Trends in Biotechnology

CellPress OPEN ACCESS

Review

Escherichia coli selection strains for growth-coupled metabolic engineering

Helena Schulz-Mirbach [1,*], Beau Dronsella [1], and Tobias J. Erb [1,2,*]

Synthetic metabolism has the potential to transform carbon capture, bioremediation, or bioproduction strategies. To transfer metabolic designs from an *in vitro* context to living model systems such as the bacterium *Escherichia coli*, metabolic engineers incentivize the maintenance and use of the introduced metabolic module by making cell survival dependent on it (growth-coupled selection). However, creating and characterizing appropriately rewired selection strains is nontrivial and requires labor-intensive growth phenotyping in various conditions. To enhance the community use of extant selection strains, we compiled designs covering the central, amino acid, and energy metabolism of *E. coli* for this review, and we revisit the key concepts of growth-coupled selection.

Highlights

'Designer' metabolism could enhance carbon capture, bioremediation, and bioproductions.

Transferring synthetic metabolism into living cells is challenging.

Coupling cell growth to the pathway of interest in metabolically rewired selection strains advances the implementation of synthetic metabolism.

Growth rates and biomass yields can be used to approximate pathway turnover or compare pathway efficiencies.

A selection strain must be thoroughly validated before use.

Selection strains covering *E. coli*'s central, amino acid, and energy metabolism are readily available in the community.

Growth-coupled selection brings synthetic metabolism to life

Synthetic biology aims at developing biological systems and organisms with improved or new functions compared with their naturally evolved counterparts. These (re)designed systems provide new opportunities in health and medicine, food production, environmental protection, and sustainable manufacturing [1–3]. At the core of many of these efforts are 'designer' metabolic pathways that include new-to-nature reactions and synthetic designs, which allow access to new feed streams and resources and/or produce new compounds [4]. Prominent applications of synthetic metabolism are the assimilation of greenhouse gases and one-carbon (C1) feedstocks, such as CO_2, methane, methanol or formate [1–3], and their conversion into commodity chemicals [5–7] or the degradation of environmentally harmful substances [8,9].

One crucial step in realizing new-to-nature enzymes and synthetic pathways is their successful implementation in the context of the living cell. This is typically achieved by creating a cellular dependence (i.e., an 'auxotrophy') of the host of interest on the novel designs. In other words, the corresponding host cell (i.e., the 'auxotroph') is forced to rely on the newly introduced enzyme or metabolic pathway (segment) for growth and survival (Figure 1). In such growth-coupled selections, the growth phenotype is a readout for the flux through the synthetic pathway of interest, where ideally growth rate correlates with pathway performance [10,11].

Growth-coupled selections have been used successfully in the recent past. For example, several metabolic designs for greenhouse gas and C1 use have been drafted and prototyped *in vitro* [12–14], and some pathways have been either fully or partially transplanted *in vivo* [2,15–18] through the help of growth-coupled selections. This has allowed engineering the heterotrophic microbial platform *Escherichia coli* to use C1 molecules as sole feedstock, namely via transplantation of the natural Calvin Benson-Bassham (CBB) [15] and ribulose monophosphate cycles [2], as well as the implementation of the synthetic reductive glycine pathway [17] and serine threonine cycle [19].

In the above examples, growth-coupled selection not only was used as a tool for *in vivo* demonstration of the feasibility of a design but also was, in many cases, necessary to realize the design

[1]Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany
[2]Center for Synthetic Microbiology (SYNMIKRO), Karl-von-Frisch-Straße 14, 35043 Marburg, Germany

*Correspondence:
helena.schulz-mirbach@mpi-marburg.mpg.de (H. Schulz-Mirbach) and
toerb@mpi-marburg.mpg.de (T.J. Erb).

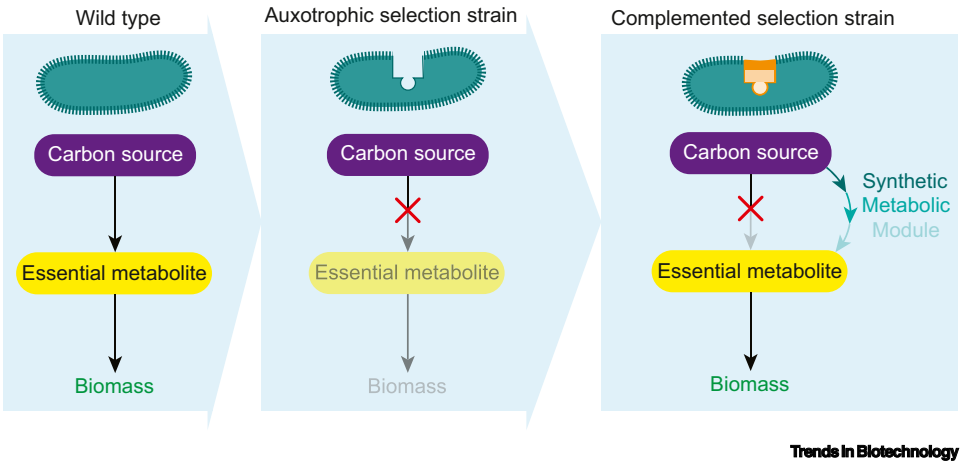

Figure 1. Growth-coupled selection for synthetic metabolic modules. A selection strain is auxotrophic for certain biomass precursors and can be complemented by specific metabolic modules.

in vivo in the first place. Thus, the selection strains did not directly show growth but had to be subjected to prolonged (or fluctuating) selecting conditions and were used to screen for mutants that exhibited pathway-dependent growth. By identifying the underlying mutations and resulting changes in metabolic fluxes, bottlenecks in the metabolic architecture of the host could be retrospectively rationalized. Such 'alleviating' or 'enabling' mutations typically include removal of intracellular sinks for pathway intermediates [20] or elimination of promiscuous activities of native enzymes interfering with intermediates of the newly introduced pathway [21]. Because such incompatibilities are rarely annotated in the databases, let alone known beforehand, they could neither rationally nor computationally have been predicted. Thus, selection strains can serve as a convenient 'black box' (untargeted) approach for metabolic engineering and thus allow a much larger (and unbiased) exploration of solution space than fully rational metabolic engineering strategies.

As the above examples show, growth-coupled selection is a powerful strategy to address the complexity of metabolic pathway engineering. At the same time, it is precisely the complexity of metabolism that makes selection strain construction a labor-intensive and complicated process. Being able to predict the deletions required to create auxotrophies requires detailed knowledge of the underlying native metabolic network of the host cell. This knowledge is in principle accessible via databases such as KEGG [22] or EcoCyc [23] and can be incorporated into metabolic models, which allow the computational prediction of deletions required to create an auxotrophy [24–27]. Such predictions become more reliable the better the host's metabolism is known, which explains why most engineering efforts have been successfully reported in the well-characterized model organisms, such as *E. coli*.

However, even in well-established models, metabolic bypasses can easily emerge (or evolve), which can alleviate (or 'break') a given auxotrophy upon longer incubation [28–30]. Even for the best-studied model systems, our metabolic knowledge does not suffice to reliably rule out the emergence of metabolic bypasses [31]. Therefore, the construction of selection strains still is an iterative process that requires in-depth characterization, testing, and further refinement under selection conditions. Although previous efforts to engineer synthetic metabolism in *E. coli* resulted in a broad array of selection strains that are ready to use, finding those selection strains requires a thorough literature search. To make growth-coupled selection more accessible for future metabolic engineers, we provide a comprehensive overview of *E. coli* selection strains

constructed in recent years, covering central, amino acid, and energy metabolism (Figure 2; Tables S1–S7 and Figure S1 in the supplemental information online).

Strain design and validation

Database search and computational predictions allow identifying relevant deletion targets

Growth-coupled selection relies on a connection between the pathway of interest and the host metabolism. Therefore, the first step in creating a selection strain is to identify overlaps (or 'interfaces') between the synthetic pathway and the host metabolism.

Generally, a conceptual difference has to be made between growth-coupled production strategies and growth coupling for *in vivo* pathway engineering (which this review focuses on). For production, it is of interest to minimize biomass formation while maintaining a link between cell survival and maximal product yield [32]. This strategy is most straightforward for pathways entailing lyase reactions, in which two metabolites are formed in stoichiometric amounts [33,34] but can also be achieved by coupling product secretion to cofactor regeneration [35,36]. For pathway engineering, the objective is to create a direct coupling between pathway performance and cell growth without the condition for the latter to be minimal [10,37].

In the simplest case for the latter strategy, the product of the synthetic pathway of interest is already an essential intermediate of the host organism (Figure 1). To select for the activity of the entire pathway, only the pathway product (and no other pathway intermediate upstream) is connected to the host metabolism. If this is not possible (i.e., upstream metabolites break the selection scheme), these pathway intermediates need to be metabolically isolated from the remaining metabolism to ensure that the complete synthetic pathway is under selection. To isolate a given intermediate from the remaining metabolism, all metabolic pathways allowing its biosynthesis can be identified through the help of databases that cover the host's metabolism. The databases KEGG [2024-07-11 covering 10 076 genomes (eukaryotes 1054; bacteria: 8585; archaea: 437)], MetaCyc [2024-07-11 representing >1900 organisms (3153 pathways, 19 020 reactions, and 19 372 metabolites)] [22], and BioCyc (2024-07-11 covering 20 000 organisms and various subdatabases) [38] allow filtering for an organism of interest. EcoCyc is the *E. coli* K12 MG1655–specific part of the BioCyc databases and covers information on the organism's metabolome, proteome, genome, gene essentiality, and regulation, which was curated from more than 44 142 publications (February 2024) [23].

Genes involved in the synthesis pathways leading to the auxotrophy target metabolite, ideally those with the least number of isozymes, are then deleted to create the auxotrophic selection strain (Figure 3). Given the size of such databases, metabolic engineers increasingly transition from manual selection identification to employing algorithms such as those used by Feist *et al.* [39] or Schneider *et al.* [40] to couple growth to product formation, Antonowski *et al.* to find knockouts coupling growth to a reaction of interest [24], OptAux (finding knockouts coupling growth to the presence of a metabolite) [41], or OptEnvelope (finding minimal knockout strategies for growth coupling) [42], which can assist the search process and propose deletion combinations to create target auxotrophies [2,31,43] (Figure 3, panel 1). Auxotrophies are here not limited to metabolites but can also concern cofactors [35], which is a strategy that has especially been pursued for growth-coupled biotechnological productions [44]. To limit the computational workload, core metabolic models covering only central metabolism have mostly been employed as a search base for OptKnock, which, however, limits the solution space and overlooks amino acid or cofactor metabolism.

This limitation was recently tackled by Corrao *et al.* [45] with their medium-scale iCH360 model, which was manually curated to reduce the highly complex genome-scale iML1515 model to

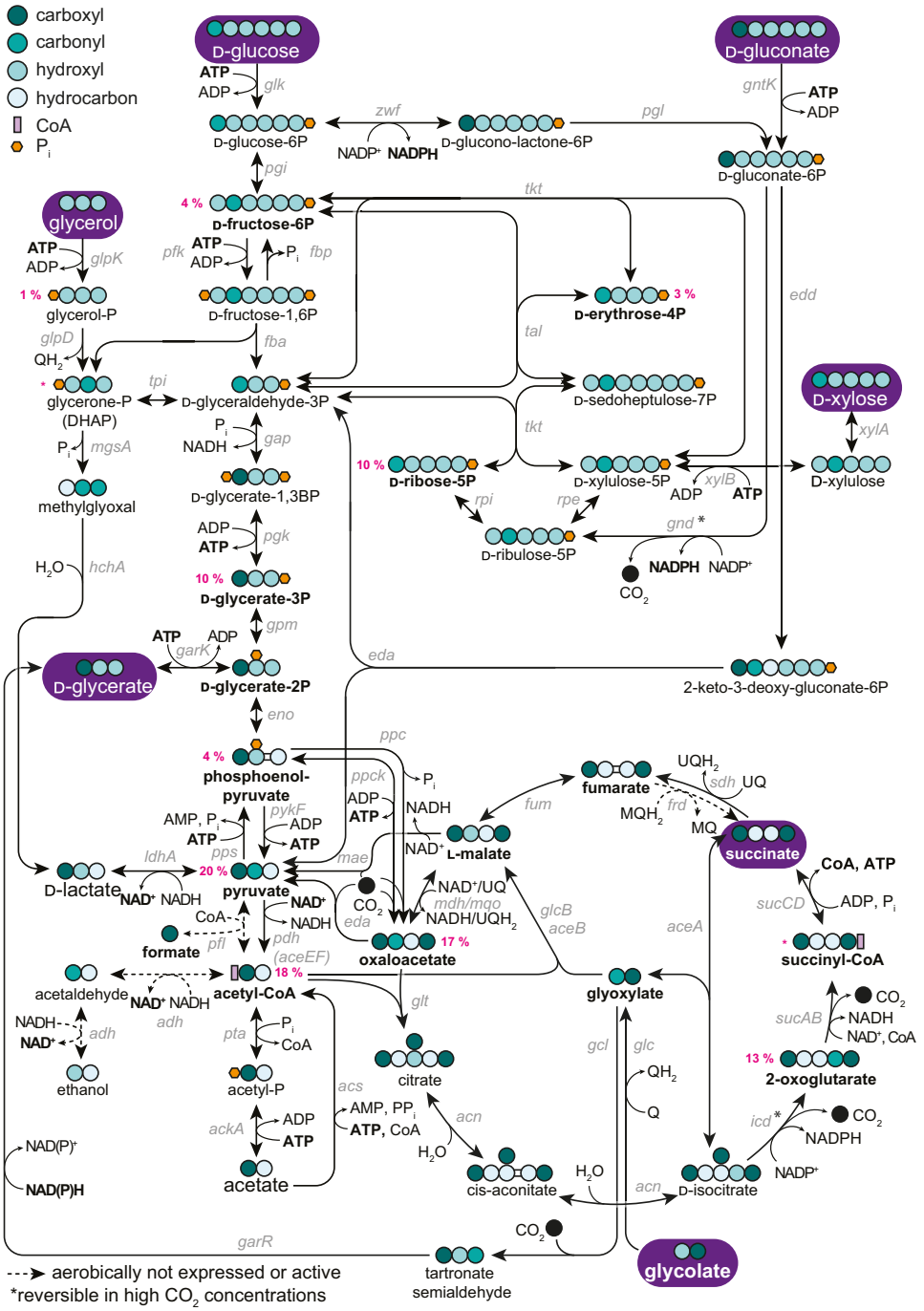

Figure 2. Updated central metabolic scheme for *Escherichia coli*. Adapted, with permission, from Noor *et al.* [84]. Metabolites for which selections were created are shown in bold. Commonly fed C-sources have a violet box. Pink numbers indicate biomass percentages derived from the intermediate, inferred from demand values in μmol/g dry weight provided for each metabolite by Neidhardt *et al.* [85], normalized to the number of carbons (done by Noor *et al.* [84]). Note that glucose-6-phosphate was previously denoted as essential but is in fact not strictly needed for lipopolysaccharide synthesis and thus *E. coli* survival.

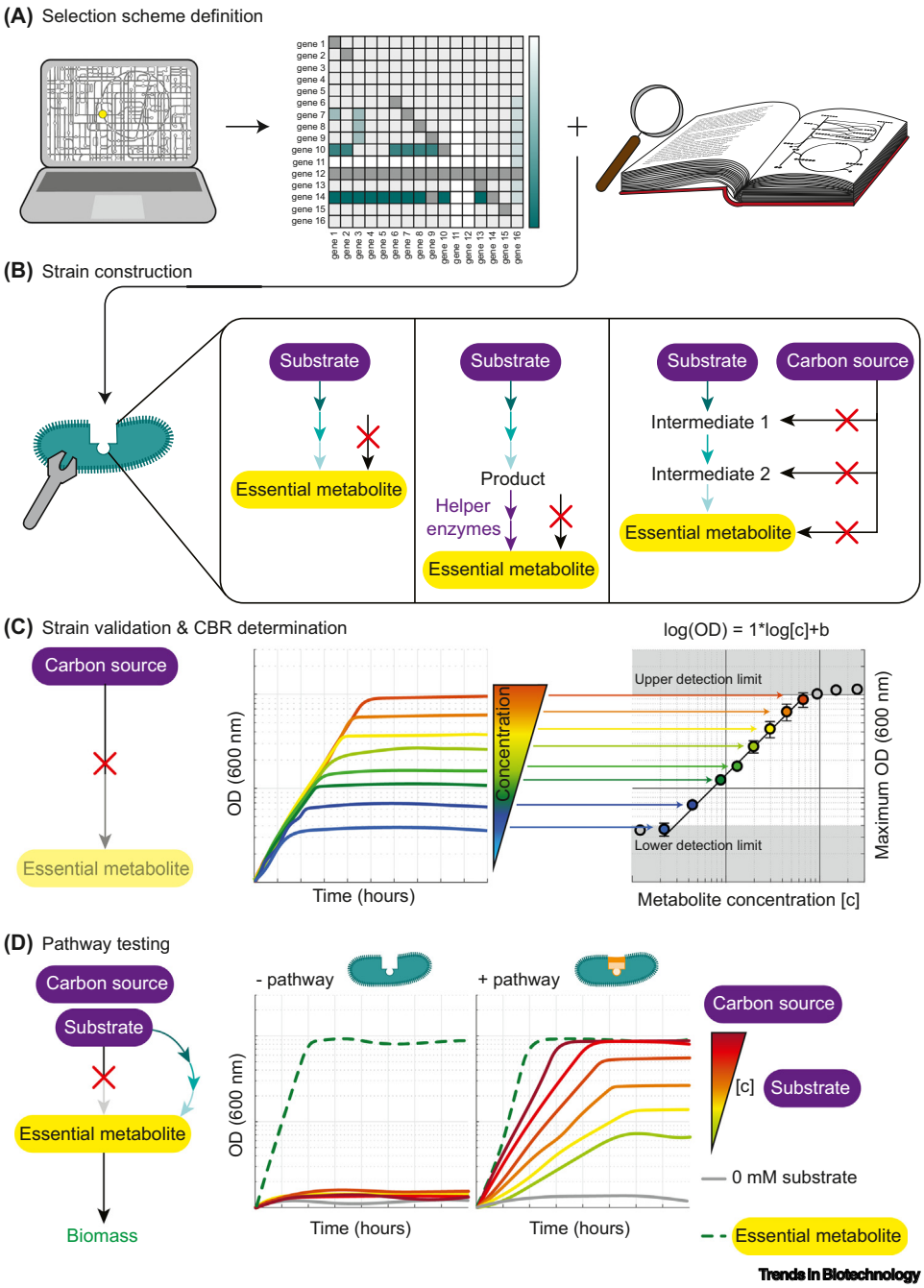

Figure 3. Workflow for selection identification and strain characterization. (A) Selection schemes can be determined by combined computational search and literature screening. (B) Ideally, the product of the synthetic pathway to be tested is an essential metabolite of the host (left). If that is not the case, non-native 'helper enzymes' can facilitate the conversion of product to a native metabolite (middle). To select for an entire pathway, all intermediates need to be metabolically isolated (right). (C) As the first part of strain characterization, dependence on the auxotrophy target should be verified, followed by determining the compound biomass ratio (CBR) and thus the demand of a selection experimentally. The selection strain is cultivated with several substrate concentrations to determine the maximum OD achieved with each. The equation describing the linear dependency of log(biomass yields) on the log(substrate concentration) (right) can be rearranged to determine the CBR = exp(–b). (D) Then, the pathway of interest can be tested, where growth on the pathway substrate should occur only in the presence of the pathway genes and should be substrate concentration-dependent.

central metabolism as well as amino acid biosynthesis and energy metabolism. Although the model was already employed to create a set of glyoxylate/glycolate sensors [46], the workflow still required manual model curation and thus some bioinformatic understanding. Making metabolic modeling accessible to scientists with little to no informatics skillset, for example, by combining analysis tools with an easy-to-use graphical user interface would likely enhance the relevance of metabolic models for metabolic engineering efforts. However, the creation of such end-user–friendly platforms requires effort and constant maintenance, which is currently less rewarded in the scientific community.

All efforts to identify suitable selection schemes are limited by how far the target host has been genetically characterized (thus how reliably deletion targets can be identified). For example, the number of uncharacterized proteins and genetic redundancies is still impairing metabolic engineering of specialized microbes such as the CO_2- and H_2-consuming 'Knallgas' bacterium *Cupriavidus necator,* the plastic degrading microbe *Ideonella sakaiensis,* or the methylotroph *Methylorubrum extorquens*. Additionally, because the current computational models do not take complex regulatory patterns into account, even in *E. coli*, every selection scheme requires complementation by literature search, experimental validation, and, if needed, further adaptation and fine-tuning (additional knockouts, etc.).

In some cases, the synthetic route is completely orthogonal to the host metabolism – in this case, a link can be artificially created by adding non-native 'helper enzymes' to the pathway of interest, which connect an essential biomass component to the pathway product (Figure 3, panel 2). For example, β-alanine and thus coenzyme A (CoA) auxotrophic strains were used to demonstrate the 3-hydroxypropionyl-CoA–forming activity of a novel lactyl-CoA mutase *in vivo* [47]. Because 3-hydroxypropionyl-CoA is not an intermediate of the native metabolism of *E. coli*, a CoA transferase, a 3-hydroxypropionate dehydrogenase, and a β-alanine:oxaloacetate transaminase were heterologously expressed as helper enzymes, allowing the formation of β-alanine from 3-hydroxypropionyl-CoA. Until recently, the computational prediction of selection schemes employing helper enzymes was highly limited because non-native metabolic reactions had to be manually added to the respective metabolic model. Alter *et al.* [25] tackled this problem by defining enzyme groups according to substrate-product pair similarity and incorporating them in the iML1515 model of the metabolism of *E. coli* to build a heterologous reaction database.

Strain construction and thorough selection validation

Metabolic interventions causing specific auxotrophies are canonically achieved by gene deletions [16], which can isolate either individual metabolites or entire metabolite pools [31]. In *E. coli*, the λ-red recombineering system is frequently employed to generate such deletions, such as with the SIJ488 strain as engineering background [48]. However, the system comes at the cost of flippase recognition site scars in the strain genome, which can result in undesired reshuffling when chaining many deletions. Scarless methods such as CRISPR or base editors [49–51] circumvent such reshuffling but require an initial setup in the target strain background. Inducible knockdown systems such as CRISPRi are currently gaining attention for their ability to allow high-throughput strain modifications [52] but need to be robust to maintain selection pressure over long periods of time to be employed for selection strain creation and adaptive laboratory evolution.

To experimentally validate a selection scheme requires one first to confirm the expected behavior in the target conditions. First, the selection demand, measured in millimoles of the auxotrophy target required per gram of cell dry weight [often inferred from optical density $(OD)_{600}$ measurements], is to be determined (Figure 3, panel 3). By knowing the demand, the biomass yields achieved by such a complemented auxotroph harboring the pathway of interest can be used

as a proxy to estimate the required total pathway product formation *in vivo*. To determine the selection demand [compound biomass ratio (CBR)], the auxotroph is cultivated with varying concentrations of the auxotrophy target (Figure 3, Box 1). Here, increases in growth rate, which are proportional to the substrate concentration, could indicate limitations in substrate uptake (e.g., transporter affinities), whereas decreasing growth rates or lag phases at high substrate concentrations could indicate toxicity.

Generally, a low CBR indicates a highly sensitive selection in which typically only small biomass fractions need to be synthesized from the supplied compound. In contrast, high CBR selections are often auxotrophic for 'entire pathways' of metabolism [e.g., the tricarboxylic acid cycle, glycolysis intermediates, amino acid groups synthesized from one intermediate (such as the glutamate family amino acids being made from 2-oxoglutarate), pentose phosphate shunt intermediates]. Hence, the smaller the metabolic fraction made from a metabolite is, the more sensitive the selection strain will be. Thus, removing metabolic reactions using the auxotroph target metabolite ('sink' reactions) by deleting the corresponding genes and supplying the downstream metabolites to the medium is a common strategy to increase selection sensitivities (Figure 4A). The workflow of (computational) selection definition based on database search and subsequent strain characterization has so far been pursued in two studies that yielded entire arrays of glycerate [31] or glyoxylate biosensors [46] with varying selective demands.

Growth-based pathway testing

In the next step, the pathway of interest can be introduced. Because many synthetic metabolic designs are highly complex, they are typically broken down into smaller (selection) 'modules,' which are first tested individually [16,53]. In case of module dependence, growth should occur

Box 1. The CBR describes the demand of a selection

Plotting the maximum OD versus concentration of the auxotrophic target in log scale for both axes will yield a sigmoidal curve (see Figure 3 in the main text) with a linear part, in which biomass formation and substrate concentration are correlated. Ideally, the slope of this linear part is 1, indicating that biomass yield is linearly correlated with the amount of substrate provided and not limited by uptake, toxicity, or substrate sinks. The linear function describing the correlation between biomass (indicated by OD) and substrate concentration (c) in log-log format can be used to infer the selective demand, CBR, on the basis of which pathway output can be approximated from the OD:

$$log_{10}(OD) = log_{10}(c) + b \quad [I]$$

$$-b = log_{10}(c) - log_{10}(OD) = log_{10}\left(\frac{c}{OD}\right) \quad [II]$$

$$10^{-b} = \frac{c}{OD} = CBR \quad [III]$$

With c as compound concentration in mM, OD as standard OD at 600 nm, and *b* as the curve offset.

With the assumption that OD 1 (10^8 cells) corresponds to 0.36 g dry weight of cells (CDW) per liter [86] (BNID 109837), CBR can be normalized to CDW:

$$CBR = \frac{c}{OD} = \frac{c}{OD\ 0.36}\left[\frac{\frac{mmol}{L}}{\frac{g\ CDW}{L}}\right] = \frac{10^{-b}}{0.36}\left[\frac{mmol}{gCDW}\right] \quad [IV]$$

As an example, for an acetyl-CoA auxotrophic strain, a CBR of 1.47 mmol acetate per OD (4.08 mmol acetate per g of CDW) was determined experimentally, which means that making 1 g of CDW requires 0.24 g of acetate (inferred with the molecular weight of acetate being 59.04 g/mol). If it is a carbon-only selection and no acetate is burned to gain energy, that would mean that about 24% of the biomass is dependent on the compound. If a strain's CBR cannot be determined experimentally (e.g., prior to strain construction), intracellular metabolite concentrations (e.g., Bennett *et al.* [87]) can give a first clue of the compound quantities needed for steady-state cell maintenance.

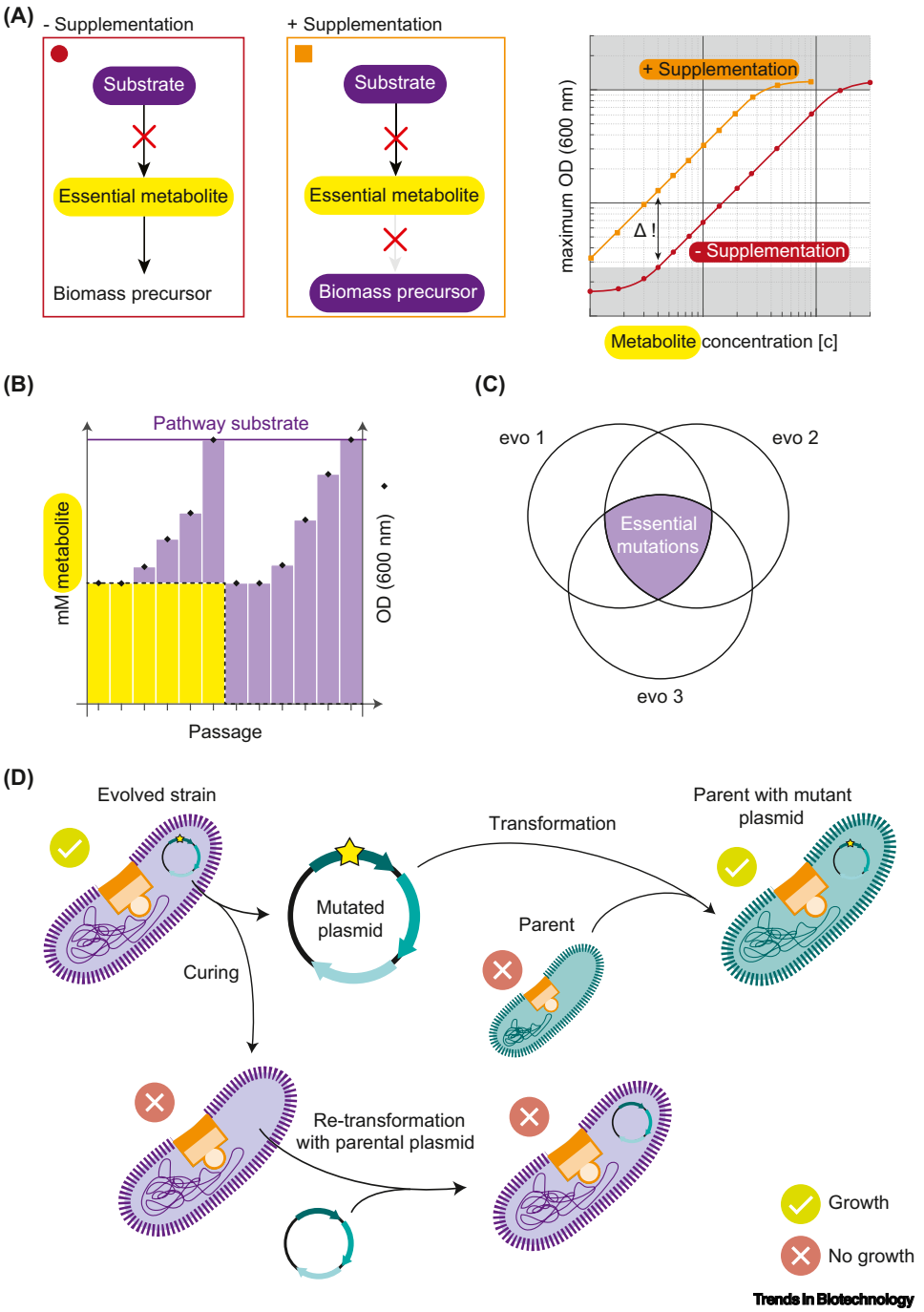

Figure 4. Supplementing biomass precursors made from the metabolite of interest can increase the selection sensitivity. (A) Left: Selection without additional supplementation. All of the auxotrophy target needs to be converted to biomass precursors. Middle: By additionally supplementing biomass precursors, the biomass fraction made from the auxotrophy target decreases. Right: Mock graph of how the same strain with and without supplementation behaves, depending on the metabolite concentration. The Δ indicates the increasing selection sensitivity upon precursor addition. (B) Scheme of a semirelaxing evolution in which a limiting concentration of auxotrophy complementing metabolite (dotted line) is supplied initially together with an unlimiting amount of pathway substrate (violet line). At first, all biomass (indicated

(Figure legend continued at the bottom of the next page.)

only if (i) all module enzymes are sufficiently expressed, (ii) the module substrate and all required cofactors are present, and (iii) the individual enzymes are sufficiently active. If basal growth via the module is observed, it can be improved using adaptive laboratory evolution [54]. This 'stepwise' testing – using multiple selection strains – increases the chances of identifying individual bottlenecks within a pathway segment and thus ultimately facilitates successful engineering of the entire pathway of interest.

Once all individual modules have been demonstrated to be functional, the entire pathway can be assembled. To exclude unanticipated auxotrophy complementation by the intermediates of the synthetic pathway of interest, an initial 'no growth' verification should be done when providing each of the pathway intermediates while not expressing the pathway genes. Upon introducing all pathway genes, the most thorough approach would be to repeat the selection strain characterization to see whether jointly expressing the pathway enzymes interferes with native metabolism in any way by promiscuously acting on endogenous metabolites. Then, the entire pathway can be tested by growing the selection strain carrying all pathway genes with the pathway substrate. As for individual modules, growth should occur only if all pathway genes are expressed and the respective enzymes convert the pathway substrate into the auxotrophy target metabolite and not if the pathway substrate is omitted (and no alternate synthesis pathway is provided) (Figure 3, panel 4).

Available selection strains

To maximize the use of previously constructed *E. coli* selection strains for the community, we provide a comprehensive overview of published designs, their cultivation conditions, and the relevant author contact information in Tables S1–S7 in the supplemental information online.

We grouped the designs by 'metabolic areas' as defined in Figure S1 in the supplemental information online. For each selection strain, we summarize the underlying gene deletions, cultivation conditions, the 'main' auxotrophy (i.e., the trait selected for), and 'additional' auxotrophies (i.e., secondary auxotrophies from strain construction, supporting the main selection and/or expanding/tuning selection sensitivity or selection strength), as well as the corresponding authors these strains can be requested from (Tables S1–S7 in the supplemental information online). Because many designs isolate entire metabolic areas, they can in principle be used to select for a variety of compounds. Therefore, we provide a matrix with all biomass precursors that are dependent on the respective metabolic intermediate.

The core challenge of a growth-based readout: too little pathway flux for growth

One of the major problems of growth-coupled selection is that not all pathways initially support selection strain growth. If growth of a strain via a pathway does not occur, this does not equal absence of pathway activity. Potentially, the pathway is active, but the flux does not suffice to satisfy the selection demand and thus support growth. In this case, depending on the synthetic pathway product, transcription-based sensors can allow screening for lower product concentrations in a high-throughput fashion. Here, by transcribing a fluorescent protein in a pathway product-dependent manner and sorting cells by fluorescence, libraries of enzyme candidates can be screened for improved activity. Either these libraries can be generated *a priori* to be transformed

by bars) is made from the rescue metabolite (yellow fractions), but over time, increasing fractions are made from the pathway substrate (violet fractions). (C) Identifying the essential set of mutations that are shared between independently evolved strain isolates can help in understanding how growth evolved. (D) Workflow to check whether mutations on the plasmid carrying pathway genes are essential to allow growth. Growth and no-growth indications are given for the case that plasmid mutations are allowing pathway-dependent growth.

into cells for one screening round or the pathway enzymes can be mutagenized *in vivo. In vitro*, if the active site residues involved in catalysis are known, they can be modified in a targeted manner by site saturation mutagenesis. Alternatively, untargeted mutagenesis can be performed by error-prone PCR. *In vivo*, multiplex automated genome engineering can be used to modify individual residues, whereas inducible mutagenesis systems can be employed for multiple rounds of either targeted mutagenesis (e.g., employing the MutaT7 system) [55] or global mutagenesis [using hypermutators such as MP6 (abbreviating mutagenesis plasmid 6, containing the engineered DNA polymerase variant III, mutagenesis enhancing enzymes such as cytosine deaminase, and components reducing mutagenesis repair)] [56] (see Note S1 in the supplemental information online for mutation rate comparisons). Iterative mutagenesis and fluorescence-based cell sorting can be used to enrich improved enzyme variants as an alternative to adaptive laboratory evolution (instead of selecting for growth).

Another potential strategy in case of apparent lack of growth is supplying downstream metabolites to increase selection sensitivity (e.g., supplying tyrosine to an E4P auxotrophic strain; Figure 4A). Alternatively, switching to a more sensitive selection strain (with a lower CBR) could solve the problem and lead to pathway-dependent growth. If this is not the case, 'semirelaxing' adaptive laboratory evolution can allow generating pathway variants supporting sufficient flux for growth. During this cultivation, a limiting concentration of the substrate relieving the strain's auxotrophy (termed 'relaxing substrate' following the nomenclature of swapostat cultivations in which an auxotrophy resolving 'relaxing medium' is fed in alternation with a 'stressing medium' [57]) is supplied along with the substrate of the target pathway (Figure 4B). On the basis of CBR, the expected OD with the amount of relaxing substrate fed can be estimated. Any increases in OD beyond this expectation could indicate the use of the synthetic metabolic module for the synthesis of a fraction of biomass. As observed when evolving autotrophic growth of *E. coli via* the Calvin–Benson–Bassham cycle (CBB), cultivating the strain with a limiting amount of relaxing substrate (xylose in case of the CBB) over a prolonged time allowed growth in selective conditions (formate and CO_2 for the CBB). The minimum set of mutations allowing growth can be identified by whole-genome sequencing of several independently evolved isolates and identifying their commonalities (Figure 4C), as shown by Ben-Nissan and Milstein *et al.* for the CBB-dependent *E. coli* [20].

A blessing and a curse: metabolic flexibility and metabolic bypasses

Metabolic flexibility and enzyme promiscuity are the main advantages of *in vivo* optimization, because unpredictable adaptations allowing pathway-dependent growth can be obtained during adaptive laboratory evolutions and characterized retrospectively [19,20,58]. The activation of such 'dormant' metabolic activities or promiscuous activities of native enzymes during evolution can be of benefit for the pathway of interest, as seen for the GED (Gnd-Entner-Doudoroff) cycle (a CO_2 fixation cycle composed of entirely native metabolic activities [21]) or the THETA cycle (relying on the native succinate dehydrogenase promiscuously converting methylsuccinate to mesaconate [58]). However, such underground metabolism can also bypass the selection, resulting in pathway-independent growth [28,59]. For example, once overexpressed through mutations, the promiscuous activity of a succinate semialdehyde dehydrogenase complemented deletions of erythrose-4-phosphate dehydrogenase or phosphoglycerate kinase and thus bypassed the respective selections upon prolonged strain incubation [29]. Beyond metabolic bypasses arising during a semi-relaxing evolution, obtained growth improvements could also be caused by higher uptake rates for the relaxing substrate or a reduction of metabolic sinks for it and so further characterising obtained growth is essential.

Verifying pathway dependency is essential after evolving growth

Because of metabolic plasticity, growth obtained upon pathway overexpression and especially after adaptive laboratory evolution should be confirmed to be pathway dependent by (i) curing

the strain of the pathway genes and demonstrating absence of growth (Figure 4D) and (ii) isotopic labeling experiments. For the latter, isotopically labeled substrates are fed to the strain – because the amino acid biosynthesis routes are known in *E. coli*, tracing the occurrence of labeled carbons in amino acids allows understanding which central metabolism intermediates are metabolically connected to the fed substrate (Figure S2 in the supplemental information online). Further engineering can narrow down the origin of the growth improvements: if transforming the unevolved parent with the plasmid harboring the pathway genes of the evolved clones allows growth, and *vice versa* introducing the unevolved pathway genes to cured mutants does not allow growth, there likely are pathway-specific mutations involved (Figure 4D). If the unevolved parent cannot grow with the pathway genes from the evolved background, whereas the cured mutants grow when transformed with the parental plasmid, adaptations of the surrounding metabolic architecture apparently enabled growth. In addition to whole-genome sequencing, proteomics of evolved strains compared with the unevolved parent can give mechanistic insight into the metabolic changes allowing cell growth. However, the retransformations of parental strains with mutant plasmids and subsequent growth assays can allow identifying copy number changes, such as by plasmid multimerization [60], which can be hard to determine with whole-genome sequencing.

Harnessing native metabolism or reducing interference?

Two opposing dogmas in the field are that (i) pathways with a high degree of overlap with native metabolism have higher success changes because many of the reactions 'work' already [58] or that (ii) highly orthogonal designs are better suited, because they reduce interference with essential native metabolic fluxes [61]. On the basis of recent works in the field, it seems that exploiting pathway overlap with native metabolism supports novel designs *in vivo* [5,18–20,58] and that endogenous fluxes can be rewired to support novel architectures. However, especially with cyclic CO_2 fixation pathways, mutations tend to emerge at efflux nodes for cycle intermediates and native metabolism, because flux needs to be carefully balanced between cycle operation and cycle efflux, preventing unbalanced intermediate drainage and thus cycle instability [5,20]. This stability problem can be bypassed, to some extent, by linear metabolic designs such as the reductive glycine pathway [17]. Nonetheless, even here, operation of the pathway relies on native metabolic enzymes such as the glycine cleavage system in case of the reductive glycine pathway, where modified expression levels could hamper endogenous metabolism.

Reducing such cross-connections is the main argument for orthogonal designs such as the FORCE [61] or ReForm [62] pathways. However, although those have been demonstrated *in vitro*, they could not yet be engineered *in vivo* by growth coupling. Because even those orthogonal designs are still employing the same cofactor pools as natural metabolism, their driving force *in vivo* is still governed by intracellular metabolite pool sizes. In the future, harnessing noncanonical cofactors could allow operating synthetic metabolism entirely separately from host metabolite pools and thus could be less constrained by physiological metabolite concentrations [63,64].

Robust pathway activity readouts allow high-throughput metabolic engineering

In principle, growth rates of a selection strain should correlate with the specific activity of the rescuing metabolic module if nothing else is rate-limiting (Figure 5). This correlation can be exploited for *in vivo* pathway comparisons and selection of better enzyme variants (Figure 5B, panel i), as was done for various methanol dehydrogenases using a set of formaldehyde sensors [65]. From the correlation between growth rate, yield, and substrate uptake rates, the specific activity of the growth-limiting pathway (or enzyme) can be approximated in a back-of-the envelope calculation (Box 2; Note S2 in the supplemental information online). However, if either substrate uptake (Figure 5B, panel ii) or downstream conversions into essential metabolites (Figure 5B, panel iii)

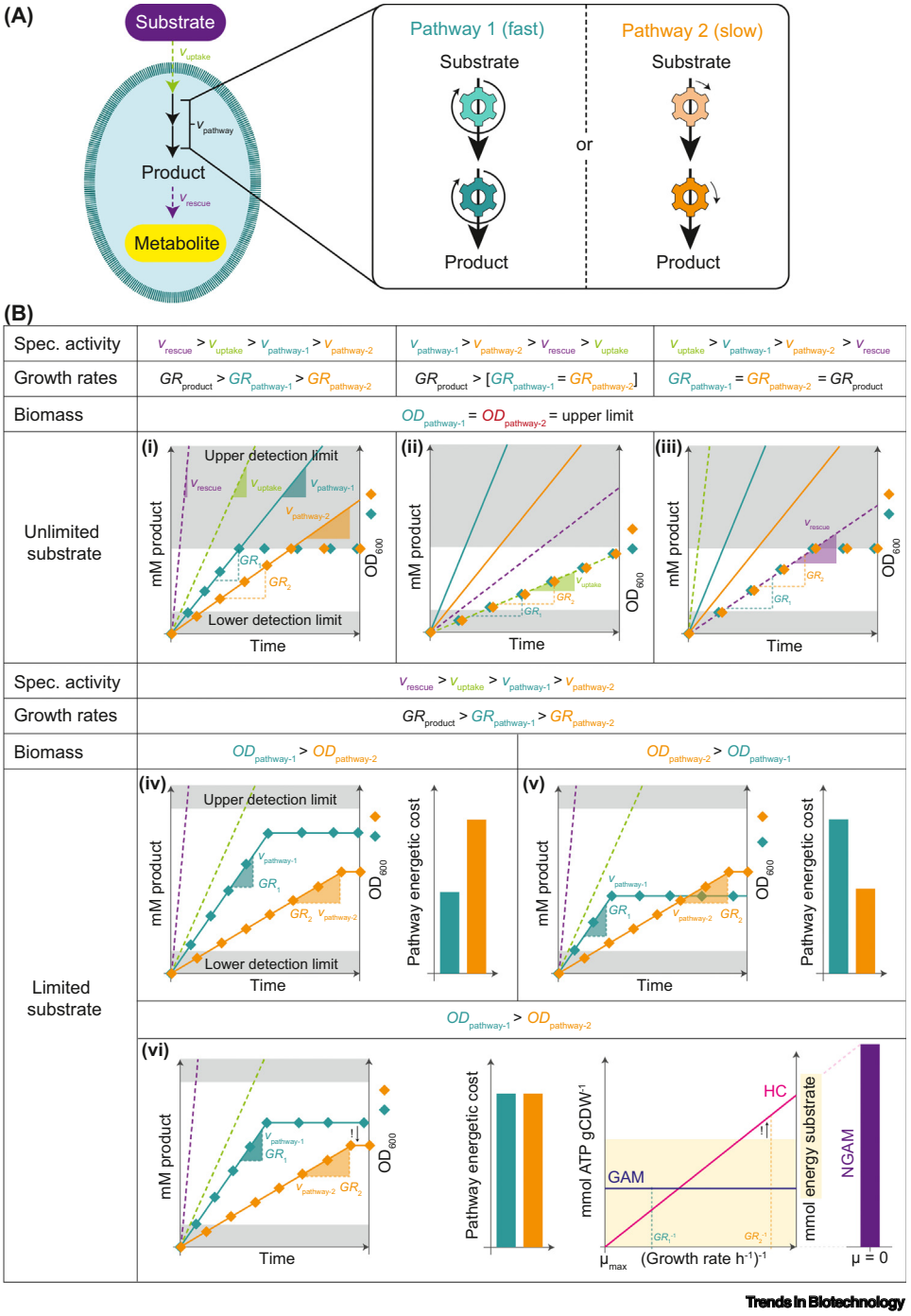

Figure 5. Comparison of fast and slow pathways for growth-coupled selection. (A) Schematic representation of the parameters compared in panel (B). (B) Case comparison for the correlations of pathway-specific activity (governed by the specific activity of the rate-limiting enzyme), growth rates, and biomass formation (indicated by OD). Cases i–iii assume an unlimited amount of pathway substrate. (i) The use of rescue metabolite (v_{rescue}, violet) and pathway substrate uptake (v_{uptake}, green) are not rate-limiting. The growth rate correlates with the rate-limiting step, in this case the pathway of interest. However, if v_{uptake} (panel ii) or v_{rescue} (panel iii) are slower than the pathway of interest, the growth rate will

(Figure legend continued at the bottom of the next page.)

are rate-limiting, growth rates will not correlate with heterologous pathway rates. The OD achieved by the strain is expected to increase proportionally to the amount of substrate fed if the pathway output is between the upper and lower detection limits of the selection strain (Figure 3) [46,65,66]. However, the biomass yield obtained via two pathways of different rates can differ depending on their energetic cost (Figure 5B, panel iv). For example, C1 assimilation via the reductive glycine pathway (2 ATP per pyruvate) is expected to result in higher biomass yields than can be achieved with the energetically less efficient CBB cycle (7 ATP per pyruvate), which was recently proven in *Cupriavidus necator* [67]. A fast pathway that is less energy-efficient will still result in a faster growth rate but lower yield (Figure 5B, panel v). Last, even for pathways with equivalent resource costs, the faster pathway will have a higher biomass yield. This is because the effect of maintenance cost not associated with growth (non–growth-associated maintenance, upholding membrane gradients, etc.) becomes increasingly larger the slower the growth rate (and thus pathway rates; Figure 5B, panel vi). If the specific activity of the rate-limiting pathway enzyme(s) and the target selection sensitivity are known, the achievable growth rate can be inferred when estimating the protein content of the rate-limiting enzyme per cell. In principle, if the turnover rate and concentration of the rate-limiting enzyme constrain growth to very low growth rates, the increased maintenance cost could result in an increased need for energy, which might require increased carbon source supplementation or the addition of an energy module and substrate to the cultivation (Box 3; Figure S3D–F in the supplemental information online).

Exploiting the correlation between growth rate and pathway turnover as well as the dependence of yields on the energetic efficiency of a pathway for high-throughput pathway comparisons is a likely future development in the metabolic engineering field. Coupling growth to pathways and mutagenizing them *in vivo* using novel tools such as eMutaT7 [55] or OrthoRep [68] can allow efficient and high-throughput screening for improved variants (Note S1 in the supplemental information online). On a similar note, machine learning (ML)-assisted protein design is on the rise [69–71] and requires high-throughput screens for novel protein variants [72]. Improving the speed and selectivity of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) [73,74], such as by using the glycerate selection strains [31], and increasing the turnover of PETases [75,76] are major challenges that could be targeted by ML-based engineering efforts in combination with growth coupling.

Concluding remarks and future perspectives

Growth-coupled selection is an important tool for the implementation of synthetic metabolism and holds great potential for further efforts. A selection of *E. coli* selection strains is available that cover large areas of central, amino acid, and energy metabolism. However, even in *E. coli*, emerging metabolic bypasses still pose a threat to selection stability. Therefore, strains with minimized genomes will gain attention due to higher predictability by removal of unknown-function genes [77]. However, this approach would come at the cost of potentially removing side activities that might be beneficial to the engineering effort. Similarly to efforts aiming at the construction of synthetic yeast genomes [78], incorporating recombinase recognition sites could allow or facilitate faster genetic engineering of these cells. Inducibly expressing core metabolism genes

Outstanding questions

How far can metabolic models be refined to minimize the need for manual model curation for selection design predictions?

How can highly sensitive readouts of pathway activity (e.g., transcription factors) be more seamlessly interfaced with growth-coupled selections to achieve a maximal feasible selection range?

Can synthetically assembled minimal genomes provide a basis to combine multiple selections in a switch-based manner?

How can selection construction throughput be enhanced for less characterized, more specialized microbes?

Can synthetic communities reduce metabolic burden through pathway overexpression on the individual cell and thereby assist the implementation of large metabolic designs?

Which strain collection and data curation guidelines are required to maximize the community benefit for growth-coupled designs?

correlate with those. The biomass yield is defined by the upper detection limit of the selection strain. The remaining cases assume a limited amount of pathway substrate, so the pathway output is within the detection range of the selection strain. Again, growth rates are dependent only on pathway-specific activities if nothing else is rate-limiting. In addition, biomass yields depend on the pathway energetic cost: A fast and energy-efficient pathway allows higher growth rates and higher biomass yields (panel iv). In contrast, if the slower pathway variant is more energy-efficient, the growth rate will be lower, but the final biomass yield might exceed that of the faster pathway (panel v). Last, if a slow pathway enforces very slow growth, the non–growth-associated maintenance cost (NGAM) can increase the energy demand beyond what the supplied energy substrate (often glucose or glycerol for *Escherichia coli* selections) can supply. In that case, the biomass yield via the slow pathway might be lowered due to limited energy availability (indicated by the arrow with an exclamation mark).

Box 2. Growth rates can be used to approximate the minimum specific activity required for selection strain growth

A rough estimate of the minimal specific activity of a pathway required to achieve a certain growth rate can be inferred from the relationship between substrate consumption (dS/dt), the growth rate (μ, in h^{-1}), mass carbon yield [Y, in g dry weight of cells (CDW) per mol substrate consumed], and cell dry mass (for full biomass formation or, as shown later for selections, the mass fraction of the limiting pathway) (X, in g CDW) ([88], Eq. 6.30):

$$\frac{dS}{dt} = \frac{\mu}{Y} X \quad [I]$$

When assuming that the activity of enzymes in the carbon assimilation pathway is rate-limiting to growth, their specific activity [A, mol substrate consumed per time, = dS/dt per g of CDW formed (X, in g)] should directly correlate to substrate consumption ([88], Eq. 6.31):

$$\frac{A}{X}\left[\frac{mol}{h\ gCDW}\right] = \frac{\mu}{Y} = \left[\frac{mol\ carbon}{h\ gCDW}\right] \quad [II]$$

Assuming that 1 gram of CDW corresponds to ~0.5 g of protein, this allows us to derive the specific activity per g of protein:

$$\frac{A}{X\ 0.5}\left[\frac{mol}{h\ g_{protein}}\right] = \frac{\mu}{Y\ 0.5}\left[\frac{mol\ carbon}{h\ g_{protein}}\right] = Spec.act. \quad [III]$$

If carbon is the sole growth substrate and a cell is composed of 50% carbon per gram CDW, the mass carbon yield Y is defined as follows:

$$Y = \frac{1\ gCDW}{0.5\ g\ carbon} \quad [IV]$$

With a molecular weight of carbon of 12.0107 g per mol, Y becomes:

$$Y = \frac{1\ gCDW}{\left(\frac{0.5\ g\ carbon}{12.0107\ \frac{g\ carbon}{mol}}\right)} = \frac{1\ gCDW}{0.0416\ mol\ carbon} \quad [V]$$

By replacing Y with above expression, it follows:

$$Spec.act. = \frac{\mu\ 0.0416}{0.5}\left[h^{-1}\ mol\ carbon\ g_{protein}^{-1}\right] = \frac{\mu\ 0.0416\ 10^9}{500}\left[h^{-1}\ nmol\ carbon\ mg_{protein}^{-1}\right] \quad [VI]$$

This approximation, among others, was used by Herter *et al.* [89] to estimate the minimal required specific carbon fixation rate during autotrophic growth. Šmejkalová *et al.* [90] used the same approach to estimate the specific carbon fixation rate of serine cycle–dependent methanol assimilation by *Methylorubrum extorquens* for a given growth rate of 0.231 h^{-1}:

$$Spec.act = \frac{0.231}{60} min^{-1}\ \frac{0.0416\ 10^9\ nmol\ carbon}{500\ mg\ protein} = 320.5\ min^{-1}\ nmol\ mg^{-1} \quad [VII]$$

In this example, the pathway of interest was expected to synthesize all cellular carbon and protein from methanol only. Taking into account that two molecules of carbon were fixed per cycle turn, the authors concluded that the specific carbon fixation activity was 160 min^{-1} nmol mg^{-1}. We further apply this calculation to estimate the specific carbon fixation activity of the Calvin cycle in *Cupriavidus necator* to be 1741 min^{-1} × nmol × mg^{-1}, and that of the reductive glycine pathway in engineered *C. necator* strains 540 min^{-1} × nmol × mg^{-1} (Note S2 in the supplemental information online). Although our estimation of enzyme turnover based on proteome fractions, yield, and growth rates serves for back-of-the envelope calculations, Davidi *et al.* refined the approach to estimate *in vivo* kinetics on the basis of growth rates [91], which was then harnessed for machine-learning–guided kinetic parameter predictions for wild-type enzymes in 2023 [92].

In case of targeted growth coupling to synthetic metabolism, the fraction of biomass to be made from the pathway of interest might be much smaller. The exact amount of pathway product needed to form 1 g of CDW is described by the CBR, which can replace the mass carbon yield term:

$$Spec.act = \frac{\mu}{Y\ 0.5}\left[\frac{mol\ carbon}{h\ g_{protein}}\right] = \frac{\mu}{\frac{1\ g\ CDW}{mol\ pathway\ product}}\ \frac{1\ g\ CDW}{0.5\ g\ protein} = \mu\ \frac{mol\ pathway\ product}{1\ g\ CDW}\ \frac{1\ g\ CDW}{0.5\ g\ protein}$$
$$= \mu\ CBR\left[\frac{mol}{g\ CDW}\right]\ \frac{1\ g\ CDW}{0.5\ g\ protein} \quad [VIII]$$

With this equation, we can approximate the minimum specific activity a metabolic pathway needs to have to allow selection strain growth at a defined growth rate or how much pathway protein needs to be in the cell to achieve growth with a defined specific activity and growth rate.

Note that this is a back-of-the-envelope calculation, which aims to deliver ballpark estimates, assumes no intermediate efflux, and does not take relevant factors such as enzyme K_m values, intermediate toxicity, expression stoichiometries, or metabolic burden into account. For example, in the work describing various glyoxylate sensors, a malate cleavage module employing malate thiokinase and malyl-CoA lyase was tested with various expression strengths in a glycine auxotrophic strain (CBR 1.28 mmol glyoxylate g CDW^{-1}). The turnover of MtkAB as the slowest enzyme of the pathway is 2.7 µmol × min^{-1} × mg^{-1} [93]. When expressing MtkAB with rbsA [94], the specific growth rate of the strain was 0.169 h^{-1}. Plugging those values into the formula above should allow approximating the fraction of total protein that would need to be the rate-limiting enzyme:

$$2700\ nmol\ malate\ min^{-1}\ mg^{-1} = \frac{0.169}{60}\ min^{-1}\ 0.00128\left[\frac{mol\ glyoxylate}{1\ g\ CDW}\right]\ \frac{1\ g\ CDW}{X\ mg\ protein} \quad [IX]$$

However, the specific enzyme activity is expressed in mol substrate (malate), whereas the selective demand is expressed in mol auxotrophy target (in this case, glyoxylate). To account for this, we calculate with the mol carbon:

$$2700\ 4\ nmol\ carbon\ min^{-1}\ mg^{-1} = \frac{0.169}{60}\ min^{-1}\ 0.00128\ 2\ 10^9\left[\frac{nmol\ carbon}{1\ g\ CDW}\right]\ \frac{1\ g\ CDW}{X\ mg\ protein} \quad [X]$$

With this normalization, we would, however, suggest that one malate molecule (C4) could yield two glyoxylates (C2), which is not the case for the MtkAB-Mcl module (yielding one glyoxylate and one acetyl-CoA). To account for this, we add the stoichiometry factor X (1 C4*X = 1 C2, in this case X = 0.5) to the specific activity:

$$2700\ 0.5\ 4\ nmol\ carbon\ min^{-1}\ mg^{-1} = \frac{0.169}{60}\ min^{-1}\ 0.00128\ 2\ 10^9\left[\frac{nmol\ glyoxylate}{1\ g\ CDW}\right]\ \frac{1\ g\ CDW}{X\ mg\ protein} \quad [XI]$$

Now, the formula can be rearranged to yield the protein content per cell:

$$\frac{2700\ 0.5\ 4\ nmol\ min^{-1}\ mg^{-1}}{0.00282\ min^{-1}\ 0.00128\ 2\ 10^9\left[\frac{nmol}{g\ CDW}\right]} = 0.748\ mg^{-1} = \frac{1}{X\ mg\ protein} \quad [XII]$$

When assuming that a cell contains 0.5 g of protein per g CDW, 1.34 mg pathway protein would be equivalent to 0.26% of the total cell protein.

With the information of how much pathway protein was approximately made when expressing the gene with rbsA, we can also rearrange the formula to determine the maximum CBR that is expected to allow growth via the specific activity of interest of at least 0.05 h^{-1} (equivalent to 13.9 h doubling time) with the amount of protein calculated for rbsA (1.3 mg g CDW^{-1}):

$$2700\ 10^{-9}\ mol\ min^{-1}\ mg^{-1}\ \frac{60}{0.05}min\ \frac{1.34\ mg\ protein}{1\ g\ CDW} = 0.00434\left[\frac{mol}{1\ g\ CDW}\right] = 4.34\left[\frac{mmol}{1\ g\ CDW}\right] \quad [XIII]$$

Indeed, when testing this module in the UPPAUX sensor strain with a higher CBR of 6.2 mmol g CDW^{-1}, no growth could be achieved via the MtkAB-Mcl module, indicating that rough sensitivity requirements can be calculated on the basis of this formula.

Box 3. Low specific activities enforce energy-intensive slow growth of high-demand selection strains

If the specific activity (Sp. act.) and concentration of the slowest pathway enzyme as well as the selective demand (CBR) are known, the expected growth rate of a selection strain can be estimated.

$$Sp.act.\left[nmol\ h^{-1}\ mg^{-1}\right] = \mu\left[h^{-1}\right] CBR\left[\frac{nmol}{1\ g\ CDW}\right] \frac{1\ g\ CDW}{X\ mg\ protein} \quad [I]$$

$$\frac{Sp.act.\left[nmol\ h^{-1}\ mg^{-1}\right]}{CBR\left[\frac{nmol}{1\ g\ CDW}\right]} \frac{X\ mg\ protein}{1\ g\ CDW} = \mu\left[h^{-1}\right] \quad [II]$$

Sticking with the example of MtkAB-Mcl implementation, we can look at the possibly achievable growth rate in the most demanding glyoxylate sensor, the LOWAUX strain (CBR = 47.25 mmol glyoxylate g CDW^{-1}) – please note that this would require metabolically isolating the pathway substrate malate from the remaining tricarboxylic acid cycle and is simply used as a calculation example here:

$$\frac{2700\ 60\ 10^{-9}\left[mol\ h^{-1}\ mg^{-1}\right]}{0.047\left[\frac{mol}{1\ g\ CDW}\right]} \frac{1.34\ mg\ protein}{1\ g\ CDW} = 0.0046\left[h^{-1}\right] \quad [III]$$

The lower a cell's growth rate is, the higher is the homeostasis cost (HC, see Figure 5B in the main text), such as the energy the cell has to invest to maintain all maintenance-associated costs that do not concern growth. This cost is maximal at no growth, so μ = 0, which is defined as NGAM = non–growth-associated maintenance (Figure S3 in the supplemental information online). On the basis of previously determined energy requirements that are integrated in metabolic models of *E. coli*, such as the iCH360 model, we can calculate the expected glucose uptake for various growth rates (Figure S3D in the supplemental information online). On the basis of this requirement, we can predict the mmol glucose required to grow to an OD of 1 (or 0.36 g L^{-1} CDW) at various growth rates (Figure S3E in the supplemental information online; see source code in https://github.com/helenaschulzmirbach/FBA_Glc-uptake) and how long it would take to reach this OD (Figure S3F in the supplemental information online). This estimation shows that for growth rates lower than 0.035 (equivalent to a doubling time of 19.8 h), more than 10 mM glucose (which should be equivalent carbon to 20 mM glycerol) is required. For the example of MtkAB-Mcl, this means that more carbon needs to be supplied to support the hypothetical growth rate of 0.0046 h^{-1}.

Alternatively, an additional energy module could be introduced (e.g., formate dehydrogenase) to allow the use of an orthogonal energy substrate. In this case, the energy module is required not only to make biomass from C1 assimilation pathways or to power energetically demanding routes but also to maintain cells growing very slowly via low-flux routes, which is relevant to couple growth to slow catalysts of new-to-nature reactions such as HACL with formyl-CoA [95] or lactyl-CoA mutase with 3-hydroxypropionyl-CoA [47]. To the best of our knowledge, such a use of extra energy has not been shown before and could be an interesting approach to establish some thus far unsuccessful synthetic designs.

could in principle even allow switching the expression of multiple genes on and off in parallel, thus creating multiple selections in one cell (essential for this is tightness of an 'off' switch to prevent selection bypasses!).

A known challenge in engineering synthetic metabolism is the burden associated with the expression of additional non-native genes. Distributing this burden between strains in cocultivations of robust selections that complement one another is an interesting and novel strategy to achieve high product titers [79]. Here, community stability can be ensured by coupling cell growth to the community members to the presence of and secretion of defined products by one another. Multispecies communities, and especially autotroph-heterotroph cocultivations, hold promise for optimized productions from C1 compounds [80].

Specialized traits such as toxin tolerance, energy generation modules with rare cofactors (e.g., formate dehydrogenases, hydrogenases) or traits such as very fast growth (e.g., *Vibrio natriegens* with a 10-min doubling time [81]) are highly interesting to synthetic biologists but proved difficult to transplant to *E. coli*. Therefore, with increasing metabolic characterization, such specialized

organisms are becoming popular engineering chassis: *Pseudomonas putida* for its high metabolic flexibility; chemical robustness, including the ability to produce fluorinated compounds [82]; or *Cupriavidus necator* for the ability to use the greenhouse gas CO_2 and H_2 to produce the bioplastic PHB [83]. Expanding the engineering repertoire for those organisms and characterizing their metabolism further will be future research directions (see Outstanding questions). Maximizing the use of existing strains for the community requires thorough documentation of CBR, strain stability, emerging bypasses that are observed, and other phenotypic peculiarities. We hope that the ongoing creation of novel selection strains, their deposition in freely available repositories, and our information, summarized in this review, will help to accelerate the use of growth-coupled selections to develop new solutions and applications for synthetic metabolism in the future.

Author contributions

H.S.-M.: Conceptualization, calculation, literature review, writing – original draft, review and editing, visualization; B.D.: conceptualization, review and editing; T.J.E.: conceptualization, writing – review and editing.

Acknowledgments

This work was funded by the Max Planck Society. H.S.-M. acknowledges funding by the Bosch Research Foundation and by the Joachim Herz Foundation in form of an Add-On Fellowship for Interdisciplinary Life Science. The authors thank Dr. Elad Noor for critical reading and feedback on the manuscript and thank Prof. Dr Maria Suarez Diez and the BioSB Course 'Constraint-based modeling: Introduction and advanced topics' by Wageningen University for discussions and feedback on the representation of growth-associated maintenance and non–growth-associated maintenance. The authors are grateful to Dr. Hai He for helpful discussions on the prediction of carbon source uptake during the revision process.

Declaration of interests

H.S.-M. is coauthor of a patent involving the NNmini strain (WO2024188980A1). The other authors have no interests to declare.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tibtech.2025.06.015.

References

1. Yishai, O. *et al.* (2016) The formate bio-economy. *Curr. Opin. Chem. Biol.* 35, 1–9
2. Keller, P. *et al.* (2022) Generation of an *Escherichia coli* strain growing on methanol via the ribulose monophosphate cycle. *Nat. Commun.* 13, 5243
3. Claassens, N.J. *et al.* (2019) Making quantitative sense of electromicrobial production. *Nat. Catal.* 2, 437–447
4. Erb, T.J. *et al.* (2017) Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol.* 37, 56–62
5. Reiter, M.A. *et al.* (2024) A synthetic methylotrophic Escherichia coli as a chassis for bioproduction from methanol. *Nat. Catal.* 7, 560–573
6. Kim, S. *et al.* (2023) Optimizing E. coli as a formatotrophic platform for bioproduction via the reductive glycine pathway. *Front Bioeng Biotechnol.* 11, 1091899
7. Diehl, C. *et al.* (2023) Synthetic anaplerotic modules for the direct synthesis of complex molecules from CO2. *Nat. Chem. Biol.* 19, 168–175
8. Sevilla, M.E. *et al.* (2023) Degradation of PET Bottles by an Engineered Ideonella sakaiensis PETase. *Polymers* 15, 1779
9. Schada von Borzyskowski, L. *et al.* (2023) Implementation of the β-hydroxyaspartate cycle increases growth performance of Pseudomonas putida on the PET monomer ethylene glycol. *Metab. Eng.* 76, 97–109
10. Orsi, E. *et al.* (2021) Growth-coupled selection of synthetic modules to accelerate cell factory development. *Nat. Commun.* 12, 5295
11. Li, Z. *et al.* (2023) Growth-coupled high throughput selection for directed enzyme evolution. *Biotechnol. Adv.* 68, 108238
12. Schwander, T. *et al.* (2016) A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 354, 900–904
13. McLean, R. *et al.* (2023) Exploring alternative pathways for the in vitro establishment of the HOPAC cycle for synthetic CO2 fixation. *Sci. Adv.* 9, eadh4299
14. Luo, S. *et al.* (2022) A cell-free self-replenishing CO2-fixing system. *Nat. Catal.* 5, 154–162
15. Gleizer, S. *et al.* (2019) Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell* 179, 1255–1263.e12
16. Wenk, S. *et al.* (2018) An Engineering Approach for Rewiring Microbial Metabolism. In *Methods in Enzymology* (608), pp. 329–367, Elsevier
17. Kim, S. *et al.* (2020) Growth of E. coli on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* 16, 538–545
18. Wu, T. *et al.* (2023) Engineering a synthetic energy-efficient formaldehyde assimilation cycle in Escherichia coli. *Nat. Commun.* 14, 8490
19. Wenk, S. *et al.* (2022) Evolution-assisted engineering of *E. coli* enables growth on formic acid at ambient CO2 via the Serine Threonine Cycle. *Metab. Eng.* 88, 14–24 ISSN 1096-7176.
20. Ben-Nissan, R. *et al.* (2023) Autotrophic growth of E. coli is achieved by a small number of genetic changes. *eLife* 12, RP88793
21. Satanowski, A. *et al.* (2020) Awakening a latent carbon fixation cycle in Escherichia coli. *Nat. Commun.* 11, 5812
22. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30
23. Keseler, I.M. *et al.* (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45, D543–D550
24. Burgard, A.P. *et al.* (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657

25. Alter, T.B. *et al.* (2025) Metabolic growth-coupling strategies for in vivo enzyme selection systems. *Metab. Eng. Commun.* 20, e00257 ISSN 2214-0301
26. Alter, T.B. and Ebert, B.E. (2019) Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics* 20, 447
27. Legon, L. *et al.* (2022) gcFront: a tool for determining a Pareto front of growth-coupled cell factory designs. *Bioinformatics* 38, 3657–3659
28. Iacometti, C. *et al.* (2022) Activating Silent Glycolysis Bypasses in *Escherichia coli*. *Biodes. Res.* 2022, 1–17
29. He, H. *et al.* (2024) Adaptive laboratory evolution recruits the promiscuity of succinate semialdehyde dehydrogenase to repair different metabolic deficiencies. *Nat. Commun.* 15, 8898
30. Cotton, C.A. *et al.* (2020) Underground isoleucine biosynthesis pathways in *E. coli*. *Elife* 9, e54207
31. Aslan, S. *et al.* (2020) Design and engineering of E. coli metabolic sensor strains with a wide sensitivity range for glycerate. *Metab. Eng.* 57, 96–109
32. Klamt, S. and Mahadevan, R. (2015) On the feasibility of growth-coupled product synthesis in microbial strains. *Metab. Eng.* 30, 166–178
33. Lu, N. *et al.* (2024) Growth-coupled production of L-isoleucine in *Escherichia coli* via metabolic engineering. *Metab. Eng.* 86, 181–193
34. An, N. *et al.* (2023) Establishing a growth-coupled mechanism for high-yield production of β-arbutin from glycerol in *Escherichia coli*. *Bioresour. Technol.* 369, 128491
35. Nielsen, J.R. *et al.* (2023) Growth-coupled enzyme engineering through manipulation of redox cofactor regeneration. *Biotechnol. Adv.* 63, 108102
36. von Kamp, A. and Klamt, S. (2017) Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat. Commun.* 8, 15956
37. Chen, J. *et al.* (2022) Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends Biotechnol.* 40, 773–776
38. Karp, P.D. *et al.* (2019) The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* 20, 1085–1093
39. Feist, A.M. *et al.* (2010) Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab. Eng.* 12, 173–186
40. Schneider, P. *et al.* (2021) Systematizing the different notions of growth-coupled product synthesis and a single framework for computing corresponding strain designs. *Biotechnol. J.* 16, 2100236
41. Lloyd, C.J. *et al.* (2019) The genetic basis for adaptation of model-designed syntrophic co-cultures. *PLoS Comput. Biol.* 15, e1006213
42. Motamedian, E. *et al.* (2023) OptEnvelope: A target point guided method for growth-coupled production using knockouts. *PLoS ONE* 18, e0294313
43. Antonovsky, N. *et al.* (2016) Sugar Synthesis from CO2 in *Escherichia coli*. *Cell* 166, 115–125
44. Boecker, S. *et al.* (2023) Growth-coupled anaerobic production of isobutanol from glucose in minimal medium with *Escherichia coli*. *Biotechnol. Biofuels Bioprod.* 16, 148
45. Corrao, M. *et al.* (2024) A compact model of *Escherichia coli* core and biosynthetic metabolism. *arXiv*, Published online June 24, 2024. https://doi.org/10.48550/arXiv.2406.16596
46. Orsi, E. *et al.* (2025) Computation-aided designs enable developing auxotrophic metabolic sensors for wide-range glyoxylate and glycolate detection. *Nat. Commun.* 16, 2168
47. Schulz-Mirbach, H. *et al.* (2024) New-to-nature CO2-dependent acetyl-CoA assimilation enabled by an engineered B12-dependent acyl-CoA mutase. *Nat. Commun.* 15, 10235
48. Jensen, S.I. *et al.* (2016) Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. *Sci. Rep.* 5, 17874
49. Kozaeva, E. *et al.* (2024) The pAblo·pCasso self-curing vector toolset for unconstrained cytidine and adenine base-editing in Gram-negative bacteria. *Nucleic Acids Res.* 52, e19
50. Volke, D.C. *et al.* (2022) Modular (de)construction of complex bacterial phenotypes by CRISPR/nCas9-assisted, multiplex cytidine base-editing. *Nat. Commun.* 13, 3026
51. Kundu, B.B. *et al.* (2025) Extracellular Respiration is a Latent Energy Metabolism in *Escherichia coli*. *Cell* 188, 2907–2924.e23
52. Rong, Y. *et al.* (2024) CRISPRi-mediated metabolic switch enables concurrent aerobic and synthetic anaerobic fermentations in engineered consortium. *Nat. Commun.* 15, 8985
53. Claassens, N.J. *et al.* (2019) Synthetic Methanol and Formate Assimilation Via Modular Engineering and Selection Strategies. *Curr. Issues Mol. Biol.* 33, 237–248
54. Lin, P.P. *et al.* (2018) Construction and evolution of an *Escherichia coli* strain relying on nonoxidative glycolysis for sugar catabolism. *Proc. Natl. Acad. Sci.* 115, 3538–3546
55. Seo, D. *et al.* (2023) A dual gene-specific mutator system installs all transition mutations at similar frequencies *in vivo*. *Nucleic Acids Res.* 51, e59
56. Badran, A.H. and Liu, D.R. (2015) Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* 6, 8425
57. Marlière, P. *et al.* (2011) Chemical Evolution of a Bacterium's Genome. *Angew. Chem. Int. Ed.* 50, 7109–7114
58. Luo, S. *et al.* (2023) Construction and modular implementation of the THETA cycle for synthetic CO2 fixation. *Nat. Catal.* 6, 1228–1240
59. Schulz-Mirbach, H. *et al.* (2024) Engineering new-to-nature biochemical conversions by combining fermentative metabolism with respiratory modules. *Nat. Commun.* 15, 6725
60. Satanowski, A. *et al.* (2025) Design and implementation of aerobic and ambient CO2-reduction as an entry-point for enhanced carbon fixation. *Nat. Commun.* 16, 3134
61. Chou, A. *et al.* (2021) An orthogonal metabolic framework for one-carbon utilization. *Nat. Metab.* 3, 1385–1399
62. Landwehr, G.M. *et al.* (2024) A synthetic cell-free pathway for biocatalytic upgrading of one-carbon substrates. *bioRxiv*, 2024.08.08.607227. https://doi.org/10.1101/2024.08.08.607227. PMID: 39149402; PMCID: PMC11326285
63. Orsi, E. *et al.* (2024) Harnessing noncanonical redox cofactors to advance synthetic assimilation of one-carbon feedstocks. *Curr. Opin. Biotechnol.* 90, 103195
64. Black, W.B. *et al.* (2023) Design, construction, and application of noncanonical redox cofactor infrastructures. *Curr. Opin. Biotechnol.* 84, 103019
65. Schann, K. *et al.* (2024) Design, construction and optimization of formaldehyde growth biosensors with broad application in biotechnology. *Microb. Biotechnol.* 17, e14527
66. Schulz-Mirbach, H. *et al.* (2022) On the flexibility of the cellular amination network in *E. coli*. *eLife* 11, e77492
67. Dronsella, B.B. *et al.* (2025) One-carbon fixation via the synthetic reductive glycine pathway exceeds yield of the Calvin cycle. *Nat. Microbiol.* 10, 646–653
68. Tian, R. *et al.* (2024) Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. coli*. *Science* 383, 421–426
69. Ferruz, N. *et al.* (2022) ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* 13, 4348
70. Madani, A. *et al.* (2023) Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* 41, 1099–1106
71. Beck, J. *et al.* (2024) Diversifying de novo TIM barrels by hallucination. *Protein Sci.* 33, e5001
72. Lanster, D.L. *et al.* (2024) A Genetically Encoded System to Quantify and Evolve RuBisCO-Catalyzed Carbon Fixation. *bioRxiv*, 2024.05.15.594374. https://doi.org/10.1101/2024.05.15.594374
73. Flamholz, A.I. *et al.* (2019) Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* 58, 3365–3376
74. Prywes, N. *et al.* (2025) A map of the rubisco biochemical landscape. *Nature* https://doi.org/10.1038/s41586-024-08455-0
75. Deng, B. *et al.* (2023) Improving the activity and thermostability of PETase from *Ideonella sakaiensis* through modulating its post-translational glycan modification. *Commun. Biol.* 6, 1–10
76. Wang, T. *et al.* (2024) Molecular engineering of PETase for efficient PET biodegradation. *Ecotoxicol. Environ. Saf.* 280, 116540
77. Gherman, I.M. *et al.* (2024) Accelerated design of *Escherichia coli* genomes with reduced size using a whole-cell model and machine learning. *bioRxiv*, 2023.10.30.564402; https://doi.org/10.1101/2023.10.30.564402

78. Schindler, D. *et al.* (2023) Design, construction, and functional characterization of a tRNA neochromosome in yeast. *Cell* 186, 5237–5253.e22
79. Scarinci, G. *et al.* (2024) Enhanced metabolic entanglement emerges during the evolution of an interkingdom microbial community. *Nat. Commun.* 15, 7238
80. Chen, W. *et al.* (2024) Synthetic, marine, light-driven, autotroph-heterotroph co-culture system for sustainable β-caryophyllene production. *Bioresour. Technol.* 410, 131232
81. Weinstock, M.T. *et al.* (2016) *Vibrio natriegens* as a fast-growing host for molecular biology. *Nat. Methods* 13, 849–851
82. Wirth, N.T. and Nikel, P.I. (2021) Combinatorial pathway balancing provides biosynthetic access to 2-fluoro-*cis*,*cis*-muconate in engineered *Pseudomonas putida*. *Chem. Catal.* 1, 1234–1259
83. Lim, J. *et al.* (2023) Biohybrid CO2 electrolysis for the direct synthesis of polyesters from CO2. *Proc. Natl. Acad. Sci.* 120, e2221438120
84. Noor, E. *et al.* (2024) A metabolic map of *E. coli*'s central metabolism. Zenodo https://doi.org/10.5281/zenodo.11472550
85. Neidhardt, F.C. *et al.* (1990) *Physiology of the bacterial cell: a molecular approach*, Sinauer Associates ISBN 0878936084
86. Milo, R. *et al.* (2010) BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38, D750–D753
87. Bennett, B.D. *et al.* (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5, 593–599
88. Lengeler, J.W. *et al.*, eds (1999) *Biology of the prokaryotes*, Thieme
89. Herter, S. *et al.* (2001) Autotrophic CO2 Fixation by *Chloroflexus aurantiacus*: Study of Glyoxylate Formation and Assimilation via the 3-Hydroxypropionate Cycle. *J. Bacteriol.* 183, 4305–4316
90. Šmejkalová, H. *et al.* (2010) Methanol Assimilation in *Methylobacterium extorquens* AM1: Demonstration of All Enzymes and Their Regulation. *PLoS ONE* 5, e13001
91. Davidi, D. *et al.* (2016) Global characterization of *in vivo* enzyme catalytic rates and their correspondence to in vitro kcat measurements. *Proc. Natl. Acad. Sci.* 113, 3401–3406
92. Kroll, A. *et al.* (2023) Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nat. Commun.* 14, 4139
93. Yu, H. *et al.* (2018) Augmenting the Calvin–Benson–Bassham cycle by a synthetic malyl-CoA-glycerate carbon fixation pathway. *Nat. Commun.* 9, 2008
94. Zelcbuch, L. *et al.* (2013) Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* 41, e98
95. Chou, A. *et al.* (2019) 2-Hydroxyacyl-CoA lyase catalyzes acyloin condensation for one-carbon bioconversion. *Nat. Chem. Biol.* 15, 900–906