoding for ribulose 5-phosphate 3-epimerase (Δrpe). This strain cannot grow on ribose as a sole carbon source, as ribose-5-phosphate (R5P) cannot be converted to xylulose 5-phosphate, thus blocking the pentose phosphate pathway (Fig. 2a). The activity of Gnd, Edd, and Eda should restore growth by enabling

![Diagram of metabolic pathways](image_url)

**Fig. 2 Activity of the GED shunt in a Δrpe strain.** a) Design of the Δrpe selection scheme. Ribose can be assimilated only via the activity of the GED shunt, where the biosynthesis of almost all biomass building blocks is dependent on the pathway (marked in yellow). Growth on gluconate (violet) is not dependent on reductive carboxylation via Gnd and thus serves as a positive control. Reaction directionality is shown as predicted by flux balance analysis. b) Growth of a Δrpe strain on ribose (20 mM) as a sole carbon source is dependent on elevated CO2 concentration (20%, i.e., 200 mbar) and overexpression of gnd, edd, and eda (pGED). Overexpression of only gnd (pG) or only edd and eda (pGED) failed to establish growth (less than two doublings). Cultivation at ambient CO2 also failed to achieve growth. Values in parentheses indicate doubling times. Curves represent the average of technical duplicates, which differ from each other by <5%. Growth experiments were repeated independently three times to ensure reproducibility. c) Cultivation on 13CO2 confirms the operation of the GED shunt. On the left, a prediction of the labeling pattern of key amino acids is shown. The observed labeling fits the prediction and differs from the WT control cultivated under the same conditions. Labeling of amino acids in the WT strain stems from the natural occurrence of 13C as well as from reactions that exchange cellular carbon with CO2, e.g., the glycine cleavage system and anaplerotic/cataplerotic cycling. Values represent averages of two independent cultures that differ from each other by <10%. 3PG 3-phosphoglycerate, ALA Alanine, GAP glycereraldehyde-3-phosphate, GLY Glycine, HIS Histidine, PYZ pyruvate, SER Serine, VAL Valine. Source data underlying b and c are provided as a Source Data file.
the conversion of RSP to GAP and pyruvate, from which all cellular building blocks can be derived (Fig. 2a). This would enable direct selection for the activity of the GED shunt.

We found that overexpression of Gnd, Edd, and Eda from a plasmid (pGED) enabled the growth of the Δrpe strain on ribose only under elevated CO2 concentration (green lines in Fig. 2b). The observed growth rate via the GED shunt is almost half of that obtained with gluconate, which requires no carboxylation by Gnd and thus serves as a positive control (doubling times 6.9 and 2.8 h, respectively). While Gnd, Edd, and Eda are all present in the genome of E. coli, their native expression level is too low to enable sufficient activity of the GED shunt: overexpression of Gnd alone (pG) or of Edd and Eda alone (pED) did not support growth (less than two doublings, brown lines in Fig. 2b).

To confirm that growth indeed proceeds via the GED shunt, we performed a 13C-labeling experiment. We cultivated the Δrpe strain with unlabeled ribose and 13CO2, and measured the labeling pattern of five proteinogenic amino acids—serine, glycine, alanine, valine, and histidine. The results confirm the activity of the GED shunt (Fig. 2c): (i) since 13CO2 is incorporated as the carboxylic carbon of 6PG, GAP is completely unlabeled and hence serine and glycine that are derived from it are unlabeled; (ii) pyruvate is generated both from GAP and directly from Eda activity (Fig. 2c), such that about half of the pyruvate molecules are unlabeled and half are labeled once at their carboxylic carbon—as a result, half of the alanine and valine molecules are labeled (during valine biosynthesis two pyruvate molecules are condensed and one carboxylic carbon is lost as CO2).

These results confirm that, upon overexpression of Gnd, Edd, and Eda, the GED shunt is sufficiently active to provide the cell with almost all cellular building blocks as well as energy (by complete oxidation of pyruvate via the TCA cycle). As mentioned above, the growth of this strain requires the simultaneous activity of most enzymes on the GED cycle, including those of glycolysis and the pentose phosphate pathway; for example, net production of erythrose 4-phosphate (E4P) from ribose requires the combined activity of Gnd, the ED pathway, and enzymes of the pentose phosphate pathway.

A ΔtktAB context requires additional metabolic adaptations to enable growth via the GED shunt. To check whether the operation of the GED shunt is robust, we decided to use another metabolic background to select for its activity. We deleted the genes encoding for both isozymes of transketolase (ΔtktAB). This strain, in which the non-oxidative pentose phosphate pathway is effectively abolished, cannot grow when provided with a pentose as the sole carbon source37,41. Furthermore, as E4P cannot be synthesized in this strain, small amounts of essential cellular components, the biosynthesis of which is E4P-dependent, need to be added to the media37,41: phenylalanine, tyrosine, tryptophan, shikimate, pyridoxine, 4-aminobenzoate, 4-hydroxybenzoate, and 2,3-dihydroxybenzoate (referred to as E4P supplements, see “Methods” section).

As with the Δrpe strain, we expected overexpression of Gnd, Edd, and Eda to enable growth on a pentose substrate such as xylose (supplemented with E4P supplements) (Fig. 3a). However, we failed to obtain growth even at an elevated CO2 concentration (less than two doublings, green lines in Fig. 3b). This is in line with previous findings that seemingly small differences in the design of metabolic growth selection schemes (e.g., the choice of deleted enzymes) can lead to substantially dissimilar metabolic behaviors37,42.

Following the failure to obtain GED shunt-dependent growth within the ΔtktAB context, we sought to harness adaptive evolution to provide us with information on further cellular adaptations required for the activity of the synthetic route. Toward this aim, we inoculated the ΔtktAB+pGED strain into multiple test-tubes with xylose and E4P supplements and incubated them for an extended period of time at 37 °C and 20% CO2. After ~2 weeks, the culture in several of these test-tubes showed apparent growth. When the cells from the growing cultures were transferred to a fresh selective medium, we observed immediate growth, indicating that genetic adaptation had occurred. Yet, only one of these cultures showed robust growth on the selective medium, whereas that of the others was less reproducible and highly sensitive to the exact conditions, preculture, and inoculation. An isolated single clone from the robust culture grew, under elevated CO2 concentration, with a doubling time of 5.3 h (red lines in Fig. 3b).

We sequenced the genome of the mutated strain and identified a single mutation (compared to the parental strain): the mobile element IS543 was inserted 104 bp upstream of the pntAB operon (Supplementary Data 1), which encodes for the membrane-bound transhydrogenase that plays a key role in supplying the cell with NADPH44,45. Insertion of the IS5 mobile element is well-known to occur in adaptive evolution experiments33, increasing the expression levels of the downstream genes46,47. Indeed, we found that the transcription of pntA increased ~3-fold in the mutated strain compared to the non-mutated parent and WT strains (Fig. 3c). The contribution of this mutation to the activity of the GED shunt can be easily explained, as it increases the generation of NADPH required for the reductive carboxylation of Ru5P by Gnd (Fig. 1). To confirm that increased pntAB expression indeed enables GED shunt-dependent growth, we replaced the native promoter of pntAB (within the unmaturated strain) with three previously characterized constitutive promoters: weak (W-pntAB), medium (M-pntAB), and strong (S-pntAB), with relative strengths of 1:10:20, respectively48. We found that while the weak promoter failed to support growth (less than two doublings, purple line in Fig. 3d), the medium and strong promoters supported growth with a similar doubling time to that of the mutated strain, ~5.4 h, at elevated CO2 concentrations (orange and green lines in Fig. 3d). This indicates that sustaining a sufficiently high expression of pntAB suffices to enable the activity of the GED shunt within a ΔtktAB metabolic context.

We wondered whether other metabolic manipulations that target NADPH homeostasis could also enable the growth of the ΔtktAB+pGED strain. We tested the deletion of the gene encoding for the soluble transhydrogenase (Δsth), as this enzyme is known to provide a strong sink for NADPH44,45. Yet, the ΔtktAB Δsth+pGED strain did not grow on xylose even at an elevated CO2 concentration (less than two doublings, brown lines in Fig. 3d). Furthermore, the deletion of sth in the ΔtktAB S-pntAB+pGED strain improved its growth only marginally (light blue line in Fig. 3d). Taken together, it seems that the soluble transhydrogenase has only a minor effect on NADPH availability within this metabolic context.

We hypothesized that alongside NADPH availability, competing sources of 6PG could play a key role in determining the feasibility of the GED shunt. Specifically, 6PG is natively produced by the oxidative pentose phosphate pathway, which thus provides a metabolic push against the reductive activity of Gnd. Moreover, the activity of glucose 6-phosphate 1-dehydrogenase (encoded by zwf), the first enzyme of the oxidative pentose phosphate pathway, has been reported to increase under conditions of high NADPH demand (i.e., upon depletion of cellular NADPH)9,30. Hence, we wondered whether the deletion of zwf could remove a barrier for reductive carboxylation by Gnd and thus assist the activity of the GED shunt. We found that this is indeed the case, where the ΔtktAB Δzwf+pGED strain was able to grow under elevated CO2 concentrations with a doubling
Fig. 3 Activity of the GED shunt in a ΔtktAB strain. a Design of the ΔtktAB selection scheme. Xylose can be assimilated only via the GED shunt. E4P supplements are provided as the ΔtktAB strain cannot synthesize erythrose 4-phosphate. Growth on gluconate is not dependent on reductive carboxylation by Gnd and thus serves as a positive control. Reaction directionalities are shown as predicted by flux balance analysis. b Growth on xylose upon overexpression of gnd, edd, and eda (pGED) was achieved only after mutation and was dependent on elevated CO2 concentration. Values in parentheses indicate doubling times. Curves represent the average of technical duplicates, which differ from each other by <5%. Growth experiments were repeated independently three times to ensure reproducibility. c Expression analysis by quantitative RT-PCR revealed that the transcript level of pntA increased ~3-fold in the mutated strain. Bars correspond to the average of two independent experiments, which are shown as circles. Gluconate and xylose indicate carbon sources used. d Genomic overexpression of pntAB using medium (M) or strong (S) promoter, but not weak (W) promoter, supported growth of a ΔtktAB Δzwf ΔpntAB mutant strain on xylose (legend to the left). e Deletion of glucose 6-phosphate dehydrogenase (Δzwf) supported the growth of a ΔtktAB ΔpntAB strain on xylose (legend to the left). f 13C-labeling experiments confirm the operation of the GED shunt. Cells were cultivated with xylose (1-13C) and 13CO2. Observed labeling fits the expected pattern and differs from that of a WT strain cultured under the same conditions. Results from additional labeling experiments are shown in Supplementary Fig. 2. 3PG 3-phospho-glycerate, ALA Alanine, GAP glyceraldehyde-3-phosphate, GLY Glycine, HIS Histidine, PYR pyruvate, SER Serine, VAL Valine. Source data underlying b-f are provided as a Source Data file.
time of 5.4 h (blue line in Fig. 3e). Deleting zwf in the ΔtktAB S-pmtAB + pGED strain did not improve growth (dark yellow line in Fig. 3e), suggesting that the effects of NADPH and 6PG availability are not additive or that a different bottleneck is limiting growth.

To confirm the activity of the GED shunt within the ΔtktAB metabolic context we performed several 13C-labeling experiments (Fig. 3f and Supplementary Fig. 2). When the ΔtktAB Δzwf + pGED strain was fed with both 13CO2 and 1-13C-xylene, we expected the GED shunt to produce unlabeled GAP and twice labeled pyruvate (Fig. 3f). Hence, serine and glycine, which are derived directly from GAP, should be unlabeled while about half of pyruvate (derived from GAP metabolism) should be unlabeled and the other half (generated directly by Eda activity) twice labeled. This should lead to half of the alanine being unlabeled and half twice labeled while the labeling of valine should roughly follow a 1:1:1:1 pattern (unlabeled: once labeled: twice labeled: thrice labeled). The observed labeling confirms these expected patterns (Fig. 3f). Histidine, the carbons of which originate from RSP and the β-carbon of serine is labeled once as expected (Fig. 3f). The labeling patterns we observe upon feeding with unlabeled xylene and 13CO2 (Supplementary Fig. 2A) as well as upon feeding with 5-13C-xylene and unlabeled CO2 (Supplementary Fig. 2B) further confirm that growth of the ΔtktAB Δzwf + pGED strain indeed takes place exclusively via the GED shunt.

Growth via the GED shunt in a strain that could support cyclic flux. While the Δtkt and ΔtktAB strains were useful selection platforms to test the activity of the GED shunt, they are, in a sense, metabolic dead-ends. This is because the activities of both ribulose-phosphate 3-epimerase (Rpe) and transketolase (Tkt) are essential for the operation of the full GED cycle, that is, for the regeneration of Ru5P from GAP. To address this problem, we aimed to construct a strain which keeps all necessary enzymes of the GED cycle intact, while still allowing to select for the activity of the GED shunt, i.e., preventing utilization of a pentose substrate as sole carbon source via the canonical pentose phosphate pathway. Such a strain would enable a smooth transition from GED shunt-dependent growth on a pentose substrate towards autotrophic growth via the GED cycle.

We, therefore, constructed a strain deleted in all enzymes that can metabolize fructose 6-phosphate (F6P), directly or indirectly, into a downstream glycolytic intermediate (ΔpkA ΔfsaAB ΔfruK) or channel it into the oxidative pentose phosphate pathway (Δzwf). The latter gene deletion should also support the activity of the GED shunt, as was shown above within the ΔtktAB context. The strain containing all of these deletions, which we term ΔPZF, establishes a uni-directional block within the pentose phosphate pathway. That is, growth on a pentose substrate is not possible due to the accumulation of F6P that prevents further conversion of pentose phosphates into GAP (Fig. 4a)8. In contrast, flux in the opposite direction, as required for the GED cycle, can still occur, since fructose 1,6-biphosphosphate can be dephosphorylated to F6P which is then used to regenerate Ru5P.

To establish the growth of the ΔPZF strain on xylene via the GED shunt, we overexpressed Gnd, Edd, and Eda. However, transforming the ΔPZF strain with pGED failed to support growth on xylene even at elevated CO2 (less than two doublings, the red line in Fig. 4b). Hence, we again harnessed natural selection and performed short-term evolution by incubating the strain in multiple test-tubes for an extended period of time in xylene minimal medium at 37 °C and 20% CO2. Within 6–8 days, three parallel cultures started growing. Isolated clones from two of these mutant cultures displayed a fairly high growth rate (doubling time of 8–10 h, green and blue lines in Fig. 4b), while clones from the third culture showed considerably slower growth (doubling time >50 h, orange line in Fig. 4b).

A recent study reported that a similar E. coli deletion strain (deleted in pfkA, zwf, and the glucose uptake system) was able to grow on a xylose minimal medium, but was accompanied by the secretion of a substantial amount of glucose (34% of consumed xylose)31. Such secretion of a dephosphorylated sugar could relieve the inhibitory accumulation of F6P and thus theoretically enable the growth of the ΔPZF strain even without the activity of the GED shunt. However, the growth of the ΔPZF mutants we identified cannot be explained by such a phenomenon since (i) growth at ambient CO2 was not observed (Fig. 4b), confirming strict dependency on the activity of the GED shunt; and (ii) no glucose could be detected in the supernatants of the growing cells (see “Methods” section). This excludes the possibility that growth was even partially supported by the conversion of xylose into glucose.

To provide unequivocal confirmation that the xylose assimilation in the ΔPZF strain mutants proceeds via the GED shunt, we conducted 13C-labeling experiments using the fastest-growing strain (ΔPZF + pGED mutant “C”). We found that this strain displayed a labeling pattern almost identical to that of the ΔtktAB Δzwf + pGED strain described above, thus confirming growth via the GED shunt (Fig. 4c and Supplementary Fig. 3). We sequenced the genomes of the mutant strains, compared them to the parental strain, and discovered several mutations (Supplementary Table 1). All isolated colonies from the two fast-growing cultures shared an identical mutation at the start of an L-leucyl-tRNA (leuX) and, in most colonies, avtA, encoding for valine-pyruvate aminotransferase, had mutated. While the exact contribution of these mutations to the growth phenotype remains elusive, the isolated strains provide a promising starting point for the evolution of the full GED cycle.

Discussion

Our computational analysis identified multiple carbon fixation pathways that are based solely on endogenous E. coli enzymes. A key factor for the success of this analysis was to ignore the rather arbitrary dichotomic classification of reactions as reversible or irreversible as suggested by metabolic models. Instead, we first identified potential pathways based on pure stoichiometric analysis and then calculated the thermodynamic feasibility and driving force of each of the candidate routes. This enabled us to uncover potential carbon fixation pathways that were not identified before13. Indeed, the GED cycle itself was previously ignored as Gnd was considered to be an irreversible decarboxylating enzyme. As we have shown here, however, Gnd can catalyze the carboxylation reaction quite efficiently, with a kcat almost double that of a typical plant Rubisco36. This finding is similar to a recent study that found that citrate synthase—which is usually thought to be irreversible—can catalyze citrate cleavage, thus enabling carbon fixation via a unique variant of the reductive TCA cycle52,53. These examples indicate that we should revise our dogmatic interpretation of enzyme reversibility and instead adopt a more quantitative approach to understand reaction directionality.

Previous studies have suggested multiple synthetic carbon fixation pathways that could surpass the natural routes in terms of resource use efficiency, thermodynamics, and/or kinetics45,56. The most advanced of these pathways is the CETCH cycle55 that combines segments of the 3-hydroxypropionate/4-hydroxybutyrate cycle56 and the ethylmalonyl-CoA pathway57. The CETCH cycle was assembled in vitro using enzymes from nine organisms and optimized in several rounds of enzyme engineering55,58. However, the in vivo implementation of this synthetic pathway, as well as of
Observed labeling reproducibility. The average of technical quadruplicates, which differ from each other by <5%. Growth experiments were repeated independently three times to ensure reproducibility.

other previously suggested routes, is highly challenging due to its complexity and requirement for the considerable rewiring of central metabolic fluxes. The GED cycle provides a favorable alternative to these routes, as the fluxes it requires mostly correspond to native gluconeogenesis and the pentose phosphate pathway. Hence, the establishment of carbon fixation via the GED cycle might be less demanding and more likely to succeed.

We demonstrated the feasibility of carbon fixation via the GED cycle by establishing growth via the GED shunt—a linear pathway variant that requires the key pathway reactions to provide (almost) all biomass building blocks and cellular energy without regenerating the substrate Ru5P (Figs. 2a and 3a). In line with previous studies, we found that changing the metabolic context can have a dramatic effect on the activity of a metabolic module under selection: the GED shunt was able to directly support the growth of a Δpe strain but not of a ΔtktAB strain or a ΔPZF strain, even though in all of them the pentose phosphate pathway is disrupted. Despite this, a short-term adaptation was able to restore the growth of the latter two strains. Within the ΔtktAB strain, we demonstrated that either an increase in the supply of the substrate (e.g., NADPH, via pntAB overexpression) or a decrease in the availability of the product (e.g., 6PG, via zwf deletion) is sufficient to enable Gnd-dependent carboxylation and growth via the GED shunt.

The GED shunt might have biotechnological applications on its own. Previous studies have demonstrated that co-assimilation of CO2 can increase production yields from common feedstocks such as sugars. This is attributed to the fact that the biosynthesis of certain compounds from sugars results in the production of excess reducing power, which can be utilized to fix...
CO₂ and thereby generate more product. Such assimilation of CO₂ can also serve to compensate for the carbon released during the oxidation of pyruvate to acetyl-CoA, thus addressing a common challenge in the production of value-added chemicals derived from acetyl-CoA\(^6\). Indeed, we applied flux balance analysis to simulate production in non-growing cells (see "Methods" section) and found that rerouting the utilization of sugar substrates via the GED shunt is expected to increase the yield of various commercially interesting products, such as acetate, pyruvate, acetone, citrate, and itaconate (Fig. 5 and Supplementary Fig. 4). A further supply of reducing power by adding auxiliary substrates such as hydrogen or formate\(^6\) can make the GED shunt advantageous over glycolysis for even more reduced products, such as ethanol, lactate, 1-butanol, and fatty acids (Fig. 5 and Supplementary Fig. 4). While the RuBP shunt—a linear version of the RuBP cycle, which channels Ru5P via Rubisco—can also increase the fermentative yield of some products\(^5\), the GED shunt always outperforms it due to a lower ATP requirement (Fig. 5 and Supplementary Fig. 4).

A previous study has established the RuBP cycle in \textit{E. coli}, demonstrating that this heterotrophic bacterium can be modified to grow autotrophically with CO₂ as a sole carbon source\(^7\). However, to our knowledge, the current study is the first one in which the capacity for net carbon fixation was explored in vivo using only endogenous enzymes of a heterotrophic host, thus shedding light on the emergence of carbon fixation pathways. Importantly, the establishment of the RuBP cycle in \textit{E. coli} required long-term adaptive evolution of the microbe under selective conditions, which modulated the partitioning of metabolic fluxes between carbon fixation and biosynthetic pathways\(^2\). We expect that autotrophic growth via the GED cycle can be achieved in a similar manner. The ΔPZF strain serves as an ideal starting point for such a future evolution experiment, as its growth is dependent on the activity of the GED shunt while it still harbors all necessary enzymes to run the GED cycle. The gradual evolution of autotrophic growth via the GED cycle could be achieved via the additional expression of a formate dehydrogenase as an energy-supplying module and long-term cultivation with limiting amounts of xylose and saturating amounts of CO₂ and formate\(^2\).

As the GED cycle is composed of ubiquitous enzymes that are widespread throughout major bacterial phyla, it is tempting to speculate that this route naturally operates in yet unexplored microorganisms. Especially promising are \textit{α}-proteobacteria, \textit{γ}-proteobacteria, and actinobacteria, which contain many species that harbor all key enzymes of the pathway (Fig. 1d). Such bacteria could evolve autotrophic growth by recruiting the enzymes of the GED cycle if exposed to the appropriate selective conditions—for example, lack of organic carbon sources and availability of energy sources such as inorganic electron donors (e.g., hydrogen).

The GED cycle could be used to replace the RuBP cycle in plants, algae, and bacteria\(^5\), requiring relatively modest changes to the endogenous metabolic structure of carbon fixation. Replacing the RuBP cycle in chemolithotrophic bacteria of biotechnological significance, e.g., \textit{Cupriavidus necator}, would be relatively straightforward as cultivating these microorganisms on elevated CO₂ is a common practice, thus avoiding the rate limitation associated with the low affinity of Gnd to CO₂. Such an engineered microorganism may support higher product yields when cultivated under autotrophic conditions given that most value-added chemicals are derived from pyruvate and acetyl-CoA and the biosynthesis of these metabolites via the GED cycle requires less
ATP equivalents than via the RuBP cycle. Furthermore, as the carboxylating activity of Gnd would be enhanced by the carbon-concentrating mechanisms of algae and cyanobacteria, engineering these organisms to use the GED cycle could be advantageous. However, to facilitate the establishment of the GED cycle in higher plants, the affinity of Gnd towards CO₂ would have to be improved, e.g., via the rational engineering of CO₂ binding sites, as successfully demonstrated recently in a proof-of-principle study. Alternatively, replacing E. coli Gnd with a variant that has a considerably higher k_cat (~100 s⁻¹) could compensate for the low affinity towards CO₂, (i.e., achieving k_cat/K_M at least as high as that of plant Rubisco). As some variants of similar reductively carboxylating enzymes—isocitrate dehydrogenase and the malic enzyme—incorporate CO₂ with k_cat surpassing 100 s⁻¹, identifying a Gnd variant supporting such a high carboxylation rate might be feasible. Engineering such an optimized GED cycle into crop plants could boost agricultural productivity, thus addressing one of our key societal challenges.

Methods

Identifying carbon fixation cycles in E. coli using a constraint-based model. In order to find all possible carbon fixation cycles using E. coli endogenous enzymes, we used an approach similar to the one we have previously developed. In this previous study, all reactions found in the KEGG database, denoted the universal stoichiometric matrix, were used to design potential CO₂ fixation pathways using a Mixed-Integer Linear Programming (MILP) approach. Here, we focused only on enzymes present in the most recent genome-scale metabolic reconstruction of E. coli: iML1514. We further added thermodynamic constraints in order to exclude infeasible pathways and to rank the feasible ones based on their MDF.

We used the COBRApy package (version 0.17) to identify carbon fixation pathways. First, we removed all exchange and transport reactions and kept only the strictly cytoplasmic ones. Then, we added reactions that allow the free flow of electrons and energy (by regenerating ATP, NADH, and NADPH) as well as inorganic compounds (protons, water, oxygen, and ammonia). Finally, we defined an optimization problem, where the set of reactions in each pathway should overall convert 3 moles of CO₂ to one mole of pyruvate. The objective function of this optimization is a combination of the MDF and the minimum number of reactions in each pathway. The standard approach for multi-objective optimization is the maximization of a linear combination of the two functions. Here, we maximized the MDF (in units of RT) minus the number of reactions. We note that changing the relative weight between these two objectives did not change our main result, namely that the GED pathway is Pareto-optimal. A more detailed list of changes to the model and the formal description of the optimization problem can be found in Supplementary Method 1.

Phylogenetic analysis. In order to assess how many bacterial genomes contain the genes necessary for GED, we used AnnoTree—a web tool for visualization of genome annotations in prokaryotes (http://annotree.uwaterloo.ca/). AnnoTree generates a phylogenetic tree and highlights genomes that include all of the KEGG orthologues selected in the query. Since most annotations are homology based, and therefore not always precise, we restricted our search to the enzymes that are involved in the GED pathway and cannot be easily replaced by a metabolic bypass (e.g., transaldolase could be replaced by sedoheptulose 1,7-bisphosphate aldolase and phosphatase):

- K0033—6-phosphogluconate dehydrogenase (EC:1.1.1.44 1.1.1.343): the enzymatic keystone of the GED cycle.
- K01690—6-phosphogluconate dehydratase [EC:4.2.1.12]: indicative of the Enterobacteriaceae.
- K00615—transketolase [EC:2.2.1.1]: indicative of the pentose phosphate pathway.
- K00927—phosphoglycerate kinase (EC:2.7.2.3): indicative of gluconeogenesis (as glycogen can be synthesized by a non-phosphorylating glyceroldehyde-3-phosphate dehydrogenase).
- K01688—malate dehydrogenase (EC:1.1.1.37): indicative of the D-lactate dehydrogenase and phosphatase.
- K01912—fumarate hydratase (EC:4.2.1.2): indicative of the GED pathway.
- K02218—pyruvate synthase (EC:2.7.3.2): indicative of the RuBP cycle/shunt.
- K02505—NADH dehydrogenase (EC:1.6.99.3): indicative of the respiratory electron transport chain.

We chose the family level as the tree resolution since it provides a good balance between the number of leaves and branching diversity. For the sake of readability of Fig. 1d, we applied the auto collapse clades option in iTOL with the setting BRL = 1.

Yield estimation via flux balance analysis. Theoretical yields were estimated via flux balance analysis, conducted in Python with COBRApy. We used the most updated E. coli genome-scale metabolic network (iML1514) with curations for the (THDP) translocase (iML1515) in our model. As some reactions can operate simultaneously without violating the second law of thermodynamics, we added reactions of Fe²⁺, H₂, H₂S, methionine, nadh, and succinate instead of iron free or S as electron acceptors in simulated anaerobic conditions. The default value of the model for the non-growth-associated ATP maintenance reaction was used (ATPM; 6.86 mmol/gDW/h) to predict maximal theoretical yields in non-growing cells.

The Gnd reaction was changed to be reversible as part of the GED cycle/shunt; phosphoribulokinase (PRK) and RuBisCO (RBCP) reactions were added to create the RuBP cycle/shunt. In order to use hydrogen as a proxy electron donor, hydrogen dehydrogenase reaction was added, correcting the Gnd reaction. We used the model with these modifications as a “wild-type” reference.

To analyze yields of the Gnd shunt and RuBP shunt, these linear pathway variants were created by blocking the reactions of phosphofructokinase (PFK), STP-reacting phosphofructokinase (PFK;3), fructose 6-phosphate aldolase (F6PA), glucose 6-phosphate dehydrogenase (G6PDH), and fructose-bisphosphatase (FBP). Then, xylene or glucose was assumed as a constrained carbon source together with unconstrained CO₂ (similar to Hadlicek et al.) and unconstrained hydrogen (when noted). The uptake rates for xylose and glucose were set to experimentally determined values for anaerobically growing E. coli cultures (xylose: 10.8 mmol/gDW/h; glucose: 13.1 mmol/gDW/h). The full code, including changes to the model and the new reactions of each production route, can be found at https://github.com/elad-noor/ged-cycle-/tree/master/FBA.

Strains and genomic modifications. All strains used in this study are listed in Table 2. Gene deletions and growth experiments were performed in strains derived from E. coli SH4889, a strain derived from wild-type E.coli MG1655. The SH series strain contains inducible λ-Red recombinase and (fructose-1,6-biphosphate and fructose) integrated into its genome to increase ease-of-use for multiple genomic modifications. Gene deletion strains were either taken from previous studies (Table 2) or generated by P1 phage transduction. Strains from the Keio collection carrying single gene deletions with a kanamycin-resistance gene (KmR) as the selectable marker were used as donor strains. Strains that had acquired the desired deletion were selected by plating on appropriate antibiotics (Kanamycin, Km) and confirmed by determining the size of the respective genomic locus via PCR (oligonucleotide sequences are shown in Supplementary Table 2). To remove the selective marker, flipase was induced in a fresh culture grown to OD₆₀₀ ~ 0.2 by adding 50 mM l-Rhamnose and cultivating for 4–8 h at 30°C. Loss of the antibiotic resistance was confirmed by identifying individual colonies that only grew on LB in absence of the respective antibiotic and by PCR of the genomic locus.

For genomic overexpression of the ptaB operon, its promoter region was edited using a method based on λ-Red recombination. We replaced the native ptaB promoter region (spanning 443 bp upstream of the ptaB start codon) by a weak (pgp1), moderate (pgp10), or strong (pgp20) constitutive promoter and a medium-strength ribosome binding site (RBS “C”); AAGTTAAGGCAATAGAAGAATG downstream of a Cmr cassette for selection (for resulting sequence, see Supplementary Data). For this purpose, the Cmr cassette was amplified from plasmid pKD38 by the primers Cmr-Cmr1 and Cmr-2 followed by overlap extension PCR to combine it with the promoter and RBS amplified from plasmid pGED with the primers PromW-Fwd, PromM-Fwd or Proms-Fwd, respectively, and pntA-Prom2. The construct was inserted into a pJEH1.2 cloning-vector (ThermoScientific, Dreieich, Germany) and confirmed by S1 nuclease screening. Linear dsDNA donors for λ-Red recombination were generated by amplification with primers pntA-Prom1 and pntA-Prom2, DNA (400 ng) was transformed into the desired strains by electroporation after fresh culturing to OD₆₀₀ ~ 0.3 and induction of recombination enzymes by the addition of 15 mM l-Arabinose for 45 minutes. Confirmation of the recombination event was performed via genomic PCR of the pntA-Prom1 promoter region, and removal of the selective marker was performed as described above for the P1 transduction method. The pntA promoter locus was additionally

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-020-19564-5 | www.nature.com/naturecommunications
verified by Sanger sequencing (LGC Genomics, Berlin, DE) after amplification with pntA-V1 and pntA-V2.

**Construction of pGED, pG, and pED vectors.** Cloning was carried out in E. coli DH5α. The native *E. coli* genes encoding 6-phosphogluconate dehydrogenase (*gnd*, UniProt: P00300, HKGO/KDGPD aldolase (*eda*, UniProt: P0A8K5), and phos- phogluconate dehydratase (*edd*, UniProt: P0A24D) were amplified from *E. coli* MG1655 genomic DNA with high-fidelity Phusion Polymerase (ThermoScientific, Dreieich, Germany) using primers listed in Supplementary Table 2. Silent mutations were introduced to remove relevant restriction sites in *gnd* (C292T to remove a PvuI site and C202T to remove a PstI site). For synthetic operons, multiple genes were assembled in pNivC (C292T to remove a PvuI site and C202T to remove a PstI site). Assembly of synthetic operons and expression plasmids was performed as described before. In brief, genes were first inserted individually into a pNivC vector downstream of a ribosomal binding site (RBS “C”), AAGTTA-AGGCAAGA. For synthetic operons, multiple genes were assembled in pNivC vectors using BioBrick restriction enzymes (Fast-Digest: Bsal, Xhol, Sall, Nhel; ThermoScientific, Dreieich, Germany). The generated operons were excised from the pNivC vector by restriction with EcoRI and Nhel (Fast Digest, ThermoScientific, Dreieich, Germany) and inserted into a p-Z ASS vector (p15A medium-copy origin of replication, streptomycin resistance for expression under the control of the constitutive strong promoter pg1). The order of genes in the operons was *gnd, eda, edd, for* for pGED; and *eda, edd, for* for pED. Constructed vectors were confirmed by Sanger sequencing (LGC Genomics, Berlin, DE). The software Geneious 8 (Biomatters, New Zealand) was used for in silico cloning and sequence analysis.

**Culture conditions and growth experiments.** For routine culturing of *E. coli* strains, LB medium was used (5 g/L yeast extract, 10 g/L tryptone, 10 g/L NaCl). Antibiotics were added when appropriate at the following concentrations: Kanamycin 50 µg/mL; Chloramphenicol 30 µg/mL; Ampicillin 100 µg/mL; Streptomycin 100 µg/mL. Growth assays were performed in M9 minimal medium (47.8 mM Na2HPO4, 22 mM KH2PO4, 8.6 mM NaCl, 18.7 mM NH4Cl, 2 mM MgSO4 and 100 µM CaCl2), supplemented with traces element (134 mM EDTA, 31 mM FeCl3·6H2O, 6.2 µM ZnCl2, 0.76 µM CuCl2·2H2O, 0.42 µM CoCl2·2H2O, 1.62 µM H3BO3, 0.081 µM MnCl2·4H2O). Carbon sources were added as described in the text at a concentration of 20 mM. No antibiotics were used in growth experiments, except in precultures. When elevated CO2 was required, cultures were grown in an orbital shaker set to maintain 37 °C and an atmosphere of 20% CO2 mixed with air. Growth on strain deleted in tktAB required further supplementation of E4P (1 mM shikimic acid, 1 µM pyruvate, 250 µM tyrosine, 500 µM phenylalanine, 200 µM tryptophan, 6 µM 4-aminobenzoic acid, 6 µM 4-hydroxybenzoic acid, and 50 µM 2,3-dihydroxybenzoic acid.

Precultures for growth experiments were generally grown in M9 medium with 20 mM glucose as carbon source (relaxing conditions). Antibiotics were added to the precultures if appropriate but omitted for growth experiments. Cells from the preculture were washed three times in M9 medium without carbon source and inoculated to a starting OD600 of 0.02 into M9 media with the final carbon sources as detailed in the text. 96-well plates (Nuncilon Delta Surface, ThermoScientific, Dreieich, Germany) were filled with 150 µL culture and covered with 50 µL mineral oil (Merck, Darmstadt, Germany) to avoid evaporation while allowing gas exchange. Aerobic growth was monitored in technical duplicates, triplicates, or quadruplicates at 37 °C in a BioTek Epoch 2 Microplate Spectrophotometer (BioTek, Bad Friedrichshall, Germany) by absorbance measurements (600 nm) of each well every ~10 min with intermittent orbital and linear shaking. Blank measurements were subtracted and OD600 values were converted to OD600 values by multiplying with a factor of 4.35, as previously established empirically for the instruments. When elevated CO2 was required, the atmosphere was maintained at 20% CO2 mixed with 80% air by placing the plate reader inside a Kuhner ISFI-X incubator shaker (Kuhner, Birsfelden, Switzerland).

**Isolation and sequencing of a ΔtktAB + pGED mutant capable of growing via the GED shunt.** Tube cultures (batch growth) of 4 mL selective minimal medium (M9 + E4P supplements + 20 mM xylose) were inoculated to an OD600 of 0.05 (~1.5 × 10^7 cells) and monitored during prolonged incubation at 37 °C and 20% CO2 (up to 3 weeks). Several cultures reached OD600 values above 1.0 after ~2 weeks. Single colonies were isolated from these cultures by dilution streak from liquid cultures onto LB medium with chloramphenicol (to maintain the pGED plasmid). Individual clones were then re-assayed for immediate growth (observable OD600 increase within 48 h) on selective liquid minimal medium (M9 + E4P + Xylose + 20% CO2).

Genomic DNA was extracted using the GeneJET genomic DNA Purification Kit (ThermoScientific, Dreieich, Germany) from 2 × 10^9 cells of overnight culture in LB medium supplemented with chloramphenicol (to maintain the pGED plasmid). Construction of PCR-free libraries for single-nucleotide variant detection and generation of 150 bp paired-end reads on an Illumina HiSeq 3000 platform were performed by the Max-Planck Genome Centre (Cologne, Germany). Reads were mapped to the reference genome of *E. coli* MG1655 (GenBank accession no. U00096.3) using the software Geneious 8 (Biomatters, New Zealand). Using algorithms supplied by the software package, we identified single-nucleotide variants (with >50% prevalence in all mapped reads) and searched for regions with coverage deviating more than 2 standard deviations from the global median coverage. Confirmation of the pntA promoter locus in the Δtkt + pGED mutant was performed by Sanger Sequencing of a PCR product from amplification of the respective locus with high-fidelity Phusion Polymerase (ThermoScientific, Dreieich, Germany). Sanger sequencing was performed by LGC Genomics (Berlin, DE).
Expression analysis by reverse transcriptase quantitative PCR. In order to determine mRNA levels, total RNA was extracted from growing cells in the exponential phase (OD600 0.5–0.6) on M9 minimal medium with 20 mM carbon source (gluconate or xylose, and 44 µM supplements) in presence of 20% CO2. Total RNA was purified with the RNeasy Mini Kit (Qiagen, Hilden, Germany) as instructed by the manufacturer. In brief, ~5 × 10^8 cells (1 mL of OD600 0.5) were mixed with 2 volumes of RNAProtect Bacteria Reagent (Qiagen, Hilden, Germany) and treated as described above. The RNAs were measured in triplicates and each Michaelis–Menten curve was determined using at least 15 measurements.

Stable ^13C isotopic labeling of proteogenic amino acids. For isotope tracing, cells were cultured in a 3 mL M9 medium supplied with the labeled/unlabeled carbon sources described in the main text. For ^13CO2 labeling, the experiment was performed in a 10 L bioreactor that was first purged twice of the contained ambient air with a vacuum pump and refilled with an atmosphere of 80% air and 20% ^13CO2 (Cambridge Isotope Laboratories Inc., MA USA). All cultures were inoculated to OD600 of ~0.1, grown at 37 °C until the stationary phase. Then, 10^9 cells (1 mL of culture with OD600 = 1) were pelleted, washed once with ddH2O, and hydrolyzed in 1 mL hydrochloric acid (6 M) at 95 °C for a duration of 24 h. Subsequently, the acid was evaporated by heating at 95 °C and the hydrolyzed biomass was re-suspended in ddH2O. Hydrolyzed amino acids were separated using ultra-performance liquid chromatography (Acquity, Waters, Milford, MA, USA) using a C18-reversed-phase column (Waters, Eschborn, Germany). Mass spectra were acquired using an Exactive mass spectrometer (ThermoScientific, Dreieich, Germany). Data analysis was performed using Xcalibur (ThermoScientific, Dreieich, Germany). Prior to analysis, amino-acid standards (Merck, Darmstadt, Germany) were analyzed under the same conditions in order to determine typical retention times.

Purification and kinetic characterization of E. coli Gnd. Proteins were expressed from E. coli BL21-AI strains (Invitrogen) carrying appropriate plasmids for expression of E. coli Gnd or E. coli RpiA (ribose-5-phosphate isomerase), which were taken from the ASKA collection. Expression was induced overnight at 30 °C in LB medium (24 g/L tryptone, 12 g/L yeast extract, 12 g/L NaCl) supplemented with 100 µg/mL of kanamycin. The cell culture was dissolved with 1 mM of KHPO4, 72 mM K2HPO4 by addition of 0.5 mM IPTG and 2.5 mM arabinose upon reaching an OD600 of 1. Cells were lysed in 500 mM NaCl, 20 mM Tris-HCl pH 6.9 by sonication. After centrifugation (1 h at 30,000 × g), proteins were purified on an AKTA start system (GE Healthcare) by HitTrap Purification (GE Healthcare, Illinois, USA) as instructed by the manufacturer, using a wash step with 18% Buffer B (500 mM NaCl, 20 mM Tris-HCl pH 6.9, 500 mM imidazole). Desalting was performed in 100 mM NaCl, 20 mM Tris-HCl pH 6.9, and enzymes were stored at −20 °C in desalting buffer with 20% glycerol.

Kinetic assays were carried out on a Cary-60 UV–vis spectrometer (Agilent, Ratingen, Germany) at 30 °C using a 1 mm quartz cuvette (Hellma). All assays were carried out in 100 mM Tris-HCl buffer at pH 8 following consumption or production, respectively, of NADPH at 340 nm (ε = 6.2 cm⁻¹ μM⁻¹). The reductive carboxylation parameters for Gnd were determined with assays containing 2.4 mM NADPH, 16 mM ribose-5-phosphate, 1.5 M KHCO3, and 140 µM RpiA (with varying concentrations of the substrate under investigation). The assays were preincubated for 2 min and started with the addition of 750 nM of freshly diluted Gnd. Carbonic anhydrase was used to confirm that CO2 equilibration was not rate-limiting in these assays. Isomerization of ribose-5-phosphate to ribulose-5-phosphate was confirmed not to be rate-limiting. The concentration of ribulose-5-phosphate was calculated from the equilibrium constant of the isomerization reaction: K_θ = 0.458 (eQuilibrator; http://eQuilibrator.weizmann.ac.il) (9). The kinetic parameters of the oxidative decarboxylation were determined with assays containing either 800 µM NADP⁺ (for 6-phosphogluconate parameters) or 200 µM 6-phosphogluconate (for NADP + parameters). Assays were started with the addition of 7.5 mM Gnd. All assays were measured in triplicates and each Michaelis–Menten curve was determined using at least 15 measurements.

Determination of extracellular glucose concentrations via enzymatic assay. Glucose concentrations in supernatants of ΔPZΔF mutant cultures were determined using a commercial glucose oxidase-based assay kit following the manufacturer’s instructions (Merck, Darmstadt, Germany; Catalog No. GA020). In brief, three independent cultures each of ΔPZΔF mutant “B” and “C” were grown in 3 mL M9 medium with 20 mM xylose, and samples were taken in exponential phase (OD600 0.4–0.8) and in early stationary phase (OD600 = 1.1), centrifuged for 3 min at 20,000 g and supernatants frozen at −20 °C for later use. A standard curve with varying xylose concentrations confirmed negligible background signal from the xylose contained in the medium (20 mM xylose resulting in a signal corresponding to 0.163 mM (i.e., 0.029 mg/mL) glucose). The lower detection limit for glucose in the supernatant was thus assumed to be 0.17 mM, i.e., such concentrations or higher would be detected even in the case of complete consumption of all xylose in the media by the growing cells (i.e., an order of magnitude below relevant reported values for glucose excretion: 1.8 mM per unit increase in OD600) (1). Glucose standards were prepared in triplicate in the growth medium at the following concentrations (mg/mL): 0, 0.02, 0.04, and 0.08. Assays were performed by mixing 400 µL of standard or supernatant sample with 800 µL reagent mix, prepared following the manufacturer’s instructions, and incubated at 37 °C for 30 min. The reaction was stopped by adding 800 µL of sulfuric acid (6 M). Absorbance was measured at 540 nm and sample glucose concentrations determined by means of a standard curve. No glucose signal above background was detected in any culture sample.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this Article is available as a Supplementary Information file. The strains reported here are available from the corresponding authors upon request. The complete sequence of the pGEd plasmid has been deposited to GenBank under accession MW509923. Where indicated, data from the following public repositories were used for the findings in this study: KEGG (https://www.kegg.jp/); BiGG (http://bigg.ucsd.edu/); and eQuilibrator (http://eQuilibrator.weizmann.ac.il). Source data are provided with this paper.

Code availability. The code used in this study can be found at GitLab (https://gitlab.com/elad.noor/gedcycle) or Zenodo (https://doi.org/10.5281/zenodo.4066983).

Received: 23 October 2019; Accepted: 15 October 2020; Published online: 16 November 2020

References.

1. Hugler, M. & Sievert, S. M. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Annu. Rev. Mar. Sci. 3, 261–289 (2011).
2. Berg, I. A. Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. Appl. Environ. Microbiol. 77, 1925–1936 (2011).
3. Bar-Even, A., Noor, E. & Milo, R. A survey of carbon assimilation and bacterial diversity. ISME J. 4, 2109–2117 (2010).
4. Daims, H. & Schleper, C. Diversity of autotrophic microorganisms. Nat. Rev. Microbiol. 8, 57–66 (2010).
5. Sánchez-Andrea, I. et al. The reductive glycine pathway allows autotrophic growth of Desulfitobrio desulfuricans. Nat. Commun. 11, 5090 (2020).
6. Classens, N. J., Sousa, D. Z., Dos Santos, V. A., de Vos, W. M. & van der Oost, J. Harnessing the power of microbial autotrophy. Nat. Rev. Microbiol. 14, 692–706 (2016).
7. Satonowski, A. & Bar-Even, A. A one-carbon path for fixing CO2. EMBO Rep. 21, e50273 (2020).
8. Antorensky, N. et al. Sugar synthesis from CO2 in Escherichia coli. Cell 166, 115–126 (2016).
9. Gleitzer, S. et al. Conversion of Escherichia coli to generate all biomass carbon from CO2. Cell 179, 1255–1263 e1219 (2019).
10. Gassler, T. et al. The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2. Nat. Biotechnol. 38, 210–216 (2020).
11. Mattei, Z., Ziesack, M., Voges, J. M., Silver, P. A. & Way, J. C. Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: toward horizontal transfer of autotrophic growth. Metab. Eng. 16, 130–139 (2013).
12. Braakman, R. & Smith, E. The emergence and early evolution of biological carbon-fixation. Perspect. Biol. Med. 56, 1002455 (2012).
13. Erb, T. J. & Zarzycki, J. A short history of Rubisco: the rise and fall (?) of nature’s predominant CO2 fixing enzyme. Curr. Opin. Biotechnol. 49, 100–107 (2018).
Kwon, Y. D., Kwon, O. H., Lee, H. S. & Kim, P. The effect of NADP-
71. Stoffel, G. M. M. et al. Four amino acids define the CO₂ binding pocket of enoyl-CoA carboxylases/reductases. Proc. Natl Acad. Sci. USA 116, 13964–13969 (2019).
72. Bernhardsgrubter, I. et al. Awakening the sleeping carboxylase function of enzymes: engineering the natural CO₂-binding potential of reductases. J. Am. Chem. Soc. 141, 9778–9782 (2019).
73. Cotton, C. A., Edlich-Muth, C. & Bar-Even, A. Reinforcing carbon fixation: CO₂ reduction catalysts supporting carboxylation. Curr. Opin. Biotechnol. 49, 49–56 (2018).
74. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30 (2000).
75. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduek, D. R. COBRApy: COConstraints-based reconstruction and analysis for Python. BMC Syst. Biol. 7, 74 (2013).
76. Mendler, K. et al. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. Nucleic Acids Res. 47, 4442–4448 (2019).
77. Bizzozar, T., van Boxel, G. I., Bhakta, T. & Jackson, J. B. Nucleotide binding affinities of the intact proton-translocating transhydrogenase from Escherichia coli. Biochim. Biophys. Acta 1708, 404–410 (2005).
78. He, H., Höper, R., Dodenhöft, M., Marlière, P. & Bar-Even, A. An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in E. coli. Metab. Eng. 60, 1–13 (2020).
79. Gonzalez, J. E., Long, C. P. & Antoniewicz, M. R. Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by (13)C metabolic flux analysis. Metab. Eng. 39, 9–18 (2017).
80. Jensen, S. L., Lennen, R. M., Herrgard, M. J. & Nielsen, A. T. Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress. Sci. Rep. 5, 17874 (2015).
81. Thomason, L. C., Costantino, N. & Court, D. L. E. coli genome manipulation by PI transduction. Curr. Protoc. Mol. Biol. Chapter 1, 17 (2007).
82. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 2006–2008 (2006).
83. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl Acad. Sci. USA 97, 6640–6645 (2000).
84. Braatsch, S., Helmark, S., Kranz, H., Koebmann, B. & Jensen, P. R. Escherichia coli strains with promoter libraries constructed by Red/ET recombination pave the way for transcriptional fine-tuning. Biotechniques 45, 335–337 (2008).
85. Zelbuch, L. et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res. 41, e98 (2013).
86. Hayashi, K. et al. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. Mol. Syst. Biol. 2, 0007 (2006).
87. Rocha, D. J., Santos, C. S. & Pacheco, L. G. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. Antonie Van Leeuwenhoek 108, 685–693 (2015).
88. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25, 402–408 (2001).
89. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C T method. Nat. Protoc. 3, 1101 (2008).
90. Giavalisco, P. et al. Elemental formula annotation of polar and lipophilic metabolites using 15C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. Plant J. 68, 364–376 (2011).
91. Kitagawa, M. et al. Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res. 12, 291–299 (2005).
92. Krusemann, J. L. et al. Artificial pathway emergence in central metabolism from three recursive phosphoketolase reactions. FEBS J. 285, 4367–4377 (2018).

Acknowledgements
We thank Anne Michaels for assistance with LC-MS analysis of amino acids; Nicole Piazzi, Stefano Donati, and Hannes Link for metabolite analysis; Selcuk Aslan for assistance with expression analysis; Lorenz Heck for assistance in molecular biology work; and Charlie Cotton and Nico Claessen for critical reading of the manuscript. This study was funded by the Max Planck Society.

Author contributions
S.N.L. and A.B.-E. conceived the study. A.S., B.D., S.N.L., and A.B.-E. designed the experiments. A.S., B.D., P.W., and S.N.L. performed the in vivo experiments. E.N. and H. H. performed the in silico experiments. B.V. and T.E. performed the in vitro experiments. A.S., B.D., S.N.L., and A.B.-E. analyzed the results. A.S., B.D., E.N., S.N.L., and A.B.-E. wrote the manuscript with contributions from all authors.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-19564-5.
Correspondence and requests for materials should be addressed to S.N.L.

Peer review information Nature Communications thanks Rogier Braakman, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.