



Rational Metabolic Pathway Prediction and Design: Computational Tools and Their Applications for Yeast Systems and Synthetic Biology

Pedro A. Saa

Abstract

Continuous progress in metabolic engineering of microbial cell factories like yeast requires the support of computational tools for finding novel unintuitive biotransformations routes. In this chapter, a succinct overview is provided of the most relevant computational tools for pathway prediction by retro-biosynthesis, and pathway design through stoichiometry-based optimization methods. Illustrative case studies are also presented showcasing different strategies for pathway optimization in yeast, namely redox cofactor balancing, improved precursor supply, and heterologous expression of carbon fixation pathways. Finally, challenges and limitations hindering the broad adoption and implementation of these tools for metabolic engineering will be discussed.

Keywords

Pathway design • Pathway analysis • Metabolic engineering • Yeast

1 Introduction

Advancements in metabolic engineering and synthetic biology have enabled accelerated engineering of microbial factories for the production of valuable chemicals (Smolke and Tyo 2012; Lee and Kim 2015; Isaacs et al. 2011), realizing the

P. A. Saa (✉)

Department of Chemical and Bioprocess Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

e-mail: pnsaa@ing.puc.cl

Institute for Mathematical and Computational Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

F. Darvishi Harzevili (ed.), *Synthetic Biology of Yeasts*,

https://doi.org/10.1007/978-3-030-89680-5_1

promise of a more sustainable (bio)economy (Voigt 2020). To keep pace with these expectations, pathway prediction and design play a crucial role in finding novel pathways for various applications like drug discovery (Galanie et al. 2015; Moura et al. 2016; Hafner et al. 2021) and value-added biochemical production (Yim et al. 2011; Tokic et al. 2018; Henry et al. 2010a, b). In this scenario, metabolic workhorses like yeast could be greatly benefited by broadening their product spectrum and improving their metabolic capabilities and performance in terms of their yields, titers, and productivities (Nielsen and Keasling 2016; Ko et al. 2020). For this task, progress in computational tools and methods capable of guiding experimental efforts is crucial for the optimization of cellular metabolism and incorporation of synthetic designs for the production of unnatural heterologous compounds.

In this chapter, an overview of the most relevant retro-biosynthesis and pathway optimization methods is provided with a focus on tools with direct application in metabolic engineering tasks. Starting with the reconstruction of a comprehensive reaction network from public databases and resources, both retro-biosynthesis tools for de novo pathway prediction and stoichiometry-based pathway optimization methods for metabolic redesign are described. Particularly in the latter case, convenient engineering objectives taking into consideration product yield, cofactor use, thermodynamic plausibility, and enzyme cost are discussed. Additionally, several relevant pathway engineering case studies in yeast are also presented, highlighting the improvement potential from the implementation of rational pathway designs. Finally, perspectives on the increasing adoption of these tools for metabolic engineering as well as limitations reducing their effectiveness are discussed.

2 In Silico Pathway Prediction and Design

Retro-biosynthesis and stoichiometry-based optimization methods have been established for pathway design and prediction, differing mostly in their scope and methodology. While both tools generate metabolic pathways producing the target metabolite, they do so by applying fundamentally different reaction network representations (i.e., graph or stoichiometric matrix) and search algorithms (i.e., optimization-based enumeration or retro-synthetic search) (the reader is referred to Wang et al. (2017) for a comprehensive review). Furthermore, their computational complexity and efficacy can vary significantly depending on the product of interest, and thus, careful selection of the appropriate tool for the case at hand is a must (Saa et al. 2019). In the following, the most relevant data- and knowledge-bases for reconstructing and parameterizing reaction networks are presented, which constitute the starting point for the application of any of these tools (Fig. 1a). Then, the most relevant retro-biosynthesis (Fig. 1b) and stoichiometry-based optimization methods for pathway prediction and design (Fig. 1c) are described.

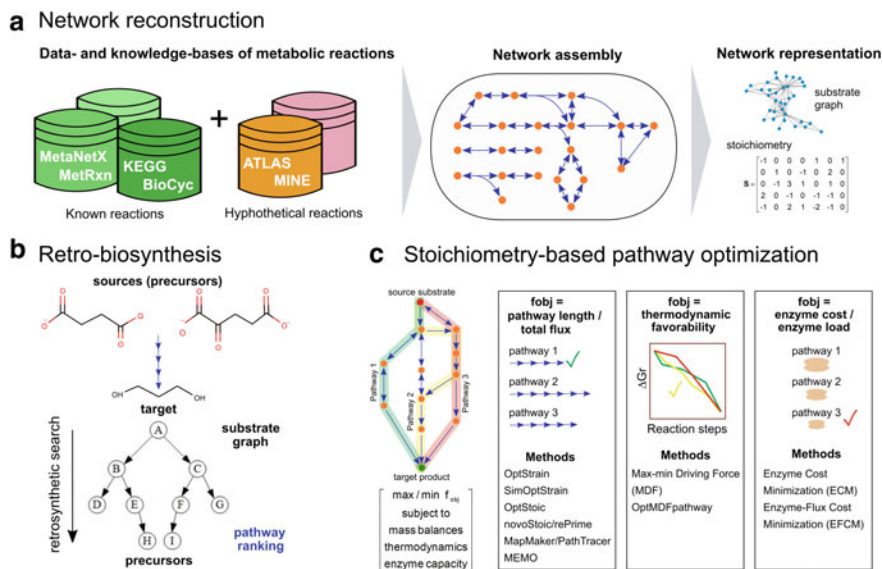


Fig. 1 Workflow for reaction network reconstruction and application of metabolic pathway prediction and design tools. **a** Assembly of accumulated metabolic reaction data into a comprehensive reaction network is a requirement for the application of the reviewed tools. Depending on the application objective, different network representations are employed for either predicting de novo pathways (i.e., retrosynthesis typically using a graph representation), or (re)designing pathways for higher metabolic performance (i.e., optimization-based pathway design using an stoichiometric representation). **b** Retro-biosynthetic tools explore a substrate graph seeking to connect the target molecule with some predefined precursors. Starting with the target molecule and moving backwards, these tools can generate several possible pathways that are typically ranked using different criteria (e.g., length, enzyme availability, thermodynamics, among others). **c** Stoichiometry-based pathway prediction methods employ a reaction network with known and fixed reactions to enumerate mass-balanced pathways that optimize a desired objective such as product yield, pathway length, thermodynamic favorability, and enzyme cost. In this case, different types of constraints can be defined to restrict the feasible solution space and narrow the search upfront

2.1 Data- and Knowledge-Bases for Metabolic Reaction Network Reconstruction

Databases for pathway search are an absolute requirement for exploring the feasible reaction space, as they contain the critical information of how metabolites are connected to others through biochemical reactions. There are numerous public data- and knowledge-bases populated with metabolic reaction data. Among the most popular, KEGG (Kanehisa et al. 2016), MetaCyc (Caspi et al. 2016), BIGG (King et al. 2016), KBase (Arkin et al. 2018), ModelSEED (Henry et al. 2010a, b), MetRxn (Kumar et al. 2012), and MetaNetX (Ganter et al. 2013; Moretti et al. 2021) stand out to name a few (for a more details refer to Wang et al. (2017)). Some of these databases (KEGG, MetaCyc, and Kbase) integrate multiple sources of biological information, e.g., genetic, molecular, physicochemical,

and experimental, which makes them not only useful for metabolic pathway prediction purposes but also data integration (Lewis et al. 2012). The rest of the databases are mostly devoted to metabolic network reconstruction, offering either highly curated reconstructions for specific organisms (e.g., BIGG) or broader albeit possibly less curated biochemical reaction networks (e.g., MetRxn, ModelSEED, and MetaNetX). Ultimately, the modeling purpose will dictate the most convenient source of information considering their specific scope, breadth, and information quality. Complementary databases like BRENDA (Jeske et al. 2019) (kinetic information) and eQuilibrator (Flamholz et al. 2012) (thermodynamic information) also constitute valuable resources for parameterizing different optimization formulations.

The aforementioned databases contain information for known reactions, which may restrict the pathway search considering the current enzymatic knowledge gaps. Resources like the ATLAS of Biochemistry (Hadadi et al. 2016) (derived from the BNICE tool (Hatzimanikatis et al. 2005)) and MINE (Jeffryes et al. 2015) offer larger networks including hypothetical reactions and metabolites that can expand the reachable chemical space and allow higher complexity. Briefly, these resources exploit user-defined reaction rules that can act on chemically similar compounds, thereby yielding new hypothetical reactions. The latter reactions have recently been shown to enable filling some of the gaps in current enzyme-reaction associations (Hadadi et al. 2019). Lastly, another significant and recent tool for proposing hypothetical reactions that has been employed for pathway prediction is rePrime used by the novoStoic tool (Kumar et al. 2018) (see subsection 2.3 for more details). The former method extracts reaction rules from molecular signatures found in annotated reactions—defined by the presence of a set of chemical ‘moieties’—for proposing hypothetical enzymatic transformations with a high structural encoding fidelity. Unfortunately, this tool currently lacks an associated open database for its use.

2.2 Pathway Prediction Using Retro-Biosynthesis Tools

Firstly, a distinction is made between retro-biosynthesis and classical retrosynthesis, as the latter is focused on the design of chemical reaction pathways, typically without relying on enzyme catalysis (Lin et al. 2019). Retro-biosynthesis tools seek to identify *de novo* biosynthetic pathways for the production of valuable compounds from inexpensive precursors using known and hypothetical enzyme activities (Wang et al. 2017; Lin et al. 2019). Another—though less explored—application of these tools involves the opposite, that is, the prediction of novel enzymatic routes for the degradation of recalcitrant compounds, e.g., for bioremediation purposes (Finley et al. 2009, 2010; Ellis et al. 2006). For pathway prediction, retro-biosynthesis tools explore the full chemical space for synthetic pathways toward the target compound. For this task, these tools typically represent the network as a (substrate) graph that can be readily traversed using known enumeration algorithms. Graph traversal is possible by connecting the substrates using

various criteria based on structural (chemical) similarity, reaction promiscuity, and defined reaction rules. In the following, a relevant subset of retro-biosynthesis tools employed for metabolic engineering/synthetic biology applications is described (Table 1).

One of the most established tools for de novo pathway retro-biosynthesis is BNICE (Hatzimanikatis et al. 2005). This framework employs predefined ‘generalized enzymatic reaction rules’ (encoded in a bond-electron matrix) that are applied to precursor molecules on their reactive sites to yield new product molecules. BNICE uses a substrate graph representation of the chemical network, which can be traversed using graph search algorithms starting from the target compound and moving backwards until connecting with one of the defined precursors. Different pruning criteria are employed to keep the search breadth computationally tractable.

Table 1 Retro-biosynthesis tools for metabolic pathway prediction

Tool	Database	Pathway scoring	Applications and references	Source
BNICE	KEGG, ATLAS	Pruning criteria assessment (thermodynamics, pathway length, etc.)	In silico identification of novel pathways and enzyme candidates validated experimentally for the production of (S)-tetrahydropalmitine derived from intermediates of the known benzyloisoquinoline alkaloid biosynthetic pathway (Hafner et al. 2021) In silico evaluation of putative feasible biodegradation pathways for 1,2,4-trichlorobenzene (Finley et al. 2010) Enzyme annotation for orphan and novel reactions from KEGG (Hadadi et al. 2019) Discovery of novel metabolic pathways for the biosynthesis of 3-hydroxypropanoate (Henry et al. 2010a)	https://lcsb-databases.epfl.ch/pathways/GraphList
PathPred	KEGG, RPAIR	Compound similarity and pathway scores	Prediction of putative biodegradation pathways for xenobiotics (e.g., 1,2,3,4-tetrachlorobenzene) as well as biosynthetic pathways for plants secondary metabolites (e.g., gentiodelphin)	https://www.genome.jp/tools/pathpred/

(continued)

Table 1 (continued)

Tool	Database	Pathway scoring	Applications and references	Source
SimPheny (BioPathway Predictor)	BiGG	Pathway length, thermodynamics, product yield, number of known metabolites/enzymes	In silico prediction and subsequent experimental validation of an unprecedented synthetic pathway for 1,4-butanediol production in <i>E. coli</i> (Yim et al. 2011)	-
RetroPath	MetaCyc, BioCyc	Pathway thermodynamics, sequence similarity, pathway length, number of putative enzyme steps, and product yield	Prediction of successfully-implemented biosynthetic pathways for 146 compounds in metabolic engineering projects. 81% of the compounds were predicted with at least one pathway (Delépine et al. 2018)	https://www.myexperiment.org/workflows/4987.html

At the end of the algorithm, pathways are ranked by features such as pathway length, thermodynamics, among others. Methodologically close to BNICE, PathPred (Moriya et al. 2010) uses instead RDM patterns consisting of reaction center atoms (R), atoms of different regions (D), and atoms of the matched region (M) for exploring the substrate graph. Pruning of the network is executed using structural similarity criteria, and pathway ranking is performed using compound similarity and pathway scores. SimPheny (Yim et al. 2011; Schilling et al. 2005) uses reactions rules from the third Enzyme Commission (EC) number level for generating reaction rules that enable reaction promiscuity for broader explorations. In this case, a retro-synthetic search is employed for enumerating feasible routes that produced intermediates of reasonable size (i.e., below a predefined size), and later they are ranked based on various criteria. Another retro-biosynthesis tool with recent important applications is RetroPath (Delépine et al. 2018). This tool uses a retro-synthetic search, albeit combined with MILP formulations for the application of various ranking criteria, such as thermodynamics, gene prediction, pathway length, number of putative steps, and product yield. In contrast to BNICE, RetroPath maintains a stoichiometric representation of the network (as opposed to a substrate graph) that enables computation of various scores. Moreover, molecular signatures are used to generate reaction rules based on a substructure of adjacent atoms, enabling the generation of substantially more and flexible reaction rules (Duigou et al. 2018). A recent implementation of reinforcement learning in RetroPath (RetroPath RL) has yielded promising results in the retro-biosynthetic prediction of biologically relevant pathways (Koch et al. 2020).

2.3 Stoichiometry-Based Optimization Methods for Pathway (Re)design

Given a universal metabolic reaction network, the ‘pathway design’ problem seeks to identify ‘optimal’ route(s) for the production of the target compound. As opposed to the retro-biosynthesis problem, possible connecting reactions are fixed and known upfront. Construction of the reaction network knowledge base is achieved by combining metabolic data from curated databases and/or from databases that also include putative reactions derived, for example, from generalized reaction rules (Hadadi et al. 2016; Hatzimanikatis et al. 2005; Jeffryes et al. 2015) or molecular signatures (Kumar et al. 2018). Regardless of the source of the data, reaction data must be charge- and mass-balanced to yield correct results, which typically is ensured in a manual curation step. Metabolite/reaction name inconsistencies are also an important source of issues that affect network connectivity and consistency, which often have to be resolved manually. While there have been attempts to standardize reaction and metabolite identifiers (King et al. 2016; Kumar et al. 2012; Alcántara et al. 2012), name reconciliation is challenging due to the incessant annotation of new metabolites and enzymatic activities, albeit important progress has been made in recent database versions (Moretti et al. 2021).

Once the reaction network has been assembled and mathematically formulated into a stoichiometric matrix, prediction of different pathway designs can be readily computed using optimization-based methods provided a convenient objective function. Among the most relevant objectives, one can name the minimization of the pathway length ensuring a minimum product yield (e.g., by fixing the overall stoichiometry) (Pharkya et al. 2004), maximization of the product yield observing thermodynamic constraints (Kumar et al. 2018; Kamp and Klamt 2020; Chowdhury and Maranas 2015), maximization of the thermodynamic favorability of the pathway (Flamholz et al. 2013; Noor et al. 2014; Hädicke et al. 2018; Yang et al. 2020; Ng et al. 2019), and minimization of the pathway’s enzymatic cost (Flamholz et al. 2013; Ng et al. 2019; Court et al. 2015; Bar-Even et al. 2010). For each of these objectives, different optimization problems must be formulated and solved, often requiring various parameters (e.g., thermodynamic and kinetic) from other sources for computing the optimal solution(s). In the following, the most relevant stoichiometry-based optimization methods for metabolic pathway prediction are presented. Further details about the methods and applications can be found in Table 2.

2.3.1 Pathways with Desired Stoichiometric Properties

Constraint-based modeling methods (Edwards and Palsson 2000) can be readily adapted for the computation of pathways exhibiting a desired stoichiometry (i.e., yield) (Kumar et al. 2018; Chowdhury and Maranas 2015; Ng et al. 2019), shortest length (Chowdhury and Maranas 2015; Ng et al. 2019), convenient precursor use (Kamp and Klamt 2020), and if using an organism’s reaction network as metabolic chassis, minimum addition of exogenous reactions (Pharkya et al. 2004; Kim et al.

Table 2 Optimization-based methods for metabolic pathway design

Pathway design objective	Method	Goal(s)	Applications and references
Desired stoichiometric properties (e.g., yield, length, among others)	OptStrain	Maximization of product yield using the minimum number of heterologous reactions in a production host	Prediction of in silico production pathways for hydrogen in <i>E. coli</i> , <i>C. acetobutylicum</i> , and <i>M. extorquens</i> AM1, as well as exploration of interventions strategies for achieving growth-coupling (Pharkya et al. 2004) In silico prediction of heterologous vanillin production pathways in <i>E. coli</i> and computation of their maximum theoretical yields and number of non-native enzymes to be included (Pharkya et al. 2004)
	SimOptStrain	Maximization of growth-coupled production including heterologous reactions in a production host	Computation of optimal succinate and glycerol production pathways using <i>E. coli</i> as metabolic chassis in terms of gene deletions and non-native reaction additions (Kim et al. 2011)
	OptStoic	Optimization of overall reaction stoichiometry (e.g., max. product yield, co-factor use, etc.), and minimization of the length or total flux through the pathway capable of sustaining the desired stoichiometry	In silico verification of the non-oxidative glycolysis pathway design (Chowdhury and Maranas 2015) Computation of optimal conversions and novel pathways for co-utilization of carbon dioxide and methanol for the production of C2+ metabolites (e.g., malonate, citrate, succinate, acetate, among others) (Chowdhury and Maranas 2015) Evaluation of feasible glycolytic alternatives and optimality analysis of trade-offs in AIP yield, enzymatic cost, and maximum-minimum driving force (Ng et al. 2019)

(continued)

Table 2 (continued)

Pathway design objective	Method	Goal(s)	Applications and references
	MEMO	Minimization of the size of the regeneration pathway (module) needed for sustaining cofactor(s) requirement(s) of a production pathway	Enumeration of optimal regeneration modules capable of sustaining the operation of the synthetic crotonyl-CoA/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle for in vitro carbon dioxide fixation in cell-free systems (Kamp and Klamt 2020)
	GEM-Path	Maximization of growth-coupled production including putative (promiscuous) enzymatic reactions in a production host	Computation of 245 unique synthetic pathways for the production of bulk chemicals and integration of these pathways into <i>E. coli</i> metabolism for yield analysis and metabolic optimization (Campodonico et al. 2014)
	novoStoic	Minimization of the pathway length Or Maximization of profit	In silico verification of synthetic pathways for 1,4-butanediol (BDO) production, identification of putative pathways for phenylephrine synthesis and benzolalpyrene degradation (Kumar et al. 2018)
	MapMaker/ PathTracer	Minimization of carbon transfers between reactants and products using the shortest pathway	Identification of possible conversion pathways to putrescine from glutamate, and CO ₂ fixation pathways in <i>E. coli</i> (Tervo and Reed 2016)

(continued)

Table 2 (continued)

Pathway design objective	Method	Goal(s)	Applications and references
Thermodynamic favorability	OptMDFpathway	Maximization of the minimum driving force of the production pathway	<p>Computation of all substrate-product combinations in <i>E. coli</i> that enable net CO₂ assimilation via thermodynamically feasible pathways (Hädicke et al. 2018)</p> <p>In silico design of multi-strain communities of a single-species (<i>E. coli</i>) that achieve maximum thermodynamic driving force for product synthesis under the assumption of pathway specialization (i.e., parts of the production pathway are performed by different community members) (Bekiaris and Klamt 2021)</p>
Enzymatic cost or burden	ECM	Minimize enzyme cost or metabolic burden	<p>In silico computation of putative efficient CO₂ fixation pathways into C2-C4 compounds (Bar-Even et al. 2010)</p> <p>Optimality analysis of enzyme cost minimization in central carbon metabolism of <i>E. coli</i> (Noor et al. 2016)</p> <p>Prediction of rate/yield trade-offs in metabolic pathways under oxygen-limited conditions in <i>E. coli</i> (Wortel et al. 2018)</p>

2011). In practice, these pathway enumeration methods rely on the solution of various LP and/or MILP optimization problems that optimize some of the aforementioned objectives subject to not only stoichiometric constraints but possibly to thermodynamic and/or economic constraints.

OptStrain, SimOptStrain, and OptStoic are classical tools for pathway prediction, although they differ in their scope. While the first two seek to predict optimal pathways and metabolic interventions for the production of a target metabolite leveraging the microbial host reaction network (Pharkya et al. 2004; Kim et al. 2011), the latter aims to find complete mass-balanced conversion pathways that yield a desired stoichiometry from precursors to product(s) using metabolic databases as the input reaction network (Chowdhury and Maranas 2015). Additional constraints related to a minimum guaranteed product yield, thermodynamic plausibility of the pathway, and/or substrate costs can be readily included to obtain more convenient designs. Recently, a computational method called MEMO (Kamp and Klamt 2020) has been proposed for identifying the smallest metabolic modules with specified stoichiometric and thermodynamic properties. For instance, this approach has been employed to find small cofactor regeneration (e.g., ATP/ADP, NAD(P)H, NAD(P), among others) modules that can sustain bioconversions in the context of cell-free applications under defined thermodynamic conditions.

The aforementioned methods rely on existing annotated enzymatic reactions for metabolic conversions. However, as mentioned in the previous subsection, promiscuous enzymatic activities are characteristic features of metabolic reaction networks, likely playing an evolutionary role as a starting point in enzyme functions (Khersonsky and Tawfik 2010). Importantly, the existence of this feature suggests that there is still untapped potential for a broader chemical reaction space to be explored. By using various extraction techniques for learning putative reactions from known enzymatic reactions, it is possible to populate and assemble larger databases for pathway prediction. An example of these methods is MapMaker/PathTracer (Tervo and Reed 2016), which employs precomputed carbon transfer maps (CTMs) based on chemical and stoichiometric information (MapMaker) for the prediction of short, carbon-balanced pathways from substrates to products (PathTracer). GEM-Path (Campodonico et al. 2014) is another framework that, using a genome-scale metabolic reconstruction of *E. coli* as base reaction network, combines heterologous pathway integration (similar to OptStoic) with constraint-based growth-coupled methods for the computation of metabolic designs. Increased biochemical reaction exploration is achieved through the introduction of a chemical similarity measure to assess enzyme-catalyzed reaction promiscuity. Lastly, the novoStoic/rePrime framework (Kumar et al. 2018) enables exploration of a far greater chemical transformation space through the imposition of chemical ‘moiety’ conservation (refer to Sect. 2.1) that is particularly suited for the prediction of optimal pathways with maximum yield or length. Importantly, this mathematical treatment avoids chemical reaction information loss (e.g., stereoselectivity) as opposed to other approaches like MapMaker/PathTracer.

2.3.2 Pathways with Maximum Thermodynamic Favorability

Pathway thermodynamics exerts a fundamental control in metabolic flux with seemingly important consequences for microbial fitness (Du et al. 2018). While there have been different methods for combining stoichiometric-based analysis with reaction thermodynamics (Henry et al. 2007; Kummel et al. 2006), it has not been until recently that thermodynamic favorability has been mathematically formalized. For this task, the Max-min Driving Force (MDF) index (Noor et al. 2014) has been proposed for quantifying the smallest absolute Gibbs free energy (or driving force) of a given pathway under the most favorable metabolic conditions. As the latter captures the driving force of the most unfavorable conversion step (i.e., thermodynamic bottleneck), its maximization yields the most favorable operating conditions for a *given* pathway. More recently, the OptMDFpathway method (Hädicke et al. 2018) was introduced to identify the most thermodynamically favorable pathways in a *given* reaction network, thereby enabling exploration of thermodynamically plausible production pathways in the context of microbial metabolism (Hädicke et al. 2018; Yang et al. 2020), and more recently, in microbial communities (Bekiaris and Klamt 2021).

2.3.3 Pathways with Minimum Enzymatic Cost

Cellular metabolism incurs a metabolic cost when committing to the synthesis of a particular set of proteins (enzymes). As seemingly similar enzymes can still display large differences in their catalytic properties (Bar-Even et al. 2011), it is natural to seek pathways that can yield the maximum return of investment (flux) per protein (enzyme) mass synthesized. For this task, the Enzyme Cost Minimization (ECM) (Noor et al. 2016)—and later termed the Enzyme-Flux Cost Minimization (EFCM) (Wortel et al. 2018)—formulation computes the minimum enzyme load (i.e., the aggregated enzyme mass allocated) required for a metabolic pathway to yield a *given* flux (Flamholz et al. 2013; Bar-Even et al. 2010). While this formulation originally required a thermodynamically consistent, fully parameterized kinetic model for this calculation (Saa and Nielsen 2017), increasingly enzymatically-constraint GSMMs (Sánchez et al. 2017) and ME-models (metabolic and expression) (Lerman et al. 2012) are being considered and employed for these calculations under the optimistic scenario of enzymatic catalysis at capacity. Finally, the ECM/EFCM does not support performing pathway enumeration, although it can be readily employed as a ranking index when combined with the previous approaches.

3 Case Studies of Metabolic Pathway Prediction and Optimization in Yeast

In this section, selected case studies illustrate different pathway engineering aspects required for improving metabolic performance overall, and particularly, in yeast. These examples showcase strategies for redox cofactor balancing, increased precursor supply, and engineering of central pathways for carbon fixation. The

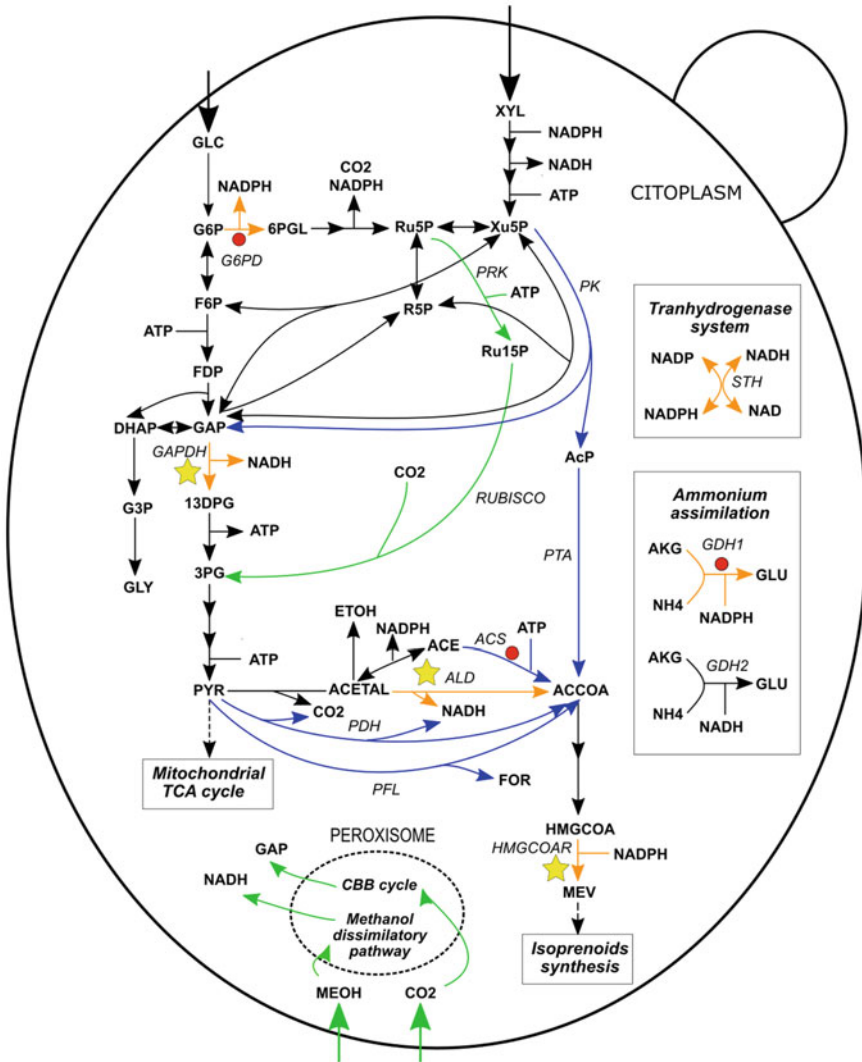
impact of the latter applications is especially highlighted in the context of harnessing the metabolic potential of yeast for industrial bioproduction. Figure 2 illustrates the details of the revised strategies.

3.1 Balancing Redox Cofactor Supply for Improving Substrate Utilization and Isoprenoids Production

Regeneration of either redox and/or energy cofactors often limits the production of high-value metabolites. In order to increase the availability of the required cofactor(s), central carbon metabolism must be intervened and engineered in such a way that it favors bioproduction without extremely affecting microbial growth (Lee et al. 2013). This challenge is particularly relevant for many NAD(P)H-expensive valuable compounds that are being produced in yeast (Cataldo et al. 2020; López et al. 2020, 2019) and other microbes (Ko et al. 2020).

Increased supply of redox cofactors can be achieved by either overexpressing key enzymes involved in cofactor generation (Lee et al. 2007; San et al. 2002; Lim et al. 2002) or by increasing the expression of alternative redox partner systems. A recent application of the latter has proved effective for enhancing the unprecedented heterologous production of violaxanthin in *S. cerevisiae* by approx. two-fold (Cataldo et al. 2020). However, the success of these approaches is likely limited due to the presence of different intrinsic balancing mechanisms for maintaining homeostasis in yeast (Hou et al. 2010). An illustrative example of the latter can be found in the study of Nissen et al. (2001). Here, heterologous expression of the pyridine nucleotide transhydrogenase system (*sth* gene, absent in yeast) that transfers reducing equivalents from NADPH to NADH (and vice versa), did not improve ethanol formation in anaerobic conditions. On the contrary, ethanol production was reduced concomitantly with the increase of fermentation by-products (glycerol and 2-oxoglutarate) required for redox rebalancing. Another, less intuitive and possibly more effective, strategy for (re)balancing redox cofactors supply and demand involves cofactor swapping (Verho et al. 2003; Martínez et al. 2008). Computational studies in *S. cerevisiae* and *E. coli* support this strategy as a promising intervention for forcing higher metabolic performance (King and Feist 2014). Simply put, this approach seeks to replace native (redox-consuming) enzymes with heterologous counterparts with a different cofactor specificity (e.g., NAD(P)H—for a NA(D)H-dependent enzyme).

The first application of the latter strategy involved the optimization of D-xylose utilization for ethanol production in *S. cerevisiae* (Verho et al. 2003). This carbon source is assimilated through the pentose phosphate pathway (PPP) as D-xylulose-5-phosphate and then incorporated as glyceraldehyde 3-phosphate in glycolysis. In theory, D-xylose should produce CO₂ and ethanol in a 1:1 molar ratio under redox-neutral anaerobic conditions (Kötter and Ciriacy 1993). However, D-xylose assimilation requires extra NADPH and NAD⁺ that must be regenerated by other separate processes, which are very inefficient in yeast, rendering D-xylose fermentation slow. To overcome this bottleneck and force higher



Key

- Carbon fixation intervention (reaction addition)
- Redox balancing intervention (reaction addition)
- Increase of acetyl-CoA pool intervention (reaction addition)
- Replacement with enzyme with different cofactor specificity (cofactor swap)
- Gene knock out

◀**Fig. 2** Illustration of selected reported strategies for achieving improved cofactor balancing, increased acetyl-CoA supply and engineering CO₂ fixation in yeast. The details of each strategy are discussed in Sect. 3. Relevant metabolite names are represented by uppercase bold fonts, whereas enzyme names are represented by uppercase *italics* fonts. Abbreviations: 13DPG, 1,3-diphosphoglycerate; 3PG, 3-phosphoglycerate; 6PGL, 6-phospho-D-glucono-1,5-lactone; ACCOA, acetyl-CoA; ACE, acetate; ACETAL, acetaldehyde; AcP, acetyl phosphate; ACS, acetyl-CoA synthetase; AKG, alpha-ketoglutarate; ALD, aldehyde dehydrogenase; DHAP, dihydroxyacetone phosphate; ETOH, ethanol; F6P, D-fructose 6-phosphate; FDP, D-fructose 1,6-disphosphate; FOR, formate; G3P, glycerol 3-phosphate; G6P, D-glucose 6-phosphate; G6PD, glucose 6-phosphate dehydrogenase; GAP, glyceraldehyde 3-phosphate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GDH, glutamate dehydrogenase; GLC, D-glucose; GLU, glutamate; GLY, glycerol; HMGCOA, 3-hydroxy-3-methyl-glutaryl-CoA; HMGCOAR, HMG-CoA reductase; MEOH, methanol; MEV, mevalonate; NH₄, ammonia; PDH, pyruvate dehydrogenase; PFL, pyruvate formate lyase; PK, phosphoketolase; PRK, phosphoribulokinase; PTA, phosphotransacetylase; PYR, pyruvate; R5P, D-ribose 5-phosphate; Ru15P, ribulose 1,5-disphosphate; Ru5P, D-ribulose 5-phosphate; STH, transhydrogenase; Xu5P, D-xylulose 5-phosphate; XYL, D-xylose

NADPH supply and flux through lower glycolysis, the native NAD-dependent GAP dehydrogenase (GAPDH) was replaced by an NADP-dependent GAPDH and the NADPH-dependent glucose-6-phosphate dehydrogenase (G6PD) was knocked out, which also prevented carbon loss as CO₂ (Verho et al. 2003). This strategy almost doubled the ethanol yield on D-xylose (from 18 to 41%) and reduced the CO₂/ethanol molar ratio close to the theoretical 1:1 (from 2.5 to 1.3). Later, expression of the heterologous phosphotransacetylase (PTA) and phosphoketolase (PK) for improving NADH reoxidation in the D-xylose utilization pathway generated an increase in ethanol yield (25% higher) without affecting the growth rate (Sonderregger et al. 2004).

Cofactor rebalancing and swapping strategies for the synthesis of NADPH-expensive isoprenoid-derived compounds have shown to be particularly effective in yeast. For instance, α -santalene production yields a net production of NADH and consumption of NADPH, which calls for the rebalancing of the cofactor supply (Scalcinati et al. 2012). By deleting known reactions involved in glutamate metabolism (ammonium assimilation) that consume NADPH (GDH1) and activating NAD-dependent counterparts (GDH2) (Nissen et al. 2000), the production of α -santalene was substantially improved. Similarly in a different study of protopanaxadiol production—another isoprenoid-derived compound—the availability of NADPH was enhanced by replacing the native NADH-generating acetaldehyde dehydrogenase (ALD2) with a functionally equivalent NADPH-generating enzyme (ALD6), resulting in a 11-fold increase in titer (Kim et al. 2018). Lastly, swapping of the native NADP-dependent 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMG-CoA reductase)—third committed step of the mevalonate pathway responsible for the production of isoprenoid precursors—has also shown to increase the overall pathway flux in *E. coli* (Ma et al. 2011). This result was leveraged by Meadows et al. (2016) whereby an NADH-consuming HMG-CoA reductase from *Silicibacter pomeroyi* was employed for the overproduction of the sesquiterpene

farnesene. Implementation of other computationally predicted major metabolic cofactor swaps like the alcohol dehydrogenase (ALCD2) and GAPD for the improved production of isoprenoids remains to be tested experimentally (King et al. 2016), as they could be potentially beneficial for boosting production as shown in other microorganisms (Martínez et al. 2008).

3.2 Increasing Cytosolic Acetyl-CoA Availability for Metabolic Production

Cytosolic acetyl-CoA is a key metabolite for the production of a range of valuable compounds in yeast (Rossum et al. 2016). Native production of this compound requires 2 mol of ATP and yields 2 mol of acetyl-CoA and 4 mol of NAD(P)H per mol of glucose (Rossum et al. 2016). To improve the availability of this precursor and lower the ATP cost, different heterologous enzymes have been introduced to either bypass the native aldehyde dehydrogenase (ALD) and acetyl-CoA synthetase (ACS) system using bacterial counterparts, i.e., A-ALD and PFL, that do not incur such high cost (Kozak et al. 2014a, b), or to enable acetyl-CoA biosynthesis in situ by expressing the pyruvate dehydrogenase (PDH) complex in the cytoplasm (Kozak et al. 2014a, b). While the former application showed mixed results in terms of growth and yield (mainly due to by-product accumulation), the second approach along with a knock-out of the native ACS reaction exhibited similar metabolic performance to the control, but at a lower ATP cost.

A different approach for improving acetyl-CoA availability relies on increasing its yield. For this task, the phosphoketolase pathway (PKP) was early suggested for the conversion of 1 mol of F6P into 3 mol of acetyl-P without carbon loss (Schramm and Racker 1957). Conversion of acetyl-P to acetyl-CoA can be later achieved by the reversible phosphotransacetylase (PTA) reaction (Rossum et al. 2016). This was initially implemented in yeast for improving D-xylose fermentation (Sonderegger et al. 2004) (refer to previous section). More recently, Bogorad et al. (2013) implemented the full PKP in *E. coli* and demonstrated almost stoichiometric conversion of C5 and C6 sugars into acetate under anaerobic conditions. A similar approach was replicated in yeast accompanied by several genetic interventions to increase acetyl-CoA-derived farnesene (Meadows et al. 2016). This non-native pathway increased carbon utilization by 25%, decreased oxygen consumption by 75%, and reached 15% v/v of farnesene. As illustrated here, increasing acetyl-CoA availability may be critical not only for maximizing production but also for overall improving metabolic performance.

3.3 Engineering a Heterologous CBB Cycle for CO₂ Fixation

There is a growing interest in the field for engineering carbon assimilation pathways in heterotrophs for improving product yields—e.g., by reducing carbon loss as CO₂ –, and most notably, for implementing one-carbon (C1) compounds (e.g., CO₂) fixation pathways to develop more sustainable fermentation bioprocesses.

In an early effort from Guadalupe-Medina et al. (2013), a heterologous Calvin–Benson–Bassham (CBB) cycle was implemented in *S. cerevisiae* seeking to improve ethanol yield by reducing carbon loss under anaerobic conditions. The authors noted that by expressing the CBB enzymes phosphoribulokinase (PRK) and ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO), a working pathway could be realized where CO₂ is effectively used as an electron acceptor for NADH oxidation, thereby coupling CO₂ fixation by RuBisCO with the fermentation redox balance. Importantly, this mechanism rendered NADH reoxidation through the native glycerol formation pathway unnecessary (90% reduction in glycerol titer), increasing ethanol yield by 14% (Guadalupe-Medina et al. 2013).

While the latter strategy was successful in increasing product yield, it did so by reducing carbon loss as glycerol and not by significantly increasing CO₂ assimilation (Guadalupe-Medina et al. 2013). A more radical approach is to engineer a CO₂ assimilation pathway capable of sustaining growth and production. In a pioneer work from Antonovsky et al. (2016), *E. coli* was transformed and evolved to grow solely on CO₂ as a carbon source and pyruvate as an electron source. Again, expression of the missing CBB enzymes PRK and RuBisCO, and knock-out of the phosphoglycerate mutase (PGM)—revealed by Flux Balance Analysis (Lewis et al. 2012)—forced CBB operation by decoupling carbon fixation from energy production. This metabolic phenotype was termed hemi-autotrophic growth, and it has since been implemented in other bacteria like the methanol-consuming bacteria *Methylobacterium extorquens* AM1 through the expression of the previous CBB enzymes and deletion of essential genes for methanol assimilation (Borzyskowski et al. 2018). Building on these strategies, a recent study reported the conversion of the yeast *P. pastoris* into an autotroph that grows on CO₂ as the sole carbon source and methanol as an energy source (Gassler et al. 2020). Briefly, *P. pastoris* can use methanol as both energy and carbon sources. By decoupling the formaldehyde—the assimilated product of methanol oxidation—dissimilatory (carbon-fixating) and assimilatory (energy-producing) pathway branches, one can force CO₂ assimilation by blocking the dissimilatory branch through the deletion of the dihydroxyacetone synthase (DAS1 and DAS2) and alcohol oxidase 1 (AOX1). Then, complementation of the native peroxisomal xylose monophosphate (XuMP) cycle with six enzymatic steps enables operation of the CBB cycle allowing growth on CO₂. In stoichiometric terms, 1 mol of oxidized methanol produces 2 mol of NADH, which can be used to fuel the CBB cycle though not in stoichiometric proportions with CO₂ (3 mol of ATP and 2 mol of NADH are needed to fix 1 mol of CO₂). The resulting mutant strain reached a maximum specific growth rate of 0.018 h⁻¹ (Gassler et al. 2020) and constitutes an unprecedented advance for compartmentalized C1 carbon fixation in yeast differing from seemingly similar efforts in bacteria (Antonovsky et al. 2016; Bang and Lee 2018).

4 Challenges and Outlook

During the past decade, yeast metabolic engineering has shown great progress and promise (Smolke and Tyo 2012; Nielsen and Keasling 2016), quickly becoming one of the preferred microbial factories for realizing the bioproduction of new chemicals or improving the production of traditional ones. This success has been largely driven by the continuous advances in the development of genetic and molecular tools (Smolke and Tyo 2012), as well as novel computational frameworks for pathway discovery and optimization (Wang et al. 2017; Saa et al. 2019). The latter has brought not only new possibilities for the evaluation of novel biochemical synthesis routes but also has provided more rational methods for designing metabolic pathways with superior performance by rewiring metabolism at a whole-cell scale (Saa et al. 2019). In time, such capabilities will become increasingly essential for arriving at designs that scale industrially and meet commercial expectations.

Pathway discovery is supported by the use of retro-synthesis tools that generate putative routes connecting substrates to products. A comprehensive exploration of the chemical space typically rests on the availability of reaction rules, which fills the gaps between the metabolic precursors and target chemical(s). Generation and application of such rules must be carefully performed, as they may provide infeasible pathways that may obscure results interpretation (Wang et al. 2017). Atom mapping information can be of great aid for validating the application and generation of certain reaction rules, see e.g., RouteSearch (Latendresse et al. 2014) and ReTrace (Pitkänen et al. 2009), which can be further completed with enzyme promiscuity knowledge if available (Mazurenko et al. 2020). Another incipient alternative for learning novel chemical reaction routes rests on machine learning techniques (Koch et al. 2020), which can potentially increase exponentially the size of the reachable chemical space as shown elsewhere (Coley et al. 2019; Mikulak-Klucznik et al. 2020). Efficient navigation of such vast space would necessarily have to rely on the introduction of pathway scores and rankings to focus the attention on the most promising and realizable designs. For this task, evaluation of the objectives reviewed here along with others—e.g., use of enzymes with known promiscuous activity or cofactor specificity—constitutes a natural way for prioritizing and selecting desired pathways. Rational integration of the various objectives can be achieved by leveraging mature multi-decision multi-criteria techniques (Bonissone et al. 2009), which remains largely unexplored in the field. Notably, the latter techniques are also transferable to optimization-based methods for pathway prediction, which could enable a more holistic evaluation of pathway performance and robustness.

While the revised computational methods and tools for pathway prediction have provided unintuitive and useful insights, their experimental application and validation remain still limited. Although there have been recent applications in yeast (Hafner et al. 2021) and other model organisms (Yim et al. 2011) where some of the tools have proven to be critical for finding effective *in vivo* metabolic

designs, there is still resistance to their broad adoption. Indeed, in vivo implementation of complex in silico metabolic designs is not trivial, typically demanding great amounts of experimentation time before arriving at a working pathway (Antonovsky et al. 2016; Schwander et al. 2016; Savile et al. 2010). Such efforts could gain from recent computational frameworks for kinetic model construction (Saa and Nielsen 2017) that could help to predict a priori the expected performance of the pathway (see, for example, (Theisen et al. 2016)), greatly reducing the time and resources needed. As the metabolic prediction capabilities of current models increase (Foster et al. 2021), it is expected that the use of these tools for rational pathway engineering in yeast and other microbial factories will progressively become part of the basic toolbox for metabolic engineering.

Acknowledgements The author acknowledges the financial support of ANID Fondecyt Iniciacion Grant No 11190871 and FONDAF Grant No 15090007 of the Center for Genome Regulation (CGR).

References

- Alcántara R, Axelsen KB, Morgat A, Belda E et al (2012) Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res* 40:D754–D760
- Antonovsky N, Gleizer S, Noor E, Zohar Y et al (2016) Sugar Synthesis from CO₂ in *Escherichia coli*. *Cell* 166:115–125
- Arkin AP, Cottingham RW, Henry CS, Harris NL et al (2018) KBase: the united states department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569
- Bang J, Lee SY (2018) Assimilation of formic acid and CO₂ by engineered *Escherichia coli* equipped with reconstructed one-carbon assimilation pathways. *Proc Natl Acad Sci* 115:E9271
- Bar-Even A, Noor E, Lewis NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci* 107:8889
- Bar-Even A, Noor E, Savir Y, Liebermeister W et al (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry-US* 50:4402–4410
- Bekiaris PS, Klamt S (2021) Designing microbial communities to maximize the thermodynamic driving force for the production of chemicals. *Plos Comput Biol* 17:e1009093
- Bogorad IW, Lin TS, Liao JC (2013) Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* 502:693–697
- Bonissone PP, Subbu R, Lizzi J (2009) Multicriteria decision making (mcdm): a framework for research and applications. *IEEE Comput Intell Mag* 4:48–61
- Campononico MA, Andrews BA, Asenjo JA, Palsson BO, Feist AM (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm GEM-Path. *Metab Eng* 25C:140–158
- Caspi R, Billington R, Ferrer L, Foerster H et al (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 44:D471–480
- Cataldo VF, Arenas N, Salgado V, Camilo C et al (2020) Heterologous production of the epoxy-carotenoid violaxanthin in *Saccharomyces cerevisiae*. *Metab Eng* 59:53–63
- Chowdhury A, Maranas CD (2015) Designing overall stoichiometric conversions and intervening metabolic reactions. *Sci Rep* 5
- Coley CW, Thomas DA, Lummiss JAM, Jaworski JN et al (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* 365:eaax1566
- Court SJ, Waclaw B, Allen RJ (2015) Lower glycolysis carries a higher flux than any biochemically possible alternative. *Nat Commun* 6:8427

- Delépine B, Duigou T, Carbonell P, Faulon J-L (2018) RetroPath2.0: a retrosynthesis workflow for metabolic engineers. *Metab Eng* 45:158–170
- Du B, Zielinski DC, Monk JM, Palsson BO (2018) Thermodynamic favorability and pathway yield as evolutionary tradeoffs in biosynthetic pathway choice. *Proc Natl Acad Sci* 115:11339
- Duigou T, du Lac M, Carbonell P, Faulon J-L (2018) RetroRules: a database of reaction rules for engineering biology. *Nucleic Acids Res* 47:D1229–D1235
- Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *P Natl Acad Sci USA* 97:5528–5533
- Ellis LBM, Roe D, Wackett LP (2006) The University of Minnesota biocatalysis/biodegradation database: the first decade. *Nucleic Acids Res* 34:D517–D521
- Finley SD, Broadbelt LJ, Hatzimanikatis V (2009) Computational framework for predictive biodegradation. *Biotechnol Bioeng* 104:1086–1097
- Finley SD, Broadbelt LJ, Hatzimanikatis V (2010) In silico feasibility of novel biodegradation pathways for 1,2,4-trichlorobenzene. *Bmc Syst Biol* 4:7
- Flamholz A, Noor E, Bar-Even A, Milo R (2012) eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res* 40:770–775
- Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci* 110:10039
- Foster CJ, Wang L, Dinh HV, Suthers PF, Maranas CD (2021) Building kinetic models for metabolic engineering. *Curr Opin Biotech* 67:35–41
- Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD (2015) Complete biosynthesis of opioids in yeast. *Science* 349:1095–1100
- Ganter M, Bernard T, Moretti S, Stelling J, Pagni M (2013) MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* 29:815–816
- Gassler T, Sauer M, Gasser B, Egermeier M et al (2020) The industrial yeast *Pichia pastoris* is converted from a heterotroph into an autotroph capable of growth on CO₂. *Nat Biotechnol* 38:210–216
- Guadalupe-Medina V, Wisselink HW, Luttkik MA, de Hulster E et al (2013) Carbon dioxide fixation by Calvin-cycle enzymes improves ethanol yield in yeast. *Biotechnol Biofuels* 6:125
- Hadadi N, Hafner J, Shajkofci A, Zisaki A, Hatzimanikatis V (2016) ATLAS of Biochemistry: a repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies. *ACS Synth Biol* 5:1155–1166
- Hadadi N, MohammadiPeyhani H, Miskovic L, Seijo M, Hatzimanikatis V (2019) Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites. *Proc Natl Acad Sci* 116:7298
- Hädicke O, von Kamp A, Aydogan T, Klamt S (2018) OptMDFpathway: Identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO₂ fixation potential of *Escherichia coli*. *Plos Comput Biol* 14:e1006492
- Hafner J, Payne J, MohammadiPeyhani H, Hatzimanikatis V, Smolke C (2021) A computational workflow for the expansion of heterologous biosynthetic pathways to natural product derivatives. *Nat Commun* 12:1760
- Hatzimanikatis V, Li C, Ionita JA, Henry CS et al (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* 21:1603–1609
- Henry CS, Broadbelt LJ, Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. *Biophys J* 92:1792–1805
- Henry CS, Broadbelt LJ, Hatzimanikatis V (2010a) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnol Bioeng* 106:462–473
- Henry CS, DeJongh M, Best AA, Frybarger PM et al (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28:977–982
- Hou J, Scalcinati G, Oldiges M, Vemuri GN (2010) Metabolic impact of increased NADH availability in *Saccharomyces cerevisiae*. *Appl Environ Microb* 76:851–859

- Isaacs FJ, Carr PA, Wang HH, Lajoie MJ et al (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* 333:348
- Jeffreyes JG, Colastani RL, Elbadawi-Sidhu M, Kind T et al (2015) MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *Journal of Cheminformatics* 7:44
- Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D (2018) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* gky1048-gky1048
- Kamp AV, Klant S (2020) MEMO: A Method for computing metabolic modules for cell-free production systems. *ACS synthetic biology* 9:556–566
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462
- Khersonsky O, Tawfik DS (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu Rev Biochem* 79:471–505
- Kim J, Reed JL, Maravelias CT (2011) Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *Plos One* 6:e24162
- Kim J-E, Jang I-S, Sung BH, Kim SC, Lee JY (2018) Rerouting of NADPH synthetic pathways for increased protopanaxadiol production in *Saccharomyces cerevisiae*. *Sci Rep* 8:15820
- King ZA, Feist AM (2014) Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*. *Metab Eng* 24:117–128
- King ZA, Lu J, Dräger A, Miller P et al (2016) BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 44:D515–D522
- Ko Y-S, Kim JW, Lee JA, Han T et al (2020) Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem Soc Rev* 49:4615–4636
- Koch M, Duigou T, Faulon J-L (2020) Reinforcement learning for bioretrosynthesis. *ACS Synth Biol* 9:157–168
- Kötter P, Ciriacy M (1993) Xylose fermentation by *Saccharomyces cerevisiae*. *Appl Microbiol Biot* 38:776–783
- Kozak BU, van Rossum HM, Benjamin KR, Wu L et al (2014a) Replacement of the *Saccharomyces cerevisiae* acetyl-CoA synthetases by alternative pathways for cytosolic acetyl-CoA synthesis. *Metab Eng* 21:46–59
- Kozak BU, van Rossum HM, Luttik MAH, Akeroyd M et al (2014) Engineering acetyl coenzyme a supply: Functional expression of a bacterial pyruvate dehydrogenase complex in the cytosol of *Saccharomyces cerevisiae*. *mBio* 5:e01696–01614
- Kumar A, Suthers PF, Maranas CD (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics* 13:6
- Kumar A, Wang L, Ng CY, Maranas CD (2018) Pathway design using de novo steps through uncharted biochemical spaces. *Nat Commun* 9:184
- Kummel A, Panke S, Heinemann M (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* 2(2):0034
- Latendresse M, Krummenacker M, Karp PD (2014) Optimal metabolic route search based on atom mappings. *Bioinformatics* 30:2043–2050
- Lee SY, Kim HU (2015) Systems strategies for developing industrial microbial strains. *Nat Biotechnol* 33:1061–1072
- Lee WH, Park JB, Park K, Kim MD, Seo JH (2007) Enhanced production of epsilon-caprolactone by overexpression of NADPH-regenerating glucose 6-phosphate dehydrogenase in recombinant *Escherichia coli* harboring cyclohexanone monooxygenase gene. *Appl Microbiol Biotechnol* 76:329–338
- Lee WH, Kim MD, Jin YS, Seo JH (2013) Engineering of NADPH regenerators in *Escherichia coli* for enhanced biotransformation. *Appl Microbiol Biotechnol* 97:2761–2772
- Lerman JA, Hyduke DR, Latif H, Portnoy VA et al (2012) In silico method for modelling metabolism and gene product expression at genome scale. *Nat Commun* 3
- Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10:291–305

- Lim SJ, Jung YM, Shin HD, Lee YH (2002) Amplification of the NADPH-related genes *zwf* and *gnd* for the oddball biosynthesis of PHB in an *E. coli* transformant harboring a cloned *phbCAB* operon. *J Biosci Bioeng* 93:543–549
- Lin G-M, Warden-Rothman R, Voigt CA (2019) Retrosynthetic design of metabolic pathways to chemicals not found in nature. *Curr Opin Syst Biol* 14:82–107
- López J, Bustos D, Camilo C, Arenas N, Saa PA (2020) Engineering *Saccharomyces cerevisiae* for the overproduction of β -ionone and its precursor β -carotene. *Front Bioeng Biotechnol* 8:1–13
- López J, Cataldo VF, Peña M, Saa PA et al (2019) Build your bioprocess on a solid strain— β -carotene production in recombinant *Saccharomyces cerevisiae*. *Front Bioeng Biotechnol* 7
- Ma SM, Garcia DE, Redding-Johanson AM, Friedland GD et al (2011) Optimization of a heterologous mevalonate pathway through the use of variant HMG-CoA reductases. *Metab Eng* 13:588–597
- Martínez I, Zhu J, Lin H, Bennett GN, San K-Y (2008) Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from *Clostridium acetobutylicum* facilitates NADPH dependent pathways. *Metab Eng* 10:352–359
- Mazurenko S, Prokop Z, Damborsky J (2020) Machine learning in enzyme engineering. *ACS Catal* 10:1210–1223
- Meadows AL, Hawkins KM, Tsegaye Y, Antipov E et al (2016) Rewriting yeast central carbon metabolism for industrial isoprenoid production. *Nature* 537:694–697
- Mikulak-Kluczniak B, Gołbiewska P, Bayly AA, Popik O et al (2020) Computational planning of the synthesis of complex natural products. *Nature* 588:83–88
- Moretti S, Tran, Van Du T, Mehl F, Ibberson M, Pagni M (2021) MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models. *Nucleic Acids Res* 49:D570–D574
- Moriya Y, Shigemizu D, Hattori M, Tokimatsu T et al (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res* 38:W138–W143
- Moura M, Finkle J, Stainbrook S, Greene J et al (2016) Evaluating enzymatic synthesis of small molecule drugs. *Metab Eng* 33:138–147
- Ng CY, Wang L, Chowdhury A, Maranas CD (2019) Pareto optimality explanation of the glycolytic alternatives in nature. *Sci Rep* 9:2633
- Nielsen J, Keasling JD (2016) Engineering cellular metabolism. *Cell* 164:1185–1197
- Nissen TL, Kielland-Brandt MC, Nielsen J, Villadsen J (2000) Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation. *Metab Eng* 2:69–77
- Nissen TL, Anderlund M, Nielsen J, Villadsen J, Kielland-Brandt MC (2001) Expression of a cytoplasmic transhydrogenase in *Saccharomyces cerevisiae* results in formation of 2-oxoglutarate due to depletion of the NADPH pool. *Yeast* 18:19–32
- Noor E, Bar-Even A, Flamholz A, Reznik E et al (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. *Plos Comput Biol* 10:e1003483
- Noor E, Flamholz A, Bar-Even A, Davidi D et al (2016) The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *Plos Comput Biol* 12: e1005167
- Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. *Genome Res* 14:2367–2376
- Pitkänen E, Jouhten P, Rousu J (2009) Inferring branching pathways in genome-scale metabolic networks. *Bmc Syst Biol* 3:103
- Saa PA, Nielsen LK (2017) Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. *Biotechnol Adv* 35:981–1003
- Saa PA, Cortés MP, López J, Bustos D et al (2019) Expanding metabolic capabilities using novel pathway designs: computational tools and case studies. *Biotechnol J* 14:1800734
- San KY, Bennett GN, Berrios-Rivera SJ, Vadali RV et al (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. *Metab Eng* 4:182–192

- Sánchez BJ, Zhang C, Nilsson A, Lahtvee P-J et al (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol* 13:935
- Savile CK, Janey JM, Mundorff EC, Moore JC et al (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* 329:305
- Scalcinati G, Partow S, Siewers V, Schalk M et al (2012) Combined metabolic engineering of precursor and co-factor supply to increase α -santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact* 11:117
- Schilling C, Thakar R, Travník E, Dien S, Wiback S (2005) SimPheny: a Computational Infrastructure for Systems Biology. In: US Department of Energy, Genomic Science Program publications
- Schramm M, Racker E (1957) Formation of erythrose-4-phosphate and acetyl phosphate by a phosphorylytic cleavage of fructose-6-phosphate. *Nature* 179:1349
- Schwander T, von Schada Borzyskowski L, Burgener S, Cortina NS, Erb TJ (2016) A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 354: 900–904
- Smolke CD, Tyo KEJ (2012) Synthetic biology: Emerging methodologies to catalyze the metabolic engineering design cycle. *Metab Eng* 14:187–188
- Sonderegger M, Schumperli M, Sauer U (2004) Metabolic engineering of a phosphoketolase pathway for pentose catabolism in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 70:2892–2897
- Tervo CJ, Reed JL (2016) MapMaker and PathTracer for tracking carbon in genome-scale metabolic models. *Biotechnol J* 11:648–661
- Theisen MK, Lafontaine Rivera JG, Liao JC (2016) Stability of ensemble models predicts productivity of enzymatic systems. *Plos Comput Biol* 12:e1004800
- Tokic M, Hadadi N, Ataman M, Neves D et al (2018) Discovery and evaluation of biosynthetic pathways for the production of five methyl ethyl ketone precursors. *ACS Synth Biol* 7:1858–1873
- van Rossum HM, Kozak BU, Pronk JT, van Maris AJA (2016) Engineering cytosolic acetyl-coenzyme A supply in *Saccharomyces cerevisiae*: pathway stoichiometry, free-energy conservation and redox-cofactor balancing. *Metab Eng* 36:99–115
- Verho R, Londesborough J, Penttilä M, Richard P (2003) engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*. *Appl Environ Microb* 69:5892–5897
- Voigt CA (2020) Synthetic biology 2020–2030: six commercially-available products that are changing our world. *Nat Commun* 11:6379
- von Schada Borzyskowski L, Carrillo M, Leupold S, Glatter T et al (2018) An engineered Calvin-Benson-Bassham cycle for carbon dioxide fixation in *Methylobacterium extorquens* AM1. *Metab Eng* 47:423–433
- Wang L, Dash S, Ng CY, Maranas CD (2017) A review of computational tools for design and reconstruction of metabolic pathways. *Synth Syst Biotechnol* 2:243–252
- Wortel MT, Noor E, Ferris M, Bruggeman FJ, Liebermeister W (2018) Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *Plos Comput Biol* 14:e1006010
- Yang X, Mao Z, Zhao X, Wang R et al (2020) Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *bioRxiv* 2020.2011.2030.403519
- Yim H, Haselbeck R, Niu W, Pujol-Baxley C et al (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol* 7:445–452