Biotechnology Advances 54 (2022) 107806

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Enhancing in silico strain design predictions through next generation metabolic modeling approaches

Adil Alsiyabi [a], Niaz Bahar Chowdhury [a,1], Dianna Long [b,1], Rajib Saha [a,b,*]

[a] Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, United States of America
[b] Complex Biosystems, University of Nebraska-Lincoln, United States of America

ARTICLE INFO

Keywords:
Metabolic modeling
Strain design
Thermodynamic analysis
Resource allocation
Kinetic modeling
Metabolic engineering

ABSTRACT

The reconstruction and analysis of metabolic models has garnered increasing attention due to the multitude of applications in which these have proven to be practical. The growing number of generated metabolic models has been accompanied by an exponentially expanding arsenal of tools used to analyze them. In this work, we discussed the biological relevance of a number of promising modeling frameworks, focusing on the questions and hypotheses each method is equipped to address. To this end, we critically analyzed the steady-state modeling approaches focusing on resource allocation and incorporation of thermodynamic considerations which produce promising results and aid in the generation and experimental validation of numerous predictions. For smaller networks involving more complex regulation, we addressed kinetic modeling techniques which show encouraging results in addressing questions outside the scope of steady-state modeling. Finally, we discussed the potential application of the discussed frameworks within the field of strain design. Adoption of such methodologies is believed to significantly enhance the accuracy of in silico predictions and hence decrease the number of design-build-test cycles required.

1. Introduction

Since the advent of the first genome-scale metabolic model (GEM) of *Haemophilus influenzae* RD in 1999 (Edwards and Palsson, 1999), GEM reconstruction has become an established protocol to perform systems-level investigation of metabolic networks of different species (Thiele and Palsson, 2010). These reconstructions constitute a mathematical representation of the actual metabolic network and enable in silico determination of intracellular metabolic activity and hence phenotypic behavior under different contexts (Lewis et al., 2012; Price et al., 2004). GEMs also serve as a platform for the integration and analysis of various types of omics data such as transcriptomics, proteomics, metabolomics, and fluxomics (Kim et al., 2015; O'Brien et al., 2015; Ryu et al., 2015). This external data-integration has transformed genome-scale metabolic modeling into a data-driven discipline by enabling scientists to simultaneously measure and incorporate data on large numbers of molecular components (e.g., nucleic acids, proteins, and metabolites) into a single coherent framework that can then be analyzed to generate testable predictions (Zhang et al., 2009).

Due to the underdetermined nature of GEMs, principles from optimization and constraint-based modeling (CBM) are often utilized to analyze such models (Islam and Saha, 2018). Given a defined cellular objective, CBM frameworks can identify a feasible metabolic state(s) corresponding to the optimal objective value (e.g. maximal growth rate) (Islam and Saha, 2018). Following CBM principles, computational strain design tools based on GEMs aim to identify the minimal set of genetic interventions resulting in optimal production of a desired metabolite (Paulo et al., 2021). These computational techniques can be broadly divided into two classes. In the first class, the identified set of genetic interventions aims to couple product formation to biomass generation (Machado and Herrgård, 2015). The most widely used frameworks in this class are based on the bi-level programming framework OptKnock (Burgard et al., 2003). Other tools have been developed that use elementary mode analysis to identify genetic interventions without having to assume a predefined biological objective (Trinh and Thompson, 2012). More recently, another class of strain design tools based on orthogonality principles has been developed that identifies genetic interventions aimed at de-coupling the activity of the pathway of interest from biomass generation (Pandit et al., 2017). A case study predicting succinate overproduction strategies utilizing this approach showed that

* Corresponding author at: Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, United States of America.
E-mail address: rsaha2@unl.edu (R. Saha).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.biotechadv.2021.107806
Received 29 March 2021; Received in revised form 22 June 2021; Accepted 15 July 2021
Available online 21 July 2021

it was advantageous compared to growth-coupled approaches (Pandit et al., 2017). With the aid of these tools, the field of model-guided strain design has resulted in remarkable improvements in the production of various industrially relevant products (Simeonidis and Price, 2015). Most recently, GEMs of the industrially relevant bacteria *Thermotoga* sp. and *Geobacillus icigianus* were reconstructed and used to identify strategies to improve production of an array of commercially viable metabolites including 2,3-butandiol (Simeonidis and Price, 2015). Furthermore, a GEM of *Saccharomyces cerevisiae* was used to identify genetic interventions that would enhance its capability to overproduce succinate, ethanol, and 2,3-butanediol (Shen et al., 2019). A number of excellent review articles describing current state of the art strain design methodologies and their applications have been published (Burgard et al., 2003; Chen et al., 2020; Chowdhury et al., 2015; Feist et al., 2010; Tian et al., 2017).

However, due to the computational complexity of incorporating thermo-kinetic constraints or expanding the metabolic network to account for other cellular processes, current strain design tools are primarily based on steady-state mass balance constrains and neglect to consider other physical and biological aspects that reduce the actual allowable solution space of potential strain candidates. Therefore, they generate a large set of potential genetic interventions, a significant portion of which are not physiologically feasible due to inherent thermodynamic, regulatory, or kinetic constraints. Furthermore, certain bioproducts can only be made by specific species when cells are in the stationary growth phase such as the malynol-coA-derived metabolites 3-hydroxypropionic acid and naringenin produced in *E. coli* (Tokuyama et al., 2019) and poly-β-hydroxybutyrate produced in *Synechocystis* sp. PCC 6803 (Koch et al., 2020). Considerations including thermodynamic feasibility of a reaction/pathway and the change in required resources (i.e., enzymes) under different conditions have yet to find widespread use in strain design. Incorporation of these constraints in other contexts has resulted in significant improvements in prediction accuracy compared to purely stoichiometric methods (Fleming et al., 2009; Goelzer et al., 2015; Jankowski et al., 2008; Mori et al., 2016; Sánchez et al., 2017).

Despite the wide array of successful applications achieved using CBMs, these methods suffer from two inherent drawbacks, which can potentially lead to inaccurate strain design predictions. First, these methods rely on optimizing an objective function such as maximizing the rate through a particular reaction. Therefore, the accuracy of the solutions obtained is dependent on how representative the chosen objective is of the metabolic state. Experimental studies showed that the cell's metabolic activity is a result of multiple objectives (Schuetz et al., 2012). Furthermore, the weight of each objective varies depending on the growth condition (Schuetz et al., 2012). Moreover, optimization of these objectives occurs over evolutionary timescales and is therefore most apparent in wild-type species (Segrè et al., 2002). It is therefore not realistic to assume that the cell will re-optimize its metabolic activity once a genetic intervention is introduced (Segrè et al., 2002). Instead, the thermo-kinetic attributes of the enzymes will naturally cause the metabolic activity to reach a new steady state. Consequently, the greater the introduced genetic intervention disrupts the "native" metabolic activity, the less accurate the optimization-based approach becomes. Second, constraint-based methods developed up to date are not capable of directly accounting for regulatory interactions in the network (Strutz et al., 2019). Hence it is difficult for such methods to predict the effect that such interactions may play once a genetic intervention is introduced. For these reasons, the modeling community have been interested in developing frameworks that can address these drawbacks. One such approach which has been garnering increasing attention is kinetic modeling. In this approach, metabolic and regulatory processes are described through kinetic expressions such as mass-action or Michaelis-Menten (Kim et al., 2018). Starting with a set of initial conditions, the temporal behavior of the metabolic and regulatory network can be determined. However, until recently, practical application of this framework has been dependent on the availability of measured enzyme kinetic parameters for all enzymes in the network, which is usually not feasible (Srinivasan et al., 2015). In the past decade, a significant effort has gone into developing methodologies capable of constructing large-scale kinetic models using more attainable omics data (i.e. fluxomics, metabolomics, etc.). These efforts have cultivated in the development of unified and self-containing frameworks such as MASSpy that allow the generation of standardized kinetic models (Haiman et al., 2021). Several excellent reviews have described the different kinetic frameworks available and the limitations associated with each (Chowdhury et al., 2015; Foster et al., 2021; Link et al., 2014; Saa and Nielsen, 2017; Strutz et al., 2019). These methods have been used to hypothesize reaction and regulatory mechanisms (Alsiyabi et al., 2021; Link et al., 2013; Schroeder and Saha, 2019), suggest engineering interventions (Foster et al., 2021; Kim et al., 2018; Tan and Liao, 2012), and analyze network sensitivity to predict rate-limiting reactions in a pathway (Theisen et al., 2016). While several kinetic modeling frameworks have been developed, this review focuses primarily on ensemble modeling (EM) (Greene et al., 2017; Lee et al., 2014; Rizk and Liao, 2009a; Tan and Liao, 2012; Tran et al., 2008; Zomorrodi et al., 2013) which is one of the most widely implemented methods and requires relatively minimal experimental data compared to the other frameworks. Incorporation of kinetic information during the process of strain design is expected to result in more refined and higher confidence engineering suggestions that subsequently require lesser resources to validate (Islam et al., 2021).

This review describes a number of recently developed approaches that can be incorporated into the strain design process. The following three sections describe how concepts from thermodynamics, protein allocation, and enzyme kinetics can be applied to metabolic networks. Each section provides a simplified description of the biological underpinnings underlying each concept and discusses a number of recently developed approaches that incorporate such frameworks. The reviewed approaches were chosen based on analysis of current trends in the modeling community which reveal their widespread adoption across many different modeling applications. Recent successful applications of each framework are also discussed. Finally, each of the described approaches is discussed in terms of the utility for engineering applications. This discussion aims to provide a general guideline regarding the type of approach most useful based on the organism of interest and the available data. Following such a systematic approach during the computational strain design phase is expected to significantly accelerate the rate of enhanced strain development.

2. Thermodynamic based approach

Despite the convoluted nature of metabolic networks in living organisms, metabolism, just like any other natural process, follows the laws of thermodynamics. Therefore, it is important to understand the fundamental principles that connect thermodynamic properties to the metabolic networks (Dai and Locasale, 2018). As laws of thermodynamics are applicable to metabolic networks, including thermodynamic properties of reactions greatly increases the predictability of stoichiometry-based methods (SBMs) by reducing the feasible solution space of the system. The second law of thermodynamics states that a positive net flux through a reaction always corresponds to a negative change in the Gibb's free energy of the reaction and vice versa (Hoppe et al., 2007). Therefore, incorporation of thermodynamic information is useful to determine reaction directionality (Henry et al., 2006), avoid infeasible cycles (Henry et al., 2007; Schellenberger et al., 2011), and identify different regulatory locations in the metabolic network (Henry et al., 2006; Kümmel et al., 2006). Here we focus on a number of widely used thermodynamics-based modeling frameworks including Thermodynamics-based Metabolic Flux Analysis (TMFA) and Max-min driving force (MDF) analysis to illustrate the predictive capability of such methods.

In TMFA (Henry et al., 2007), a set of linear thermodynamic

constraints ensuring proper reaction directionality is added to the conventional steady state mass-balance constraints of Flux Balance Analysis (FBA). In the TMFA algorithm, all reversible reactions are decomposed into a forward and backward direction to determine whether a reaction can carry a nonzero flux in either direction. This is achieved using the standard Gibbs free energy values either measured or calculated by the group-contribution method (Constantinou and Gani, 1994). Additionally, if the standard Gibbs free energy change of certain reactions is not known and cannot be reliably estimated using group contribution methods, those reactions are lumped into a new set of metabolic transformations for which the standard Gibbs free energy change can be estimated (Henry et al., 2007). TMFA has been incorporated as part of the computational framework called the Biochemical Network Integrated Computational Explorer (BNICE) (Henry et al., 2010) which can be used for automated design and evaluation of novel biosynthesis routes. BNICE relies on the TMFA framework to calculate Gibbs free energy changes to determine the thermodynamic feasibility of reactions in the given network. With these capabilities, BNICE was used to evaluate routes for production of 3-hydroxypropanoate (3-HP) from pyruvate within *Escherichia coli*. Through this approach, researchers were able to calculate maximum theoretical yield of cellular production of 3-HP and predicted that pyruvate and succinate stand out as the most efficient intermediates to produce 3-HP (Henry et al., 2010). Although the TMFA approach is useful in strain design, it is limited by its inability to quantify the extent to which free energy restricts flux through a reaction and therefore cannot predict the most thermodynamically efficient route.

To expand the predictive capability of thermodynamic-based modeling approaches, the MDF framework was developed to quantify the effect of free energy on the activity of a metabolic pathway (Noor et al., 2014). MDF assumes that activity through a linear pathway is limited by a thermodynamic bottleneck in one of the participating reactions. Therefore, the MDF algorithm identifies a concentration profile which maximizes the thermodynamic driving force through this rate-limiting reaction. The biological interpretation underlying the framework is that reactions operating close to equilibrium (low thermodynamic driving force) require significantly more enzyme, and hence carbon and energy resources, to achieve the same forward rate as a reaction with a high driving force (Fig. 1). Therefore, in order to minimize the total enzyme requirement, pathways evolve to maximize the metabolic driving force associated with the rate-limiting reaction.

A recent framework named OptMDFpathway (Hädicke et al., 2018) which is an extension of MDF has been developed to generalize the use of this method to metabolic networks instead of single pathways. The Mixed Integer Linear Programming (MILP) based OptMDFpathway calculates both the optimal MDF for a desired phenotypic behavior and the respective pathways that support the optimal driving force. To show its efficacy, the OptMDFpathway approach was used to systematically identify all substrate-product combinations in *E. coli* where product synthesis allows for net CO_2 assimilation through thermodynamically feasible pathways. Since CO_2 fixation requires overcoming high thermodynamic potentials, this analysis required a method to search for pathways that are not only stoichiometrically but also thermodynamically feasible. The OptMDFpathway method enabled the identification of thermodynamically viable CO_2 fixation pathways on a genome-scale (Hädicke et al., 2018). Such applications illustrate how thermodynamics-based approaches can potentially be used to identify and objectively rank pathways based on their respective metabolic driving forces.

Although the application of thermodynamic considerations in metabolic modeling has opened a new direction for strain design, better methods are still required to constrain metabolite concentrations. Accurate estimation of intracellular driving forces is dependent on a prespecified set (or range) of metabolite concentrations. This limitation can be mitigated by using experimentally measured values for highly connected metabolites such as cofactors and central metabolites to further reduce the solution space and obtain more accurate predictions (Noor et al., 2014). Furthermore, some frameworks incorporate objectives to minimize the cellular concentration of metabolites to predict physiological concentrations (Tepper et al., 2013). In addition, thermodynamics-based methods inherently assume that metabolic flux is not restricted by other factors such as enzyme kinetics, saturation, or regulation. Therefore, it is apparent that the accuracy of such methods decreases in cases in which such factors are relevant. For example, as will be discussed in the following section, there are cases where enzyme availability is the limiting factor in a specific pathway and hence, other modeling approaches, such as resource allocation modeling, can provide more information.

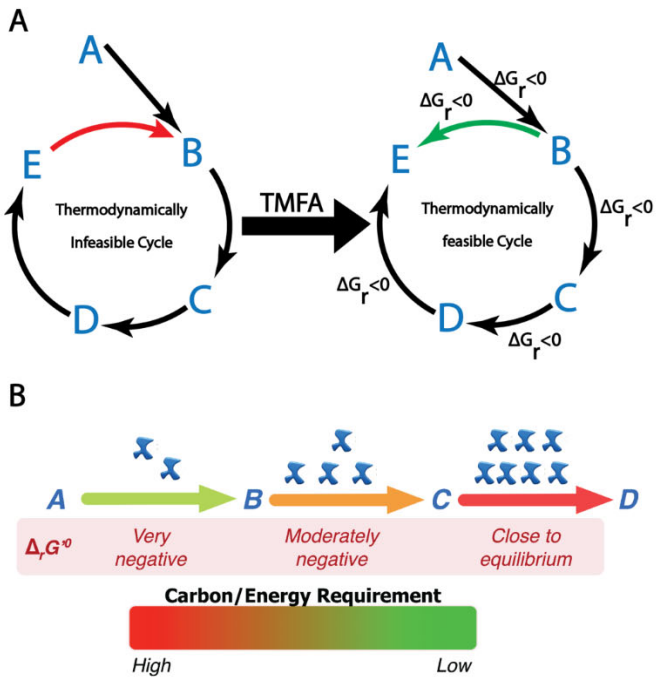

Fig. 1. Common approaches for incorporating thermodynamic constraints into metabolic models. (A) In TMFA, thermodynamic constraints are used to determine thermodynamically feasible directionality as well as flux and metabolite activity profiles. In this formulation, free energy does not directly constrict the rate of a reaction. (B) In MDF, the activity of a pathway is assumed to be limited by the thermodynamically limiting reaction. In this formulation, the enzymatic requirement of a reaction is assumed to be proportional to the thermodynamic driving force. Therefore, the maximal rate through a pathway is achieved by maximizing the metabolic driving force through the thermodynamically limiting reaction. In this illustration, the pathway converting metabolite A to D is limited by the highly carbon and energy intensive third reaction. The low thermodynamic through this reaction means that more enzyme is required to maintain the same flux as other reactions in the pathway.

3. Resource allocation-based approach

Resource allocation is another important aspect of metabolism to consider during strain design. Cells are self-replicating and require not only energy, but also machinery such as ribosomes and metabolic enzymes to duplicate and grow (Basan, 2018). Cells allocate resources for the machinery through cross-talk between metabolism and gene product expression, thus allowing for shifts in strategies in response to environmental stimuli (Basan, 2018; Donati et al., 2018). The "cost" of the enzymatic machinery plays a role in determining which strategy is followed and therefore helps determine metabolic processes and phenotypic profiles. For instance, when considering growth, purely stoichiometric approaches such as flux balance analysis may reveal two pathways resulting in the same flux of biomass precursors. However, the pathways may differ in the extent of enzymatic machinery required.

Consideration of the machinery cost would elucidate which pathway results in faster growth. Furthermore, conventional stoichiometry-based modeling frameworks account for non-metabolic energetic costs by incorporating ad hoc parameters such as growth and non-growth associated maintenance requirements (Thiele and Palsson, 2010). These parameters, often obtained through experimental data, are usually assumed to remain constant across different growth conditions. The inclusion of the expression matrix in resource allocation based methods allows direct incorporation of the cellular processes incurring a significant portion of such costs (O'Brien et al., 2013), allowing for direct and condition-specific predictions of the overall energy requirements. Understanding how resource allocation impacts cell phenotypes is important for advancements in understanding phenomena such as the Crabtree effect in yeast cells and the Warburg effect in cancers cells, which are both instances where fast growing cells are utilizing low yield pathways (Goel et al., 2012). Take, for example, the Crabtree effect described in yeast cells (Pfeiffer and Morley, 2014). This effect describes the phenomenon in which some yeast cells produce ethanol via fermentation in aerobic environments with high glucose concentrations. This is interesting because typically, when oxygen is present, the cells produce biomass via respiration, which is a more energy efficient metabolic pathway than fermentation. The shift to less efficient metabolism seen in the Crabtree effect can be explained by taking into account the limited membrane space that enzymes can occupy (Goel et al., 2012). In order to maintain cell membrane integrity, the ratio of membrane-bound proteins to lipids cannot get too high, When the limit is reached for respiratory proteins, that is when fermentation kicks in and acts as an alternative pathway for ATP production. The shift to fermentation explained by membrane space limitations is a clear example of how understanding resource limitations and allocation strategies can elucidate unintuitive phenotypes. Such strategies can be explained by looking at resource allocation and enzyme cost. Therefore, resource allocation should be considered during strain design since it is important for predicting behaviors such as catabolite repression and overflow metabolism (Basan, 2018) which are two important aspects to consider during strain design. Metabolic models that can capture resource allocation within a cell can allow for strain design with more accurate phenotypic predictions. Additionally, accounting for enzyme cost would allow for strain design to avoid pathway structures with too high enzyme per unit flux, which could be outcompeted during evolution or allow for ranking of strain design suggestions based on enzyme cost (Lerman et al., 2012). While several methods have been developed to address resource allocation, two of the main methodologies, Metabolic and Expression (ME)-modeling and enzyme cost minimization (ECM), will be discussed in this paper.

ME-models are genome-scale optimality models of metabolism that not only include the inputs of metabolic stoichiometry, but also include gene expression information (O'Brien et al., 2013) (See Fig. 2). This is similar to Resource Balance Analysis (Goelzer et al., 2011), a methodology developed around the same time as ME-modeling. RBA incorporates metabolic fluxes as well as protein concentrations and in doing so allows for cells to be modeled as a set of subsystems which all share resources. RBA and ME-modeling both improve upon simple metabolic models in similar ways (Goelzer and Fromion, 2011), and in this review we will discuss only ME-models in more depth.. As input, ME-models require the conditions of a steady-state environment and can then output predictions for maximum growth rate, substrate uptake and by-product secretion, metabolic fluxes, and gene expression levels. This method utilizes a growth optimization function along with coupling constraints that tie flux to transcriptional and translational reactions in the model. These constraints are functions of the growth rate and are included as rows in the stoichiometric matrix. The three types of coupling constraints include those to approximate the passage of intact transcription units to daughter cells, limit the number of times mRNA can be translated before degradation, and approximate enzyme abundance and activity. These coupling constraints are based on the effective

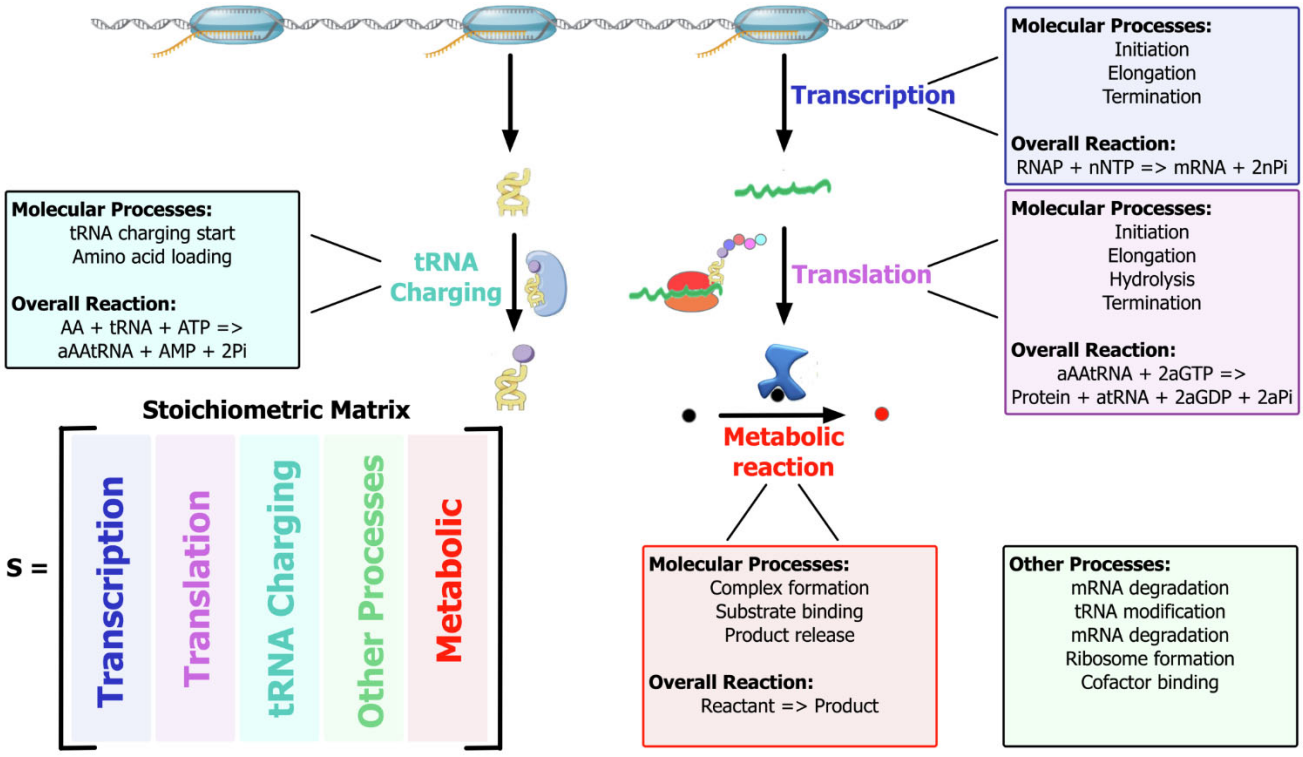

Fig. 2. Overview of metabolic processes incorporated into metabolic and expression modeling. Molecular processes involved in the expression of enzymes are incorporated as reactions similar to metabolic transformations. These reactions can then be appended to the stoichiometric matrix of GEMs, where each macromolecular component constitutes an additional row in the matrix and each process corresponds to an independent row. The stoichiometric coefficients of catalytic macromolecules (e.g. enzymes) correspond to their turnover rate, which corresponds to the number of reactions catalyzed per doubling time.

catalytic rate, which is the rate at which the enzyme-substrate complex dissociates into free enzyme and product, and the degradation rate, which is the rate at which the molecular machinery breaks down. By including these constraints, ME-models set limitations on fluxes based on transcription, translation, and the dilution of cellular machinery to daughter cells. Since ME-models account for dilution of structural materials, it is possible to utilize a structural equation that can account for composition changes based on environmental variations and growth rate (Lerman et al., 2012). For an illustrative diagram explaining ME-modeling, we recommend Fig. 1 of Lerman et al., 2012.

The ME-modeling approach was used in *E. coli* to accurately predict non-linear growth patterns and secretion profiles that account for both nutrient limiting and proteome limiting conditions (O'Brien et al., 2013) and to identify mutations linked to increased growth rate (LaCroix et al., 2015). ME-modeling was also applied in other organisms including *Thermotoga maritima* and *Clostridium ljungdahlii*. In *Thermotoga maritima* ME-modeling provided accurate predictions of growth and secretions as well as improved upon the known genome and transcriptome annotations by identifying new regulons (Lerman et al., 2012). In *Clostridium ljungdahlii,* an ME-model demonstrated improvement over an M-model by providing predictions in both batch and nutrient limited conditions, providing simultaneous predictions of carbon uptake and maximal growth rate, and intrinsically elucidating alternative fermentation pathways leading to overflow metabolism (Liu et al., 2019).

ECM (Noor et al., 2016) is another modeling methodology that captures resource allocation; however, unlike ME-modeling, ECM can incorporate enzyme kinetics and metabolite concentrations. This is a tiered approach that utilizes an algorithm based on convex optimization to predict steady-state enzyme costs and allows implementation with scalable amounts of input. Since the exact metabolite concentrations are not often known and measuring these in each possible growth condition would be impractical, ECM works by setting ranges for metabolite levels, imposing thermodynamic constraints, and then choosing the solution by minimizing enzyme cost. The enzyme cost function can be based on different possible burdens such as enzyme mass or amino acid composition. In scenarios where less information is available, lower ECM tiers allow for more simplified rate laws to be computed. The simplest tier, ECM0, is equivalent to parsimonious FBA (Lewis et al., 2010) and utilizes only a metabolic network with steady-state fluxes. ECM0 sets all enzymes to the same catalytic constants and burdens, thus setting the enzyme cost proportional to the sum of fluxes. However, there can be vastly different catalytic constants and enzyme masses, making ECM0 oversimplified. ECM1 adds an additional layer of complexity by including catalytic constants and enzyme burden, allowing for individual flux burden for each enzyme. ECM1, however, is still simplified by setting all enzymes to their maximal rates and assumes irreversibility and substrate saturation. ECM2 includes terms for reversibility and ECM3 includes terms for both reversibility and substrate saturation, thus allowing for more complexity to be included. ECM4, the highest tier, includes the most realistic rate law which can include specific mechanisms for substrate binding and allows for the most complexity. This scalable method thus allows for flexibility based on desired complexity of the model and available data. For an illustrative diagram explaining ECM, we suggest Figure 8 of Noor et al., 2016.

Comparisons of model-based predictions of enzyme levels in *E. coli* demonstrated that ECM performed better than a time-dependent kinetic model of *E. coli*'s central metabolism, thus supporting the reasoning that enzyme cost is important in cell function (Noor et al., 2016). In another study, ECM was used to solve across all possible elementary flux modes to specifically study growth/yield trade-offs in *E. coli* (Wortel et al., 2018). This study elucidated that when oxygen was being consumed, the reduced oxygen levels lowered the growth rate by requiring higher enzyme levels in oxidative phosphorylation. This is an example showing how ECM can shed light on very specific mechanisms.

Both ME-models and ECM provide modeling frameworks for incorporating resource allocation considerations which allow for elucidation of instances where metabolic processes are limited by machinery availability. These two strategies provide options for researchers to model a system based on available data using two different approaches: ME-models are an extension of conventional GEMs and can therefore predict steady-state reaction rates of all modeled molecular processes. However, they do not consider metabolite concentrations and therefore cannot incorporate thermodynamic constraints or predict metabolite levels. This framework is ideal for incorporating transcriptomic or proteomic data as well as validating model predictions against such data. On the other hand, ECM relies on experimental reaction rate data to predict optimal enzyme and metabolite levels. Therefore, metabolomics data can be directly incorporated into the framework. Moreover, both proteomic and metabolomic data can be used to validate model predictions. Either approach allows for improved predictions on certain phenotypes such as overflow metabolism (Basan, 2018), which can be an important aspect to consider during strain design. These modeling frameworks do, however, come with limitations. While shown to be useful, ME-models are limited by the necessary set of assumptions, which are comparable to those made in the ECM1 tier mentioned previously. It has been shown that the effective catalytic rates of enzymes vary under nutrient limitation, with the trend being a decrease in activity (O'Brien et al., 2013). ME-models account for this by making two assumptions; 1) under nutrient limitation, proteins content is maximized, and 2) the catalytic rates of all proteins are below maximum (O'Brien et al., 2013). With these assumptions, all catalytic rates are decreased, which could result in inaccurate values for certain enzymes that are important in specific nutrient limiting conditions. A model that accounts for the relationship between metabolite concentrations and enzymatic activities would improve the accuracy of the predictions generated similar to what is found in enzyme cost minimization. However, ECM also suffers a major limitation due to the scarcity of the kinetic data required. Therefore, parameter predictions and simplifications must be used for lower ECM tiers, which might result in inaccurate predictions. In order to address this limitation, there is a need for easier ways to generate "kinetomes," libraries of all kinetic parameters, for organisms of interest, either through in vitro characterization or reliable model predictions (Nilsson et al., 2017). Even with all the data available, ECM still has some limitations. Unlike ME-models, ECM does not explicitly incorporate transcription and translation reactions which allows for less fine-tuned exploration of cellular resource allocation. Sometimes minimizing enzyme cost cannot predict cellular behavior (Nilsson et al., 2017). There may be benefits to cells maintaining some unused level of proteins in order to adapt to new environments or chemical storage and ECM does not account for such situations. Furthermore, both ME-models and ECM are limited by the assumption that the systems are operating at steady-state. Since, steady state assumptions do not allow for dynamic modeling of the system, this limits the ability to incorporate cellular regulation. Next, we discuss kinetic models, which can capture dynamic behaviors over time and allow more direct incorporation of regulatory mechanisms.

4. Kinetic modeling approach

Kinetic modeling allows for a detailed description of the metabolic and regulatory processes through non-linear kinetic expressions (Chowdhury et al., 2015; Foster et al., 2021; Islam et al., 2021). In this section, we focus on the EM approach as a tool to describe the kinetics of metabolic networks using readily available experimental data (Khodayari and Maranas, 2016; Tan and Liao, 2012; Tran et al., 2008). The EM framework was developed to sample through the entire allowable kinetic solution space to generate an ensemble of kinetic models that describe the system (Tan et al., 2011). By doing so, the EM approach overcomes the limitation of data sparsity regarding kinetic parameters. Furthermore, only standard Gibbs free energy of the modeled reaction and wild-type flux distributions are required to parameterize the initial ensemble of models. Gibbs free energies are now routinely determined

using either group or component contribution methods (Jankowski et al., 2008; Noor et al., 2013). Furthermore, the use of ^{13}C MFA methods to elucidate flux distributions has also become prevalent in systems biology (Long and Antoniewicz, 2019). When such in vivo flux data is unavailable, uptake and secretion rates can be measured and incorporated into FBA methods to predict the in silico flux distribution (Rizk et al., 2011). This generated ensemble is then filtered using prior knowledge of the system's response to different genetic perturbations. Models passing all applied filtration steps are considered to be highly representative of the actual system kinetics (Rizk and Liao, 2009b). The EM approach has been used to capture the inherent non-linearity of metabolic systems and to identify metabolic bottlenecks in the production of various industrially relevant compounds including ethanol (Greene et al., 2019), aromatics (Rizk and Liao, 2009b), and L-lysine (Contador et al., 2009). In addition, EM has been used to predict the presence of regulatory interactions occurring in biochemical pathways (Alsiyabi et al., 2021; Link et al., 2013). It is of note that these predictions would not have been possible through constraint-based steady-state approaches.

Kinetic modeling approaches, including EM rely on a number of assumptions that result in some inherent limitations. Firstly, most kinetic modeling algorithms rely on solving a set of stiff ordinary differential equations (ODEs) and are therefore computationally costly (Srinivasan et al., 2015). However, recent advances significantly reduce the burden of such computations by converting the system of ODEs to a set of algebraic equations (Foster et al., 2019, 2021; Gopalakrishnan et al., 2020). Furthermore, model reduction techniques can often be implemented to reduce the number of parameters without sacrificing predictive power (Greene et al., 2017). The validity of the predictions made by these methods is also reliant on the accuracy of the network structure (Strutz et al., 2019). Therefore, missing reactions or regulatory interactions may result in discrepancies between prediction and observation. Although approaches to test different hypothetical network structures have been implemented (Alsiyabi et al., 2021; Steuer et al., 2006), it is often infeasible to test the entire solution space of metabolite-enzyme interactions due to the current computational cost of simulating each regulatory network. Furthermore, regulatory interactions in which the regulator is not part of the metabolic network such as enzyme phosphorylation or protein-protein interactions are currently outside the scope of the described frameworks. In addition, the accuracy of the fitted kinetic parameters is highly dependent on the experimental datasets used to train the model. For example, models trained on data obtained during aerobic growth only tend to have lower accuracy when tested under anaerobic conditions (Khodayari and Maranas, 2016). Moreover, the EM approach assumes that a genetic perturbation does not affect the expression of non-perturbed enzymes. However, this assumption is not necessary if proteomic data is available for both the reference and perturbed states. In cases where proteomic data is not available, resource allocation methods such as ME modeling or ECM can be implemented to predict how genetic or environmental changes affect the protein expression of each enzyme in the network. These predictions can subsequently be used to parameterize the kinetic model. Finally, current kinetic modeling methods do not incorporate transcriptional and translational regulation. Since the effect of such regulation is ultimately on protein levels, this limitation is related to the assumption described above, where the effect of changes in protein expression on metabolic activity is assumed to be negligible compared to the metabolic effect resulting from the applied perturbation. Therefore, implementation of these methods without incorporating proteomic data (experimental or predicted) may not be appropriate under conditions in which the effect of such regulation greatly affects the metabolic state. These limitations highlight the need for incorporating experimental or computational proteomic measurements into kinetic models to improve their accuracy.

In addition to the limitations and inherent assumptions highlighted above, EM and other kinetic modeling frameworks have other challenges that constrain their use in strain design. Most notably are the long solve times associated with solving a set of nonlinear stiff ordinary differential equations (Strutz et al., 2019). Although recently developed methods such as K-FIT (Gopalakrishnan et al., 2020) have made significant improvements in this regard, incorporating large-scale kinetic models into a strain design optimization framework remains infeasible. To circumvent this issue, methods such as k-OptForce (Chowdhury et al., 2014) integrate small-scale kinetic models with genome-scale stoichiometric models to improve strain design predictions. Alternatively, an exhaustive set of potential genetic interventions can be generated using conventional methodologies and subsequently tested on a kinetic model to filter out non-viable solutions. As the computational burden of solving such models decreases, future approaches may utilize concepts such as optimal control (Tsiantis and Banga, 2020) or cybernetic modeling (Varner and Ramkrishna, 1999) to systematically predict regulatory interactions on a genome-scale.

5. Perspectives: incorporating thermo-kinetic approaches into strain design

A general goal of metabolic modeling is to simulate and accurately predict an organism's phenotype under any growth condition. Within metabolic engineering, this translates to predicting how genotypic changes (e.g., gene knockouts) affect the desired product's titer, productivity, and yield. Numerous strain design frameworks were developed to suggest such genetic interventions (Hendry et al., 2020; St John and Bomble, 2019). Such methods were extensively reviewed (Gu et al., 2019; Machado and Herrgård, 2015; Mienda, 2017) with the majority of which evolved from the widely utilized OptKnock algorithm (Burgard et al., 2003; Machado and Herrgård, 2015). Briefly, these tools rely on multi-level optimization formulations that search for genetic interventions in the form of gene insertions, knockouts or modulations that lead to improved production metrics. However, application of such methods usually results in multiple optimal solutions with all predicted to have the same outcome (i.e., product yield). Furthermore, the suggested interventions often do not lead to the predicted results in reality, due to the lack of the incorporation of the thermo-kinetic factors that often turn out to play a significant role in the mutant strain. Moreover, metabolic engineering efforts often resort to overexpression of all genes involved in a given pathway to brute-force productivity. However, in addition to the time required to construct the required synthetic biology tools, such an approach can have unintended outcomes such as a decline in growth rate or increased by-product formation (Gasser et al., 2008; Schalén et al., 2016). Incorporation of the approaches discussed in the previous sections will significantly enhance the strain design process by addressing some of the inherent shortcomings in current procedures.

The decision of which of the discussed approaches to incorporate relies on the availability of the required data and on bibliomic knowledge of the organism of interest. The following example demonstrates a number of commonly encountered cases during strain design where implementation of the discussed approaches can be used to refine genetic engineering suggestions (Fig. 3). The pathway depicted in Fig. 3A illustrates a simplified metabolic network containing alternate pathways that lead to the growth-associated product of interest X_p. In this depiction, breakdown of the substrate generates the reduced cofactor *CH2* required in each of the two alternate product-forming pathways. A significant number of bioengineering objectives are hindered by cofactor availability (Akhtar and Jones, 2014; de Arroyo Garcia and Jones, 2020). Due to the participation of cofactors in a large number of reactions in metabolic networks, genetic perturbations often lead to imbalances in cofactor pools which manifests as a disruption in their homeostatic redox state. In fact, even substitution of the main carbon source may cause a significant effect on the redox state of various electron carriers (Liu et al., 2018). As discussed previously, this disturbance directly affects the thermodynamic driving force and hence the overall rate of any reaction incorporating such cofactors (see

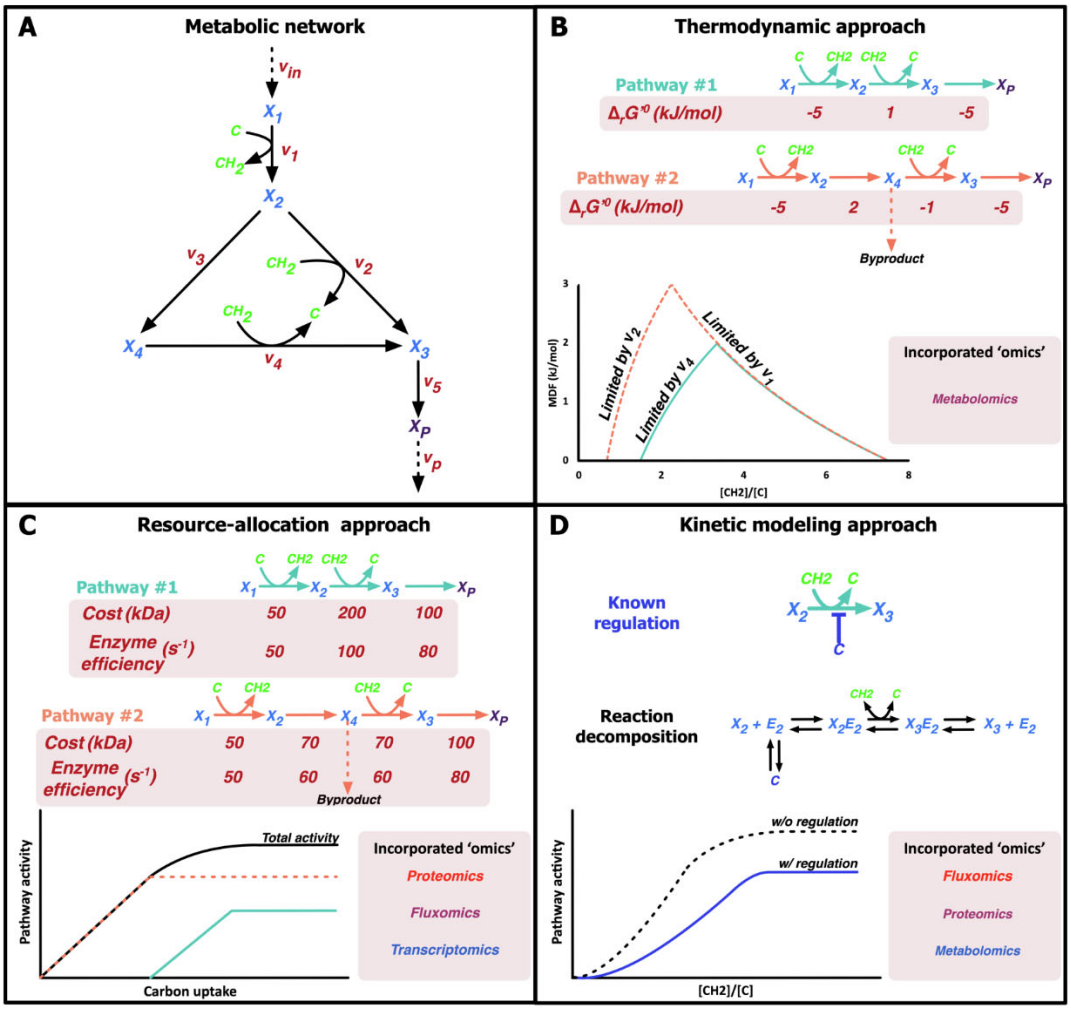

Fig. 3. Applicability of different modeling approaches within the context of strain design. (A) Simplified depiction of a metabolic network utilizing X_1 as the main carbon source and producing the growth-coupled product X_p. Two alternate pathways exist that can transform the carbon source to product. (B) Illustration of how a thermodynamic modeling approach can identify the change in metabolic driving force as a function of cofactor redox state. The plot illustrates how the driving force of both independent production pathways is dependent on the cofactor redox state. At low redox ratios, both pathways are constricted by the low driving force of v_1. At high redox ratios, the two pathways are constricted by the driving force through v_2 and v_4, respectively. Metabolomics data can be used to identify the physiological ranges of metabolite concentrations and redox states to identify the organism specific range of MDFs for each pathway. (C) Illustration of how resource-allocation can lead to differential pathway expression under varying carbon uptake rates. The plot illustrates that at low carbon uptake rates, the cell prefers the "cheaper" pathway since the production rate is limited by carbon uptake. As the uptake rate increases, the cell activates the more expensive but highly efficient pathway to enable high production rates. Proteomic and fluxomic data can be used to improve the accuracy of enzyme turnover rates (k_{eff}). Furthermore, transcriptomic data can be used to further constrict the solution space by constraining transcription reactions. (D) Illustration of how kinetic modeling can be used to predict the effect of regulatory interactions on pathway activity. The plot demonstrates how incorporation of regulatory interactions in the network can affect yield predictions. Fluxomic data is required to parameterize the kinetic model. Proteomic data can be used to improve the parameterization procedure by accounting for changes in protein concentrations across different conditions. Finally, metabolomic data can be used to calculate denormalized kinetic parameters, which can subsequently be compared to in vitro measured values. Omics data types written in red are required by the modeling approach, those in purple are not required but greatly improve accuracy, and those in blue are optional. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thermodynamic-based approach). It is therefore necessary to determine the effect of redox state on the metabolic driving force of the pathway of interest (Fig. 3B). The development of methods such as the group contribution method (Jankowski et al., 2008) and component contribution analysis (Noor et al., 2013) enables the determination of standard free energy values on a genome-scale (Niebel et al., 2019; Noor et al., 2013). Using this data, the thermodynamic-based approaches discussed can be used to investigate the effect of varying redox-state on the driving-force through the product-generating pathway. When limitations concerning cofactor redox-state arise, strategies as cofactor substitution, or the overexpression of certain enzymes in the pathway may be required. Frameworks such as MDF are ideal for identifying such potential bottlenecks and for testing the effect of different genetic interventions on the thermodynamic viability of metabolic pathways.

Another widely used metabolic engineering strategy relies on the overexpression of certain enzymes, or the incorporation of heterologous enzymes/pathways to expand the metabolic functionality of the organism (Fong, 2014). Such strategies confer additional carbon and energy costs that can lead to system-wide changes in metabolic activity (Chen and Nielsen, 2019). Conventional strain design frameworks based primarily on stoichiometric constraints cannot account for such effects as they inherently assume unlimited enzyme capacity. Therefore, it is often useful to utilize the discussed resource-allocation approaches to predict the effect of such perturbations (gene insertion or overexpression) to ensure that the designed growth conditions (e.g. substrate concentration) are sufficient to manage the added cost. Moreover, such

an approach is also advantageous when dealing with alternate production pathways (Fig. 3C). It has been observed that under such cases, metabolism has evolved to activate the most "cost-effective" pathway (Basan, 2018). Therefore, a resource-allocation based analysis can be used to ensure that the applied genetic interventions do not result in undesired carbon loss through by-product excretion. It can also be used to generate and test design strategies that ensure activity through the most carbon-efficient route. However, resource-allocation methods require the input of enzyme catalytic efficiencies (k_{eff}). Such parameters are not available on a genome-scale for most non-model organisms. Therefore, one approach that can be used to circumvent this bottleneck is to scale the catalytic efficiencies of model organisms by the enzyme specific solvent accessible surface area (SASA) (Miller et al., 1987). Recent approaches combining metabolic modeling with population genetics models have also demonstrated reasonable accuracy in predicting organism-specific catalytic turnover rates (Heckmann et al., 2018).

In addition to the metabolic considerations discussed so far, regulatory aspects of the network may also affect the production of desired compounds (Foster et al., 2019; Strutz et al., 2019). This could especially be true in the case of central metabolites such as fermentation products (Foster et al., 2019). Due to the tight regulation of central metabolism, allosteric metabolite-enzyme interactions can cause unexpected discrepancies between predicted and observed production metrics (Chowdhury et al., 2015). Recently, attempts were made to use kinetic modeling to account for such interactions (Foster et al., 2019; Gopalakrishnan et al., 2020; Khodayari et al., 2015; Khodayari and Maranas, 2016). Incorporation of allosteric interactions into the metabolic network allows prediction of how known (or postulated) regulatory interactions effect production metrics under different growth conditions or genetic interventions (Fig. 3D). Development of such models requires multiple sets of flux distributions and is therefore best suited for small to medium sized metabolic networks. Once the kinetic model is developed, the effect of allosteric interactions on the rate of production can be predicted. Results from such an analysis can be used to determine whether protein engineering efforts are required to eliminate the inherent regulation or whether other strategies can be used to circumvent such effects.

The decision of which approach to implement is dependent on the types of omics data available and on the specific pathway being optimized. In general, thermodynamic approaches such as MDF are not data intensive and can serve as useful tools in identifying the thermodynamic feasibility of different genetic interventions. In addition to readily available data on reaction standard Gibbs free energy values, incorporation of metabolomic data on highly connected metabolites such as cofactors measured under different conditions or obtained from closely related organisms can significantly constrict the solution space and hence omit infeasible strain designs. Moreover, in cases where transcriptomic or proteomic data is available, various modeling techniques based on resource allocation principles can be implemented to predict how a genetic intervention will affect overall metabolic activity. As discussed earlier, such techniques are especially useful in testing the effect of gene insertions or upregulations on productivity of the metabolite of interest. Finally, availability of fluxomic data can be leveraged to construct detailed kinetic models. These models can be further parameterized through metabolomic and proteomic data to increase prediction accuracy (Gopalakrishnan et al., 2020). As metabolic flux analysis (MFA) and isotope labeling techniques improve, the coverage of such models will increase and allow for the construction of near genome-scale kinetic models (Foster et al., 2019).

As computational resources increase, the approaches discussed here can be more directly implemented into the strain design process. Such efforts have already begun on a smaller scale and have indeed proven to yield significant improvements in terms of accuracy and identification of non-intuitive solutions (Chowdhury et al., 2014). To accelerate this process, improvements need to be made in several fronts. First, model reduction techniques (Greene et al., 2017) need to be improved and routinely implemented to reduce the number of variables without compromising model accuracy. Second, improvements in model formulation that minimize non-linearity and ensure convexity (Gopalakrishnan et al., 2020) will significantly reduce solve times and allow for larger scale models to be analyzed. Finally, implementation of sophisticated numerical techniques to solve stiff ordinary equations will facilitate the process of combining larger scale kinetic models with optimization-based strain design formulations.

6. Conclusion

The increase in omics data availability has allowed development of more complex metabolic modeling frameworks resulting in higher prediction accuracy. However, these techniques have yet to find widespread use in the field of computational strain design. This review describes three of the major metabolic modeling approaches (i.e., thermodynamic, resources allocation, and kinetic modeling) which have garnered the interest of the metabolic modeling community in recent years. The biological underpinnings of these approaches and how these can prove to be useful tools for metabolic engineering are discussed. We hope this review motivates the development of modeling frameworks that facilitate the incorporation of the discussed approaches within the context of strain design. We believe that such tools will significantly enhance the prediction accuracy and hence decrease the number of subsequent design-build-test cycles.

Declaration of Competing Interest

The authors declare no conflict of interests.

Acknowledgments

We gratefully acknowledge funding support from National Science Foundation (NSF) CAREER grant (25-1106-0039-001), NSF EPSCoR Center for Root and Rhizobiome Innovation grant (25-1215-0139-025), NSF MCB grant (25-1215-0175-001), Nebraska Center for Energy Science and Research grant (26-1217-0020-403), and Nebraska Corn Board Award (26-6248-0059-001).

References

Akhtar, M.K., Jones, P.R., 2014. Cofactor engineering for enhancing the flux of metabolic pathways. Front. Bioeng. Biotechnol. 2, 30. https://doi.org/10.3389/fbioe.2014.00030.

Alsiyabi, A., Solis, A.G., Cahoon, E.B., Saha, R., 2021. Dissecting the regulatory roles of ORM proteins in the sphingolipid pathway of plants. PLoS Comput. Biol. 17, e1008284.

Basan, M., 2018. Resource allocation and metabolism: the search for governing principles. Curr. Opin. Microbiol. 45, 77–83. https://doi.org/10.1016/j.mib.2018.02.008.

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.

Chen, Y., Nielsen, J., 2019. Energy metabolism controls phenotypes by protein efficiency and allocation. Proc. Natl. Acad. Sci. 116 https://doi.org/10.1073/pnas.1906569116, 17592 LP – 17597.

Chen, G.-Q., Chen, X.-Y., Wu, F.-Q., Chen, J.-C., 2020. Polyhydroxyalkanoates (PHA) toward cost competitiveness and functionality. Adv. Ind. Eng. Polym. Res. https://doi.org/10.1016/j.aiepr.2019.11.001.

Chowdhury, A., Zomorrodi, A.R., Maranas, C.D., 2014. K-OptForce: integrating kinetics with flux balance analysis for strain design. PLoS Comput. Biol. 10 e1003487.

Chowdhury, A., Khodayari, A., Maranas, C.D., 2015. Improving prediction fidelity of cellular metabolism with kinetic descriptions. Curr. Opin. Biotechnol. 36, 57–64. https://doi.org/10.1016/j.copbio.2015.08.011.

Constantinou, L., Gani, R., 1994. New group contribution method for estimating properties of pure compounds. AICHE J. 40, 1697–1710. https://doi.org/10.1002/aic.690401011.

Contador, C.A., Rizk, M.L., Asenjo, J.A., Liao, J.C., 2009. Ensemble modeling for strain development of l-lysine-producing Escherichia coli. Metab. Eng. 11, 221–233. https://doi.org/10.1016/j.ymben.2009.04.002.

Dai, Z., Locasale, J.W., 2018. Thermodynamic constraints on the regulation of metabolic fluxes. J. Biol. Chem. 293, 19725–19739. https://doi.org/10.1074/jbc.RA118.004372.

de Arroyo Garcia, L., Jones, P.R., 2020. In silico co-factor balance estimation using constraint-based modelling informs metabolic engineering in Escherichia coli. PLoS Comput. Biol. 16, e1008125.
Donati, S., Sander, T., Link, H., 2018. Crosstalk between transcription and metabolism: how much enzyme is enough for a cell? WIREs Syst. Biol. Med. 10, e1396 https://doi.org/10.1002/wsbm.1396.
Edwards, J.S., Palsson, B.O., 1999. Systems properties of the haemophilus influenzaeRd metabolic genotype*. J. Biol. Chem. 274, 17410–17416. https://doi.org/10.1074/jbc.274.25.17410.
Feist, A.M., Zielinski, D.C., Orth, J.D., Schellenberger, J., Herrgard, M.J., Palsson, B.Ø., 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.
Fleming, R.M.T., Thiele, I., Nasheuer, H.P., 2009. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. Biophys. Chem. 145, 47–56. https://doi.org/10.1016/j.bpc.2009.08.007.
Fong, S.S., 2014. Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. Comput. Struct. Biotechnol. J. 11, 28–34. https://doi.org/10.1016/j.csbj.2014.08.005.
Foster, C.J., Gopalakrishnan, S., Antoniewicz, M.R., Maranas, C.D., 2019. From Escherichia coli mutant 13C labeling data to a core kinetic model: a kinetic model parameterization pipeline. PLoS Comput. Biol. 15, e1007319.
Foster, C.J., Wang, L., Dinh, H.V., Suthers, P.F., Maranas, C.D., 2021. Building kinetic models for metabolic engineering. Curr. Opin. Biotechnol. 67, 35–41. https://doi.org/10.1016/j.copbio.2020.11.010.
Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., Porro, D., Ferrer, P., Tutino, M.L., Mattanovich, D., Villaverde, A., 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb. Cell Factories 7, 11. https://doi.org/10.1186/1475-2859-7-11.
Goel, A., Wortel, M.T., Molenaar, D., Teusink, B., 2012. Metabolic shifts: a fitness perspective for microbial cell factories. Biotechnol. Lett. 34, 2147–2160. https://doi.org/10.1007/s10529-012-1038-9.
Goelzer, A., Fromion, V., 2011. Bacterial growth rate reflects a bottleneck in resource allocation. Biochim. Biophys. Acta, Gen. Subj. 1810, 978–988. https://doi.org/10.1016/j.bbagen.2011.05.014.
Goelzer, A., Fromion, V., Scorletti, G., 2011. Cell design in bacteria as a convex optimization problem. Automatica 47, 1210–1218. https://doi.org/10.1016/j.automatica.2011.02.038.
Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadassou, M., Aymerich, S., Hecker, M., Noirot, P., Becher, D., Fromion, V., 2015. Quantitative prediction of genome-wide resource allocation in bacteria. Metab. Eng. 32, 232–243. https://doi.org/10.1016/j.ymben.2015.10.003.
Gopalakrishnan, S., Dash, S., Maranas, C., 2020. K-FIT: an accelerated kinetic parameterization algorithm using steady-state fluxomic data. Metab. Eng. 61, 197–205. https://doi.org/10.1016/j.ymben.2020.03.001.
Greene, J.L., Wäechter, A., Tyo, K.E.J., Broadbelt, L.J., 2017. Acceleration strategies to enhance metabolic ensemble modeling performance. Biophys. J. 113, 1150–1162. https://doi.org/10.1016/j.bpj.2017.07.018.
Greene, J., Daniell, J., Köpke, M., Broadbelt, L., Tyo, K.E.J., 2019. Kinetic ensemble model of gas fermenting Clostridium autoethanogenum for improved ethanol production. Biochem. Eng. J. 148, 46–56. https://doi.org/10.1016/j.bej.2019.04.021.
Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., Lee, S.Y., 2019. Current status and applications of genome-scale metabolic models. Genome Biol. 20, 121. https://doi.org/10.1186/s13059-019-1730-3.
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.
Haiman, Z.B., Zielinski, D.C., Koike, Y., Yurkovich, J.T., Palsson, B.O., 2021. MASSpy: building, simulating, and visualizing dynamic biological models in Python using mass action kinetics. PLoS Comput. Biol. 17, e1008208.
Heckmann, D., Zielinski, D.C., Palsson, B.O., 2018. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. Nat. Commun. 9, 5270. https://doi.org/10.1038/s41467-018-07649-1.
Hendry, J.I., Bandyopadhyay, A., Srinivasan, S., Pakrasi, H.B., Maranas, C.D., 2020. Metabolic model guided strain design of cyanobacteria. Curr. Opin. Biotechnol. 64, 17–23. https://doi.org/10.1016/j.copbio.2019.08.011.
Henry, C.S., Jankowski, M.D., Broadbelt, L.J., Hatzimanikatis, V., 2006. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. Biophys. J. 90, 1453–1461. https://doi.org/10.1529/biophysj.105.071720.
Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. Biophys. J. 92, 1792–1805. https://doi.org/10.1529/biophysj.106.093138.
Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2010. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol. Bioeng. 106, 462–473. https://doi.org/10.1002/bit.22673.
Hoppe, A., Hoffmann, S., Holzhütter, H.-G., 2007. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. BMC Syst. Biol. 1, 23. https://doi.org/10.1186/1752-0509-1-23.
Islam, M.M., Saha, R., 2018. In: Jensen, M.K., Keasling, J.D. (Eds.), Computational approaches on stoichiometric and kinetic modeling for efficient strain design BT - synthetic metabolic pathways: methods and protocols. Springer New York, New York, NY, pp. 63–82. https://doi.org/10.1007/978-1-4939-7295-1_5.
Islam, M.M., Schroeder, W.L., Saha, R., 2021. Kinetic modeling of metabolism: present and future. Curr. Opin. Syst. Biol. 26, 72–78. https://doi.org/10.1016/j.coisb.2021.04.003.
Jankowski, M.D., Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys. J. 95, 1487–1499. https://doi.org/10.1529/biophysj.107.124784.
Khodayari, A., Maranas, C.D., 2016. A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nat. Commun. 7 https://doi.org/10.1038/ncomms13806.
Khodayari, A., Chowdhury, A., Maranas, C.D., 2015. Succinate overproduction: a case study of computational strain design using a comprehensive *Escherichia coli* kinetic model. Front. Bioeng. Biotechnol. 2, 76. https://doi.org/10.3389/fbioe.2014.00076.
Kim, B., Kim, W.J., Kim, D.I., Lee, S.Y., 2015. Applications of genome-scale metabolic network model in metabolic engineering. J. Ind. Microbiol. Biotechnol. 42, 339–348. https://doi.org/10.1007/s10295-014-1554-9.
Kim, O.D., Rocha, M., Maia, P., 2018. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. Front. Microbiol. 9, 1690. https://doi.org/10.3389/fmicb.2018.01690.
Koch, M., Bruckmoser, J., Scholl, J., Hauf, W., Rieger, B., Forchhammer, K., 2020. Maximizing PHB content in Synechocystis sp. PCC 6803: a new metabolic engineering strategy based on the regulator PirC. Microb. Cell Factories 19, 231. https://doi.org/10.1186/s12934-020-01491-1.
Kümmel, A., Panke, S., Heinemann, M., 2006. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Mol. Syst. Biol. 2, 2006.0034. https://doi.org/10.1038/msb4100074.
LaCroix, R.A., Sandberg, T.E., Brien, E.J., Utrilla, J., Ebrahim, A., Guzman, G.I., Szubin, R., Palsson, B.O., Feist, A.M., 2015. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. Appl. Environ. Microbiol. 81, 17. https://doi.org/10.1128/AEM.02246-14.
Lee, Y., Lafontaine Rivera, J.G., Liao, J.C., 2014. Ensemble modeling for robustness analysis in engineering non-native metabolic pathways. Metab. Eng. 25, 63–71. https://doi.org/10.1016/j.ymben.2014.06.006.
Lerman, J.A., Hyduke, D.R., Latif, H., Portnoy, V.A., Lewis, N.E., Orth, J.D., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Zengler, K., Palsson, B.O., 2012. In silico method for modelling metabolism and gene product expression at genome scale. Nat. Commun. 3, 929. https://doi.org/10.1038/ncomms1928.
Lewis, N.E., Hixson, 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.Ø., 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6, 390. https://doi.org/10.1038/msb.2010.47.
Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10, 291–305. https://doi.org/10.1038/nrmicro2737.
Link, H., Kochanowski, K., Sauer, U., 2013. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nat. Biotechnol. 31, 357–361. https://doi.org/10.1038/nbt.2489.
Link, H., Christodoulou, D., Sauer, U., 2014. Advancing metabolic models with kinetic information. Curr. Opin. Biotechnol. 29, 8–14. https://doi.org/10.1016/j.copbio.2014.01.015.
Liu, J., Li, H., Zhao, G., Caiyin, Q., Qiao, J., 2018. Redox cofactor engineering in industrial microorganisms: strategies, recent applications and future directions. J. Ind. Microbiol. Biotechnol. 45, 313–327. https://doi.org/10.1007/s10295-018-2031-7.
Liu, J.K., Lloyd, C., Al-Bassam, M.M., Ebrahim, A., Kim, J.-N., Olson, C., Aksenov, A., Dorrestein, P., Zengler, K., 2019. Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen. PLoS Comput. Biol. 15, e1006848 https://doi.org/10.1371/journal.pcbi.1006848.
Long, C.P., Antoniewicz, M.R., 2019. High-resolution 13C metabolic flux analysis. Nat. Protoc. 14, 2856–2877. https://doi.org/10.1038/s41596-019-0204-0.
Machado, D., Herrgård, M.J., 2015. Co-evolution of strain design methods based on flux balance and elementary mode analysis. Metab. Eng. Commun. 2, 85–92. https://doi.org/10.1016/j.meteno.2015.04.001.
Mienda, B.S., 2017. Genome-scale metabolic models as platforms for strain design and biological discovery. J. Biomol. Struct. Dyn. 35, 1863–1873. https://doi.org/10.1080/07391102.2016.1197153.
Miller, S., Lesk, A.M., Janin, J., Chothia, C., 1987. The accessible surface area and stability of oligomeric proteins. Nature 328, 834–836. https://doi.org/10.1038/328834a0.
Mori, M., Hwa, T., Martin, O.C., De Martino, A., Marinari, E., 2016. Constrained allocation flux balance analysis. PLoS Comput. Biol. 12, e1004913.
Niebel, B., Leupold, S., Heinemann, M., 2019. An upper limit on Gibbs energy dissipation governs cellular metabolism. Nat. Metab. 1, 125–132. https://doi.org/10.1038/s42255-018-0006-7.
Nilsson, A., Nielsen, J., Palsson, B.O., 2017. Metabolic models of protein allocation call for the kinetome. Cell Syst. 5, 538–541. https://doi.org/10.1016/j.cels.2017.11.013.
Noor, E., Haraldsdóttir, H.S., Milo, R., Fleming, R.M.T., 2013. Consistent estimation of Gibbs energy using component contributions. PLoS Comput. Biol. 9, e1003098. https://doi.org/10.1371/journal.pcbi.1003098.
Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., Milo, R., 2014. Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput. Biol. 10, e1003483.

Noor, E., Flamholz, A., Bar-Even, A., Davidi, D., Milo, R., Liebermeister, W., 2016. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. PLoS Comput. Biol. 12, e1005167 https://doi.org/10.1371/journal.pcbi.1005167.

O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9, 693. https://doi.org/10.1038/msb.2013.52.

O'Brien, E.J., Monk, J.M., Palsson, B.O., 2015. Using genome-scale models to predict biological capabilities. Cell 161, 971–987. https://doi.org/10.1016/j.cell.2015.05.019.

Pandit, A.V., Srinivasan, S., Mahadevan, R., 2017. Redesigning metabolism based on orthogonality principles. Nat. Commun. 8, 15188. https://doi.org/10.1038/ncomms15188.

Paulo, M., Miguel, R., Isabel, R., 2021. In silico constraint-based strain optimization methods: the quest for optimal cell factories. Microbiol. Mol. Biol. Rev. 80, 45–67. https://doi.org/10.1128/MMBR.00014-15.

Pfeiffer, T., Morley, A., 2014. An evolutionary perspective on the Crabtree effect. Front. Mol. Biosci. 1, 17. https://doi.org/10.3389/fmolb.2014.00017.

Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. 2, 886–897. https://doi.org/10.1038/nrmicro1023.

Rizk, Matthew L., Liao, J.C., 2009a. Ensemble modeling for aromatic production in *Escherichia coli*. PLoS One 4, e6903. https://doi.org/10.1371/journal.pone.0006903.

Rizk, Matthew L., Liao, J.C., 2009b. Ensemble modeling and related mathematical modeling of metabolic networks. J. Taiwan Inst. Chem. Eng. 40, 595–601. https://doi.org/10.1016/j.jtice.2009.05.003.

Rizk, M.L., Laguna, R., Smith, K.M., Tabita, F.R., Liao, J.C., 2011. Redox homeostasis phenotypes in RubisCO-deficient Rhodobacter sphaeroides via ensemble modeling. Biotechnol. Prog. 27, 15–22. https://doi.org/10.1002/btpr.506.

Ryu, J.Y., Kim, H.U., Lee, S.Y., 2015. Reconstruction of genome-scale human metabolic models using omics data. Integr. Biol. 7, 859–868. https://doi.org/10.1039/c5ib00002e.

Saa, P.A., Nielsen, L.K., 2017. Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. Biotechnol. Adv. 35, 981–1003. https://doi.org/10.1016/j.biotechadv.2017.09.005.

Sánchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P.-J., Kerkhoven, E.J., Nielsen, J., 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13, 935. https://doi.org/10.15252/msb.20167411.

Schalén, M., Anyaogu, D.C., Hoof, J.B., Workman, M., 2016. Effect of secretory pathway gene overexpression on secretion of a fluorescent reporter protein in Aspergillus nidulans. Fungal Biol. Biotechnol. 3, 3. https://doi.org/10.1186/s40694-016-0021-y.

Schellenberger, J., Lewis, N.E., Palsson, B.Ø., 2011. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys. J. 100, 544–553. https://doi.org/10.1016/j.bpj.2010.12.3707.

Schroeder, W.L., Saha, R., 2019. KOPTIC: A novel approach for in silico prediction of enzyme kinetics and regulation. bioRxiv 807628. https://doi.org/10.1101/807628.

Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., Sauer, U., 2012. Multidimensional optimality of microbial metabolism. Science (80-.) 336. https://doi.org/10.1126/science.1216882, 601 LP – 604.

Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. 99 https://doi.org/10.1073/pnas.232349399, 15112 LP – 15117.

Shen, F., Sun, R., Yao, J., Li, J., Liu, Q., Price, N.D., Liu, C., Wang, Z., 2019. OptRAM: in-silico strain design via integrative regulatory-metabolic network modeling. PLoS Comput. Biol. 15, e1006835.

Simeonidis, E., Price, N.D., 2015. Genome-scale modeling for metabolic engineering. J. Ind. Microbiol. Biotechnol. 42, 327–338. https://doi.org/10.1007/s10295-014-1576-3.

Srinivasan, S., Cluett, W.R., Mahadevan, R., 2015. Constructing kinetic models of metabolism at genome-scales: a review. Biotechnol. J. https://doi.org/10.1002/biot.201400522.

St John, P.C., Bomble, Y.J., 2019. Approaches to computational strain design in the multiomics era. Front. Microbiol. 10, 597. https://doi.org/10.3389/fmicb.2019.00597.

Steuer, R., Gross, T., Selbig, J., Blasius, B., 2006. Structural kinetic modeling of metabolic networks. Proc. Natl. Acad. Sci. 103 https://doi.org/10.1073/pnas.0600013103, 11868 LP – 11873.

Strutz, J., Martin, J., Greene, J., Broadbelt, L., Tyo, K., 2019. Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain. Curr. Opin. Biotechnol. 59, 24–30. https://doi.org/10.1016/j.copbio.2019.02.005.

Tan, Y., Liao, J.C., 2012. Metabolic ensemble modeling for strain engineers. Biotechnol. J. 7, 343–353. https://doi.org/10.1002/biot.201100186.

Tan, Y., Rivera, J.G.L., Contador, C.A., Asenjo, J.A., Liao, J.C., 2011. Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. Metab. Eng. 13, 60–75. https://doi.org/10.1016/j.ymben.2010.11.001.

Tepper, N., Noor, E., Amador-Noguez, D., Haraldsdóttir, H.S., Milo, R., Rabinowitz, J., Liebermeister, W., Shlomi, T., 2013. Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. PLoS One 8, e75370.

Theisen, M.K., Lafontaine Rivera, J.G., Liao, J.C., 2016. Stability of ensemble models predicts productivity of enzymatic systems. PLoS Comput. Biol. 12, e1004800.

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat. Protoc. 5, 93–121. https://doi.org/10.1038/nprot.2009.203.

Tian, M., Kumar, P., Gupta, S.T.P., Reed, J.L., 2017. Metabolic Modeling for Design of Cell Factories. Syst. Biol. (Stevenage). Wiley Online Books. https://doi.org/10.1002/9783527696130.ch3.

Tokuyama, K., Toya, Y., Matsuda, F., Cress, B.F., Koffas, M.A.G., Shimizu, H., 2019. Magnesium starvation improves production of malonyl-CoA-derived metabolites in *Escherichia coli*. Metab. Eng. 52, 215–223. https://doi.org/10.1016/j.ymben.2018.12.002.

Tran, L.M., Rizk, M.L., Liao, J.C., 2008. Ensemble modeling of metabolic networks. Biophys. J. 95, 5606–5617. https://doi.org/10.1529/biophysj.108.135442.

Trinh, C.T., Thompson, R.A., 2012. In: Wang, X., Chen, J., Quinn, P. (Eds.), Elementary mode analysis: a useful metabolic pathway analysis tool for reprograming microbial metabolic pathways BT - reprogramming microbial metabolic pathways. Springer Netherlands, Dordrecht, pp. 21–42. https://doi.org/10.1007/978-94-007-5055-5_2.

Tsiantis, N., Banga, J.R., 2020. Using optimal control to understand complex metabolic pathways. BMC Bioinformatics 21, 472. https://doi.org/10.1186/s12859-020-03808-8.

Varner, J., Ramkrishna, D., 1999. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. Biotechnol. Prog. 15, 407–425. https://doi.org/10.1021/bp990017p.

Wortel, M.T., Noor, E., Ferris, M., Bruggeman, F.J., Liebermeister, W., 2018. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. PLoS Comput. Biol. 14, e1006010 https://doi.org/10.1371/journal.pcbi.1006010.

Zhang, Y., Thiele, I., Weekes, D., Li, Z., Jaroszewski, L., Ginalski, K., Deacon, A.M., Wooley, J., Lesley, S.A., Wilson, I.A., Palsson, B., Osterman, A., Godzik, A., 2009. Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. Science (80-.) 325. https://doi.org/10.1126/science.1174671, 1544 LP – 1549.

Zomorrodi, A.R., Lafontaine Rivera, J.G., Liao, J.C., Maranas, C.D., 2013. Optimization-driven identification of genetic perturbations accelerates the convergence of model parameters in ensemble modeling of metabolic networks. Biotechnol. J. 8, 1090–1104. https://doi.org/10.1002/biot.201200270.