

Metabolic growth-coupling strategies for *in vivo* enzyme selection systems

Tobias B. Alter^{a,*}, Pascal A. Pieters^a, Colton J. Lloyd^b, Adam M. Feist^{a,b}, Emre Özdemir^a, Bernhard O. Palsson^{a,b}, Daniel C. Zielinski^{a,b,**}

^a The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800, Kgs. Lyngby, Denmark

^b Department of Bioengineering, University of California, 9500 Gilman Dr. #0412, La Jolla, San Diego, CA, 92093-0412, USA

ARTICLE INFO

Keywords:

Constraint-based metabolic modeling
Growth-coupling
Enzyme engineering
Microbial strain design
Microbial cell factories

ABSTRACT

Whole-cell biocatalysis facilitates the production of a wide range of industrially and pharmaceutically relevant molecules from sustainable feedstocks such as plastic wastes, carbon dioxide, lignocellulose, or plant-based sugar sources. The identification and use of efficient enzymes in the applied biocatalyst is key to establishing economically feasible production processes. The generation and selection of favorable enzyme variants in adaptive laboratory evolution experiments using growth as a selection criterion is facilitated by tightly coupling enzyme catalytic activity to microbial metabolic activity. Here, we present a computational workflow to design strains that have a severe, growth-limiting metabolic chokepoint through a shared class of enzymes. The resulting chassis cell, termed enzyme selection system (ESS), is a platform for growth-coupling any enzyme from the respective enzyme class, thus offering cross-pathway application for enzyme engineering purposes. By applying the constraint-based modeling workflow, a publicly accessible database of 25,505 potential and experimentally tractable ESS designs was built for *Escherichia coli* and a broad range of production pathways with biotechnological relevance. A model-based analysis of the generated design database reveals a general design principle that the target enzyme activity is linked to overall microbial metabolic activity, not just the synthesis of one biomass precursor. It can be observed that the stronger the predicted coupling between target enzyme and metabolic activity, the lower the maximum growth rate and therefore the viability of an ESS. Consequently, growth-coupling strategies with only suboptimal coupling strengths, as are included in the ESS design database, may be of interest for practical applications of ESSs in order to circumvent overly restrictive growth defects. In summary, the computed design database, which is accessible via <https://biosustain.github.io/ESS-Designs/>, and its analysis provide a foundation for the generation of valuable *in vivo* ESSs for enzyme optimization purposes and a range of biotechnological applications in general.

1. Introduction

Microbial production of value-added molecules in integrated bioprocesses is widely recognized as a major driver towards a sustainable, circular bioeconomy. The microorganisms used in such bioprocesses typically require significant engineering to achieve economically feasible yields and production rates under fermentation conditions. Genetic engineering efforts usually involve the expression of

heterologous enzymes enabling a microbe to synthesize a target molecule from its native metabolites. The performance of heterologous enzymes, i.e., the maximum achievable rate with which substrates are converted to products, is a key factor determining the performance of the overall production pathway.

Neither heterologous nor native enzymes are typically adapted for maximum production efficiency in an engineered microbial host (Bar-Even et al., 2011). Traditional enzyme engineering techniques,

Abbreviations: ALE, adaptive laboratory evolution; ATPsc, ATP synthesis capability; CC, coupling chemistry; ESS, enzyme selection system; FVA, flux variability analysis; GCS, growth-coupling strength; GPR, gene-protein-reaction; KS, Kolmogorov-Smirnov; MILP, mixed integer linear program; pFBA, parsimonious flux balance analysis.

* Corresponding author.

** Corresponding author. The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800, Kgs. Lyngby, Denmark.

E-mail addresses: tobias.alter@rwth-aachen.de (T.B. Alter), dczielin@ucsd.edu (D.C. Zielinski).

¹ present address: Institute of Applied Microbiology, Aachen Biology and Biotechnology, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany.

<https://doi.org/10.1016/j.mec.2025.e00257>

Received 30 August 2024; Received in revised form 13 December 2024; Accepted 12 February 2025

Available online 12 February 2025

2214-0301/© 2025 The Authors. Published by Elsevier B.V. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

such as random mutagenesis or sequence-based semi-rational strategies (Cheng et al., 2015), can be applied to improve relevant enzyme properties, but require significant experimental investment. Generation and screening of enzyme mutant libraries need to be conducted at high throughput, which necessitates non-standard laboratory equipment and an assay suitable for rapid and reliable testing of the target property (Shivange et al., 2009). Effective screening of large enzyme libraries is becoming more accessible due to the worldwide advance of biofoundries (Hillson et al., 2019). However, establishing proper screening assays remains a non-trivial and laborious task. Moreover, assays are often conducted *in vitro*, thus may not resemble the native biophysical conditions enzymes are subjected to within a cell, potentially leading to unfavorable discrepancies in the performance of a selected enzyme variant between the *in vitro* screening and the *in vivo* application conditions.

In vivo enzyme screening and selection systems based on a strict metabolic coupling between a target enzyme's activity and growth, here referred to as enzyme selection systems or ESSs, have several advantages. ESSs allow for the use of growth as a simple selection criterion for target enzyme activity. This feature is exploited for enzyme engineering purposes in Adaptive Laboratory Evolution (ALE) experiments during which a microbial culture is repeatedly propagated into fresh medium and naturally selected for favorable mutations leading to higher growth rates (Sandberg et al., 2019). The necessity for screening assays or high throughput instrumentation is omitted because cells harboring favorable mutations will outgrow others with detrimental or neutral mutations. ALE is thereby not restricted to adaptations in the target enzyme's primary sequence but can simultaneously select for, as an example, an improved cofactor or precursor supply in the microbial host. As a proof-of-principle for ESSs, an *Escherichia coli* strain has recently been designed to couple methyltransferases to growth and subsequent ALE experiments have led to improved enzymes variants (Luo et al., 2019).

Due to the vast complexity and redundancy of metabolic networks, a generalized approach for growth-coupling any enzyme requires computer-aided techniques to derive appropriate genetic designs. Various computational methods that use constraint-based metabolic models as dense biochemical knowledgebases have been built to serve this purpose (Alter and Ebert, 2019; Burgard et al., 2003; Feist et al., 2010; Jensen et al., 2019; Pharkya et al., 2004; von Kamp and Klamt, 2014). While most model-based studies have focused on computing genetic designs for coupling growth to the production of a target molecule or the activity of endogenous reactions, growth-coupling of specific enzymes has not been extensively considered.

In this study, we developed chassis ESS designs in *E. coli* for a range of enzyme classes. Each single chassis ESS is applicable to all enzymes within the associated enzyme class, thus representing a "plug-and-play" *in vivo* engineering and optimization tool for multiple enzymes with a single strain design. For this purpose, we defined an appropriate enzyme class as a group of enzymes sharing a common substrate-product pair, which we term a coupling chemistry (CC). Linking a CC to growth produces an ESS for all enzymes sharing that CC due to their common substrate-product pair. We identified an exhaustive list of CCs for *E. coli* and selected for candidates which allowed for an integration into the genome-scale constraint-based metabolic model iML1515 (Monk et al., 2017). Furthermore, a computational framework was established and applied for identifying diverse sets of growth-coupled strain designs for each modeled CC considering gene knockouts, heterologous reaction insertions, carbon source, and metabolite cofeeds as engineering variables. We show that ESS strain designs exist for 33 CCs, and that a notable number of CCs hold the potential for ESSs with a strong selection pressure. The enzymes in the subset of CCs with feasible growth-coupled strain designs span various metabolic pathways. This finding supports the potential universality of *in vivo* ESSs as a tool for optimizing enzymes towards efficient, high-flux target molecule synthesis routes. Generally, CCs involved in the transfer of a high-molecular group, such as a CoA cofactor, show a decreased success rate in finding strong coupling strain

designs compared to those transferring low-molecular groups. Moreover, the target enzyme activity in most identified designs is linked to the overall microbial metabolic activity, not just the synthesis of one biomass precursor, thus implies yielding tighter growth-coupling. All identified strain designs, performance metrics such as predicted maximum growth rates and coupling strengths, and value-added target molecules associated with the designs or CCs are made available online via <https://biosustain.github.io/ESS-Designs/>.

2. Methods

2.1. A metabolic model-based workflow for computing enzyme selection system designs

An enzyme selection system (ESS) is derived from a coupling chemistry (CC) by growth-coupling the CC's generic reaction. As is the case for any heterologous reaction, CCs are not necessarily growth-coupled in a wildtype metabolic background. Finding strain designs that enforce a coupling between the catalytic activity of a non-essential or non-catabolic enzyme and biomass formation is a challenging task due to the inherent complexity of metabolic networks. Genome-scale constraint-based metabolic models enable a directed, computer-aided search for feasible growth-coupling strain designs and have been widely studied for that purpose.

A Python-based workflow was implemented for computing and evaluating enzyme selection system strain designs by exploiting the rich metabolic information content of constraint-based metabolic models. COBRApy (Ebrahim et al., 2013) functionalities were integrated to enable intuitive and community-friendly model management. The workflow can be divided into three main steps: (i) set up of the metabolic model and optimization algorithm parameters, (ii) search for growth-coupling strain design solutions, and (iii) evaluation and summarization of design results (Fig. 1).

The workflow developed for computing enzyme selection systems bases on the gcOpt optimization algorithm (Alter and Ebert, 2019). gcOpt determines relevant growth-coupling strain designs for a target reaction and was implemented in Python in the course of this work. Briefly, gcOpt solves a Mixed-Integer Linear Programming (MILP) problem that maximizes a minimally guaranteed target reaction flux for a fixed growth rate. The design variables considered are gene knockouts, heterologous reaction insertions, and changes in media conditions in the form of carbon source selection and metabolite cofeed. Besides applying various maximum numbers of design variables, the fixed growth rate was additionally varied to diversify the computed set of growth-coupling strain designs. The various sets of parameters applied to each selected CC are summarized in Table 1. Design characteristics such as the growth-coupling strength (GCS), maximum growth rate, and ATP synthesis capability (ATPsc) are computed. Detailed descriptions of the workflow for computing growth-coupling strain designs and the underlying gcOpt algorithm are given in the Supplementary Methods section.

The Python code for running the workflow, here termed the Growth-Coupling Suite, is freely available online on <https://github.com/biosustain/Growth-coupling-suite>. Examples are provided in Jupyter notebooks to guide the user through typical workflows. While this study focuses on designing and evaluating growth-coupled strains using CCs in the *E. coli* model, this workflow is applicable to any target reaction or metabolic model.

3. Results

3.1. Coupling chemistries of potential enzyme selection systems span a range of enzyme classes and metabolic sectors

We define an enzyme selection system (ESS) as a microbial chassis strain which allows for the evolutionary optimization of a group of

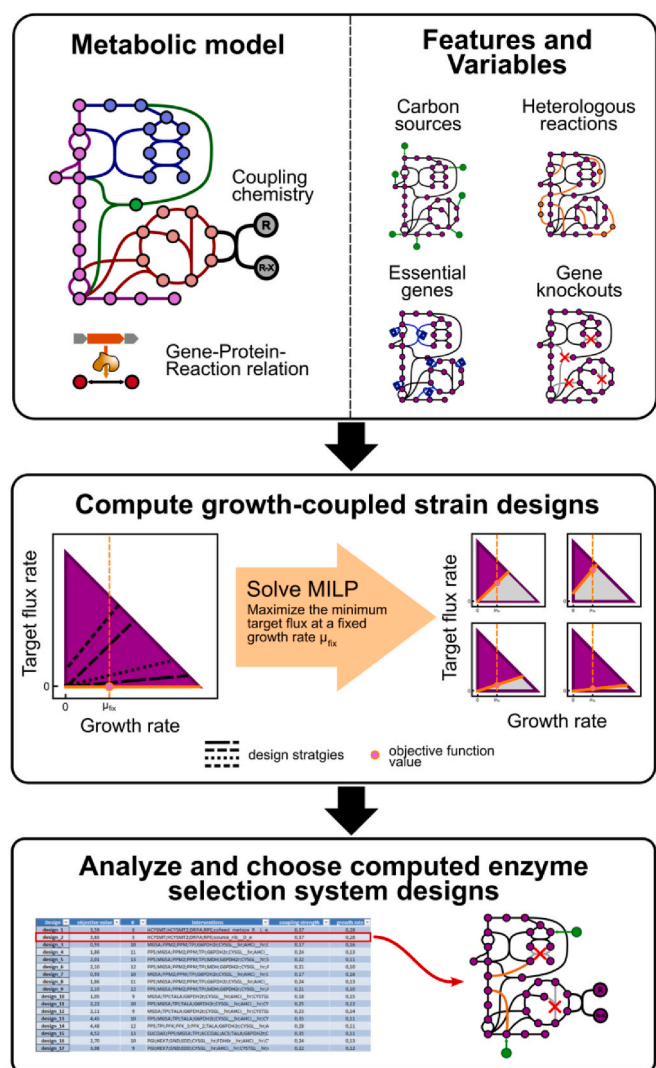


Fig. 1. Workflow for computing enzyme selection system designs from coupling chemistries using constraint-based metabolic models. At the first stage, a coupling chemistry is added to a metabolic model, constraints are applied to match the environmental conditions, essential genes are defined, and carbon sources, heterologous reactions, as well as gene knockouts are set as design variables. Secondly, the gcOpt algorithm is applied to derive growth-coupled strain designs, i.e., sets of appropriate design variables, which are being summarized in human-readable tables. At last, suitable strain designs can be chosen for constructing an enzyme selection system considering simulated metadata and experimental expert knowledge.

enzymes using ALE experiments and growth as the selection criterion. All enzymes associated with an ESS share a common metabolic function, which is coupled to growth. Such a common metabolic function can be abstracted by a generic reaction formula, here termed as a coupling chemistry (CC). A CC has an explicit substrate-product pair common to all associated enzymes and a set of generic, enzyme specific reactants (Fig. 2A). For example, a CC with S-adenosyl-L-methionine and its demethylated form S-adenosyl-L-homocysteine as the common substrate-product pair describes many methyltransferases (EC 2.1.1). The generic reactants of the CC thereby represent the methyl acceptor and the methylated product specific to each enzyme. However, CCs do not have to comprise one EC class exclusively but can represent enzymes from multiple different classes, if they share the CC's common substrate-product pair.

To investigate whether ESSs can be designed from generic CCs, CCs comprising two or more enzymes were derived from the BRENDA

Table 1

Maximum number of genetic and medium interventions used to constrain growth-coupling optimization algorithm gcOpt. Each set of parameters (lines) was applied for computing growth-coupling designs for any coupling chemistry using fixed growth rates of 0.1 h^{-1} and 0.3 h^{-1} resulting in six gcOpt runs per coupling chemistry. gcOpt was allowed to choose a primary carbon source from a pool of 48 manually preselected carbon sources (cf. Supplementary Data) and the potential cofeeds were determined following the descriptions in the Supplementary Methods section 1.2. A set of up to 1523 metabolic reactions was extracted from the BIGG model database, added to the model, and used as the source for heterologous reaction additions (Supplementary Methods section 1.4).

Total interventions	Gene deletions	Heterologous reaction additions	Primary carbon sources	Cofeed sources	Fixed growth rate $[\text{h}^{-1}]$
6	5	5	1	1	0.1/0.3
9	8	8	1	1	0.1/0.3
12	11	11	1	1	0.1/0.3

enzyme database (Jeske et al., 2019). 2564 CCs were identified representing 3089 enzymes spanning the full range of enzyme classes (Fig. 2B), where an enzyme may be associated with multiple CCs. Most CC-associated enzymes are oxidoreductases (EC 1) and transferases (EC 2) (together 69%) reflecting the similarity between the CC definition and their catalyzed reactions typically involving a transfer of functional groups, oxygen, or hydrogen atoms from a donor to an acceptor molecule. Methyltransferases (EC 2.1.1) and acyltransferases (EC 2.3.1) were most abundant and represented 10% and 8% of all enzymes assigned to a CC, respectively (Fig. S3).

We selected a subset of all identified CCs for further analysis according to the following criteria: 1) Representing at least five different enzymes to satisfy the demand for a “plug-and-play” evolutionary enzyme optimization platform, and 2) the ability to be integrated into the genome-scale metabolic model iML1515 of *E. coli* K-12 MG1655 (Monk et al., 2017), i.e., their explicit reactants need to be present in the model. As a result, the complete list of CCs was condensed to a core of 44 CCs (Table 2) encompassing 924 enzymes from all top-level enzyme classes but isomerases (EC 5) and translocases (EC 7) (Fig. 2C). As in the complete set of CCs, enzymes covered by the selected CCs primarily belong to oxidoreductases (EC 1) and transferases (EC 2) (together 90%). Despite the bias towards transferases in particular, the selected CCs and their enzymes cover a range of metabolic subsystems according to the enzymes' pathway associations in the KEGG database (Kanehisa et al., 2017) (Fig. 2D, Fig. S5). As we will show below, the broad coverage of metabolic subsystems can be attributed to the remarkable metabolic versatility and importance of methyltransferases (EC 2.1.1) accounting for a third of all enzymes within the CC subset (Fig. S4).

3.2. A database of growth-coupling strain designs for enzyme selection systems

To generate a database of ESSs across various enzyme classes, a growth-coupling strain design workflow was applied to the 44 selected CCs (Table 2). The focus was set on *E. coli* as a platform organism for ESSs, due to the availability of many established genetic engineering tools, its ease of handling in the laboratory, a demonstrated suitability as a multi-purpose microbial cell factory, and its well curated genome-scale metabolic model iML1515 (Monk et al., 2017). To diversify the ESS database, the growth-coupling workflow was applied to each selected CC using multiple optimization problem parameter sets (cf. Methods section). The parameters that were manipulated included the number of total interventions, heterologous insertions, and gene knockouts, as well as the fixed growth rate. Unless otherwise noted, the choice of one out of 48 manually selected carbon sources and the addition of a metabolite cofeed were additionally considered as design variables (cf.

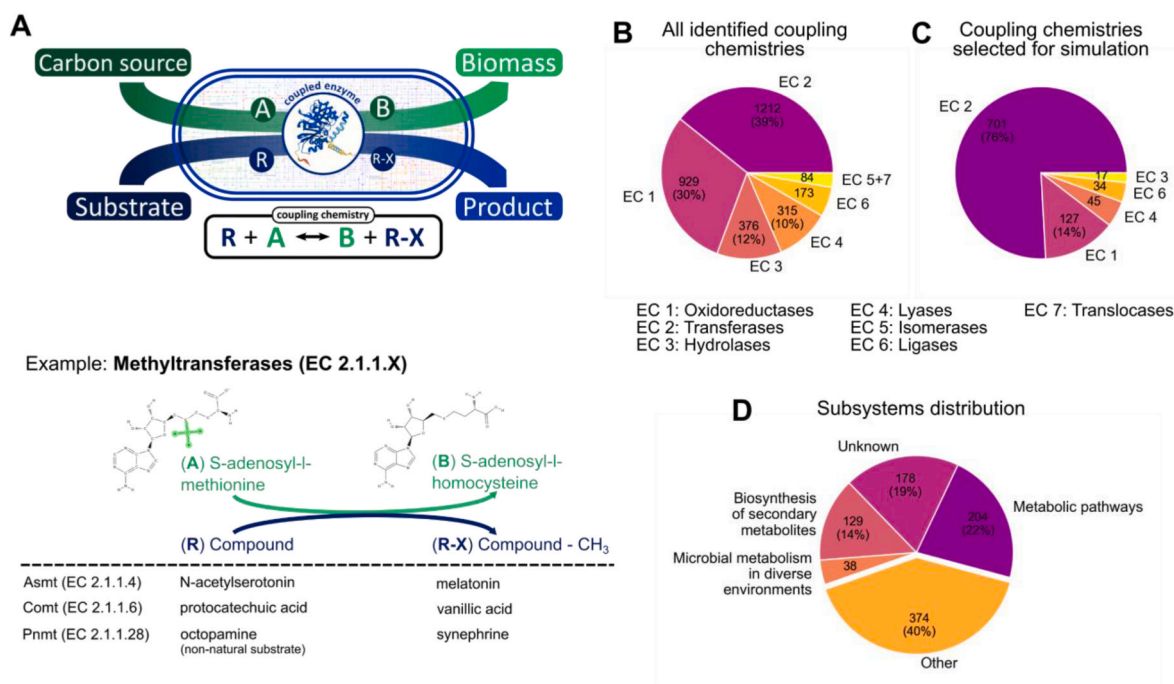


Fig. 2. Illustration and analysis of coupling chemistries being generic reaction formulas for groups of enzymes sharing a common substrate-product pair. (A) Scheme of a coupling chemistry (CC) metabolically coupled to growth in a microbial host. All enzymes represented by a CC share the common substrate A and product B which link to the host's native metabolism. R and R-X are the individual reactants for each enzyme. A methyltransferase reaction scheme is depicted as an example CC for which S-adenosyl-L-methionine and S-adenosyl-L-homocysteine is the common substrate-product pair. The enzymes acetylserotonine O-methyltransferase (Asmt), catechol O-methyltransferase (Comt), and phenylethanolamine N-methyltransferase (Pnmt) and their corresponding reactants are exemplarily shown as members of the methyltransferase EC. (B) Enzyme class distribution among enzymes from all coupling chemistries identified from a BRENDA database search (N = 3089) and a subset of those selected for the *in silico* growth-coupling study in *E. coli* (N = 924) (C). (D) Distribution of the metabolic sectors associated with enzymes within the coupling chemistries selected for the *in silico* study. Enzymes without an explicit metabolic sector association are summarized under "Unknown". The pie section "Other" summarizes enzymes with a share below 4%. A detailed breakdown of the subsystems associated with these enzymes is shown in Fig. S5. Metabolic sector associations were derived from the KEGG database.

Supplementary Methods section 1.2 and the Supplementary Data). Moreover, six variants of the base iML1515 model were employed to investigate the influence of environmental and genotypic conditions on finding growth-coupling strain designs. Details of each applied model variant are summarized in Table 3. Further information about the growth-coupling strain design workflow and its application in this study can be found in the Methods section.

47,635 unique growth-coupling strain designs, i.e., potential enzyme selection systems, were identified across 33 CCs (77%) (Fig. 3B, Supplementary Data). For the remaining 11 CCs (33%) no designs could be identified. The designs show a wide range of growth-coupling strength (GCS) values with a median of 0.18, which was used as a cut-off for separating low growth-coupling (GCS < 0.18) from high growth-coupling designs (GCS > 0.18) (Fig. 3C). The GCS measure quantifies the level of target reaction activity, here the CC reaction, needed to sustain growth and ranges between 0 (no coupling) to 1 (high target reaction rates needed for growth). Similarly, maximum growth rates of identified ESS strain designs are widely distributed with a bias towards 0.1 and 0.3 h⁻¹ (Fig. 3D). This is due to the fact that the fixed growth rate applied in the gcOpt formulation (Table 1) acts as a lower bound on growth for computed strain designs and higher observed growth-coupling strengths generally compromise on maximum growth rates.

3.3. Enzyme selection systems are achieved with a reasonable number of interventions

Most computed strain designs and ESSs can be realized with a reasonable number of genetic interventions of five or fewer knockouts (55%) and two or fewer heterologous insertions (65%) (Fig. 3E) which is in line with growth-coupling studies of endogenous reactions (Feist

et al., 2010; Jensen et al., 2019; Tervo and Reed, 2012). Thus, the implementation of most of the designs in the ESS database is expected to be experimentally feasible. Surprisingly, strain designs with high GCS values require slightly fewer genetic interventions than weakly growth-coupling designs (Fig. S6). 57% of all high growth-coupling designs, i.e., designs with GCS values greater than the median of 0.18 (cf. Fig. 3C), suggest knockout of five or less genes versus 53% of designs with low growth-coupling (GCS < 0.18). This counterintuitive discrepancy highlights the importance of the topological position of a target reaction in a metabolic network for finding growth-coupling design strategies and the metabolic burden its activity inflicts on the cellular host.

3.4. Genome reduction and medium composition variations only slightly impact the identification of growth-coupling designs

The success rate of finding growth-coupling strain designs for CCs ranges between 51% and 60% for most tested model variants and parameter sets (Fig. 3B). Only for the N-coupling case growth-coupling of CCs was largely prohibited, thus rendering the coupling of nitrogen provision from alternative, biotic sources, such as amino acids, to a target reaction as a weak and non-generalizable principle for establishing ESSs. Beside the complete lack of growth-coupling strain designs for 11 CCs, five CCs (rph1, coa3, mth5, coa1, and act1) were only successfully coupled to growth by using genome-reduced models or standard model variants without considering cofeeding of metabolites as design variables. Thus, identification of the strain designs was apparently made possible only by reducing the design solution space. Moreover, the design solutions found for these five CCs either have low GCS values, require many interventions, or both, indicating their general

Table 2

44 coupling chemistries (CCs) selected for growth-coupling simulations in the *E. coli* metabolic model iML1515 to derive enzyme selection system designs. R represents a generic compound in the reaction stoichiometries and is a placeholder for an enzyme-specific reactant. All other compound abbreviations follow the standard nomenclature from the BIGG database. Note that a given EC annotation represents the majority and not necessarily all enzymes within the corresponding CC.

CC ID	Class name	EC	Generic reaction	CC ID	Class name	EC	Generic reaction
hdg1	Dehydrogenases	1.1	R + glc_D → R-H2 + dgo15lac	dma1	Dimethylallyltransferase	2.5.1	R + dmp → R-dimethylallyl + ppi
hdg3	Dehydrogenases	1.1	R + lac_D → R-H2 + pyr	ger1	Geranyltransferase	2.5.1	R + grdp → R-geranyl + ppi
hdg4	Reductases	1.8.4	R + trdrd → R-H2 + trdox	pyr1	Pyruvate transferase	2.5.1	R + pep + h2o → R-pyruvate + pi
hdx1	Hydroxylases	1.14.11	R + ak + o2 → R-hydroxylated + succ + co2	ser1	Serine transferases	2.5.1	R + acser → R-serine + ac
mth1	Methyltransferases	2.1.1	R + amet → R-methylated + ahcys	amn1	Transaminase	2.6.1	R + glu_L → R-aminated + ak
mth2	Methyltransferases	2.1.1	5mthf + R → R-methylated2 + thf	amn1_rev	Transaminase	2.6.1	R-aminated + ak → R + glu_L
mth3	Methyltransferases	2.1.1	hcys_L + R → R-methylated3 + met_L	amn2	Transaminase	2.6.1	R + gln_L → R-aminated2 + glu_L
mth4	Methyltransferases	2.1.1	mlthf + R → R-methylated4 + dhf	amn3	Transaminase	2.6.1	R + ala_L → R-aminated3 + pyr
mth5	Methyltransferases	2.1.1	R + amet → met_L + dad_5 + R-methylated5	amn3_rev	Transaminase	2.6.1	R-aminated3 + pyr → R + ala_L
for1	Formyltransferases	2.1.2	5fthf + R → R-formylated + thf	amn4	Transaminase	2.6.1	R + gly → R-aminated4 + glx
for2	Formyltransferases	2.1.2	10fthf + R → R-formylated2 + thf	amn4_rev	Transaminase	2.6.1	R-aminated4 + glx → R + gly
hmt1	Hydroxymethyltransferases	2.1.2	R + h2o + mlthf → R-hydroxymethylated + thf	pho1	Phosphotransferases	2.7.1	R + pep → R-phosphorylated + pyr
cbm1	Carbamoyltransferases	2.1.3	R + cbp → R-carbamoyl + h + pi	slu1	Sulfurtransferases	2.8.1	R + cys_L → R-sulfonated + ala_L
amd1	Amidinotransferase	2.1.4	R + arg_L → R-amidine + orn	slu2	Sulfurtransferases	2.8.1	R + tsul → R-sulfonated2 + so3
act1	Acetyltransferases	2.3.1	R + accoa → R-acetyl + coa	slo1	Sulfotransferases	2.8.2	R + paps → R-sulfate + pap
act2	Reductases	1.21.4	R + actp + h2o → R-acetyl2 + pi	coa1	CoA-transferases	2.8.3	R + accoa → R-CoA + ac
mal1	Malonyltransferases	2.3.1	R + malcoa → R-malonyl + coa	coa2	CoA-transferases	2.8.3	R + succoa → R-CoA + succ
suc1	Succinyltransferases	2.3.1	R + succoa → R-succinyl + coa	coa3	CoA-transferases	2.8.3	R + forcoa → R-CoA + for
rph1	Phosphoribosyltransferase	2.4.2	R + prpp → R-phosphoribosylated + ppi	sel1	Selenotransferases	2.9.1	R + selnp → R-selenated + pi
acp1	3-amino-3-carboxypropyltransferases	2.5.1	R + amet → R-aminocarboxypropyl + 5mta	glu1	Glutamyltransferases	6.3.2	R + glu_L + atp → R-glutamated + adp + pi
ade1	Adenosyltransferases	2.5.1	R + amet → R-adenosyl + met_L	hmt1_rev	Hydroxymethyltransferases	2.1.2	R-hydroxymethylated + thf → R + h2o + mlthf
app1	Aminopropyltransferases	2.5.1	R + ametam → R-aminopropyl + 5mta	amd1_rev	Amidinotransferase	2.1.4	R-amidine + orn → R + arg_L

Table 3

Model variants or parameter adaptations applied to compute growth-coupling strain designs for each of the 44 selected coupling chemistries. All models base on the “Standard” parameter set and were adapted as described. The purpose of applying a model, i.e., the research question that is to be answered, is additionally given.

Model ID	Model specifics	Purpose
Standard	iML1515 under aerobic conditions with model and optimization parameters as described in the Methods section	Use the full metabolic capability of <i>E. coli</i> and all common design variables.
No cofeed	Disabled selection of a metabolite cofeed in addition to a carbon source	Does the use of an external metabolite cofeed promote the establishment of growth-coupling?
N-coupling	Inactivated ammonia exchange to remove the primary nitrogen source	Does an enforced nitrogen supply via a cofeed metabolite and the target reaction lead to better growth-coupling strategies?
Anaerobic	Inactivated oxygen exchange to model anaerobic conditions	Is growth-coupling favored under anaerobic conditions?
MS56	Deleted genes from iML1515 to match the genome-reduced strain MS56	How does genome reduction influence the design of growth-coupling strategies?
DGF-298	Deleted genes from iML1515 to match the genome-reduced strain DGF-298	
Essential genes	Allow for the deletion of known essential genes	Does the essentiality of a few genes hinder the realization of growth-coupling strain designs?

deficiency for growth-coupling.

Similarly, the use of models representing the genome-reduced *E. coli* strains MS56 (Park et al., 2014) and DGF-298 (Hirokawa et al., 2013) slightly reduces the total number of genetic interventions (KS value < 0.12, p-value < 10⁻¹⁰, Kolmogorov-Smirnov (KS) test) and knockouts

(KS value < 0.17, p-value < 10⁻¹⁰, KS test) necessary for growth-coupling CCs (Fig. 4A and B). 20% of all designs computed with the wildtype *E. coli* model (Standard case) comprise six or fewer interventions compared to 29% and 30% of the designs for the MS56 and DGF-298 model, respectively. These differences mainly result from lower numbers of knockouts in the designs for the genome-reduced models. Five or less knockouts are suggested in 49% of the designs for the wildtype model compared to 62% and 65% for the MS56 and DGF-298 model, respectively. On the other hand, the number of heterologous reaction insertions slightly increases (KS value < 0.27, p-value < 10⁻¹⁰, KS test) when the genome-reduced models are applied (Fig. 4C). Two or more heterologous reactions must be inserted in 20% of all wildtype model designs compared to 39% and 46% of the MS56 and DGF-298 model designs, respectively.

Independent of the target CC or the chosen model variant, the choice of the primary carbon source does not necessarily restrict the identification of ESS strain designs. Various designs with differing carbon sources were identified for the CCs amenable to growth-coupling. A preference for one or more carbon sources among the designs could not be identified. The choice for a carbon source also does not predetermine the metabolic intervention strategy across CCs. The majority (73%) of randomly chosen pairs of growth-coupling designs for different CCs sharing the carbon source do not have a single knockout or insertion target in common (Fig. 4D). Similarly, growth-coupling design strategies for the same CC are generally different from each other, if different carbon sources are chosen. 56% of all screened design pairs of the same CC which do not share the carbon source have either none or one genetic intervention in common. Moreover, the distance in the metabolic network between a carbon source and the substrates of a CC does not appear to be of importance for finding an optimal medium composition for growth-coupling (KS value = 0.1, p-value = 3.0 × 10⁻⁶, KS test) (Fig. 4E). The distance distribution for carbon source-CC pairs among the identified growth-coupling designs is comparable to the distribution

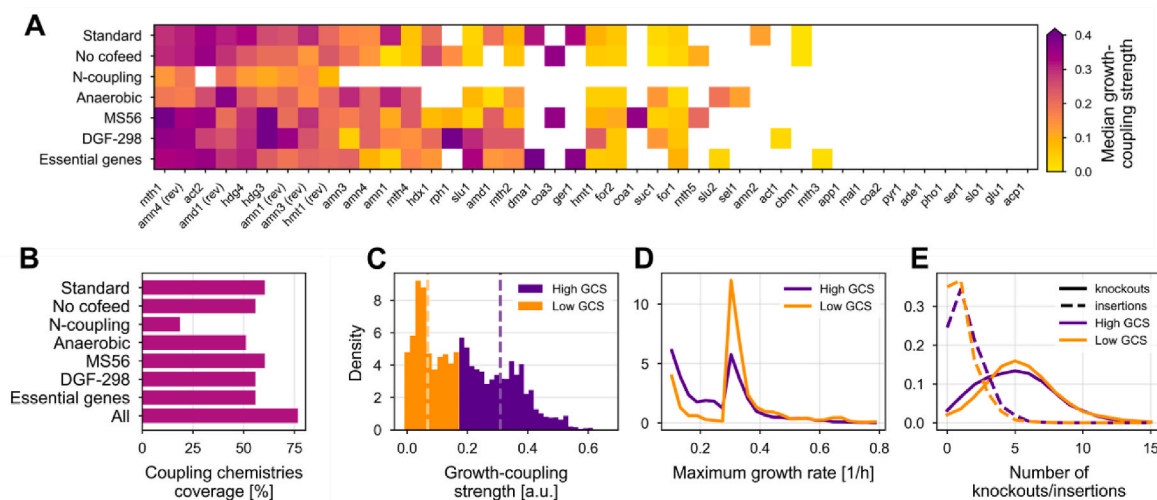


Fig. 3. Overview of the strain designs and their basic characteristics encountered for each coupling chemistry (CC). (A) The median growth-coupling strength (GCS) computed from all growth-coupling strain designs for any CC-model combination is shown. White areas denote a complete absence of valid strain design solutions. (B) Share of CCs for which at least one valid growth-coupling strain design was identified using a given model. (C) Distribution density of GCS values for all strain designs identified in this study. The threshold between high (purple) and low (orange) GCS designs was set at the GCS median of 0.18. The dashed lines separately mark the GCS median of all low and high GCS strain designs at 0.07 and 0.31, respectively. Distribution density of the (D) maximum growth rate and (E) the number of knockouts (straight lines) as well as heterologous reaction insertions (dashed lines) for high and low GCS strain designs. Distribution densities were normalized such that the integral for the whole range of each curve equals 1.

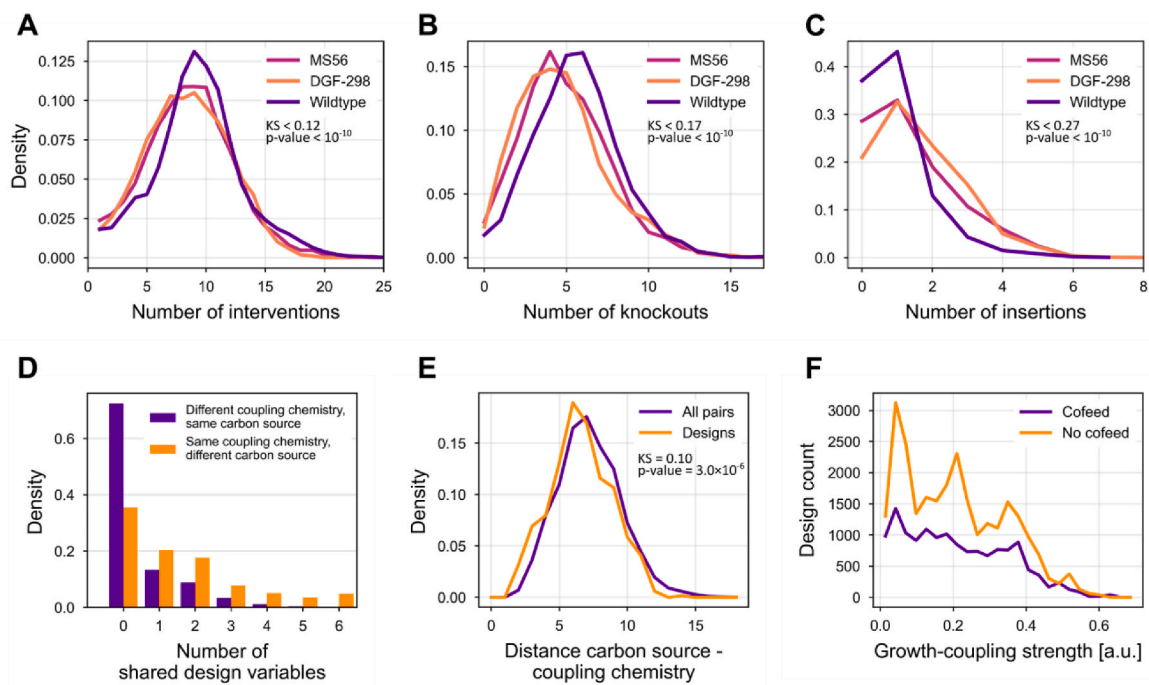


Fig. 4. Model- and design variable-specific characteristics of growth-coupling strain designs. Distributions of the number of interventions (A), knockouts (B), and heterologous insertions (C) for strain designs computed using models for the genome-reduced *E. coli* strains MS56 (pink) and DGF-298 (orange), as well as the wildtype strain (purple, Standard model). Upper bounds of KS test statistics and p-values are from Kolmogorov-Smirnov tests between genome-reduced and wildtype strains. (D) The number of shared design variables, i.e., shared knockout and heterologous insertion targets, are shown for 1,000,000 randomly chosen pairs of growth-coupling strain designs. Designs in each pair were computed for different coupling chemistries (CC) but a common carbon source (purple) or for the same CC but different carbon sources (orange). (E) Distribution densities of the distances between the carbon source (extracellular metabolite) and the substrate of a CC. The purple and orange lines show the distributions considering all possible CC-carbon source pairs and only those encountered in computed growth-coupling strain design solutions, respectively. KS test statistic and p-value are from a Kolmogorov-Smirnov test. (F) Growth-coupling strength distributions for all strain designs that consider an additional metabolite cofeed (purple) or only a single carbon source (orange). Strain designs for the model variant “No Cofeed” were excluded here.

derived from all possible carbon source-CC pairs. Thus, a possible hypothesis that growth-coupling of a target reaction is more easily established for shorter metabolic distances to the primary carbon source

cannot be confirmed based on the design set identified here. Similarly, a switch to anaerobic conditions neither facilitates nor significantly prohibits the identification ESS strain designs across CCs (Fig. 3A and B).

Besides choosing a primary carbon source, addition of a metabolite cofeed was generally allowed as a design variable except for the model variant “No Cofeed”. Examples of previous experimental realizations of growth-coupling systems have frequently applied a cofeed strategy to establish the coupling through an auxotrophy (Harder et al., 2016; Long et al., 2020; Luo et al., 2019, 2020). In principle, if a target reaction is directly involved in the catabolism of the cofed compound, growth-coupling may be established by making essential cellular functions strictly dependent on the metabolite cofeed and therefore the target reaction. This decoupling of essential functions from main metabolic activities promises to generate strong growth-coupling designs while maintaining cellular viability due to the provision of an additional carbon source and the avoidance of rerouting fluxes in central carbon metabolism. However, providing the opportunity for a second carbon source did not fundamentally enable growth-coupling of CCs or increase the GCS of identified strain designs (Figs. 3A and 4F). Even if the gcOpt algorithm is allowed to freely add a metabolite cofeed, i.e., designs for the model variant “No Cofeed” are not considered, growth-coupling strain designs without an additional cofeed predominate. Moreover, the GCS distributions are qualitatively similar for designs with and without a secondary carbon source (Fig. 4F). This observation also applies to the maximum growth rate (Fig. S8). Thus, considering or neglecting a metabolite cofeed does not have a significant effect on the success of finding appropriate growth-coupling strain designs. Addition of a secondary carbon source can still be considered to exhaust all valid strain design possibilities for the final, manual inspection and choice. In summary, a generally valid rule for choosing a carbon source for a given growth-coupling problem cannot be deduced, which highlights the importance of model-based approaches in the

decision-making for optimal medium composition and design strategies.

3.5. Growth-coupling strength correlates with flux distribution characteristics

As has been argued in the previous sections, the identification of growth-coupling designs for CCs is a challenging task and predictions about the genetic interventions or medium compositions cannot be reliably made *a priori*, e.g., based on the stoichiometry of the target reaction. On the other hand, growth-coupling strain designs follow general metabolic principles, as has been previously shown on designs coupling the secretion of various compounds to growth (Alter and Ebert, 2019). A major finding was that strong growth-coupling is enforced by making a target reaction an indispensable part of the major energy (ATP) supply route. The contribution of a target reaction to the cellular ATP supply was quantified by the ATP synthesis capability (ATPsc) measure, which describes the change in ATP production with an enforced unit change in the target reaction flux rate (cf. detailed descriptions in the Supplementary Methods section 1.7).

For almost all CC reactions, the ATPsc is negative in the *E. coli* wildtype model (Fig. 5A) indicating that activation of CCs generally exerts a metabolic burden on the cells by diverting fluxes away from optimal ATP and biomass synthesis routes. Applying the identified ESS strain designs increases the ATPsc to positive values across CCs (KS value = 0.74, p -value $< 10^{-10}$) (Fig. 5A, Supplementary Data) entailing that the metabolic network is reconfigured such that the activity of CCs positively contributes to the ATP supply. Furthermore, positive linear correlations between ATPsc and GCS are observed for a majority of CCs (91%) of which 80% are significant (p -value ≤ 0.01 , Pearson correlation)

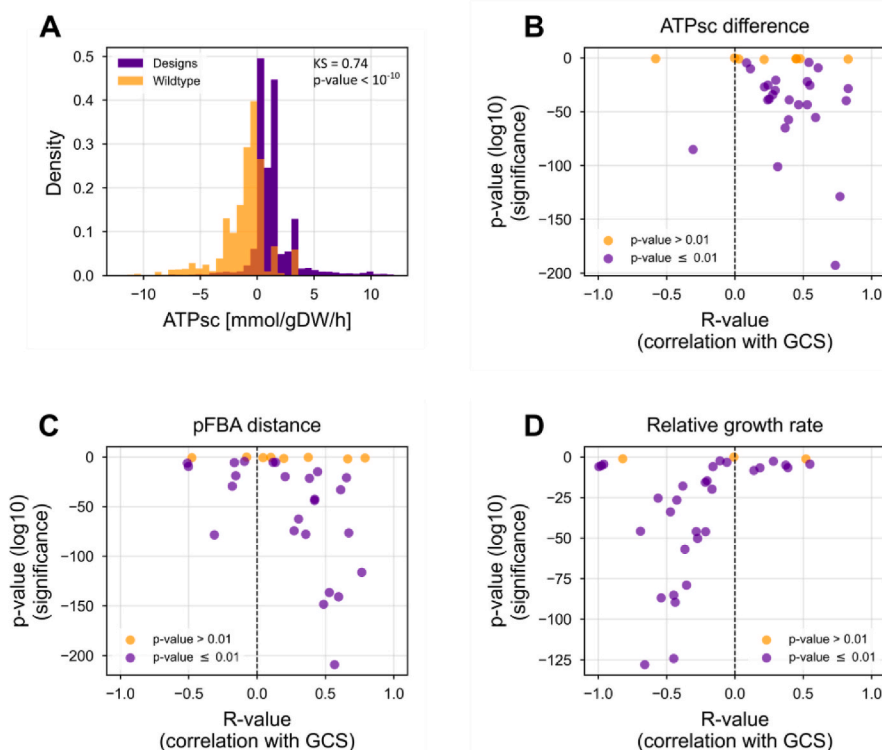


Fig. 5. (A) Absolute ATP synthesis capability (ATPsc) distributions for the generic stoichiometric reactions of all coupling chemistries (CC) in a wildtype metabolic model (orange) and in models with applied growth-coupling strain designs identified in this study (purple). KS test statistic and p -value are from a Kolmogorov-Smirnov test between the wildtype and design distribution. Several quantitative metabolic measures were computed for all identified strain design for a CC and linearly correlated with the growth-coupling strength (GCS) (B–D). The goodness-of-fit, i.e., the Pearson correlation coefficient (R-value), for each individual CC and the corresponding significance (p -value) are shown. The quantitative metabolic measures are (B) the difference of ATPsc values for a CC, (C) the distance between flux distribution vectors of pFBA solutions at optimal growth, and (D) the change in maximum growth rates between a wildtype and strain design model for a given design and CC.

according to the Pearson correlation coefficient (Fig. 5B, Fig. S9). Thus, there is a quantitative link between the coupling strength and the contribution of a CC to cellular ATP supply.

Slightly less prominent observations were made for correlations between GCS values and the distances between flux vectors of parsimonious Flux Balance Analysis solutions (pFBA) (Lewis et al., 2010) computed for wildtype and growth-coupling strain design models at optimal growth (Fig. 5C). Detailed correlations for each CC are shown in Fig. S10. For 55% of the CCs, the distance of pFBA solutions between growth-coupling designs and wildtype significantly increases with higher coupling strengths (p -value ≤ 0.01). Thus, for approximately half of the CCs, the implementation of ESS designs with increasing coupling strengths is associated with an increased need for significant metabolic flux rearrangements. On the other hand, for the majority (73%) of CCs, tighter coupling in terms of higher GCS values significantly leads to increasingly severe growth defects, as suggested by correlations between GCS and maximum growth rates simulated using models with growth-coupling strain designs (Fig. 5D, Fig. S11).

3.6. Identified enzyme selection systems cover a range of pathways relevant for bioproduction purposes

Although a robust ESS could not be identified for every CC, we still expect that a wide range of molecules can be growth-coupled using the designs proposed here. Especially since some CCs represent a multitude of unique, known reactions, there is potential for the development of biocatalysts for various distinct, industrially relevant products. To further explore this notion, we set out to map which compounds could be produced through a pathway of at most eight heterologous enzymes that ends with a viable CC, starting from *E. coli* native metabolites. The choice of native metabolites as a starting point may exclude reactions that modify complex, non-native compounds into industrially relevant

products, but it ensures a focus on whole-cell biocatalysis, where the product of interest can be synthesized from any of the carbon sources the cell is able to utilize. First, reactions matching the templates of viable CCs (Table 2) were identified in the KEGG database of biochemical reactions (Kanehisa et al., 2017). Subsequently, the KEGG database was traversed in a retrosynthetic manner, using a maximum of eight reaction steps, to arrive at metabolites endogenous to *E. coli* (Zhang et al., 2016), as determined by their presence in the iML1515 metabolic model (Monk et al., 2017). This methodology readily revealed 285 products that have the potential to be produced using an ESS. Since there was no lower bound imposed on the number of heterologous reactions, these products include 119 compounds already present in the *E. coli* metabolome. Generally, a large fraction of the amino acids and organic acids in the central carbon metabolism is used as a starting point for the heterologous biosynthetic pathways, in addition to the cofactors and more trivial compounds, such as H^+ , O_2 and H_2O . Most products are associated with transaminase and methyltransferase reactions (Fig. 6A), where the mth1 (S-adenosyl-L-methionine-dependent methyltransferase) CC accounts for 93 products. For 84% of the products, a single CC was identified, with the remaining compounds largely representing endogenous metabolites involved in a vast number of reactions. Among the products associated with just a single CC, a large fraction (95%) is non-native to *E. coli*. The redundancy of CCs for non-native compounds all occurred within the same class of chemistries (mainly transaminases), whereas compounds endogenous to *E. coli* sometimes span multiple classes.

We determined the depth of each pathway leading to a non-native product in *E. coli* and observed a notable trend wherein the number of pathways decreased as the reaction depth increased (Fig. 6B). Notably, loops were disallowed in the retrosynthetic analysis, preventing the overrepresentation of longer pathways due to, e.g., trivial back-and-forth traversals over reversible reactions. Given that almost two-thirds of the non-native products were reached only after at least two

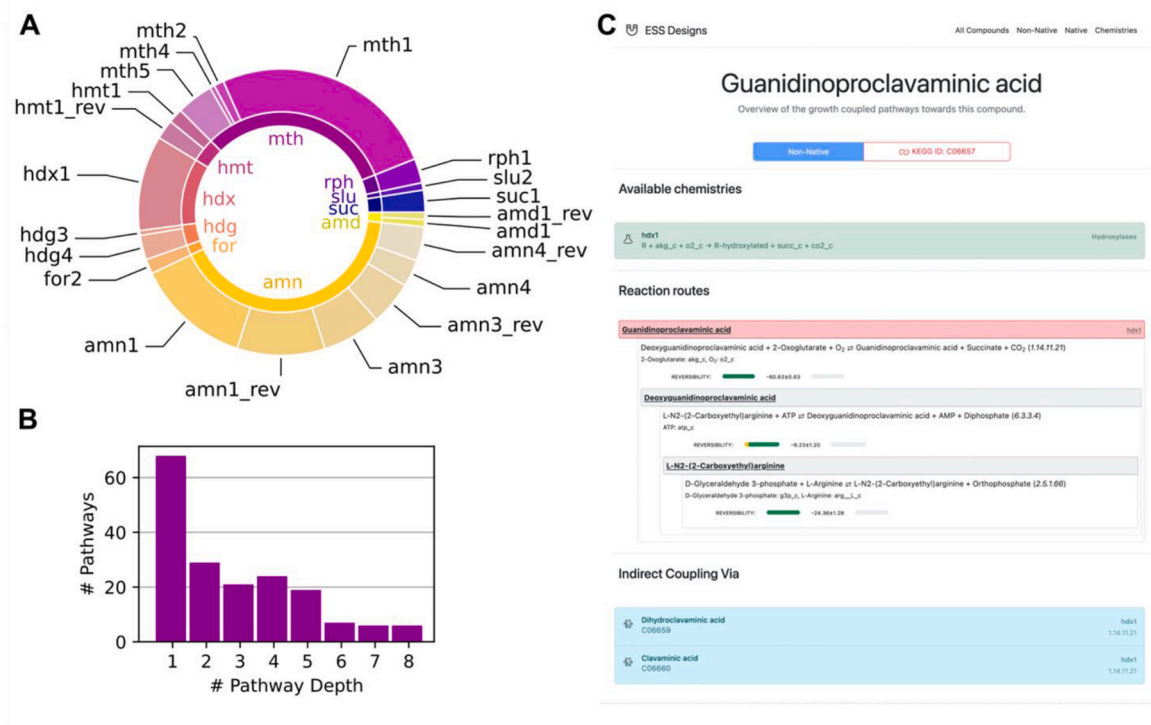


Fig. 6. (A) Classification of the pathways identified in a retrosynthetic analysis according to associated coupling chemistries (CCs). The inner ring of the donut chart represents the classification of the pathways into the various enzyme classes, whereas the outer ring specifies their subdivision into CC IDs (Table 2). (B) Distribution of pathway lengths, as measured by the maximum number of heterogenous reaction steps between a native *E. coli* metabolite and the non-native pathway product. The search depth of the retrosynthetic approach was limited to eight reaction steps. (C) Screenshot of a product information page of the ESS designs database (<https://biosustain.github.io/ESS-Designs/>). The page displays relevant CC information, the heterogenous pathway towards the product, and indirect coupling approaches, when applicable.

consecutive heterologous reaction steps, there is a wide array of downstream intermediates whose production can be optimized indirectly through selective pressure on the growth-coupled upstream product. Inclusion of this set of compounds adds 147 novel products to our analysis and provides an alternative, indirect coupling option for 45 already included products. Implementation and analysis of data obtained for indirectly growth-coupled products requires careful consideration, since not all adaptations necessarily benefit the production of the target compound and transport as well as toleration of the compound are not directly part of the growth-coupling mechanism.

An easily navigable instance of computed ESS strain designs as well as associated pathways and products is available at <https://biosustain.github.io/ESS-Designs/> (Fig. 6C). Access is provided to a subset of ESS strain designs computed in this work due to license terms. These include 25,505 unique designs for major enzyme classes such as methyltransferases and transaminases, among others. The online database can be searched by CCs or target compounds yielding suggestions for ESS strain designs computed in this study and growth-couplable pathways to the queried product, respectively. We further enriched the dataset with reaction reversibility estimates, as determined using the eQuilibrator web interface (Beber et al., 2021), and design features, such as the predicted maximum growth rate, to facilitate a direct assessment of the suitability of strain designs for a given application. Thus, a rich and accessible resource is provided for engineering cellular enzyme optimization platforms with an application focus for industrial biotechnology purposes.

4. Discussion

The development of microbial cell factories and cost-effective bioprocesses hinges on the availability of heterologous enzymes with high catalytic efficiencies towards the production of a desired compound. Here, we introduce the concept of enzyme selection systems (ESSs) as a whole-cell tool for improving enzyme properties and screening of environmental conditions on enzyme activity by using growth as a selection criterion. Besides the computational tools and workflows for individually deriving ESS engineering strategies, a database of designs for multi-enzyme engineering platforms is provided. These platforms give access to pathway engineering capabilities for the synthesis of various biotechnologically relevant compounds without the need for specialized laboratory equipment.

The first proof-of-principles for ESSs were conducted by Luo et al. (2020, 2019) and basic parts of the experimentally implemented strain designs were identified by the computational approach presented in this study. In the case of the ESS for methyltransferases (mth1, Table 2), model-based predictions suggest, among other strategies, the same medium composition (glucose and methionine as carbon sources) and heterologous insertions (cystathionine- β -synthase, cystathionine- γ -lyase), but differing gene knockouts as compared to the manually derived design. This is since the experimentally knocked out gene *cysE* (b3607) is deemed essential according to the essentiality dataset and was therefore excluded as a design variable. However, even by disregarding gene essentialities, the growth-coupling strategy from Luo et al. (2019) was not exactly reproduced, which is explained by the particularly low growth-coupling strength (GCS) of their design (0.004, Fig. S12). Due to this, the growth-coupling design algorithm skipped it in favor of strategies with higher GCS values. Given the successful application of the ESS for producing improved methyltransferase variants, weak coupling designs should not be generally dismissed as discussed in more detail below. Consequently, it is necessary to test diverse sets of model parameters and constraints to exhaustively compute growth-coupling designs. This provides the user with full flexibility in selecting an appropriate strategy. Nevertheless, rigid model parameterization, e.g., by using conditional experimental data, can ensure experimental feasibility while facilitating the computation of diverse growth-coupling strategies, including simple and tractable designs. To

generally encourage and guide the implementation of ESSs, we discuss the implications for design principles and the implementation process based on the properties of identified designs in the following.

4.1. Enzyme selection system designs show shared metabolic principles, but their identification eludes simple rules

The observed diversity of growth-coupling strain designs for coupling chemistries (CCs) yielding ESSs (Fig. 3) suggests that deriving growth-coupled systems or increasing the coupling strength is not necessarily restricted to complex genetic engineering strategies involving a multitude of interventions such as gene knockouts or heterologous reaction insertions. Ultimately, the involvement of the reactants of an enzymatic target reaction within the microbial host's metabolic network determines if a target enzyme's activity can be effectively coupled to growth. For example, the success rate of finding growth-coupling designs and the coupling strength for CCs transferring low-molecular groups, e.g., methyltransferases (mth1-5), transaminases (amn1-4), or dehydrogenases (hdg3), was relatively high compared to CCs such as CoA-transferases (coa1-3), glutamate ligases (glu1), or malonyl transferases (mal1) (Fig. 3A). In fact, among the 11 CCs (23%) for which no valid growth-coupling designs were found, CoA, malonyl, phosphate, or 3-amino-3-carboxypropyl groups are involved among others in the reaction stoichiometries highlighting the difficulty to growth-couple reactions transferring or otherwise acting on large or highly connected molecular groups and cofactors. Even if designs were found for CCs involved in the transfer of high-molecular groups, e.g., carbamoyl (cbm1) or pyruvate (pyr1), these generally require an increased number of interventions or show low GCS values (Fig. S7).

However, a strict rule for predicting the success of growth-coupling a CC or the design itself based on the reaction stoichiometry could not be deduced. For example, neither the connectivity of reactants nor their molecular weights were found to correlate with the ability to growth-couple a corresponding CC or the growth-coupling strength of identified strain designs (data not shown). Thus, designing ESSs or other growth-coupled systems eludes intuition and requires a model-based optimization approach. On the other hand, we have demonstrated that the establishment of ESSs for CCs, if amenable to growth-coupling, is flexible in the choice of the primary carbon source (Fig. 4D) or the oxygen availability (Fig. 3). This flexibility allows ESS design strategies to be tailored to a given biotechnological application and its process requirements.

For the CCs amenable to practical growth-coupling strategies, the complexity of strain designs was further decreased in genome-reduced strains (Fig. 4A–C). Thus, the implementation of ESSs can benefit from genome reduction in terms of engineering efficiency. This potential may be enhanced by creating genome reduced strains with a minimal metabolic repertoire. So far, metabolic network-agnostic approaches were used for deducing genome reduction strategies by focusing on the deletion of insertion sequences or large clusters of non-essential parts of the genome, as was the case for MS56 (Park et al., 2014) and DGF-298 (Hirokawa et al., 2013). A strategy that precisely and specifically reduces non-essential and redundant parts of the metabolic network while accounting for synthetic lethality promises to advance genome reduction in *E. coli* and reduce the number of interventions necessary to establish growth-coupling designs.

While the process of deriving ESS strain designs is non-trivial, the intertwining of the activity of a target reaction with cellular energy metabolism was identified as a general principle governing ESS strain designs (section 3.5). Although an interdependence between ATP supply and target reaction activity may seem trivial for a growth-coupled system, it has fundamental implications for the metabolic principle of growth-coupling strain designs: Since ATP is the central energy currency and is involved in many cellular processes, growth-coupling of CCs is imposed by introducing a strict link between the target reaction and metabolic activity as a whole rather than the synthesis of just one or a

few biomass precursors. This is in line with growth-coupling principles derived from designs coupling various metabolite secretion pathways to growth (Alter and Ebert, 2019).

4.2. Implementation of ESSs requires a trade-off between growth-coupling strength and growth rate

The analysis of identified enzyme selection system (ESS) strain designs in this work has implications for the implementation of ESSs and growth-coupling strategies in general. It has been recognized that high coupling strengths are desirable in growth-coupling systems to guarantee an evolutionarily robust activity of a target functionality (Legon et al., 2022), i.e., enzymatic catalytic activity in the case of ESSs. High growth-coupling strengths (GCS) imply a strong selection pressure on the growth-coupled target enzyme's activity, since it must carry high catalytic flux per growth rate unit. This enables better control of the selection pressure through the target enzyme expression strength, while avoiding an increased metabolic burden on the ESS host imposed by an over-allocation of cellular resources toward enzyme expression. Moreover, a tighter growth-coupling results in a more direct, and therefore, more accurate readout of the target enzyme's activity based on the growth rate of the ESS, which is useful in applications where the ESS is used as an enzyme activity screening system.

However, higher coupling strengths divert (carbon) flux distributions away from optimal energy, redox, and biomass precursor generation and towards the target reaction, thus compromises cellular viability (Fig. 5). Furthermore, model-based phenotype predictions are usually only reproducible in evolved strains (Fong and Palsson, 2004; Ibarra et al., 2002). Thus, the experimental cellular viability of ESSs is expected to be even lower than the *in silico* predictions, as was demonstrated by a metabolic model-based investigation of a genotype-phenotype dataset of growth-coupling strain designs extracted from literature (King et al., 2016). Immoderate growth defects prior to laboratory evolution can hamper applicability of ESSs and growth-coupling designs in general, particularly if a high initial growth rate is required for applying an ESS in an ALE experiment. Besides ALE, artificially generated enzyme mutant libraries can also be applied to ESSs, thus circumventing the need for a minimum cellular viability and allowing for augmenting the selection process with knowledge from previous structural analyses or specifically increasing the mutation frequency.

In summary, ESS strain designs enabling only suboptimal growth-coupling still have their justification for practical implementations. The diversity of computed growth-coupling strain designs in this work allows for matching the design strategy to the desired strain properties, e.g., to enable a reasonable coupling strength between the target enzyme and growth while maintaining high cellular viability.

4.3. Enzyme selection system designs can escape growth-coupling through unknown enzymatic activities

Although model microorganisms, such as *E. coli*, have been intensively studied for decades, our knowledge about their enzymatic and metabolic repertoire is generally incomplete or not fully covered by metabolic models. For an implemented ESS design, the upregulation of unknown, shallow, or otherwise suppressed cellular functions can cause a metabolic escape, allowing the cell to bypass the coupling and evolve to higher growth rates without improving the target enzyme activity. Such functions can include enzyme promiscuity toward an alternative substrate or the metabolic activity of an unknown enzyme. A previous study has shown that even low-level promiscuous enzyme activities are able to sustain novel phenotypes if evolved and accentuated in ALE experiments (Guzmán et al., 2019). Thus, a high selective pressure on engineered strains is likely to evoke metabolic escapes from growth-coupling of the CC and the corresponding enzyme during ALE experiments. An emerged metabolic escape can be mitigated by eliminating the upregulated, unwanted function or gene. For strain designs

with a strong growth-coupling strength and thus a high chance of triggering a metabolic escape from growth-coupling, repeated design-build-test-learn (DBTL) cycles are likely to be required to establish a robust ESS for a given CC or enzyme. While the process of mitigating metabolic escapes delays the construction of an ESS, it offers opportunities for metabolic discoveries and the improvement of metabolic models as a byproduct. The emergence of metabolic escapes triggered by unknown cellular functionalities may be avoided in the first place by using genome-reduced strains as ESS hosts. By applying a host strain lacking non-essential genomic parts, the chance for metabolic escapes caused by an evolved upregulation of uncharacterized metabolic functions is reduced. Consequently, the need for feedback loops between design and testing of ESS strains is reduced.

5. Conclusion

A general concept for the construction of microbial chassis strains, so called enzyme selection systems (ESS), has been introduced, allowing the targeted *in vivo* engineering of a class of enzymes using adaptive laboratory evolution (ALE) experiments. The concept is based on coupling chemistries (CC), i.e., generic reactions representing multiple enzymes of a particular class. A large-scale *in silico* study was conducted to identify growth-coupling strain designs for selected CCs and to build a database of ESSs for a broad range of enzymes. Although ESS designs could not be determined for all selected enzymes classes or CCs, the ESS database still covers a wide range of metabolic pathways and associated compounds, thus opening the possibility for in-place pathway engineering approaches for many applications. It was shown that practical, experimentally accessible growth-coupling designs generally exist, and no specific requirements are placed on the microbial base strain or the experimental conditions. Thus, there is a certain degree of freedom in choosing a favorable carbon source and metabolic engineering strategy and this study provides the database for engineering experts to make suitable choices. Moreover, the computational tools and workflow enable the comprehensive search for growth- or flux-coupling design strategies for any given application or microbial host. The decision-making process for a growth-coupling design will always involve a compromise between viability of the mutant strain or ESS and the growth-coupling strength. Future experimental efforts will reveal the influence of the coupling strength, i.e., the level of selection pressure on a target gene, on the evolution of improved target gene variants and its optimal balance with inevitable growth defects induced by design. For example, high ESS viability and prevention of metabolic escapes from growth-coupling, both assured by low coupling strength design choices, may be more advantageous for productive ESSs than strongly coupled systems with heavily rerouted metabolic fluxes, which are likely to develop unforeseeable phenotypes. Considering the power of the first experimental examples of ESSs (Luo et al., 2019, 2020), the presented concept of ESS, the corresponding design database, and the computed characteristic features of the designs within, are a cornerstone for establishing ESSs as a tool for enzyme engineering purposes.

CRedit authorship contribution statement

Tobias B. Alter: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Conceptualization. **Pascal A. Pieters:** Writing – review & editing, Writing – original draft, Software, Formal analysis. **Colton J. Lloyd:** Writing – review & editing, Software, Conceptualization. **Adam M. Feist:** Writing – review & editing, Conceptualization. **Emre Özdemir:** Writing – review & editing, Supervision. **Bernhard O. Palsson:** Writing – review & editing, Supervision, Conceptualization. **Daniel C. Zielinski:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare that a fraction of the designs produced as part of this study are part of a licensing arrangement.

Acknowledgments

The authors acknowledge funding from the Novo Nordisk Foundation (NNF20CC0035580) and thank Lei Yang and Jinbei Li for fruitful discussions of this work from an experimental point of view.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.mec.2025.e00257>.

Data availability

I have shared the link to the data and code as outlined in the manuscript

References

- Alter, T.B., Ebert, B.E., 2019. Determination of growth-coupling strategies and their underlying principles. *BMC Bioinform.* 20, 447. <https://doi.org/10.1186/s12859-019-2946-7>.
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50, 4402–4410. <https://doi.org/10.1021/bi2002289>.
- Beber, M.E., Gollub, M.G., Mozaffari, D., Shebek, K.M., Flamholz, A.I., Milo, R., Noor, E., 2021. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res.* 1–7. <https://doi.org/10.1093/nar/gkab1106>.
- Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657. <https://doi.org/10.1002/bit.10803>.
- Cheng, F., Zhu, L., Schwaneberg, U., 2015. Directed evolution 2.0: improving and deciphering enzyme properties. *Chem. Commun.* 51, 9760–9772. <https://doi.org/10.1039/C5CC01594D>.
- Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7, 74. <https://doi.org/10.1186/1752-0509-7-74>.
- Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgård, M.J., Palsson, B.O., 2010. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab. Eng.* 12, 173–186. <https://doi.org/10.1016/j.ymben.2009.10.003>.
- Fong, S.S., Palsson, B.O., 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat. Genet.* 36, 1056–1058. <https://doi.org/10.1038/ng1432>.
- Guzmán, G.I., Sandberg, T.E., LaCroix, R.A., Nyerges, Á., Papp, H., de Raad, M., King, Z. A., Hefner, Y., Northen, T.R., Notebaart, R.A., Pál, C., Palsson, B.O., Papp, B., Feist, A.M., 2019. Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.* 15, e8462. <https://doi.org/10.15252/msb.20188462>.
- Harder, B.J., Bettenbrock, K., Klamt, S., 2016. Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. *Metab. Eng.* 38, 29–37. <https://doi.org/10.1016/j.ymben.2016.05.008>.
- Hillson, N., Caddick, M., Cai, Y., Carrasco, J.A., Chang, M.W., Curach, N.C., Bell, D.J., Le Feuvre, R., Friedman, D.C., Fu, X., Gold, N.D., Herrgård, M.J., Holowko, M.B., Johnson, J.R., Johnson, R.A., Keasling, J.D., Kitney, R.I., Kondo, A., Liu, C., Martin, V.J.J., Menolascina, F., Ogino, C., Patron, N.J., Pavan, M., Poh, C.L., Pretorius, I.S., Rosser, S.J., Scrutton, N.S., Storch, M., Tekotte, H., Travnik, E., Vickers, C.E., Yew, W.S., Yuan, Y., Zhao, H., Freemont, P.S., 2019. Building a global alliance of biofoundries. *Nat. Commun.* 10, 2040. <https://doi.org/10.1038/s41467-019-10079-2>.
- Hirokawa, Y., Kawano, H., Tanaka-Masuda, K., Nakamura, N., Nakagawa, A., Ito, M., Mori, H., Oshima, T., Ogasawara, N., 2013. Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *J. Biosci. Bioeng.* 116, 52–58. <https://doi.org/10.1016/j.jbiosc.2013.01.010>.
- Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420, 186–189. <https://doi.org/10.1038/nature01149>.
- Jensen, K., Broeken, V., Hansen, A.S.L., Sonnenschein, N., Herrgård, M.J., 2019. OptCouple: joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. *Metab. Eng. Commun.* 8, e00087. <https://doi.org/10.1016/j.mec.2019.e00087>.
- Jeske, L., Placzek, S., Schomburg, I., Chang, A., Schomburg, D., 2019. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 47, D542–D549. <https://doi.org/10.1093/nar/gky1048>.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
- King, Z.A., Brien, E.J.O., Feist, A.M., Palsson, B.O., 2016. Literature mining supports a next-generation modeling approach to predict cellular byproduct secretion. *Metab. Eng.* 1–8. <https://doi.org/10.1016/j.ymben.2016.12.004>.
- Legon, L., Corre, C., Bates, D.G., Mannan, A.A., 2022. gcFront: a tool for determining a Pareto front of growth-coupled cell factory designs. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btac376>.
- Lewis, N.E., Hixon, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith, R.D., Palsson, B.O., 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6. <https://doi.org/10.1038/msb.2010.47>.
- Long, M., Xu, M., Qiao, Z., Ma, Z., Osire, T., Yang, T., Zhang, X., Shao, M., Rao, Z., 2020. Directed evolution of ornithine cyclodeaminase using an EvolvR-based growth-coupling strategy for efficient biosynthesis of <sc>L</sc>-proline. *ACS Synth. Biol.* 9, 1855–1863. <https://doi.org/10.1021/acssynbio.0c00198>.
- Luo, H., Hansen, A.S.L., Yang, L., Schneider, K., Kristensen, M., Christensen, U., Christensen, H.B., Du, B., Özdemir, E., Feist, A.M., Keasling, J.D., Jensen, M.K., Herrgård, M.J., Palsson, B.O., 2019. Coupling S-adenosylmethionine-dependent methylation to growth: design and uses. *PLoS Biol.* 17, e2007050. <https://doi.org/10.1371/journal.pbio.2007050>.
- Luo, H., Yang, L., Kim, S.H., Wulff, T., Feist, A.M., Herrgård, M., Palsson, B.O., 2020. Directed metabolic pathway evolution enables functional pterin-dependent aromatic-amino-acid hydroxylation in *Escherichia coli*. *ACS Synth. Biol.* 9, 494–499. <https://doi.org/10.1021/acssynbio.9b00488>.
- Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A.M., Palsson, B.O., 2017. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Biotechnol.* 35, 904–908. <https://doi.org/10.1038/nbt.3956>.
- Park, M.K., Lee, S.H., Yang, K.S., Jung, S.-C., Lee, J.H., Kim, S.C., 2014. Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Appl. Microbiol. Biotechnol.* 98, 6701–6713. <https://doi.org/10.1007/s00253-014-5739-y>.
- Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 2367–2376. <https://doi.org/10.1101/gr.2872004>.
- Sandberg, T.E., Salazar, M.J., Weng, L.L., Palsson, B.O., Feist, A.M., 2019. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* 56, 1–16. <https://doi.org/10.1016/j.ymben.2019.08.004>.
- Shivange, A.V., Marienhagen, J., Mundhada, H., Schenk, A., Schwaneberg, U., 2009. Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* 13, 19–25. <https://doi.org/10.1016/j.cbpa.2009.01.019>.
- Tervo, C.J., Reed, J.L., 2012. FOCAL: an experimental design tool for systematizing metabolic discoveries and model development. *Genome Biol.* 13, R116. <https://doi.org/10.1186/gb-2012-13-12-r116>.
- von Kamp, A., Klamt, S., 2014. Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput. Biol.* 10. <https://doi.org/10.1371/journal.pcbi.1003378>.
- Zhang, X., Tervo, C.J., Reed, J.L., 2016. Metabolic assessment of *E. coli* as a Biofactory for commercial products. *Metab. Eng.* 35, 64–74. <https://doi.org/10.1016/j.ymben.2016.01.007>.