Recent advances in metabolic network modeling using the concepts of stoichiometric balance, thermodynamics feasibility and kinetic law formalisms

Patrick F. Suthers[1,2,3], Charles J. Foster[1,2], Debolina Sarkar[1,2], Lin Wang[1,2], Costas D. Maranas[1,2,3*]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA
[2]Center for Bioenergy Innovation, Oak Ridge, TN, USA
[3]DOE Center for Advanced Bioenergy and Bioproducts Innovation, The Pennsylvania State University, University Park, PA, USA
*Corresponding author, costas@psu.edu

Abstract

Understanding the governing principles behind organisms' metabolism and growth underpins their effective deployment as bioproduction chassis. A central objective of metabolic modeling is predicting how metabolism and growth are affected by both external environmental factors and internal genotypic perturbations. The fundamental concepts of reaction stoichiometry, thermodynamics, and mass action kinetics have emerged as the foundational principles of many modeling frameworks designed to describe how and why organisms allocate resources towards both growth and bioproduction. This review focuses on the latest algorithmic advancements that have integrated these foundational principles into increasingly sophisticated quantitative frameworks.

1. Introduction

Metabolic engineering has long been applied to modify cellular activities in order to improve the production of metabolite or protein products by altering pathway flux distributions and rates through the manipulation of cellular enzymatic, transport and regulatory functions [1]. The field has rapidly progressed, growing from the groundbreaking development of genetic engineering of Pseudomonas species to biodegrade aromatic hydrocarbons [2] into successful commercialized operations such as the recent process to produce 1,4-butanediol in *Escherichia coli* [3]. Metabolic models have emerged as both structured repositories of information and prediction tools to support the objectives of metabolic engineering, by providing quantitative predictions of cell function in response to both biological and environmental changes and tools to direct the redesign of metabolism. The fundamental concepts of reaction stoichiometry, thermodynamics, and kinetics form the foundational principles of such modeling frameworks.

Reaction stoichiometry encapsulates the network connectivity arising from metabolite transports and biochemical conversions that take place in a system. Early efforts by Papoutsakis on acetone-butanol fermentation that provided a theoretical framework for yield analysis [4] and by Watson on computerized models of microbial central metabolism during steady-state growth [5, 6] have since blossomed into genome-scale metabolic (GSM) models and their associated analyses. Only ten years after publication of the first GSM model in 1999 for *Haemophilus influenzae* RD [7], 45 GSM models of 34 organisms had

been constructed [8]. By 2019 that number had grown dramatically to encompass as many as 6,239 organisms across all domains of life with GSM models, of which 183 have manually curated models [9].

Thermodynamic information encoded by reaction free energy of change provides insight on reaction and pathway reversibility thus constraining the feasible phenotypic space for flux balance analysis (FBA) on metabolic models [10]. Significant research has been undertaken to elucidate and apply thermodynamic constraints for metabolic reactions [11]. The standard Gibbs free energy change ($\Delta_r G^o$) is available for approximately 400 reactions in the Thermodynamics of Enzyme-Catalyzed Reactions Database [12]. This covers only a small portion of reactions in metabolic models. Early work by Burton introduced a procedure which infers the unknown apparent equilibrium constants (K') of a reaction through the linear combination of two or more reactions with known $\Delta_r G^o$ using the first law of thermodynamics [13, 14]. Although this method has continued to be applied for expanding tables of thermodynamic parameters [15], the free energy of change for most reactions remains unresolved. To this end, Benson and Buss developed a group contribution method [16] to approximate the free energy of change of biochemical reactions in aqueous solutions [17]. To this day, various versions of the group contribution method remain the most prevalent technique for estimating the $\Delta_r G^o$ [18].

Kinetic models of metabolism introduce mechanistic descriptions of enzyme kinetics into metabolic models and enable the prediction of transient responses to perturbations using a variety of formalisms. One of the first structured single-cell models, by Heinmets, consisted of 19 simultaneous differential equations with 31 rate constants [19] and was later examined via digital computer simulations [20]. Subsequently, more involved single-cell models and analyses emerged, such as the Cornell Single-Cell Model [21] that could quantitatively predict the dependence of *E. coli* growth rate and cell size, shape and composition on external concentrations of glucose. Model sizes and scopes have continued to expand, with the recent publication of a whole-cell model of *E. coli* [22] that draws data from over 1200 publications; among its 19,119 parameters involved in more than 10,000 mathematical equations in 19 modules are 639 kinetic parameters governing the activity of 404 metabolic reactions. Such comprehensive models encompassing various hierarchical levels help accelerate biological discovery and engineering.

Increases in computational power [23] were leveraged to keep up with the flood of genomic [24] and phenotypic data alongside improved algorithms and approaches. A common strategy to unraveling cellular phenotypes has been to study aspects of metabolism independently (*i.e.*, fluxome, metabolome, proteome) from one of several viewpoints (*i.e.*, stoichiometry, thermodynamics, kinetics). However, increases in computation power, improved genomic [25-27] and automated analytical tools [28], and availability of multiple datasets have finally enabled the development of holistic modeling frameworks for describing metabolism. It is now possible to weave a more coherent narrative about the fluxomic, metabolomic, and proteomic phenotypes in greater detail. In this review, therefore, we focus on recent

advancements in the stoichiometric, thermodynamic, and kinetic modeling of metabolism that have fueled discoveries in biology and metabolic engineering. We begin with current developments in the methods of metabolic model reconstruction and curation, followed by integration of models with high-throughput omics data to increase model scope and predictions. We next describe developments in the methods of thermodynamics predictions using both group contribution methods and quantum mechanics and kinetic formalisms along with parameterization techniques. We contextualize these latest modeling frameworks against earlier developments in the field, and discuss the insights they enable in metabolic engineering.

2. Stoichiometric models of metabolism

A defining feature of metabolic models is the organization of metabolites and the reactions in which they are involved into a stoichiometric matrix *S*. These stoichiometric models are typically augmented with additional information, such as gene-protein-reaction (GPR) associations [29] that are Boolean logic statements describing what gene(s) need to be expressed to assemble a protein (or assembly of protein subunits) to support the enzymatic catalysis of a particular reaction. Determining the GPR associations is a cornerstone of GSM model reconstruction; a well-established and detailed manual protocol describes how to reconstruct GSM models using genome annotation data and how to curate the models with experimentally obtained information [30], as indicated in Figure 1(a). Manual GSM reconstruction is both time and labor intensive. In this section, we describe recent tools for automating reconstruction and aiding

curation of these models, incorporating experimental data to improve phenotype predictions, and prospecting metabolic pathway design computationally.

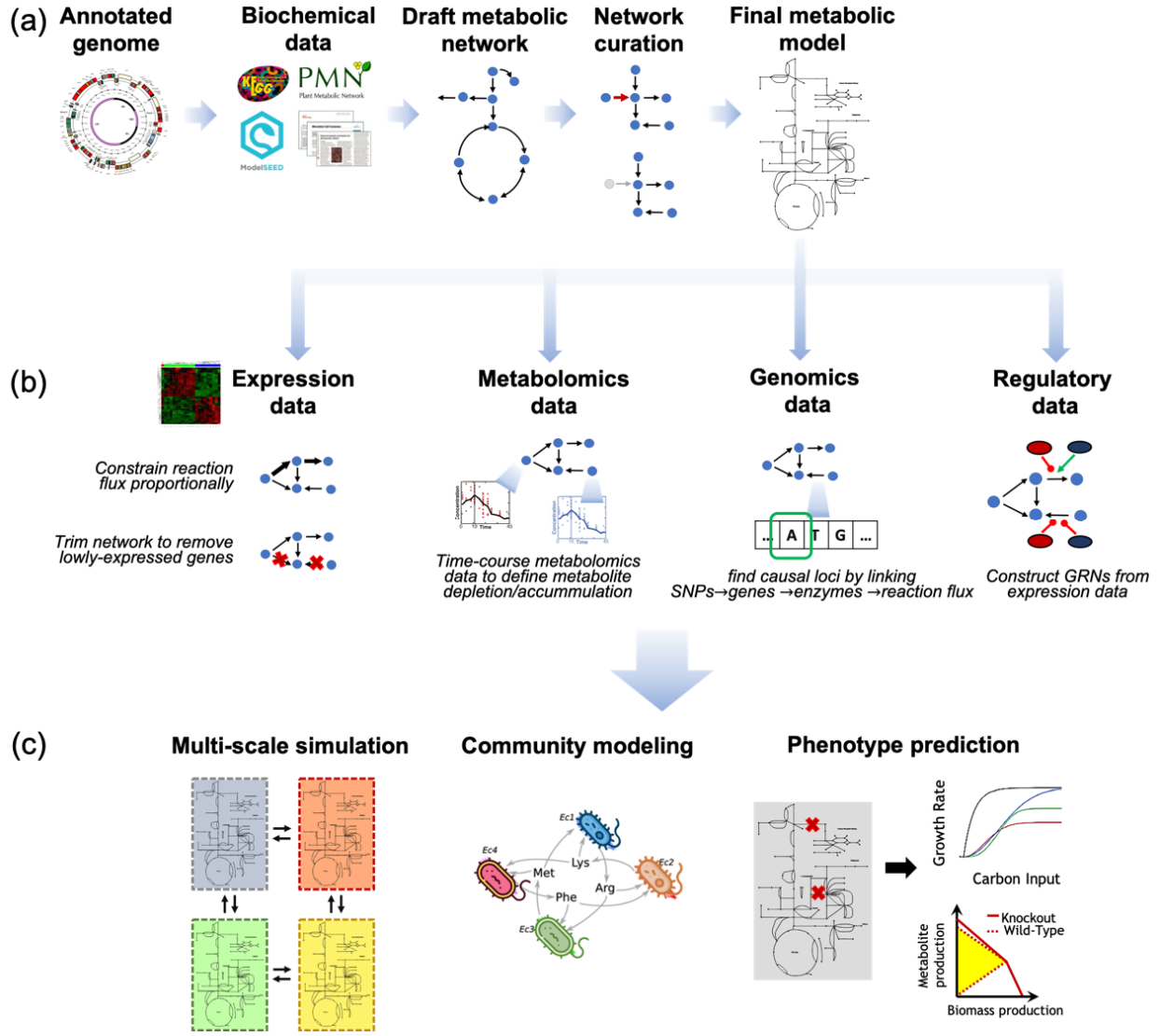

Figure 1. ***Overview of stoichiometric metabolic model construction, augmentation by incorporating large-scale experimental datasets, and applications.*** *(a) General workflow for constructing stoichiometric models of metabolism for microbes and multi-cell eukaryotes. The workflow begins with genome annotation, from which metabolic functions are extracted and cast as biochemical reactions using databases such as KEGG, ModelSEED, and MetaCyc. The resulting draft model is subjected to an iterative curation cycle to produce a final genome-scale metabolic model. (b) Summary of the types of data that can be incorporated into GSM models. Gene/protein expression (transcriptomics/proteomics), metabolite levels (metabolomics), single-nucleotide polymorphisms (genomics), and gene-regulatory networks can be integrated within GSM models to increase their prediction scope and fidelity. (c) Select applications of stoichiometric models of metabolism. GSM models can inform multi-scale simulations, evaluate interactions in a community, and predict overall phenotype by considering genetic and/or environmental perturbations.*

2.1. Tools for reconstructing and curating stoichiometric models of metabolism

A recent review of GSM models and their applications [9] provides a phylogenetic tree of 6,239 organisms for which a GSM model exists. Most of these models were generated by using software tools that automate the process of GSM model reconstruction. Recently created or updated tools include Path2Models [31],

ModelSEED [32], merlin [33], kBase [34], RAVEN 2.0 [35], CarveMe [36], AuReMe [37], AutoKEGGRec [38], PathwayTools [39] and MetaDraft [40]. In general, these tools programmatically perform genome annotation, extract subsets of metabolic genes, catalog the associated enzymes and biochemical conversions, and define GPR associations. Some may also fill network gaps to ensure connectivity. This reconstruction process is iterative, either manually or automatically, until the GSM satisfies pre-determined metabolic criteria, which typically involve biomass constituent synthesis. Because of the plethora of automated tools currently available, it is imperative to be cognizant of their respective advantages and pitfalls.

The first step of genome annotation can be performed in multiple ways and can impact the final reaction content and resulting model fidelity. AutoKEGGRec and RAVEN import metabolic functions from KEGG [41] whereas KBase and ModelSEED use RAST [42]. CarveMe and MetaDraft use the entire BiGG database [43], and the rest allow for internal annotations – merlin uses BLAST [44] or HMMER [45], Pathway Tools uses PathoLogic [46], and AuReMe allows the user to select from a variety of tools such as pantograph [47] or OrthoMCL [48]. A pivotal step where candidate reactions can be added is during gap-filling, where network connectivity is evaluated to ensure that known/expected metabolic functions are successfully captured. Although a number of gap-filling algorithms exist [49-52], only CarveMe and ModelSEED allow the user to define a media composition, and AuReMe and PathwayTools allow setting known metabolic products. The remaining automated tools rely on the initial genome annotation and source database of reactions, and thus may be unable to incorporate spontaneous or exchange reactions necessary to create a functional GSM model. Finally, software availability, licensing, and/or GUI vs command line usage are important considerations when selecting a GSM-reconstruction platform. KBase, ModelSEED, MetaDraft, Merlin, and PathwayTools offer graphical interfaces but only the first three are open-source. The rest offer command-line interfaces. Both AutoKEGGRec and RAVEN are compatible with the COBRA toolbox [53]. The interested reader is directed to Mendoza et al. [54] and Faria et al. [55], for detailed analyses and comparisons of output for automated reconstruction platforms.

As part of the curation process, subjecting the models to recent test suites such as MEMOTE [56] and minimum inconsistency under parsimony (MIP) [57] can help achieve high-quality reconstructions. MEMOTE aims to increase GSM model consistency and reuse by running a series of checks for stoichiometric inconsistencies including unbounded cycles, biomass production under different conditions, and number of blocked reactions. MEMOTE flags any elemental or charge unbalanced reactions, and scrutinizes all model elements for annotations that link to public databases. Although MEMOTE only examines the syntactic, logical and topological components of a GSM model, through use of its reports and scores, MEMOTE allows for comparisons between models creating a benchmark for model quality in this broader context. MIP [57] uses elemental balances for internal metabolic reactions to determine the

molecular weight (MW) of the biomass drain flux in a GSM model. Biomass MW discrepancies from a value of 1 gram dry weight $mmol^{-1}$ create inaccurate estimates of the substrates needed for the predicted growth. This problem becomes particularly important when modeling microbial communities, as the abundance of a microbe with an under-weighted biomass molecular weight would be over-estimated in the community; those over-weighted would be similarly under-estimated.

2.2. Incorporating experimental data increases both model prediction scope and fidelity

Optimization-based analysis frameworks such as FBA are generally used to assess GSM models, and by invoking a pseudo-steady-state assumption for the intracellular metabolites, they compute fluxes which quantify the rates of the corresponding reactions. However, FBA simply using the stoichiometric matrix and GPR associations cannot compute a unique flux distribution. Thus, a number of methods leverage additional biological information to reduce the solution space and thereby improve the precision and accuracy of predictions. Here we present the latest algorithms that seek to incorporate constraints based on experimental data such as gene or protein expression, labeled isotopes, and genetic variability.

2.2.1. Incorporating gene/protein expression

Although mRNA abundance and metabolic fluxes are only moderately correlated [58-60], incorporating transcriptomics into GSM models has been shown to increase their predictive capability [61, 62]. One of the earliest uses of transcriptomics data to inform metabolic models was to tighten reaction bounds using absolute gene expression levels (E-flux [63]). However, this approach results in variable allowable bounds leading to complications in the inference of a specific metabolic response. To lessen this limitation, E-flux2 [64] incorporates a follow-up L2-norm minimization procedure under the assumption that the cell maximizes (or minimizes) the biological objective in an energy- and resource-efficient manner, as an added step beyond simply including constraints based on transcript level. This approach is similar to the parsimonious FBA (pFBA) procedure minimizing the L1-norm, which has been shown to be quite effective at predicting flux distributions [65]. Its contemporary LBFBA [66] also constrains reaction bounds using linear functions of the expression data regressed from a training set with an even higher accuracy than pFBA. However, LBFBA requires both expression (i.e., transcriptomics and/or proteomics) and fluxomics data during training.

An early data-intensive algorithm in this domain is Metabolic Adjustment by Differential Expression (MADE), requiring multiple gene expression datasets as input [67]. MADE uses the statistical significance of change in expression between conditions to define a sequence of best-fitting binary gene states. The weighted sum of inferred gene expression states and p-value of differential expression between conditions is maximized to extract functioning metabolic models. Thus, the final models are such that the differences between successive states most-resemble those seen between mean expression levels. The recent approach Mathematical explOration of ‘Omics data on a MetabolIc Networks (MOOMIN [68])

improves upon the frequentist approach in MADE (*i.e.*, using p-values to identify significant gene expression changes) by inferring reaction weights from a differential expression analysis. A significant change in the expression of a gene carries a positive weight, whereas unchanging genes are assigned a negative weight. The sum of these weights is subsequently maximized to identify the flux distribution associated with a feasible change between growth conditions.

A notable advantage of both LBFBA and MOOMIN is the absence of *a priori* defined biological objective function, thereby broadening their applicability to an increased number of organisms and experimental conditions. This increased scope, however, comes at the cost in some cases of predicting no growth in disagreement with experimental observations [62]. Choice of modeling framework is largely dictated by the available data and assumptions. Both E-flux2 and LBFBA can be parameterized using steady-state data collected from one or more growth conditions, whereas both MADE and MOOMIN require multiple expression datasets.

Metabolic models can also be used to enhance network features gleaned from omics studies using conventional differential expression analysis. Samal et al. [69] used sparse group lasso (SGL) to find pathways associated with a given phenotype by integrating omics data with GSM models. Within this framework, elementary flux modes are first calculated using the PoCaB software [70] and mapped to gene sets across multiple expression datasets. SGL is then used to select a sparse set of genes that is the best predictor of a given phenotype.

Phenotype prediction can be further enhanced by adding Boolean type constraints to GSM models to help incorporate regulation [71] (e.g., under aerobic or anaerobic conditions), as indicated in Figure 1(b). However, reconstructing a gene regulatory network (GRN) from high-throughput data remains challenging, as elucidated by the DREAM project for over 30 network inferences methods on *E. coli*, *Saccharomyces cerevisiae*, and *Staphylococcus aureus* [72]. One of the first efforts to impose GRNs as an additional layer atop GSM models without using stringent Boolean functions was probabilistic regulation of metabolism (PROM) [73]. PROM assigns conditional probabilities to gene states and gene-TF interactions based on expression data for transcription factors (TFs) and target genes. Those conditional probabilities are then are used to scale fluxes through all metabolic reactions based on their corresponding GPR associations. Although PROM can predict organism growth rate with high accuracy (0.95 correlation coefficient), its major drawback is the requirement of a large number of transcriptomics datasets. The Integrated Deduced And Metabolism (IDREAM) formalism [74], improves on PROM by introducing GRNs constructed by EGRIN [75] with PROM to predict metabolic phenotypes across a variety of conditions. EGRIN first identifies conditionally co-regulated genes, and then uses linear regression to estimate the expression of a target gene from the mRNA levels of associated TFs. By integrating conditional co-regulation instead of relying solely on conditional probabilities inferred from gene expression profiles,

IDREAM outperformed PROM over several metabolic networks and environmental conditions. Finally, TRFBA [76] integrates GRNs with GSMs by constraining the expression of a gene to be lower than the sum of the expression levels of its regulating TFs. Reaction flux bounds in the model are constrained proportionally to the expression of genes encoding the catalyzing enzyme, with TRFBA out-performing PROM at growth rate predictions. Thus, high-throughput omics data when used in conjunction with metabolic models can help increase the prediction fidelity, scope, and purview (Figure 1(c)).

2.2.2. Incorporating isotopic labeling data into flux elucidation

^{13}C-metabolic flux analysis (MFA) uses stable isotope tracing from substrates to intracellular metabolites to elucidate intracellular carbon flow in steady-state flux estimation [77]. An atom mapping model is used to describe the carbon transitions from substrates to products for all reactions in a metabolic network. Differences in label incorporation arise from differences in carbon transitions in alternative pathways for the conversion of substrate to metabolic intermediates. MFA is formulated as a nonlinear least-squares regression problem which minimizes the difference between predicted labeling pattern (estimated as a function of metabolic flux solved through a system of algebraic equations assuming metabolic and isotopic steady-states) and experimental labeling distributions. MFA has been further expanded to include isotopically nonstationary data [78]. Several experimental techniques can be used to determine the mass distribution vector (MDV) for each labeled metabolite including NMR [79], mass spectrometry [80], and tandem mass spectrometry [81]. Flux ranges obtained through MFA are generally narrower than those obtained using purely stoichiometric FBA. Amino acid fragments from hydrolyzed proteins are the most commonly measured metabolites, but in the last few years measurement of hydrolyzed glycogen and RNA has been shown to help better-resolve upper glycolytic and pentose phosphate pathway fluxes in both *E. coli* and CHO cells [82]. Alternatives to the confidence interval estimation [83] have also been proposed to estimate flux uncertainty [84]. Numerous open source and commercial software tools exist for assembling atom mapping models and performing MFA. As indicated in Table 1, METRAN [85], WUFlux [86], OpenFlux [87], and INCA [88] are free for educational use and have graphical user interfaces useable within MATLAB. 13CFLUX2 [89], OpenMebius [90], and GS-MFA [91] are also free for educational use, but offer a command line interface within MATLAB. FluxML is an open source program which provides a formalized markup language for organizing MFA stoichiometric and atom mapping models.

Several current methodological advances have helped expand the scope of MFA and further narrow flux ranges, while parallel advances have allowed for network scale-up to the genome-scale. COMPLETE-MFA [92] exploits that different positional labeling patterns of substrates tend to better resolve flux ranges in different portions of metabolism. Thus, elucidating flux ranges from multiple labeling experiments simultaneously, each with different substrate tracer scheme, results in the narrowest possible flux ranges across metabolism [92]. The method was demonstrated using 14 parallel labeling experiments in *E. coli*,

and resolved exchange flux ranges for nine of 22 reactions across central carbon metabolism [93]. Steps have also been taken towards extending the application of MFA to include microbial consortia, as opposed to only monoculture organisms, as it has been demonstrated that model compartmentalization can be used to reliably elucidate fluxes for distinct *E. coli* mutant strains in co-culture [94].

A primary challenge to network scale-up in MFA has been the need for an organism-specific genome-scale atom mapping model. Although atom mapping can be gleaned from existing genome-scale mapping models [91, 95-97] or online repositories such as KEGG [41] and MetaCyc [98], manual curation tends to be time consuming, tedious, and prone to human error. Thus, a class of tools has been developed which identify putative mappings using parsimony [99-104]. These tools have demonstrated over 90% prediction accuracy with regard to mapping, making them useful for rapidly expediting the process of genome-scale atom mapping model construction. At the same time, algorithmic improvements have allowed for reduced computational time, facilitating MFA with large networks. Flux coupling analysis [105] was applied as pre-processing step to reduce the number of confidence intervals requiring evaluation. In the case of *E. coli*, this resulted in a 73% reduction in the number of reactions requiring confidence interval estimation in order to fully elucidate flux ranges at genome-scale [91]. This reduction in the number of calculated confidence intervals is particularly important in the case of non-stationary MFA, where the solution of large systems of ordinary differential equations (ODEs) (i.e., simultaneous evaluation of 8.4 x 10^5 ODEs in the case of *Synechocystis PCC 6803* [95, 106], for example) establishes a significant computational burden to confidence interval estimation.

2.2.3. Incorporating single nucleotide polymorphism data

Integrating genomics marker data (mainly as single-nucleotide polymorphisms or SNPs) into metabolic models is a nascent field (see Figure 1(b)). Some of the earliest examples arose in the context of medical-related studies. In the first, putative oncometabolites were predicted by analyzing loss- or gain-of-function mutations in enzymes from their metabolic pathways [107]. Nine types of cancer were analyzed and a total of 15 compounds and 23 substructures of potential oncometabolites were predicted. In the second, metabolomics data identified the functionality of SNPs in three members of *Mycobacterium tuberculosis* complex (MTBC) [108]. In this study, exometabolome data was used to constrain exchange fluxes and the activity of a SNP was computed by minimizing the flux difference between strains, the reference fluxes, and the effect of each SNP.

A similar mechanistic approach towards deciphering the activity of a SNP using metabolic models is SNPeffect [109]. SNPeffect was used to explain differential growth rates, metabolite accumulation, and phenotypes in *A. thaliana* and *P. trichocarpa* as the result of point mutations in enzyme-coding genes. A parsimonious set of functional SNPs and their final effect on plant metabolism (*i.e.*, increasing/decreasing reaction flux) was thus determined. Marker data has also been used to increase the capability of genomic

selection models. Tong et al. [110] created *A. thaliana* accession-specific models by constraining the biomass flux proportionally to marker data. The genotype-specific reaction flux predictions thus obtained were used to augment genomic selection models, resulting in a ~33% increase in accuracy.

2.3. Tools for computational metabolic pathway design

Computational metabolic pathway design tools aim at identifying energy and carbon efficient pathways and thereby enable the production of a biomolecule either using enzymatic reactions available in nature or through designing *de novo* pathways using reaction rules derived from known reactions. These pathways can then be evaluated in the context of the production host's metabolism by modifying existing or adding new entries in the model's stoichiometric coefficients. Pathway design algorithms were recently reviewed, sorted by reaction network representation and search algorithm [111]. Since that review, notable efforts have been made to improve *de novo* pathway predictions by deriving simplified reaction rules which represent pathways involving complex molecules (e.g. oligosaccharides in glycosylation) [112] and applying mixed-integer linear programming (MILP) [113] or reinforcement learning (RL) [114] to guide the search in the complex uncharted chemical space. To enable broader uptake of pathway design tools, user-friendly web applications have been developed [115]. Another emerging direction for pathway design is unifying biochemical retrosynthesis with chemical total synthesis into a single workflow [116]. Here, enzymatic reactions are used to handle regioselectivity whereas chemical reactions are preferred in the context of avoiding toxic intermediate metabolites. However, computational tools to tackle this challenge remain elusive and many industrial-scale applications are still based on expert knowledge [116].

3. Determining thermodynamic properties of metabolic reactions

Incorporating thermodynamics information such as reaction and pathway reversibility can improve predictions from stoichiometric descriptions of metabolic networks [10] and also provide novel insights for rational design. For example, the Max-min Driving Force (MDF) algorithm has been developed and applied to study the thermodynamic bottleneck of ED/EMP pathways [117], other glycolytic alternatives [118], and genome-scale metabolic networks [119]. The identified thermodynamic bottlenecks highlighted the opportunities to engineer alternative production pathways or swap co-factors for reaching a higher yield of the target product [120]. A recent study revealed the upper limits on Gibbs energy dissipation rate of *E. coli* and *S. cerevisiae* using a network-based approach [121]. The Gibbs energy dissipation rate was shown to govern the rewiring of intracellular metabolic fluxes when the cell increases glucose uptake rate. Applications such as these ones require knowledge of relevant thermodynamic properties which must be measured directly or estimated computationally. At the core of all thermodynamic analysis is the accurate prediction of the change in the Gibbs free energy of reactions $\Delta_{\mathrm{r}}G^{o}$. Thus, in this section we focus on

reviewing recent progress and tools developed within the past decade that improve the way of directly measuring or predicting $\Delta_r G^o$, as summarized pictorially in Figure 2.

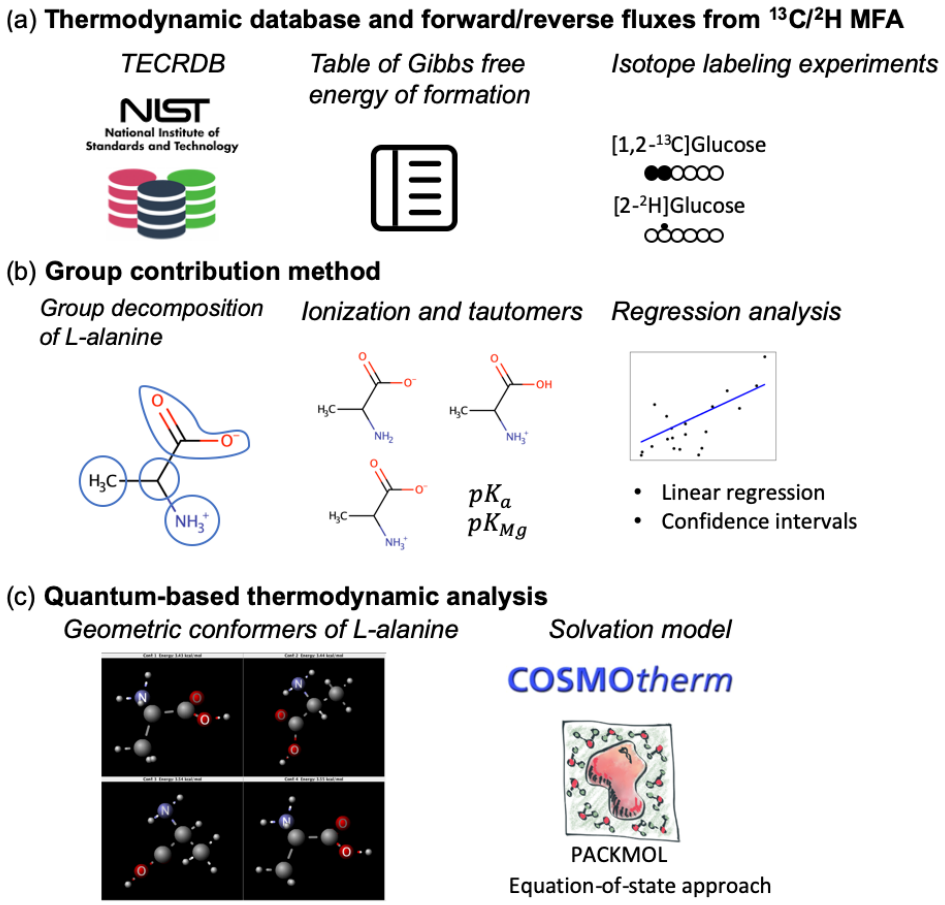

Figure 2. Overview of experimental approaches and computational tools to determine the thermodynamic properties of metabolic reactions. *(a) $\Delta_r G^o$ information can be obtained from NIST TECRDB, thermodynamics tables, and deuterium labeling metabolic flux analysis. (b) Group contribution methods expand the coverage of $\Delta_r G^o$ by decomposing metabolite into groups, capturing the mixture of pseudoisomers, and applying regression analysis to estimate the Gibbs free energy contribution of the functional groups. (c) Quantum thermochemical methods require geometric conformers and solvation models as input to extrapolate the $\Delta_r G^o$ estimation further.*

3.1. Data availability for standard Gibbs free energy change of reactions ($\Delta_r G^o$)

A collection of thermodynamic information of biochemical reaction, the Thermodynamics of Enzyme-catalyzed Reactions Database (TECRDB) was published in 2004 and includes data for approximately 400 reactions [12]. The number of reactions has since then been expanded to approximately 600 reactions (Figure 2(a)). Information on the free energy of formation of metabolites has also been compiled [122] from thermodynamics tables as additional data to expand the current breadth of knowledge. Despite the expanded TECRDB database and availability of free energy of formation for many metabolites, data availability is

still the major limitation for thermodynamics analysis of all enzymatic reactions (e.g. 11,437 reactions exist in the KEGG database alone).

Experimentally elucidated metabolic fluxes with narrow ranges offer another data type for directly characterizing reaction thermodynamics of reactions (Figure 2(a)). $\Delta_r G^o$ can be determined by the forward and reverse flux ratios elucidated from metabolic flux analysis (MFA) and absolute metabolite concentrations ($-RTln\left(\frac{v^+}{v^-}\right) = \Delta_r G^o + RTln(\Pi X_i^{S_i})$. In recent years, deuterium labeling has been incorporated into ^{13}C-MFA (i.e. ^{13}C/^{2}H MFA) to constrain ranges of exchange fluxes so as to reliably estimate $\Delta_r G^o$ of reactions across central carbon metabolism [123]. Notably, Jacobson et al. resolved forward/reverse flux ratios for 22 central carbon metabolic reactions in *Zymomonas mobilis* [123]. This is an improvement over MFA with only ^{13}C-labeled tracers, as previously using COMPLETE-MFA with 14 parallel ^{13}C-labeling experiments Crown et al. [93] were able to resolve only 9 of 22 reversible reactions in *E. coli*. Deuterium labeling in MFA has been further applied to understand the thermodynamic bottlenecks governing ethanol formation in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* [124], and study the $\Delta_r G^o$ of reactions in central carbon metabolism for mammalian cell line, *S. cerevisiae*, *E. coli*, and *Clostridium cellulolyticum* [125]. Because deuterium-incorporated MFA is used to elucidate fluxes for an entire metabolic network, it can be used with genome-scale metabolic network/atom mapping model [106] to provide additional $\Delta_r G^o$ information for reactions beyond central carbon metabolism in the future.

3.2. Thermodynamic estimation of new reactions using group contribution method

In order to expand the prediction of $\Delta_r G^o$ to the reactions that are not included in the experimental measurements, many computational tools have been developed [18]. Most of them rely on the concept of additive schemes of functional groups fitted from experimental data – group contribution method. It remains the most widely used method to extrapolate Gibbs free energy estimations.

3.2.1. Decomposing metabolites into groups

The groups defined in the commonly used group contribution methods for enzymatic reactions are based on a predefined list of chemical substructures (Figure 2(b)). Specifically, all of them rely on the functional groups assembled by Mavrovouniotis [126]. Despite the fact that these groups were observed to have small standard mean square error based on cross-validation analysis, the decomposition method still has many limitations, as recently reviewed by Du et al. [18]. For example, many metabolites are non-decomposable due to incomplete coverage of groups; large variation exists in reactions sharing the same group changes (such as dehydrogenase reactions); the method is insufficient to estimate $\Delta_r G^o$ for reactions with no net change in functional groups (such as isomerase and kinase reactions), despite significant non-zero $\Delta_r G^o$.

Development of group decomposition schemes to more effectively predict the contribution of functional groups to chemical properties is an ongoing effort in the field. Many group decomposition schemes such as Joback, Klincewicz, Lydersen, Sedlbauer-Majer, and UNIFAC [127] have been applied to group contribution methods of chemical reactions. Salmina et al. showed that an extended group decomposition scheme can improve the prediction accuracy of the chemical and biochemical properties (e.g. toxicity and solubility) of a molecule [128]. Also, an automated non-overlapping fragmentation method was recently developed [129] to rapidly evaluate the accuracy of new group decomposition schemes.

In addition to the manually derived non-overlapping group fragmentation schemes, many automated overlapping fragmentation methods have been developed, including molecular signatures [130], ISIDA fragment descriptors [131], ECFP fingerprints [132], and circular fingerprints [133]. These representations of molecules include the group interaction factors (i.e. the overlapped atoms/bonds between groups) within the group decomposition scheme, which are ignored in the component contribution method [134]. Overlapping fragmentation methods have been applied extensively in machine learning algorithms to extract features from chemical molecules [135]. Recently, Alazmi et al. [136] showed that molecular fingerprints can be used to improve $\Delta_r G^o$prediction accuracy.

3.2.2. Ionization and tautomer

The exact protonation state of a metabolite inside of a cell is typically a mixture of pseudoisomers with different protonation states (Figure 2(b)). The presence of pseudoisomers further confounds group contribution methods, as pseudoisomers tend to exhibit often quite different thermodynamic properties [137]. Group contribution methods have thus been modified by capturing the intracellular mixture of pseudoisomers using a Boltzmann distribution [138]. The difference between the standard Gibbs free energy of formation of a compound (ΔG_f^o) and the Gibbs free energy of formation of the mixture ($\Delta G_f'^o$) can be calculated using Inverse Legendre Transform as a function of acid-base dissociation constant pK_a [134].

Since pK_a is a key parameter to study pharmacokinetics of drug molecules, several datasets are publicly available to benchmark pK_a prediction accuracy from computational tools [139]. A number of studies evaluated the relative accuracy among these computational tools including ChemAxon Marvin [140], Epik [141], ACD/Labs Percepta Classic [142], and Moka [143]. Marvin has demonstrated the highest prediction accuracy. Thus, group contribution methods still rely on Marvin to predict pK_a. However, Du et al. [144] compared Marvin predictions with values available in IUPAC SC-database and found deviations resulting from charge inconsistencies and errors in calculating pK_avalues for nitrogenous moieties that were significant enough to provide inaccurate protonation states. Moving forward, a number of open source tools [145] are available that use machine learning algorithms such as neural networks and

random forests with the goal of improving the pK_a prediction accuracy. Experimental methods such as UV–vis spectroscopy, NMR spectroscopy, and X-ray crystal structure of the tautomer in protein-ligand complexes have been applied to measure the pK_a directly.

Similar to the effect of different protonation states, metal-ion binding (e.g. Mg+ binding with ATP) and tautomers (e.g. keto-enol tautomerism) also affect a metabolite's standard Gibbs free energy of formation. Recently, Du et al. [144] applied a machine learning method to predict the magnesium-ion binding constants (pK_{Mg}) to correct reaction-equilibrium constants (K_{eq}). However, tautomers have yet to be considered in the group contribution method and can thus serve as a potential research avenue since currently, only the most stable (or dominant) form of any metabolite is considered during the Gibbs free energy calculation by Jankowski et al. [146]. In principle, the distribution of tautomers can similarly be estimated based on the acid-base dissociation constant (pK_a).

3.2.3. Regression analysis and confidence intervals

Because group contribution assumes an additive schema of functional groups, linear regression is an often-applied method to estimate the Gibbs free energy contribution of groups ΔG_g^o (Figure 2(b)). The confidence interval of fitted ΔG_g^o and predicted $\Delta_r G^o$ can be calculated using the covariance matrix from model fitting. Other regression methods such as Ridge and Lasso regression have been applied by Du et al. [144] demonstrating the smallest prediction error. However, confidence interval analysis from biased regression models such as Ridge and Lasso relies on bootstrap-based methods and often produce inaccurate uncertainty estimations. Credible intervals from Bayesian inference provide an alternative metric to confidence intervals, and has been shown to provide reliable uncertainty estimation [84].

3.3 Thermodynamic estimation of new reactions using quantum mechanics

Even though group contribution methods extrapolate the Gibbs free energy estimation for more reactions, the coverage is still limited by the number of groups present in the experimental data. For example, out of the 163 groups defined in component contribution, only 92 can be estimated from linear regression in the component contribution method [134]. Quantum thermochemical methods have been proposed recently to address limitations in functional group coverage.

One quantum-based approach relies on directly modeling the metabolite–water complex. Jinch et al. [147] calculated the absolute Gibbs energy of each metabolite–water complex using the ORCA quantum chemical software [148]. Adding to the complexity of multiple protonation states addressed in group contribution method, quantum mechanics required two more inputs (Figure. 2(c)): (1) multiple geometric conformers which exist for a protonation state, and (2) solvation effect from the hydrogen bonding between metabolites and solution. Jinch et al. [147] applied the ChemAxon conformation tool [140] to generate geometric conformers that approximate the minimal energy conformations of each metabolite while applying both explicit water model (PACKMOL [149]) and implicit water model (COSMO [150]) to

account for solvation effects. Prediction errors of the reactions catalyzed by isomerases, transferases, and lyases were comparable to those obtained by group contribution method. This method was applied to compare the reduction potential of redox half-reactions without cofactors, leading to the hypothesis that NAD/NADP is the primary electron carrier because its physiological reduction potential range (i.e., -500 mV to -130 mV) matches the requirement of reversible redox reactions and irreversible redox transformations in central carbon metabolism [151].

Another quantum-based approach is to model the gas-phase thermodynamics and solvation in two independent steps. First, Hadadi et al. [152] demonstrated the calculation of the gas-phase thermodynamics of a few metabolites using the TURBOMOLE suite [153]. COSMOconfX [154] and Spartan'14 Parallel Suite [155] were used to search for geometric conformers. Next, Panayiotou et al. [156] developed a solvation model using the equation-of-state approach to account for the changes of Gibbs free energy from gas to aqueous solutions. The two-step quantum-based thermodynamic approach provided a way to calculate the Gibbs free energy of reactions at a broad range of conditions. The partial solvation parameter approach applied by Panayiotou et al. [156] can be used to estimate enthalpy and entropy of hydrations as well as the mechanism of hydration, which can be applied to model solvation at different temperatures and pressures.

Due to the significant computational cost of quantum mechanics, machine learning methods have started to be integrated with quantum mechanics to enable the high-throughput prediction. For example, in another study by Jinich et al. [157] demonstrated that Gaussian process models can be used to efficiently estimate the redox potential of 315,000 reactions. Similarly, Du et al. [144] built a machine learning model to correlate standard entropy change $\Delta_r S^o$ with the molecular properties of substrates/products of a reaction (i.e. the number of atoms, partial charge of atoms, and other molecular descriptors of the metabolites) using Lasso regression. Gibbs free energy at a different temperature is then approximated as a function of $\Delta_r S^o$ and the temperature.

4. Incorporating Kinetic Descriptions of Metabolism

Kinetic descriptions of metabolism use kinetic rate expressions to link reaction flux to the abundance of both intracellular metabolites and enzymes in order to predict metabolic response to genetic and environmental perturbation. Here we discuss two distinct sub-classes of models that leverage reaction kinetics to predict different aspects of the metabolic phenotype. The first sub-class (herein referred to as mechanistic models) captures allosteric regulations detailed within rate expressions. [158]. Networks described by mechanistic models are generally smaller because of challenges related to the paucity of available kinetic parameter data and difficulties in parameterization [159]. The second sub-class is resource allocation models, which have been developed for the prediction of genome-wide protein abundance. An extension of stoichiometric models of metabolism, these models incorporate simplified kinetic descriptions

into metabolic constraints by including rate expressions for both enzyme catalyzed reactions and macromolecular synthesis. However, they currently do not capture detailed substrate-level regulatory mechanisms [160]. Table 1 compares the network and data requirements for parameterization, forms of kinetic descriptions, and types of information predicted by mechanistic kinetic model and resource allocation models. In this section, we discuss recent tools and progress that has been made for these two classes towards the realization of comprehensive metabolic phenotype prediction through kinetics.

Table 1: Similarities and differences in network and data requirements for the construction of mechanistic kinetic models of metabolism and resource allocation models as well as commonly used kinetic descriptions and a summary of their predictive capabilities

Model Requirements	Mechanistic kinetic models	Shared Attributes	Resource allocation models
Network requirements	Reduced stoichiometric matrix Allosteric regulatory network		Genome-scale reconstruction Transcriptional regulatory network
Data used in parameterization	Metabolomics	Fluxomics	Proteomics Genome/protein sequence Protein structure Growth rate
Kinetic descriptions	Lin-log Log-lin Elementary decomposition Michaelis-Menten Hill kinetics Convenience kinetics Generalized mass action		First order approximation
Predictive capabilities	Metabolite pool	Metabolic flux	Protein abundance

4.1. Mechanistic kinetic models of metabolism

Mechanistic kinetic models of metabolism offer the promise of enhancing predictions for multiple aspects of metabolic phenotype (i.e., metabolome, fluxome, proteome) and thus have the potential to accelerate the design-build-test-learn cycles for metabolic engineering [161]. Their inclusion of mechanistic rate expressions via established rate law formalisms elevates the predictive capabilities of mechanistic kinetic models beyond those of pure stoichiometric models and facilitates quantitatively relating reaction flux to enzyme level and metabolite concentration [162]. However, improved product yield, titer, and production rate predictions are not without cost. Significant increases in experimental data requirements compared to stoichiometric models, challenges associated with model assembly and parameterization, and extensive difficulties related to follow-up analysis are all present hurdles to kinetic model development [158]. In this section, we describe the kinetic formalisms and associated data requirements available for use in kinetic models, the algorithm choices available for identifying kinetic parameters, improvements that have been made to allow for confidence interval estimation and subsequent follow-up analyses, and framework extensions that enable kinetic modeling to discover substrate-level regulations active in metabolism.

4.1.1. Selection of kinetic formalism and data requirements

A number of mechanistic or approximate kinetic formalisms exist which can be used to construct kinetic models of metabolism, each with advantages and disadvantages in terms of ease of parameterization, specific data requirements, and predictive capability. In the absence of experimental data of kinetic rate constants, identification of *in-vivo* kinetic parameters generally requires the solution of a non-linear programming problem (NLP), which fits model predictions to temporal metabolomics and/or fluxomics data across a range of genetic/environmental conditions [163]. The formalism chosen for model construction is important because it influences the types of data that are required for parameterization, the predictive capability of the model, and the parameterization methods available for use. Interested readers are referred to Saa et al. [159] for detailed descriptions of the kinetic formalisms described in brief below.

Rate expressions linearized around a reference point are convenient because they bypass the need to solve a non-convex optimization problem associated with kinetic parameterization [164]. Examples of this approach include the loglin [165] and linlog [166] formalisms, and generalized mass action kinetics (GMA) [167] which provide alternative linear approximations. Kinetic models consisting of linearized rate expressions require steady-state fluxomics and metabolomics data for parameterization [168, 169], Mass Action Stoichiometric Simulation (MASS) parameterization has been developed to parameterize GMA models from a single set of metabolomics data [164, 170]. Its capabilities to predict adequately responses to metabolite level perturbation was demonstrating using a red blood cell model [171]. A drawback to these formalisms is predictive capability tends to be localized around the reference state, making them ill-suited to predict metabolic response to enzyme level perturbation [159].

In contrast, kinetic formalisms such as Michaelis-Menten (MM) and elementary decomposition kinetics preserve the fundamental mechanism of enzymes without approximation. Kinetic models using the MM rate expressions benefit from kinetic information consolidated within databases such as BRENDA [172] and KiMoSys [173], and do not suffer from the localized predictive limitations associated with linearized rate expression. However, large numbers of kinetic parameters are unavailable in databases and still require identification, introducing the challenge of non-convex optimization using a combination of temporal metabolomics and/or steady-state fluxomics as training data. Finally, elementary decomposition kinetics offers a full mechanistic kinetic description which accounts for substrate binding, catalytic event, product release, and substrate level inhibitions and activations [174]. Elementary decomposition kinetic models require fluxomic data across a range of genetic or environmental conditions for parameterization.

4.1.2. Algorithm choices for identifying kinetic parameters

As described above, identification of *in-vivo* kinetic parameters requires the solution of a nonlinear programming problem (NLP). Overall the use of local optimization has been limited because of the non-convexity of the solution space when non-linear rate expressions are used [175]. Several algorithms using metaheuristic methods have therefore been developed and used for parameterization of kinetic models of metabolism. Scatter search [176], particle swarm [177], and genetic algorithm [178] optimization are evolutionary algorithms relying on recombination of parameters from favorable generational models to assemble a model that best recapitulates training data as evaluated by an objective function (e.g., residual sum of squares). Current implementations of particle swarm optimization [179] and genetic algorithms [180, 181] have been used to parameterize highly detailed MM kinetic models of core metabolism and demonstrate improved predictive capabilities over their predecessors [182, 183]. Scatter search optimization [184-186] and genetic algorithm optimization [187, 188] have recently been introduced within the EM paradigm. Notably, merging the EM method with metaheuristic optimization has resulted in successful parameterization of the largest mechanistic kinetic model to date (i.e., k-ecoli457, containing 457 reactions, 337 metabolites, 295 allosteric regulations, and 5,239 elementary kinetic parameters), but parameterization time for that model exceeded 1,000 hours [187].

A preprocessing step has been introduced into EM to reduce the computational expense. Greene et al. applied local stability analysis within EM to preclude unstable models from the final ensemble and evaluate training data in a logical order based on similarity to the wild-type strain [189]. This strategy resulted in a 71% speed-up in *E. coli* core model parameterization, but the method does not scale beyond core metabolism [189].

Bayesian approaches have been adopted as a means to parameterize kinetic models as well. The Approximate Bayesian Computation and General Reaction and Assembly Platform (ABC-GRASP) uses elementary decomposition kinetic expressions [190, 191], and employs Approximate Bayesian

Computation to bypass the need for explicit evaluation of a likelihood function. Although its application to construct a small kinetic model of the mammalian methionine cycle demonstrated tractability for small networks, scalability limitations of Monte-Carlo-based sampling currently prevent its use for large-scale networks [159]. Bayesian principles have also been applied to large-scale networks using linearized kinetic rate expressions with the lin-log formalism [169]. Because of a reduced computational burden associated with evaluating linearized kinetic rate expressions, the application of Bayesian inference by St. John et al. [169] was demonstrated to be scalable to large networks and useful for follow-up metabolic control analysis (MCA) [192].

One way to circumvent the challenges associated with the parameterization of large-scale kinetic models is to devise customized decomposition approaches tailored to the kinetic formalism adopted. Small scale networks have been parameterized using computationally expensive forward sensitivity analysis as a means for gradient update and optimality assessment [193], but so far only K-FIT has demonstrated network scalability to near-genome scale models [194]. The decomposition approach K-FIT circumvents computational inefficiencies associated with metaheuristic methods [187] and scalability issues of Bayesian approaches [159]. The method relies on anchoring elementary kinetic parameters to a reference state flux distribution, ensuring model feasibility as the algorithm traverses the parameter space. At each iteration, steady-state fluxes and concentrations for the perturbed networks are inferred mostly by iterating between enzyme and metabolite balances forming respectively linear systems of equations. Analytical gradient evaluation serves two additional purposes: first, it permits optimality assessment and it allows for follow-up sensitivity analysis on kinetic parameters.

4.1.3. Improvements for confidence interval analysis of kinetic parameters

Confidence interval estimation allows for follow-up analysis to identify rational design targets. Through confidence interval analysis, identifying both regulatory mechanisms which significantly limit flux and perturbation candidates for targeted overproduction strategies is possible [194]. Multivariate statistics has been adopted as a means for determining confidence intervals in ensembles of kinetic models within the EM paradigm. Hameri et al. [195] demonstrated that bootstrapping, Bonferroni, and exact normal methods used capture parameter interdependencies, and thus offer a more accurate confidence interval estimation compared to univariate confidence interval estimation. Within K-FIT, local sensitivity analysis according to the variance/covariance matrix was adopted to quantify parameter uncertainty [196].

Machine learning has also been introduced to reduce parameter uncertainty in sampling-based methods. Within the ORACLE framework [197], a decision tree algorithm was introduced to reduce kinetic parameter uncertainty ranges (iSHRUNK) [198]. Whereas most parameters remain unresolved upon application of iSHRUNK to kinetic models of *Pseudomonas putida* [199] and *E. coli* [200], reduced parameter uncertainty for a handful of reactions led to predictions for strain robustness and 1,4-butanediol

overproduction, respectively. Additionally, Inverse Metabolic Control Analysis (IMCA) has introduced matrix inversion operations to quantifying the sensitivity of enzyme level to changes in metabolite concentration, expanding the types of sensitivity analysis possible with kinetic models [201].

4.1.4. Elucidating allosteric mechanisms

For many organisms, the substrate-level regulatory network is not sufficiently described to allow for a complete kinetic description of metabolism. This difference is highlighted by the comparison of the regulatory network of the recent core kinetic model of *C. thermocellum* (i.e. k-ctherm118) [202] and the recent core kinetic model of *E. coli* metabolism (i.e. k-ecoli74) [196]. k-ctherm118 contained 22 substrate-level regulations identified in BRENDA [172] and from literature from across all *clostridia*, whereas the smaller *E. coli* network contained more (i.e., 54) *E. coli*-specific inhibitions identified in databases. Two primary methods for identifying allosteric regulation have recently emerged to address this information disparity. Hackett et al. developed systematic identification of meaningful metabolic enzyme regulation (SIMMER) as a means to hypothesize mechanisms assessing the statistical significance of fitness improvement in models fitted with regulation versus those without [203]. Although data intensive (requiring fluxomic, metabolomic and proteomics datasets across a range of growth conditions to fit rate expressions), the value of the method was demonstrated in *S. cerevisiae* metabolism [203]. In that study, three new substrate-level regulations were identified and confirmed experimentally, including alanine inhibition of ornithine transcarbamylase, phenylpyruvate inhibition pyruvate decarboxylase, and citrate inhibition of pyruvate kinase. An ensemble-based approach [204] has also been employed to elucidate regulatory mechanisms, assessing the improvement in fitness of an ensemble of models to training data upon inclusion of a substrate-level regulation. The efficacy of the method was demonstrated by Christodoulou et al., who revealed feedback inhibition of glucose-6-phosphate-1-dehydrogenase by NADPH as the primary mechanism controlling *E. coli*'s pentose phosphate pathway reserve flux response to oxidative stress [204].

4.2. Resource allocation models

Resource allocation models account for both metabolites and macromolecules and leverage approximate kinetic information for events which occur at multiple time-scales to make inferences about enzyme limitations in metabolism. Simplified kinetic descriptions allow for the estimation of kinetic parameters at the genome-scale, which are used as metabolic constraints that allow for prediction of flux, growth, and enzyme abundance. Generally, a single turnover number is estimated for each reaction in the network [160]. Two distinct levels of description have been used recently to understand resource allocation in living cells: phenomenological models that describe a snapshot of metabolism and do not account for expression machinery in their formulation, and multi-scale models that describe events spanning multiple time-scales including metabolism and macromolecular expression [160].

4.2.1. Phenomenological Methodologies

In recent years MetabOlic Modeling with ENzyme KineTics (MOMENT) [205] was introduced as an improvement over flux balance analysis with molecular crowding (FBAwMC) [206]. FBAwMC lumps information on enzyme volume and the effective catalytic rate into a single parameter, which is used in constraints to place an upper bound on reaction flux [207]. A volumetric capacity constraint places an upper bound on the total volume enzymes in a cell can occupy [207]. MOMENT simplifies the FBAwMC framework by recasting kinetic constraints in terms of only first order kinetics instead of volumetric capacity. By removing crowding information from constraints and using a single turnover rate per reaction in a metabolic network, MOMENT demonstrated a statistically significant improvement in model predicted flux and gene expression over FBAwMC in *E. coli* under exponential growth conditions [205]. The GECKO method has also been recently introduced to incorporate catalytic rate constants into enzyme usage constraints to predict an enzyme level/metabolic flux pair [208]. Rather than fitting parameters to data, GEKCO gleans all catalytic rate constants from databases [208]. The method has been shown to reduce flux ranges upon FVA by 60% compared to traditional stoichiometric FVA in *S. cerevisiae* [208].

4.2.2. Multi-scale methodologies

The expansion of resource allocation models to link the composition of the entire cellular proteome and macromolecular machinery to metabolism using kinetics, genome sequence, and biomass composition via stoichiometric constraints has enabled a more complete picture of cell behavior [160]. Two similar multi-scale methodologies have been developed in parallel: Resource Balance Analysis (RBA) [209], and multi-scale models of metabolism and macromolecular expression (ME) [210]. Through the incorporation of simplified kinetics and detailed accounting of macromolecular machinery into a constraint-based optimization problem, these methodologies have been shown to predict proteome allocation for a number of bacteria [211-214]. Dynamic ME model prediction has also been interwoven with metaheuristic parameterization, allowing for the prediction of transient proteome allocation in response to system perturbations [215]. ME models application has been further expanded towards rational strain design by being used together with the OptKnock strain design tool [216] to more accurately predict proteome and flux phenotype associated with rational design strategies [217].

The primary challenges in multi-scale model construction have been ill-conditioning of the stoichiometric matrix due to the many orders of magnitude differences in the stoichiometric coefficients of various metabolites and macromolecules. Furthermore, of *in-vivo* turnover rates are difficult to establish in the absence of experimental information. The first challenge was addressed through the use of quad-precision LP solvers [218-220]. Initially the second was addressed by identifying a single universal turnover number that could be applied to all reactions in the network to reasonably predict growth rate across a range of experimental conditions [211, 212]. However, more recently fitting parameters using steady-state flux

and metabolomics datasets [221] and use of machine learning with enzyme properties a features and turnover rate as response variables [222] have each yielded ME models with improved predictive capabilities in *E. coli*.

Notable applications of multi-scale models that extend beyond simply comparing model-predicted resource allocation with experimental proteomics include hypothesis generation of transcriptional response to iron limitation [223] and acid stress [224] in *E. coli*. Additionally, the FoldME model of *E. coli* introduced proteostasis mechanisms into protein folding constraints that enabled characterization of protein thermostability within the ME framework and captured the response of both protein expression and growth rate to temperature [225]. The success of resource allocation models in predicting non-intuitive biology and aiding in rational strain design has prompted the development of standardized toolboxes for both RBA [226] and ME [227] model construction and evaluation.

5. Discussion

In this review, we highlighted a variety of recent stoichiometric, thermodynamic, and kinetic methodologies that aim to predict of how metabolism and growth are affected by both external environmental factors and internal genotypic perturbations. . Table 2 provides a summary of the approaches and tools discussed in this review and includes the types of tasks they perform, licensing availability, and data requirements. Progress towards integrating known mechanisms onto modeling frameworks to predict new phenotypes and developing new frameworks to pinpoint undiscovered mechanisms are continuing to advance. Frameworks such as whole-cell modeling and multi-scale resource allocation models integrating multi-omics data, even though highly data intensive, are quite promising. . These frameworks are supported by developments in experimental characterization, parameterization approaches, and computational efficiency. Advances in this area to incorporate more complex kinetic frameworks and mechanisms could have wide-reaching impact.

The recent inclusion of structural metabolite and enzyme information in metabolic models provides a new layer of information that can be leveraged in metabolic engineering. Recon3D is the first example of a stoichiometric model which includes structural information for proteins and metabolites as model features [228]. This allows for modelers to understand and predict not only the effect of a gene up/down-regulation, but also contextualize the effect of a point mutation on the overall metabolic outcome. The methods of leveraging structural data are not limited to stoichiometric modeling, as Heckmann et al. have integrated structural information with machine learning to parameterization ME models with reaction-specific kinetic parameters [222]. Moving forward, the vast amounts of structural information contained within the Protein Data Bank (as of 2020 approximately 167,500 protein structures) [229] could be used to enhance metabolic models by identifying allosteric mechanisms and improving predictions for data-poor

non-model organisms. When considered with two recent studies that create whole-cell models that draw from data from thousands of publications – *Mycoplasma genitalium* [230] and *E. coli* [22] – we anticipate the trends of incorporating both omics and structural data to continue.

The significant contributions we discussed have been facilitated by algorithmic advances in mathematical optimization solvers. Solvers such as IBM ILOG CPLEX Optimization Studio [231], Gurobi Optimizer [232], SCIP [233] and other solvers are continually being improved. More specialized solvers such as SoPlex [234] and the quadruple-precision Fortran 77 optimization solver by Ma and Saunders [218] (implemented in the COBRA toolbox) are also instrumental in many of the calculations embedded in the developed tools.. SoPlex that performs iterative refinement of the solution has played an important role in enabling the resource allocation model evaluation, allowing Reimers et al. [212] to solve ill-conditioned LPs to describe diurnal cyanobacterial growth. Kinetic parameterization algorithms such as K-FIT are also potentially transformative by accelerating the reconstruction workflow. A 1,000-fold decrease in parameterization time [194] (as compared to the metaheuristic parameterization scheme used by Khodayari and Maranas [187]) was achieved for a near genome-scale kinetic model. Ultimately, the development of reliable and accelerated solution methods is of great importance and provides opportunities for collaborative work among mathematicians, computer scientists, systems biologists, and metabolic engineers.

Lastly, advances are needed in data sharing, accuracy, and annotations in databases and standardization in description of metabolic models. These considerations are critical to sustain advances in metabolic modeling. Even though the usual practice is to publish results in agreement with experiment, divergent observations can be very valuable for the discovery of missing or poorly described biology.. With the flood of data being generated by high-throughput methods and policies on data sharing, tools need to be developed to aid ease of use, quality checking and discoverability of available information. A large amount of data entered into databases unfortunately still contains omissions and errors [235] thus necessitating automated checking and gate-keeping procedures.. Finally, standardization remains a major challenge for the effective use and sharing of metabolic models. Efforts such as MIRIAM guidelines [236], tools such as MEMOTE [56], languages such as SBML [237] and BioPAX [238], and resolution services such as Identifiers.org [239] have helped significantly but increased effort and adherences to standardization is encouraged.

Table 2: Recent approaches and tools that enable metabolic modeling

Algorithm / Method	Task	License	Accessibility	Website	Data Requirement
MEMOTE	Model testing and validation	Free	openCobra	https://github.com/opencobra/memote	genome-scale model or reconstruction, experimetnal growth and genetic perturbation data

MIP	Model testing and validation	Free	MATLAB, Python	**N/A**	genome-scale model, metabolite formulas, metabolite charges
BOFdat	Generate biomass objective	Free	Python	https://github.com/jclachance/BOFdat	genome-scale model, biomass macromolecular composition
E-flux2	flux balance analysis	Free	MATLAB, java	N/A	genome-scale model, transcriptomics, fluxomics
pFBA	flux balance analysis	Free	MATLAB, Python	https://opencobra.github.io/cobratoolbox/stable/	genome-scale model
LBFBA	flux balance analysis	Free	GAMS + CPLEX/gurobi	https://academic.oup.com/bioinformatics/article/34/22/3882/5033386#supplementary-data	genome-scale model, transcriptomics
MADE	FBA to extract subnetworks	Free	MATLAB + MILP solver	https://academic.oup.com/bioinformatics/article/27/4/541/198624	genome-scale model, transcriptomics
Samal et al SGL	EFM + SGL to extract subnetworks	Free	R	http://www.abi.bit.uni-bonn.de/index.php?id=17	genome-scale model, transcriptomics
PoCaB	Compute extreme currents (ECs)	Free	Database, free to download	http://pocab.cg.cs.uni-bonn.de/gallery.html	genome-scale model
MOOMIN	Identifies 'feasible phenotypes' by analysing differential expression data	Free	MATLAB/Julia + MILP solver	github.com/htpusa/moomin	genome-scale model, transcriptomics
PROM	Constructs GRNs using omics data	Free	MATLAB	https://www.igb.illinois.edu/labs/price/downloads/	genome-scale model, transcriptomics
IDREAM	Uses GRNS to predict metabolic phenotypes	Free	Cobra toolbox	N/A	genome-scale model, transcriptomics
EGRIN	Constructs GRNs using omics data	Free	**N/A**	http://egrin2.systemsbiology.net/index/	Transcriptomics
trFBA	Integrates GRNs with GSMs	Free	MATLAB + Cobra toolbox	http://sbme.modares.ac.ir/trfba-2/	genome-scale model, transcriptomics
MASS	Dynamic modeling	Free	**N/A**	**N/A**	genome-scale model, metabolomics
iReMet-flux	Predicts phenotypes by intergating metabolomics	Code from authors on request		https://pubmed.ncbi.nlm.nih.gov/27587698/	genome-scale model, metabolomics
MetDFBA	Dynamic modeling	**N/A**	**N/A**	**N/A**	genome-scale model, metabolomics
TREM-Flux	Predicts phenotypes by intergating metabolomics and transcriptomics	Free	MATLAB + solver	**N/A**	genome-scale model, metabolomics, transcriptomics
uFBA	Predicts phenotypes by intergating metabolomics	Free	INCA https://pubmed.ncbi.nlm.nih.gov/24413674/	**N/A**	genome-scale model, metabolomics
Nam et al.	Predicts phenotypes by intergating genomics	Free	MATLAB	**N/A**	genome-scale model, transcriptomics
Oyas et al.	Predicts phenotypes by intergating metabolomics and genomics	Free	**N/A**	**N/A**	genome-scale model, metabolomics, genomics
SNPeffect	Predicts phenotypes by intergating metabolomics, transcriptomics, and genomics	Free	GAMS/Python + CPLEX/gurobi	**N/A**	genome-scale model, metabolomics, transcriptomics, genomics

Tong et al.	Genomic selection by integrating genomics	Free	MATLAB+R	https://github.com/Hao-Tong/netGS	genome-scale model, metabolomics, genomics
MBA	FBA to extract subnetworks	Free	Cobra toolbox	https://opencobra.github.io/cobratoolbox/stable/modules/dataIntegration/transcriptomics/MBA/index.html	genome-scale model, transcriptomics
mCADRE	FBA to extract subnetworks	Free	Cobra toolbox	https://opencobra.github.io/cobratoolbox/latest/modules/dataIntegration/transcriptomics/index.html	genome-scale model, transcriptomics
FASTCORE	FBA to extract subnetworks	Free	Cobra toolbox	https://opencobra.github.io/cobratoolbox/latest/modules/dataIntegration/transcriptomics/index.html	genome-scale model, transcriptomics
FASTCORMICS	FBA to extract subnetworks	Free	MATLAB	https://wwwen.uni.lu/research/fstm/dlsm/research_areas/systems_biology/software/fastcormics	genome-scale model, transcriptomics
GIMME	FBA to extract subnetworks	Free	Cobra toolbox	https://opencobra.github.io/cobratoolbox/latest/modules/dataIntegration/transcriptomics/index.html	genome-scale model, transcriptomics
GIM^{3}E	FBA to extract subnetworks	Free	CobraPy	http://opencobra.sourceforge.net/	genome-scale model, metabolomics, transcriptomics
Richelle et al.	Data-driven metabolic tasks definition	Free	Cobra toolbox	https://github.com/opencobra/cobratoolbox	genome-scale model, transcriptomics
iMAT	FBA to extract subnetworks	Free	Cobra toolbox	https://opencobra.github.io/cobratoolbox/latest/modules/dataIntegration/transcriptomics/index.html	genome-scale model, transcriptomics
INIT	FBA to extract subnetworks	Free	Cobra toolbox	https://opencobra.github.io/cobratoolbox/latest/modules/dataIntegration/transcriptomics/index.html	genome-scale model, transcriptomics
tINIT	FBA to extract subnetworks	N/A	N/A	N/A	genome-scale model, proteomics
CORDA	FBA to extract subnetworks	Free	Python	https://pypi.org/project/corda/	genome-scale model, transcriptomics
RegrEx	FBA to extract subnetworks	Free	MATLAB	N/A	genome-scale model, transcriptomics
cFBA	FBA for community modeling	Free	Python	N/A	Genome-scale model
OptCom	FBA for community modeling	Free	Python	https://resendislab.github.io/micom/micom.html#micom.community.Community.optcom	Genome-scale model
CASINO	FBA for community modeling	N/A	N/A	N/A	Genome-scale model
SteadyCom	FBA for community modeling	Free	COBRA toolbox	https://opencobra.github.io/cobratoolbox/stable/modules/analysis/multiSpecies/SteadyCom/index.html	Genome-scale model
DMMM	FBA for dynamic community modeling	N/A	N/A	N/A	Genome-scale model
dOptCom	FBA for dynamic community modeling	N/A	N/A	N/A	Genome-scale model
COMET	FBA for dynamic community modeling	Free	MATLAB/Python	https://github.com/segrelab/comets	Genome-scale model
BacArena	FBA for dynamic community modeling	Free	R	https://cran.r-project.org/package=BacArena	Genome-scale model
dynamic FBA	Dynamic flux balance analysis	N/A	N/A	N/A	Genome-scale model
Tobalina et al.	FBA for community modeling	N/A	MATLAB + CPLEX	N/A	Metaproteomics data
Henry et al.	FBA for community modeling and model reconstruction	Free	N/A	http://www.theseed.org/models/	Annotated genome

CoMiDa	FBA for community modeling	Non-commercial use Only	Python	https://github.com/borenstein-lab/CoMiDA	Genome-scale model
METRAN	^{13}C-MFA	MIT	N/A	N/A	13C-labeling data, exchange flux, growth rate
INCA	^{13}C-MFA	Free for educational use	MATLAB	https://mfa.vueinnovations.com	13C-labeling data, exchange flux, growth rate
13FLUX2	^{13}C-MFA	Free for educational use	Linux, stand alone	https://www.13cflux.net	13C-labeling data, exchange flux, growth rate
GS-MFA (stationary)	^{13}C-MFA	Free	MATLAB	https://github.com/maranasgroup/SteadyState-MFA	13C-labeling data, exchange flux, growth rate
GS-MFA (nonstationary)	^{13}C-MFA	Free	MATLAB	https://github.com/maranasgroup/Nonstationary-MFA	13C-labeling data, exchange flux, growth rate
WUflux	^{13}C-MFA	Free	MATLAB	http://www.13cmfa.org	13C-labeling data, exchange flux, growth rate
FluxML	Standardized ^{13}C-MFA markup language	Free	C++	https://github.com/modsim/FluxML	Network stoichiometry, atom transitions, labeling data
OpenMebius	^{13}C-MFA	Free	MATLAB	http://www-shimizu.ist.osaka-u.ac.jp/hp/en/software/OpenMebius.html	13C-labeling data, exchange flux, growth rate
OpenFlux	^{13}C-MFA	Free	MATLAB	http://openflux.sourceforge.net	13C-labeling data, exchange flux, growth rate
SUMOFLUX	^{13}C-MFA	Free	MATLAB	https://imsb.ethz.ch/research/zamboni/resources/fiatflux.html	13C-labeling data, exchange flux, growth rate
CLCA	Atom mapping algorithm	Free	JAVA	https://github.com/maranasgroup/MetRxn/tree/master/Alchemist	SMILES
RDT	Atom mapping algorithm	Free	JAVA	https://github.com/asad/ReactionDecoder	RXN, SMILES
DREAM	Atom mapping algorithm	Free	Webtool	http://ares.tamu.edu/dream/	RXN, SMILES
MWED	Atom mapping algorithm	Free for educational use	N/A	N/A	RXN, SMILES
ICMAP	Atom mapping algorithm	Commercial	N/A	https://www.infochem.de/chem-informatics/ic-map	RXN
AutoMapper	Atom mapping algorithm	Free for educational use	N/A	https://chemaxon.com	RXN, SMILES
Group contribution method by Jankowski et al.	GC-based dG estimation	N/A	web application	https://lcsb-databases.epfl.ch/pathways/GCM	Molfiles
Group contribution method by Noor et al.	GC-based dG estimation	MIT	Python	code.google.com/p/milo-lab	SMILES/InChI
Component contribution method	GC-based dG estimation	MIT	Python	https://gitlab.com/equilibrator/component-contribution	SMILES/InChI
Group contribution method by Du et al.	GC-based dG estimation	MIT	Python	https://github.com/bdu91/group-contribution	SMILES/InChI
Fingerprint contribution method	GC-based dG estimation	Free	Matlab	https://sfb.kaust.edu.sa/Pages/Software.aspx	N/A
Automatic molecule fragmentation method	group decompositon	MIT	Python	https://github.com/simonmb/fragmentation_algorithm_paper	SMILES

pKa estimation using machine learning	pka prediction	MIT	Python	https://github.com/czodrowskilab/Machine-learning-meets-pKa	SMILES
Quantum chemistry method by Jinch et al	Quantum-based dG estimation	N/A	N/A	N/A	SMILES
Quantum chemistry method by Hadadi et al	Quantum-based dG estimation	N/A	N/A	N/A	N/A
Equation-of-state approach by Panayiotou et al.	Quantum-based dG estimation	N/A	N/A	N/A	N/A
A mixed Quantum and machine learning method	Quantum-based dG estimation	MIT	Python	https://github.com/aspuru-guzik-group/gp_redox_rxn	SMILES
Group contribution method by Jankowski et al.	GC-based dG estimation	N/A	web application	https://lcsb-databases.epfl.ch/pathways/GCM	Molfiles
GA-based EM by Khodayari and Maranas	kinetic parameterization	Free	MATLAB	N/A	Fluxomics, gibbs free energy of reaction
EM with stability analysis By Greene et. al	kinetic parmeterization	Free	MATLAB	N/A	Fluxomics, gibbs free energy of reaction
Ensemble Modeling with Linear-Logarithmic Kinetics by St. John et al.	kinetic parmaeterization ,	Free	Python	https://github.com/pstjohn/emll	fluxomics, metabolomics
Adjoint Sensitivity by Frohlich et al.	kinetic parmaeterization	Free	MATLAB	N/A	metabolomics, proteomics, transcriptomics
K-FIT	kinetic parmaeterization	Free	MATLAB	https://github.com/maranasgroup/K-FIT	Fluxomics, reaction reversibility
SIMMER	allosteric regulation inference	free	R	https://github.com/shackett/simmer	metabolomics, proteomics, fluxomics
RBApy	resource allocation model parameterization	Free	Python	https://github.com/SysBioInra/RBApy	genome-scale reconstruction, NCBI taxonomy ID, macromolecular composition
COBRAme	ME model construction	Free	Python, COBRApy	https://github.com/SBRG/cobrame	genome-scale reconstruction, macromolecular composition
dynamicME	dynamic ME model simulation	Free	Python, COBRApy	https://github.com/SBRG/dynamicme	parameterized ME model
saCeSS	kinetic model parameterization	Free	Linux (Fortran and C)	https://bitbucket.org/DavidPenas/sacess-library/src/master/	Metabolomics, fluxomics, proteomics
VisId	Kinetic model parameterization , parameter correlation analysis	Free	MATLAB	https://github.com/gabora/visid	Metabolomics, fluxomics, proteomics
PathParser	kinetic and thermodynamic feasibility analysis	Code available from authors	N/A	N/A	Metabolic pathway
Dbsolve [240]	kinetic model parameterization , simulation, visualization	Free	Windows Installation	http://insysbio.com/en/software/db-solve-optimum	metabolomics, fluxomics, proteomics
COPASI [241]	dynamic simulation and analysis of	Free	MATLAB	http://copasi.org	metabolomics, fluxomics, proteomics

Acknowledgements

Funding provided by the Center for Bioenergy Innovation, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science under contract DE-AC05- 00OR22725. This work was partially funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Energy. Funding also provided NSF Award Number MCB-1615646.

References

[1] J. E. Bailey, "Toward a Science of Metabolic Engineering," *Science,* vol. 252, 5013, pp. 1668-1675, 1991.

[2] A. M. Chakrabarty, "Microorganisms having multiple compatible degradative energy-generating plasmids and perparation thereof," United States Patent Appl. 3813316, 1974.

[3] H. Yim *et al.*, "Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol," *Nat Chem Biol,* vol. 7, no. 7, pp. 445-52, May 22 2011, doi: 10.1038/nchembio.580.

[4] E. T. Papoutsakis, "Equations and Calculations for Fermentations of Butyric Acid Bacteria," *Biotechnol Bioeng,* vol. 26, pp. 174-187, 1984.

[5] M. R. Watson, "Metabolic maps for the Apple II," *Trans Biochem Soc,* vol. 12, pp. 1093-1094, 1984.

[6] M. R. Watson, "A discrete model of bacterial metabolism," *CABIOS,* vol. 2, no. 1, pp. 23-27, 1986.

[7] J. S. Edwards and B. O. Palsson, "Systems Properties of the Haemophilus influenzae Rd Metabolic Genotype*," *J Biol Chem,* vol. 274, no. 25, pp. 17410-17416, 1999.

[8] M. A. Oberhardt, B. O. Palsson, and J. A. Papin, "Applications of genome-scale metabolic reconstructions," *Mol Syst Biol,* vol. 5, p. 320, 2009, doi: 10.1038/msb.2009.77.

[9] C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee, "Current status and applications of genome-scale metabolic models," *Genome Biol,* vol. 20, no. 1, p. 121, Jun 13 2019, doi: 10.1186/s13059-019-1730-3.

[10] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Thermodynamics-based metabolic flux analysis," *Biophys J,* vol. 92, no. 5, pp. 1792-805, Mar 1 2007, doi: 10.1529/biophysj.106.093138.

[11] M. Ataman and V. Hatzimanikatis, "Heading in the right direction: thermodynamics-based network analysis and pathway engineering," *Curr Opin Biotechnol,* vol. 36, pp. 176-82, Dec 2015, doi: 10.1016/j.copbio.2015.08.021.
[12] R. N. Goldberg, Y. B. Tewari, and T. N. Bhat, "Thermodynamics of enzyme-catalyzed reactions--a database for quantitative biochemistry," *Bioinformatics,* vol. 20, no. 16, pp. 2874-7, Nov 1 2004, doi: 10.1093/bioinformatics/bth314.
[13] R. M. Burton, "[43] The determination of glycerol and dihydroxyacetone," (Methods in Enzymology, 1957, pp. 246-249.
[14] K. Burton and H. A. Krebs, "The Free-energy Changes Associated with the Individual Steps of the Tricarboxylic Acid cycle, Glycolysis and Alcoholic Fermentation and with the Hydrolysis of the Pyrophosphate Group of Adenosinetriphosphate," *Biochemical Journal,* vol. 54, 1, pp. 94-107, 1953.
[15] R. A. Alberty, "Biochemical thermodynamics: applications of Mathematica," *Methods Biochem Anal,* vol. 48, pp. 1-458, 2006. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/16878778.
[16] S. W. Benson and J. H. Buss, "Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties," *The Journal of Chemical Physics,* vol. 29, no. 3, pp. 546-572, 1958, doi: 10.1063/1.1744539.
[17] M. L. Mavrovouniotis, P. Bayol, T. K. M. Lam, G. Stephanopoulos, and G. Stephanopolous, "A group contribution method for the estimation of equilibrium constants for biochemical reactions," *Biotechnology Techniques,* vol. 2, no. 1, pp. 23-28, 1988.
[18] B. Du, D. C. Zielinski, and B. O. Palsson, "Estimating Metabolic Equilibrium Constants: Progress and Future Challenges," *Trends Biochem Sci,* vol. 43, no. 12, pp. 960-969, Dec 2018, doi: 10.1016/j.tibs.2018.09.009.
[19] F. Heinmets, "Cellular growth, injury and death," *Helgoländer wissenschaftliche Meeresuntersuchungen,* vol. 14, no. 1-4, pp. 168-194, 1966.
[20] E. J. Davison, "Simulation of cell behavior: normal and abnormal growth," *Bulletin of Mathematical Biology,* vol. 37, pp. 427-458, 1975.
[21] M. L. Shuler and M. M. Domach, "Mathematical Models of the Growth of Individual Cells: Tools for testing biochemical mechanisms," *ACS Symposium Series,* pp. 93-133, 1983.
[22] D. N. Macklin *et al.*, "Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation," *Science,* vol. 369, no. 6502, Jul 24 2020, doi: 10.1126/science.aav3751.
[23] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics,* vol. 38, no. 8, pp. 4-7, 1965.
[24] B. Berger, N. M. Daniels, and Y. W. Yu, "Computational Biology in the 21st Century: Scaling with Compressive Algorithms," *Commun ACM,* vol. 59, no. 8, pp. 72-80, Aug 2016, doi: 10.1145/2957324.
[25] A. D. Garst *et al.*, "Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering," *Nat Biotechnol,* vol. 35, no. 1, pp. 48-55, Jan 2017, doi: 10.1038/nbt.3718.
[26] S. Zhang *et al.*, "Recent Advances of CRISPR/Cas9-Based Genetic Engineering and Transcriptional Regulation in Industrial Biology," *Front Bioeng Biotechnol,* vol. 7, p. 459, 2019, doi: 10.3389/fbioe.2019.00459.
[27] A. C. Reis *et al.*, "Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays," *Nat Biotechnol,* vol. 37, no. 11, pp. 1294-1301, Nov 2019, doi: 10.1038/s41587-019-0286-9.
[28] M. HamediRad, R. Chao, S. Weisberg, J. Lian, S. Sinha, and H. Zhao, "Towards a fully automated algorithm driven platform for biosystems design," *Nat Commun,* vol. 10, no. 1, p. 5150, Nov 13 2019, doi: 10.1038/s41467-019-13189-z.

[29] J. L. Reed and B. O. Palsson, "Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states," *Genome Res,* vol. 14, no. 9, pp. 1797-805, Sep 2004, doi: 10.1101/gr.2546004.

[30] I. Thiele and B. Ø. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction," *Nature protocols,* vol. 5, no. 1, p. 93, 2010.

[31] F. Buchel *et al.*, "Path2Models: large-scale generation of computational models from biochemical pathway maps," *BMC Syst Biol,* vol. 7, p. 116, Nov 1 2013, doi: 10.1186/1752-0509-7-116.

[32] S. Devoid, R. Overbeek, M. DeJongh, V. Vonstein, A. A. Best, and C. S. Henry, "Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED," in *Systems Metabolic Engineering Methods in Molecular Biology (Methods and Protocols)*, vol. 985, H. Alper Ed. Totowa, NJ: Humana Press, 2013.

[33] O. Dias, M. Rocha, E. C. Ferreira, and I. Rocha, "Reconstructing High-Quality Large-Scale Metabolic Models with merlin," in *Metabolic Network Reconstruction and Modeling: Methods and Protocols*, M. Fondi Ed. New York, NY: Springer New York, 2018, pp. 1-36.

[34] A. P. Arkin *et al.*, "KBase: The United States Department of Energy Systems Biology Knowledgebase," *Nat Biotechnol,* vol. 36, no. 7, pp. 566-569, Jul 6 2018, doi: 10.1038/nbt.4163.

[35] H. Wang *et al.*, "RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor," *PLoS Comput Biol,* vol. 14, no. 10, p. e1006541, Oct 2018, doi: 10.1371/journal.pcbi.1006541.

[36] D. Machado, S. Andrejev, M. Tramontano, and K. R. Patil, "Fast automated reconstruction of genome-scale metabolic models for microbial species and communities," *Nucleic Acids Res,* vol. 46, no. 15, pp. 7542-7553, Sep 6 2018, doi: 10.1093/nar/gky537.

[37] M. Aite *et al.*, "Traceability, reproducibility and wiki-exploration for "a-la-carte" reconstructions of genome-scale metabolic models," *PLoS Comput Biol,* vol. 14, no. 5, p. e1006146, May 2018, doi: 10.1371/journal.pcbi.1006146.

[38] E. Karlsen, C. Schulz, and E. Almaas, "Automated generation of genome-scale metabolic draft reconstructions based on KEGG," *BMC Bioinformatics,* vol. 19, no. 1, p. 467, Dec 4 2018, doi: 10.1186/s12859-018-2472-z.

[39] P. D. Karp *et al.*, "Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology," *Brief Bioinform,* vol. 17, no. 5, pp. 877-90, Sep 2016, doi: 10.1093/bib/bbv079.

[40] B. G. Oliver, "SystemsBioinformatics/cbmpy-metadraft: MetaDraft release 0.9.2.," *Zendo,* 2019, doi: 10.5281/zenodo.3363099.

[41] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res,* vol. 45, no. D1, pp. D353-D361, Jan 4 2017, doi: 10.1093/nar/gkw1092.

[42] A. Wilke *et al.*, "The MG-RAST metagenomics database and portal in 2015," *Nucleic Acids Res,* vol. 44, no. D1, pp. D590-4, Jan 4 2016, doi: 10.1093/nar/gkv1322.

[43] J. Schellenberger, J. O. Park, T. M. Conrad, and B. O. Palsson, "BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions," *BMC Bioinformatics,* vol. 11, no. 1, p. 1:10, 2010.

[44] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden, "NCBI BLAST: a better web interface," *Nucleic Acids Res,* vol. 36, no. Web Server issue, pp. W5-9, Jul 1 2008, doi: 10.1093/nar/gkn201.

[45] S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez, and R. D. Finn, "HMMER web server: 2018 update," *Nucleic Acids Res,* vol. 46, no. W1, pp. W200-W204, Jul 2 2018, doi: 10.1093/nar/gky448.

[46] P. D. Karp, M. Latendresse, and R. Caspi, "The pathway tools pathway prediction algorithm," *Stand Genomic Sci,* vol. 5, no. 3, pp. 424-9, Dec 31 2011, doi: 10.4056/sigs.1794338.

[47] N. Loira, A. Zhukova, and D. J. Sherman, "Pantograph: A template-based method for genome-scale metabolic model reconstruction," *J Bioinform Comput Biol,* vol. 13, no. 2, p. 1550006, Apr 2015, doi: 10.1142/S0219720015500067.

[48] L. Li, C. J. Stoeckert, Jr., and D. S. Roos, "OrthoMCL: identification of ortholog groups for eukaryotic genomes," *Genome Res,* vol. 13, no. 9, pp. 2178-89, Sep 2003, doi: 10.1101/gr.1224503.

[49] V. S. Kumar and C. D. Maranas, "GrowMatch: an automated method for reconciling in silico/in vivo growth predictions," *PLoS Comput Biol,* vol. 5, no. 3, p. e1000308, Mar 2009, doi: 10.1371/journal.pcbi.1000308.

[50] D. Hartleb, F. Jarre, and M. J. Lercher, "Improved Metabolic Models for E. coli and Mycoplasma genitalium from GlobalFit, an Algorithm That Simultaneously Matches Growth and Non-Growth Data Sets," *PLoS Comput Biol,* vol. 12, no. 8, p. e1005036, Aug 2016, doi: 10.1371/journal.pcbi.1005036.

[51] G. L. Medlock and J. A. Papin, "Guiding the refinement of biochemical knowledgebases with ensembles of metabolic networks and semi-supervised learning," *bioRxiv,* 2019, doi: 10.1101/460071.

[52] A. S. Blazier and J. A. Papin, "Reconciling high-throughput gene essentiality data with metabolic network reconstructions," *PLoS Comput Biol,* vol. 15, no. 4, p. e1006507, Apr 2019, doi: 10.1371/journal.pcbi.1006507.

[53] L. Heirendt *et al.*, "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0," *Nat Protoc,* vol. 14, no. 3, pp. 639-702, Mar 2019, doi: 10.1038/s41596-018-0098-2.

[54] S. N. Mendoza, B. G. Olivier, D. Molenaar, and B. Teusink, "A systematic assessment of current genome-scale metabolic reconstruction tools," *Genome Biol,* vol. 20, no. 1, p. 158, Aug 7 2019, doi: 10.1186/s13059-019-1769-1.

[55] J. P. Faria, M. Rocha, I. Rocha, and C. S. Henry, "Methods for automated genome-scale metabolic model reconstruction," *Biochem Soc Trans,* vol. 46, no. 4, pp. 931-936, Aug 20 2018, doi: 10.1042/BST20170246.

[56] C. Lieven *et al.*, "MEMOTE for standardized genome-scale metabolic model testing," *Nat Biotechnol,* vol. 38, no. 3, pp. 272-276, Mar 2020, doi: 10.1038/s41587-020-0446-y.

[57] S. H. J. Chan, J. Cai, L. Wang, M. N. Simons-Senftle, and C. D. Maranas, "Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models," *Bioinformatics,* vol. 33, no. 22, pp. 3603-3609, Nov 15 2017, doi: 10.1093/bioinformatics/btx453.

[58] A. Ghazalpour *et al.*, "Comparative analysis of proteome and transcriptome variation in mouse," *PLoS Genet,* vol. 7, no. 6, p. e1001393, Jun 2011, doi: 10.1371/journal.pgen.1001393.

[59] E. J. Foss *et al.*, "Genetic basis of proteome variation in yeast," *Nat Genet,* vol. 39, no. 11, pp. 1369-75, Nov 2007, doi: 10.1038/ng.2007.22.

[60] J. Fu *et al.*, "System-wide molecular evidence for phenotypic buffering in Arabidopsis," *Nat Genet,* vol. 41, no. 2, pp. 166-7, Feb 2009, doi: 10.1038/ng.308.

[61] A. Zelezniak, S. Sheridan, and K. R. Patil, "Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes," *PLoS Comput Biol,* vol. 10, no. 4, p. e1003572, Apr 2014, doi: 10.1371/journal.pcbi.1003572.

[62] D. Machado and M. Herrgard, "Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism," *PLoS Comput Biol,* vol. 10, no. 4, p. e1003580, Apr 2014, doi: 10.1371/journal.pcbi.1003580.

[63] C. Colijn *et al.*, "Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production," *PLoS Comput Biol,* vol. 5, no. 8, p. e1000489, Aug 2009, doi: 10.1371/journal.pcbi.1000489.

[64] M. K. Kim, A. Lane, J. J. Kelley, and D. S. Lun, "E-Flux2 and SPOT: Validated Methods for Inferring Intracellular Metabolic Flux Distributions from Transcriptomic Data," *PLoS One,* vol. 11, no. 6, p. e0157101, 2016, doi: 10.1371/journal.pone.0157101.

[65] N. E. Lewis *et al.*, "Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models," *Mol Syst Biol,* vol. 6, p. 390, Jul 2010, doi: 10.1038/msb.2010.47.
[66] M. Tian and J. L. Reed, "Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis," *Bioinformatics,* vol. 34, no. 22, pp. 3882-3888, Nov 15 2018, doi: 10.1093/bioinformatics/bty445.
[67] P. A. Jensen and J. A. Papin, "Functional integration of a metabolic network model and expression data without arbitrary thresholding," *Bioinformatics,* vol. 27, no. 4, pp. 541-7, Feb 15 2011, doi: 10.1093/bioinformatics/btq702.
[68] T. Pusa *et al.*, "MOOMIN - Mathematical explOration of 'Omics data on a MetabolIc Network," *Bioinformatics,* vol. 36, no. 2, pp. 514-523, Jan 15 2020, doi: 10.1093/bioinformatics/btz584.
[69] S. S. Samal, O. Radulescu, A. Weber, and H. Frohlich, "Linking metabolic network features to phenotypes using sparse group lasso," *Bioinformatics,* vol. 33, no. 21, pp. 3445-3453, Nov 1 2017, doi: 10.1093/bioinformatics/btx427.
[70] S. S. Samal, H. Errami, and A. Weber, "PoCaB: A Software Infrastructure to Explore Algebraic Methods for Bio-chemical Reaction Networks," in *Computer Algebra in Scientific Computing*, (Lecture Notes in Computer Science, 2012, ch. Chapter 25, pp. 294-307.
[71] M. W. Covert and B. O. Palsson, "Constraints-based models: regulation of gene expression reduces the steady-state solution space," *J Theor Biol,* vol. 221, no. 3, pp. 309-25, Apr 7 2003, doi: 10.1006/jtbi.2003.3071.
[72] D. Marbach *et al.*, "Wisdom of crowds for robust gene network inference," *Nat Methods,* vol. 9, no. 8, pp. 796-806, 2012.
[73] S. Chandrasekaran and N. D. Price, "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis," *Proc Natl Acad Sci U S A,* vol. 107, no. 41, pp. 17845-50, Oct 12 2010, doi: 10.1073/pnas.1005139107.
[74] Z. Wang *et al.*, "Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast," *PLoS Comput Biol,* vol. 13, no. 5, p. e1005489, May 2017, doi: 10.1371/journal.pcbi.1005489.
[75] R. Bonneau *et al.*, "A predictive model for transcriptional control of physiology in a free living cell," *Cell,* vol. 131, no. 7, pp. 1354-65, Dec 28 2007, doi: 10.1016/j.cell.2007.10.053.
[76] E. Motamedian, M. Mohammadi, S. A. Shojaosadati, and M. Heydari, "TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data," *Bioinformatics,* vol. 33, no. 7, pp. 1057-1063, Apr 1 2017, doi: 10.1093/bioinformatics/btw772.
[77] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions," *Metab Eng,* vol. 9, no. 1, pp. 68-86, Jan 2007, doi: 10.1016/j.ymben.2006.09.001.
[78] J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, and G. Stephanopoulos, "An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis," *Biotechnol Bioeng,* vol. 99, no. 3, pp. 686-99, Feb 15 2008, doi: 10.1002/bit.21632.
[79] T. Szyperski, "Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids: An efficient analytical tool to investigate intermediary metabolism," *Eur J Biochem,* vol. 232, pp. 433-448, 1995.
[80] C. P. Long and M. R. Antoniewicz, "High-resolution (13)C metabolic flux analysis," *Nat Protoc,* vol. 14, no. 10, pp. 2856-2877, Oct 2019, doi: 10.1038/s41596-019-0204-0.
[81] J. Choi and M. R. Antoniewicz, "Tandem Mass Spectrometry for (13)C Metabolic Flux Analysis: Methods and Algorithms Based on EMU Framework," *Front Microbiol,* vol. 10, p. 31, 2019, doi: 10.3389/fmicb.2019.00031.
[82] C. P. Long, J. Au, J. E. Gonzalez, and M. R. Antoniewicz, "(13)C metabolic flux analysis of microbial and mammalian systems is enhanced with GC-MS measurements of glycogen and RNA labeling," *Metab Eng,* vol. 38, pp. 65-72, Nov 2016, doi: 10.1016/j.ymben.2016.06.007.

[83] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, "Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements," *Metab Eng,* vol. 8, no. 4, pp. 324-37, Jul 2006, doi: 10.1016/j.ymben.2006.01.004.

[84] A. Theorell, S. Leweke, W. Wiechert, and K. Noh, "To be certain about the uncertainty: Bayesian statistics for (13) C metabolic flux analysis," *Biotechnol Bioeng,* vol. 114, no. 11, pp. 2668-2684, Nov 2017, doi: 10.1002/bit.26379.

[85] H. Yoo, M. R. Antoniewicz, G. Stephanopoulos, and J. K. Kelleher, "Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line," *J Biol Chem,* vol. 283, no. 30, pp. 20621-7, Jul 25 2008, doi: 10.1074/jbc.M706494200.

[86] L. He, S. G. Wu, M. Zhang, Y. Chen, and Y. J. Tang, "WUFlux: an open-source platform for (13)C metabolic flux analysis of bacterial metabolism," *BMC Bioinformatics,* vol. 17, no. 1, p. 444, Nov 4 2016, doi: 10.1186/s12859-016-1314-0.

[87] L. E. Quek, C. Wittmann, L. K. Nielsen, and J. O. Kromer, "OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis," *Microb Cell Fact,* vol. 8, p. 25, May 1 2009, doi: 10.1186/1475-2859-8-25.

[88] J. D. Young, "INCA: a computational platform for isotopically non-stationary metabolic flux analysis," *Bioinformatics,* vol. 30, no. 9, pp. 1333-5, May 1 2014, doi: 10.1093/bioinformatics/btu015.

[89] M. Weitzel, K. Noh, T. Dalman, S. Niedenfuhr, B. Stute, and W. Wiechert, "13CFLUX2--high-performance software suite for (13)C-metabolic flux analysis," *Bioinformatics,* vol. 29, no. 1, pp. 143-5, Jan 1 2013, doi: 10.1093/bioinformatics/bts646.

[90] S. Kajihata, C. Furusawa, F. Matsuda, and H. Shimizu, "OpenMebius: an open source software for isotopically nonstationary 13C-based metabolic flux analysis," *Biomed Res Int,* vol. 2014, p. 627014, 2014, doi: 10.1155/2014/627014.

[91] S. Gopalakrishnan and C. D. Maranas, "13C metabolic flux analysis at a genome-scale," *Metab Eng,* vol. 32, pp. 12-22, Nov 2015, doi: 10.1016/j.ymben.2015.08.006.

[92] R. W. Leighty and M. R. Antoniewicz, "COMPLETE-MFA: complementary parallel labeling experiments technique for metabolic flux analysis," *Metab Eng,* vol. 20, pp. 49-55, Nov 2013, doi: 10.1016/j.ymben.2013.08.006.

[93] S. B. Crown, C. P. Long, and M. R. Antoniewicz, "Integrated 13C-metabolic flux analysis of 14 parallel labeling experiments in Escherichia coli," *Metab Eng,* vol. 28, pp. 151-158, Mar 2015, doi: 10.1016/j.ymben.2015.01.001.

[94] N. A. Gebreselassie and M. R. Antoniewicz, "(13)C-metabolic flux analysis of co-cultures: A novel approach," *Metab Eng,* vol. 31, pp. 132-9, Sep 2015, doi: 10.1016/j.ymben.2015.07.005.

[95] S. Gopalakrishnan, H. B. Pakrasi, and C. D. Maranas, "Elucidation of photoautotrophic carbon flux topology in Synechocystis PCC 6803 using genome-scale carbon mapping models," *Metab Eng,* vol. 47, pp. 190-199, May 2018, doi: 10.1016/j.ymben.2018.03.008.

[96] J. I. Hendry, S. Gopalakrishnan, J. Ungerer, H. B. Pakrasi, Y. J. Tang, and C. D. Maranas, "Genome-Scale Fluxome of Synechococcus elongatus UTEX 2973 Using Transient (13)C-Labeling Data," *Plant Physiol,* vol. 179, no. 2, pp. 761-769, Feb 2019, doi: 10.1104/pp.18.01357.

[97] D. McCloskey, J. D. Young, S. Xu, B. O. Palsson, and A. M. Feist, "Modeling Method for Increased Precision and Scope of Directly Measurable Fluxes at a Genome-Scale," *Anal Chem,* vol. 88, no. 7, pp. 3844-52, Apr 5 2016, doi: 10.1021/acs.analchem.5b04914.

[98] R. Caspi *et al.*, "The MetaCyc database of metabolic pathways and enzymes," *Nucleic Acids Res,* vol. 46, no. D1, pp. D633-D639, Jan 4 2018, doi: 10.1093/nar/gkx935.

[99] A. Kumar and C. D. Maranas, "CLCA: maximum common molecular substructure queries within the MetRxn database," *J Chem Inf Model,* vol. 54, no. 12, pp. 3417-38, Dec 22 2014, doi: 10.1021/ci5003922.

[100] E. E. Litsa, M. I. Pena, M. Moll, G. Giannakopoulos, G. N. Bennett, and L. E. Kavraki, "Machine Learning Guided Atom Mapping of Metabolic Reactions," *J Chem Inf Model,* vol. 59, no. 3, pp. 1121-1135, Mar 25 2019, doi: 10.1021/acs.jcim.8b00434.

[101] E. L. First, C. E. Gounaris, and C. A. Floudas, "Stereochemically consistent reaction mapping and identification of multiple reaction mechanisms through integer linear optimization," *J Chem Inf Model,* vol. 52, no. 1, pp. 84-92, Jan 23 2012, doi: 10.1021/ci200351b.
[102] H. Kraut, J. Eiblmaier, G. Grethe, P. Low, H. Matuszczyk, and H. Saller, "Algorithm for reaction classification," *J Chem Inf Model,* vol. 53, no. 11, pp. 2884-95, Nov 25 2013, doi: 10.1021/ci400442f.
[103] M. Latendresse, J. P. Malerich, M. Travers, and P. D. Karp, "Accurate atom-mapping computation for biochemical reactions," *J Chem Inf Model,* vol. 52, no. 11, pp. 2970-82, Nov 26 2012, doi: 10.1021/ci3002217.
[104] S. A. Rahman *et al.*, "Reaction Decoder Tool (RDT): extracting features from chemical reactions," *Bioinformatics,* vol. 32, no. 13, pp. 2065-6, Jul 1 2016, doi: 10.1093/bioinformatics/btw096.
[105] A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas, "Flux coupling analysis of genome-scale metabolic network reconstructions," *Genome Res,* vol. 14, no. 2, pp. 301-12, Feb 2004, doi: 10.1101/gr.1926504.
[106] J. I. Hendry, H. V. Dinh, C. Foster, S. Gopalakrishnan, L. Wang, and C. D. Maranas, "Metabolic flux analysis reaching genome wide coverage: lessons learned and future perspectives," *Current Opinion in Chemical Engineering,* vol. 30, pp. 17-25, 2020, doi: 10.1016/j.coche.2020.05.008.
[107] H. Nam *et al.*, "A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks," *PLoS Comput Biol,* vol. 10, no. 9, p. e1003837, Sep 2014, doi: 10.1371/journal.pcbi.1003837.
[108] O. Oyas *et al.*, "Model-based integration of genomics and metabolomics reveals SNP functionality in Mycobacterium tuberculosis," *Proc Natl Acad Sci U S A,* vol. 117, no. 15, pp. 8494-8502, Apr 14 2020, doi: 10.1073/pnas.1915551117.
[109] D. Sarkar and C. D. Maranas, "SNPeffect: identifying functional roles of SNPs using metabolic networks," *Plant J,* Mar 13 2020, doi: 10.1111/tpj.14746.
[110] H. Tong, A. Kuken, and Z. Nikoloski, "Integrating molecular markers into metabolic models improves genomic selection for Arabidopsis growth," *Nat Commun,* vol. 11, no. 1, p. 2410, May 15 2020, doi: 10.1038/s41467-020-16279-5.
[111] L. Wang, S. Dash, C. Y. Ng, and C. D. Maranas, "A review of computational tools for design and reconstruction of metabolic pathways," *Synth Syst Biotechnol,* vol. 2, no. 4, pp. 243-252, Dec 2017, doi: 10.1016/j.synbio.2017.11.002.
[112] U. Gupta, T. Le, W. S. Hu, A. Bhan, and P. Daoutidis, "Automated network generation and analysis of biochemical reaction pathways using RING," *Metab Eng,* vol. 49, pp. 84-93, Sep 2018, doi: 10.1016/j.ymben.2018.07.009.
[113] A. Kumar, L. Wang, C. Y. Ng, and C. D. Maranas, "Pathway design using de novo steps through uncharted biochemical spaces," *Nature communications,* vol. 9, no. 1, pp. 1-15, 2018.
[114] P. Carbonell, P. Parutto, C. Baudier, C. Junot, and J. L. Faulon, "Retropath: automated pipeline for embedded metabolic circuits," *ACS Synth Biol,* vol. 3, no. 8, pp. 565-77, Aug 15 2014, doi: 10.1021/sb4001273.
[115] S. Ding *et al.*, "novoPathFinder: a webserver of designing novel-pathway with integrating GEM-model," *Nucleic Acids Res,* vol. 48, no. W1, pp. W477-W487, Jul 2 2020, doi: 10.1093/nar/gkaa230.
[116] G.-M. Lin, R. Warden-Rothman, and C. A. Voigt, "Retrosynthetic design of metabolic pathways to chemicals not found in nature," *Current Opinion in Systems Biology,* vol. 14, pp. 82-107, 2019.
[117] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo, "Glycolytic strategy as a tradeoff between energy yield and protein cost," *Proc Natl Acad Sci U S A,* vol. 110, no. 24, pp. 10039-44, Jun 11 2013, doi: 10.1073/pnas.1215283110.
[118] C. Y. Ng, L. Wang, A. Chowdhury, and C. D. Maranas, "Pareto Optimality Explanation of the Glycolytic Alternatives in Nature," *Sci Rep,* vol. 9, no. 1, p. 2633, Feb 22 2019, doi: 10.1038/s41598-019-38836-9.

[119] O. Hadicke, A. von Kamp, T. Aydogan, and S. Klamt, "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,* vol. 14, no. 9, p. e1006492, Sep 2018, doi: 10.1371/journal.pcbi.1006492.
[120] S. Dash, D. G. Olson, S. H. Joshua Chan, D. Amador-Noguez, L. R. Lynd, and C. D. Maranas, "Thermodynamic analysis of the pathway for ethanol production from cellobiose in Clostridium thermocellum," *Metab Eng,* vol. 55, pp. 161-169, Jun 17 2019, doi: 10.1016/j.ymben.2019.06.006.
[121] B. Niebel, S. Leupold, and M. Heinemann, "An upper limit on Gibbs energy dissipation governs cellular metabolism," *Nature Metabolism,* vol. 1, no. 1, pp. 125-132, 2019, doi: 10.1038/s42255-018-0006-7.
[122] R. A. Alberty, A. Cornish-Bowden, R. N. Goldberg, G. G. Hammes, K. Tipton, and H. V. Westerhoff, "Recommendations for terminology and databases for biochemical thermodynamics," *Biophys Chem,* vol. 155, no. 2-3, pp. 89-103, May 2011, doi: 10.1016/j.bpc.2011.03.007.
[123] T. B. Jacobson *et al.*, "(2)H and (13)C metabolic flux analysis elucidates in vivo thermodynamics of the ED pathway in Zymomonas mobilis," *Metab Eng,* vol. 54, pp. 301-316, Jul 2019, doi: 10.1016/j.ymben.2019.05.006.
[124] T. B. Jacobson *et al.*, "In Vivo Thermodynamic Analysis of Glycolysis in Clostridium thermocellum and Thermoanaerobacterium saccharolyticum Using (13)C and (2)H Tracers," *mSystems,* vol. 5, no. 2, Mar 17 2020, doi: 10.1128/mSystems.00736-19.
[125] J. O. Park *et al.*, "Metabolite concentrations, fluxes and free energies imply efficient enzyme usage," *Nat Chem Biol,* vol. 12, no. 7, pp. 482-9, Jul 2016, doi: 10.1038/nchembio.2077.
[126] Mavrovouniotis, "Group Contrubtions for Estimating Standard Gibbs Energies of Formation of Biochemical Compounds in Aqueous Solution," *Biotechnol Bioeng,* vol. 36, pp. 1070-1082, 1990.
[127] W. D. Ihlenfeldt, E. E. Bolton, and S. H. Bryant, "The PubChem chemical structure sketcher," *J Cheminform,* vol. 1, no. 1, p. 20, Dec 17 2009, doi: 10.1186/1758-2946-1-20.
[128] E. S. Salmina, N. Haider, and I. V. Tetko, "Extended Functional Groups (EFG): An Efficient Set for Chemical Characterization and Structure-Activity Relationship Studies of Chemical Compounds," *Molecules,* vol. 21, no. 1, p. E1, Dec 23 2015, doi: 10.3390/molecules21010001.
[129] S. Müller, "Flexible heuristic algorithm for automatic molecule fragmentation: application to the UNIFAC group contribution model," *Journal of Cheminformatics,* vol. 11, no. 1, 2019, doi: 10.1186/s13321-019-0382-3.
[130] P. Carbonell, A. G. Planson, D. Fichera, and J. L. Faulon, "A retrosynthetic biology approach to metabolic pathway design for therapeutic production," *BMC Syst Biol,* vol. 5, no. 122, pp. 1-18, 2011.
[131] F. Ruggiu, G. Marcou, A. Varnek, and D. Horvath, "ISIDA Property-Labelled Fragment Descriptors," *Mol Inform,* vol. 29, no. 12, pp. 855-68, Dec 17 2010, doi: 10.1002/minf.201000099.
[132] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *J Chem Inf Model,* vol. 50, no. 5, pp. 742-54, May 24 2010, doi: 10.1021/ci100050t.
[133] D. Duvenaud *et al.*, "Convolutional Networks on Graphs for Learning Molecular Fingerprints," *arXiv,* no. 1-9, 2015.
[134] E. Noor, H. S. Haraldsdottir, R. Milo, and R. M. Fleming, "Consistent estimation of Gibbs energy using component contributions," *PLoS Comput Biol,* vol. 9, no. 7, p. e1003098, 2013, doi: 10.1371/journal.pcbi.1003098.
[135] D. Valdes-Ramirez *et al.*, "A Review of Fingerprint Feature Representations and Their Applications for Latent Fingerprint Identification: Trends and Evaluation," *IEEE Access,* vol. 7, pp. 48484-48499, 2019, doi: 10.1109/ACCESS.2019.2909497.
[136] M. Alazmi, H. Kuwahara, O. Soufan, L. Ding, and X. Gao, "Systematic selection of chemical fingerprint features improves the Gibbs energy prediction of biochemical reactions," *Bioinformatics,* vol. 35, no. 15, pp. 2634-2643, Aug 1 2019, doi: 10.1093/bioinformatics/bty1035.
[137] Y. C. Martin, "Experimental and pKa prediction aspects of tautomerism of drug-like molecules," *Drug Discov Today Technol,* vol. 27, pp. 59-64, Jul 2018, doi: 10.1016/j.ddtec.2018.06.006.

[138] E. Noor, A. Bar-Even, A. Flamholz, Y. Lubling, D. Davidi, and R. Milo, "An integrated open framework for thermodynamics of reactions that combines accuracy and coverage," *Bioinformatics,* vol. 28, no. 15, pp. 2037-44, Aug 1 2012, doi: 10.1093/bioinformatics/bts317.
[139] T. Kalliokoski and K. Sinervo, "Predicting pKa for Small Molecules on Public and In-house Datasets Using Fast Prediction Methods Combined with Data Fusion," *Mol Inform,* vol. 38, no. 7, p. e1800163, Jul 2019, doi: 10.1002/minf.201800163.
[140] B. A. Bunin, J. Bajorath, B. Siesel, and G. Morales, *Chemoinformatics: Theory, Practice, & Products*. Springer, 2007.
[141] J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin, and M. Uchimaya, "Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules," *J Comput Aided Mol Des,* vol. 21, no. 12, pp. 681-91, Dec 2007, doi: 10.1007/s10822-007-9133-z.
[142] J. Manchester, G. Walkup, and Z. You, "Evaluation of pKa Estimation Methods on 211 Druglike Compounds," *J. Chem. Inf. Model.,* vol. 50, pp. 565-571, 2010.
[143] P. Gedeck *et al.*, "Benefit of Retraining pKa Models Studied Using Internally Measured Data," *J Chem Inf Model,* vol. 55, no. 7, pp. 1449-59, Jul 27 2015, doi: 10.1021/acs.jcim.5b00172.
[144] B. Du, Z. Zhang, S. Grubner, J. T. Yurkovich, B. O. Palsson, and D. C. Zielinski, "Temperature-Dependent Estimation of Gibbs Energies Using an Updated Group-Contribution Method," *Biophys J,* vol. 114, no. 11, pp. 2691-2702, Jun 5 2018, doi: 10.1016/j.bpj.2018.04.030.
[145] M. Baltruschat and P. Czodrowski, "Machine learning meets pK a," *F1000Res,* vol. 9, 2020, doi: 10.12688/f1000research.22090.2.
[146] M. D. Jankowski, C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Group contribution method for thermodynamic analysis of complex metabolic networks," *Biophys J,* vol. 95, no. 3, pp. 1487-99, Aug 2008, doi: 10.1529/biophysj.107.124784.
[147] A. Jinich *et al.*, "Quantum chemical approach to estimating the thermodynamics of metabolic reactions," *Sci Rep,* vol. 4, p. 7022, Nov 12 2014, doi: 10.1038/srep07022.
[148] F. Neese, "Software update: the ORCA program system, version 4.0," *Wiley Interdisciplinary Reviews: Computational Molecular Science,* vol. 8, no. 1, p. e1327, 2018.
[149] L. Martinez, R. Andrade, E. G. Birgin, and J. M. Martinez, "PACKMOL: a package for building initial configurations for molecular dynamics simulations," *J Comput Chem,* vol. 30, no. 13, pp. 2157-64, Oct 2009, doi: 10.1002/jcc.21224.
[150] A. Klamt and G. Schüürmann, "COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *Journal of the Chemical Society, Perkin Transactions 2,* no. 5, pp. 799-805, 1993.
[151] A. Jinich *et al.*, "Quantum chemistry reveals thermodynamic principles of redox biochemistry," *PLoS Comput Biol,* vol. 14, no. 10, p. e1006471, Oct 2018, doi: 10.1371/journal.pcbi.1006471.
[152] N. Hadadi, M. Ataman, V. Hatzimanikatis, and C. Panayiotou, "Molecular thermodynamics of metabolism: quantum thermochemical calculations for key metabolites," *Phys Chem Chem Phys,* vol. 17, no. 16, pp. 10438-53, Apr 28 2015, doi: 10.1039/c4cp05825a.
[153] S. G. Balasubramani *et al.*, "TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations," *J Chem Phys,* vol. 152, no. 18, p. 184107, May 14 2020, doi: 10.1063/5.0004635.
[154] C. G. C. KG, "COSMOconfX version 4.1. COSMOlogic GmbH & Co.," *KG, Leverkusen,* 2016.
[155] "Spartan '14," *Wavefunction Inc.,* 2014.
[156] C. Panayiotou, S. Mastrogeorgopoulos, M. Ataman, N. Hadadi, and V. Hatzimanikatis, "Molecular thermodynamics of metabolism: hydration quantities and the equation-of-state approach," *Phys Chem Chem Phys,* vol. 18, no. 47, pp. 32570-32592, Nov 30 2016, doi: 10.1039/c6cp06281d.
[157] A. Jinich, B. Sanchez-Lengeling, H. Ren, R. Harman, and A. Aspuru-Guzik, "A Mixed Quantum Chemistry/Machine Learning Approach for the Fast and Accurate Prediction of Biochemical Redox Potentials and Its Large-Scale Application to 315000 Redox Reactions," *ACS Cent Sci,* vol. 5, no. 7, pp. 1199-1210, Jul 24 2019, doi: 10.1021/acscentsci.9b00297.

[158] J. Strutz, J. Martin, J. Greene, L. Broadbelt, and K. Tyo, "Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain," *Curr Opin Biotechnol,* vol. 59, pp. 24-30, Oct 2019, doi: 10.1016/j.copbio.2019.02.005.
[159] P. A. Saa and L. K. Nielsen, "Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks," *Biotechnol Adv,* vol. 35, no. 8, pp. 981-1003, Dec 2017, doi: 10.1016/j.biotechadv.2017.09.005.
[160] L. Yang, J. T. Yurkovich, Z. A. King, and B. O. Palsson, "Modeling the multi-scale mechanisms of macromolecular resource allocation," *Curr Opin Microbiol,* vol. 45, pp. 8-15, Oct 2018, doi: 10.1016/j.mib.2018.01.002.
[161] J. Nielsen and J. D. Keasling, "Engineering Cellular Metabolism," *Cell,* vol. 164, no. 6, pp. 1185-1197, Mar 10 2016, doi: 10.1016/j.cell.2016.02.004.
[162] A. Chowdhury, A. Khodayari, and C. D. Maranas, "Improving prediction fidelity of cellular metabolism with kinetic descriptions," *Curr Opin Biotechnol,* vol. 36, pp. 57-64, Dec 2015, doi: 10.1016/j.copbio.2015.08.011.
[163] Y. Matsuoka and K. Shimizu, "Current status and future perspectives of kinetic modeling for the cell metabolism with incorporation of the metabolic regulation mechanism," *Bioresources and Bioprocessing,* vol. 2, no. 1, 2015, doi: 10.1186/s40643-014-0031-7.
[164] N. Jamshidi and B. O. Palsson, "Formulating genome-scale kinetic models in the post-genome era," *Mol Syst Biol,* vol. 4, p. 171, 2008, doi: 10.1038/msb.2008.8.
[165] V. Hatzimanikatis and J. E. Bailey, "Effects of Spatiotemporal Variations on Metabolic Control: Approximate Analysis Using (Log)Linear kinetic Models," *Biotechnol Bioeng,* vol. 54, no. 2, pp. 91-104, 1996.
[166] J. J. Heijnen, "Approximative kinetic formats used in metabolic network modeling," *Biotechnol Bioeng,* vol. 91, no. 5, pp. 534-45, Sep 5 2005, doi: 10.1002/bit.20558.
[167] E. O. Voit, H. A. Martens, and S. W. Omholt, "150 years of the mass action law," *PLoS Comput Biol,* vol. 11, no. 1, p. e1004012, Jan 2015, doi: 10.1371/journal.pcbi.1004012.
[168] L. Wang, I. Birol, and V. Hatzimanikatis, "Metabolic control analysis under uncertainty: framework development and case studies," *Biophys J,* vol. 87, no. 6, pp. 3750-63, Dec 2004, doi: 10.1529/biophysj.104.048090.
[169] P. C. St John, J. Strutz, L. J. Broadbelt, K. E. J. Tyo, and Y. J. Bomble, "Bayesian inference of metabolic kinetics from genome-scale multiomics data," *PLoS Comput Biol,* vol. 15, no. 11, p. e1007424, Nov 2019, doi: 10.1371/journal.pcbi.1007424.
[170] N. Jamshidi and B. O. Palsson, "Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models," *Biophys J,* vol. 98, no. 2, pp. 175-85, Jan 20 2010, doi: 10.1016/j.bpj.2009.09.064.
[171] A. Bordbar, D. McCloskey, D. C. Zielinski, N. Sonnenschein, N. Jamshidi, and B. O. Palsson, "Personalized Whole-Cell Kinetic Models of Metabolism for Discovery in Genomics and Pharmacodynamics," *Cell Syst,* vol. 1, no. 4, pp. 283-92, Oct 28 2015, doi: 10.1016/j.cels.2015.10.003.
[172] S. Placzek *et al.*, "BRENDA in 2017: new perspectives and new tools in BRENDA," *Nucleic Acids Res,* vol. 45, no. D1, pp. D380-D388, Jan 4 2017, doi: 10.1093/nar/gkw952.
[173] R. S. Costa, A. Verissimo, and S. Vinga, "KiMoSys: a web-based repository of experimental data for KInetic MOdels of byological SYStems," *BMC Syst Biol,* vol. 8, no. 85, pp. 1-10, 2014.
[174] L. M. Tran, M. L. Rizk, and J. C. Liao, "Ensemble modeling of metabolic networks," *Biophys J,* vol. 95, no. 12, pp. 5606-17, Dec 15 2008, doi: 10.1529/biophysj.108.135442.
[175] R. M. Fleming, I. Thiele, G. Provan, and H. P. Nasheuer, "Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism," *J Theor Biol,* vol. 264, no. 3, pp. 683-92, Jun 7 2010, doi: 10.1016/j.jtbi.2010.02.044.
[176] R. Marti, "Scatter Search – Wellsprings and Challenges," *European Journal of Operational Research,* 2006.
[177] J. Kenedy and R. Eberhart, "Particle Swarm Optimization," *IEEE,* vol. 95, pp. 1942-1948, 1995.

[178] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing,* vol. 4, no. 2, 1994, doi: 10.1007/bf00175354.
[179] P. Millard, K. Smallbone, and P. Mendes, "Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in Escherichia coli," *PLoS Comput Biol,* vol. 13, no. 2, p. e1005396, Feb 2017, doi: 10.1371/journal.pcbi.1005396.
[180] N. Jahan, K. Maeda, Y. Matsuoka, Y. Sugimoto, and H. Kurata, "Development of an accurate kinetic model for the central carbon metabolism of Escherichia coli," *Microb Cell Fact,* vol. 15, no. 1, p. 112, Jun 21 2016, doi: 10.1186/s12934-016-0511-x.
[181] H. Kurata and Y. Sugimoto, "Improved kinetic model of Escherichia coli central carbon metabolism in batch and continuous cultures," *J Biosci Bioeng,* vol. 125, no. 2, pp. 251-257, Feb 2018, doi: 10.1016/j.jbiosc.2017.09.005.
[182] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss, "Dynamic modeling of the central carbon metabolism of Escherichia coli," *Biotechnol Bioeng,* vol. 79, no. 1, pp. 53-73, Jul 5 2002, doi: 10.1002/bit.10288.
[183] K. Peskov, E. Mogilevskaya, and O. Demin, "Kinetic modelling of central carbon metabolism in Escherichia coli," *FEBS J,* vol. 279, no. 18, pp. 3374-85, Sep 2012, doi: 10.1111/j.1742-4658.2012.08719.x.
[184] D. R. Penas, P. Gonzalez, J. A. Egea, R. Doallo, and J. R. Banga, "Parameter estimation in large-scale systems biology models: a parallel and self-adaptive cooperative strategy," *BMC Bioinformatics,* vol. 18, no. 1, p. 52, Jan 21 2017, doi: 10.1186/s12859-016-1452-4.
[185] A. Gabor and J. R. Banga, "Robust and efficient parameter estimation in dynamic models of biological systems," *BMC Syst Biol,* vol. 9, p. 74, Oct 29 2015, doi: 10.1186/s12918-015-0219-2.
[186] A. Gabor, A. F. Villaverde, and J. R. Banga, "Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems," *BMC Syst Biol,* vol. 11, no. 1, p. 54, May 5 2017, doi: 10.1186/s12918-017-0428-y.
[187] A. Khodayari and C. D. Maranas, "A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains," *Nat Commun,* vol. 7, p. 13806, Dec 20 2016, doi: 10.1038/ncomms13806.
[188] A. Khodayari, A. R. Zomorrodi, J. C. Liao, and C. D. Maranas, "A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data," *Metab Eng,* vol. 25, pp. 50-62, Sep 2014, doi: 10.1016/j.ymben.2014.05.014.
[189] J. Greene, A. Waechter, K. Tyo, and L. J. Broadbelt, "Eliminating dependent variables and guaranteeing local stability enhances metabolic ensemble modeling performance," *in review,* 2017.
[190] J. Monod, J. Wyman, and J. P. Changeux, "On the Nature of Allosteric Transitions: A Plausible Model," *J. Mol. Biol.,* vol. 12, pp. 88-118, 1965.
[191] P. A. Saa and L. K. Nielsen, "Construction of feasible and accurate kinetic models of metabolism: A Bayesian approach," *Sci Rep,* vol. 6, p. 29635, Jul 15 2016, doi: 10.1038/srep29635.
[192] H. Kacer and J. A. Burns, "The control of flux," *Soc. Exp. Biol.,* vol. 27, pp. 65-104, 1973.
[193] F. Frohlich, B. Kaltenbacher, F. J. Theis, and J. Hasenauer, "Scalable Parameter Estimation for Genome-Scale Biochemical Reaction Networks," *PLoS Comput Biol,* vol. 13, no. 1, p. e1005331, Jan 2017, doi: 10.1371/journal.pcbi.1005331.
[194] S. Gopalakrishnan, S. Dash, and C. Maranas, "K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data," *Metab Eng,* Mar 13 2020, doi: 10.1016/j.ymben.2020.03.001.
[195] T. Hameri, M. O. Boldi, and V. Hatzimanikatis, "Statistical inference in ensemble modeling of cellular metabolism," *PLoS Comput Biol,* vol. 15, no. 12, p. e1007536, Dec 2019, doi: 10.1371/journal.pcbi.1007536.
[196] C. J. Foster, S. Gopalakrishnan, M. R. Antoniewicz, and C. D. Maranas, "From Escherichia coli mutant 13C labeling data to a core kinetic model: A kinetic model parameterization pipeline," *PLoS Comput Biol,* vol. 15, no. 9, p. e1007319, Sep 2019, doi: 10.1371/journal.pcbi.1007319.

[197] L. Miskovic and V. Hatzimanikatis, "Production of biofuels and biochemicals: in need of an ORACLE," *Trends Biotechnol,* vol. 28, no. 8, pp. 391-7, Aug 2010, doi: 10.1016/j.tibtech.2010.05.003.
[198] S. Andreozzi, L. Miskovic, and V. Hatzimanikatis, "iSCHRUNK--In Silico Approach to Characterization and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks," *Metab Eng,* vol. 33, pp. 158-168, Jan 2016, doi: 10.1016/j.ymben.2015.10.002.
[199] M. Tokic, V. Hatzimanikatis, and L. Miskovic, "Large-scale kinetic metabolic models of Pseudomonas putida KT2440 for consistent design of metabolic engineering strategies," *Biotechnol Biofuels,* vol. 13, p. 33, 2020, doi: 10.1186/s13068-020-1665-7.
[200] S. Andreozzi *et al.*, "Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models," *Metab Eng,* vol. 35, pp. 148-159, May 2016, doi: 10.1016/j.ymben.2016.01.009.
[201] G. Savoglidis, A. X. da Silveira Dos Santos, I. Riezman, P. Angelino, H. Riezman, and V. Hatzimanikatis, "A method for analysis and design of metabolism using metabolomics data and kinetic models: Application on lipidomics using a novel kinetic model of sphingolipid metabolism," *Metab Eng,* vol. 37, pp. 46-62, Sep 2016, doi: 10.1016/j.ymben.2016.04.002.
[202] S. Dash *et al.*, "Development of a core Clostridium thermocellum kinetic metabolic model consistent with multiple genetic perturbations," *Biotechnol Biofuels,* vol. 10, p. 108, 2017, doi: 10.1186/s13068-017-0792-2.
[203] S. R. Hackett *et al.*, "Systems-level analysis of mechanisms regulating yeast metabolic flux," *Science,* vol. 354, no. 6311, Oct 28 2016, doi: 10.1126/science.aaf2786.
[204] D. Christodoulou, H. Link, T. Fuhrer, K. Kochanowski, L. Gerosa, and U. Sauer, "Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress," *Cell Syst,* May 2 2018, doi: 10.1016/j.cels.2018.04.009.
[205] R. Adadi, B. Volkmer, R. Milo, M. Heinemann, and T. Shlomi, "Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters," *PLoS Comput Biol,* vol. 8, no. 7, p. e1002575, 2012, doi: 10.1371/journal.pcbi.1002575.
[206] A. Vazquez *et al.*, "Impact of the solvent capacity constraint on E. coli metabolism," *BMC Syst Biol,* vol. 2, p. 7, Jan 23 2008, doi: 10.1186/1752-0509-2-7.
[207] Q. K. Beg *et al.*, "Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity," *Proc Natl Acad Sci U S A,* vol. 104, no. 31, pp. 12663-8, Jul 31 2007, doi: 10.1073/pnas.0609845104.
[208] B. J. Sanchez, C. Zhang, A. Nilsson, P. J. Lahtvee, E. J. Kerkhoven, and J. Nielsen, "Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints," *Mol Syst Biol,* vol. 13, no. 8, p. 935, Aug 3 2017, doi: 10.15252/msb.20167411.
[209] A. Goelzer and V. Fromion, "Bacterial growth rate reflects a bottleneck in resource allocation," *Biochim Biophys Acta,* vol. 1810, no. 10, pp. 978-88, Oct 2011, doi: 10.1016/j.bbagen.2011.05.014.
[210] I. Thiele, R. M. Fleming, R. Que, A. Bordbar, D. Diep, and B. O. Palsson, "Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage," *PLoS One,* vol. 7, no. 9, p. e45635, 2012, doi: 10.1371/journal.pone.0045635.
[211] E. J. O'Brien, J. A. Lerman, R. L. Chang, D. R. Hyduke, and B. O. Palsson, "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction," *Mol Syst Biol,* vol. 9, p. 693, Oct 1 2013, doi: 10.1038/msb.2013.52.
[212] A. M. Reimers, H. Knoop, A. Bockmayr, and R. Steuer, "Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth," *Proc Natl Acad Sci U S A,* vol. 114, no. 31, pp. E6457-E6465, Aug 1 2017, doi: 10.1073/pnas.1617508114.
[213] A. Goelzer *et al.*, "Quantitative prediction of genome-wide resource allocation in bacteria," *Metab Eng,* vol. 32, pp. 232-243, Nov 2015, doi: 10.1016/j.ymben.2015.10.003.
[214] J. K. Liu *et al.*, "Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen," *PLoS Comput Biol,* vol. 15, no. 3, p. e1006848, Mar 2019, doi: 10.1371/journal.pcbi.1006848.

[215] L. Yang, A. Ebrahim, C. J. Lloyd, M. A. Saunders, and B. O. Palsson, "DynamicME: dynamic simulation and refinement of integrated models of metabolism and protein expression," *BMC Syst Biol,* vol. 13, no. 1, p. 2, Jan 9 2019, doi: 10.1186/s12918-018-0675-6.
[216] A. P. Burgard, P. Pharkya, and C. D. Maranas, "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnol Bioeng,* vol. 84, no. 6, pp. 647-57, Dec 20 2003, doi: 10.1002/bit.10803.
[217] H. V. Dinh, Z. A. King, B. O. Palsson, and A. M. Feist, "Identification of growth-coupled production strains considering protein costs and kinetic variability," *Metab Eng Commun,* vol. 7, p. e00080, Dec 2018, doi: 10.1016/j.mec.2018.e00080.
[218] D. Ma and M. A. Saunders, "Solving Multiscale Linear Programs Using the Simplex Method in Quadruple Precision," in *Numerical Analysis and Optimization*, (Springer Proceedings in Mathematics & Statistics, 2015, ch. Chapter 9, pp. 223-235.
[219] D. Ma, L. Yang, R. M. Fleming, I. Thiele, B. O. Palsson, and M. A. Saunders, "Reliable and efficient solution of genome-scale models of Metabolism and macromolecular Expression," *Sci Rep,* vol. 7, p. 40863, Jan 18 2017, doi: 10.1038/srep40863.
[220] L. Yang, D. Ma, A. Ebrahim, C. J. Lloyd, M. A. Saunders, and B. O. Palsson, "solveME: fast and reliable solution of nonlinear ME models," *BMC Bioinformatics,* vol. 17, no. 1, p. 391, Sep 22 2016, doi: 10.1186/s12859-016-1240-1.
[221] A. Ebrahim *et al.*, "Multi-omic data integration enables discovery of hidden biological regularities," *Nat Commun,* vol. 7, p. 13091, Oct 26 2016, doi: 10.1038/ncomms13091.
[222] D. Heckmann *et al.*, "Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models," *Nat Commun,* vol. 9, no. 1, p. 5252, Dec 7 2018, doi: 10.1038/s41467-018-07652-6.
[223] S. W. Seo, D. Kim, H. Latif, E. J. O'Brien, R. Szubin, and B. O. Palsson, "Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli," *Nat Commun,* vol. 5, p. 4910, Sep 15 2014, doi: 10.1038/ncomms5910.
[224] S. W. Seo, D. Kim, E. J. O'Brien, R. Szubin, and B. O. Palsson, "Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in Escherichia coli," *Nat Commun,* vol. 6, p. 7970, Aug 10 2015, doi: 10.1038/ncomms8970.
[225] K. Chen, Y. Gao, N. Mih, E. J. O'Brien, L. Yang, and B. O. Palsson, "Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation," *Proc Natl Acad Sci U S A,* vol. 114, no. 43, pp. 11548-11553, Oct 24 2017, doi: 10.1073/pnas.1705524114.
[226] A. Bulovic *et al.*, "Automated generation of bacterial resource allocation models," *Metab Eng,* vol. 55, pp. 12-22, Sep 2019, doi: 10.1016/j.ymben.2019.06.001.
[227] C. J. Lloyd *et al.*, "COBRAme: A computational framework for genome-scale models of metabolism and gene expression," *PLoS Comput Biol,* vol. 14, no. 7, p. e1006302, Jul 2018, doi: 10.1371/journal.pcbi.1006302.
[228] E. Brunk *et al.*, "Recon3D enables a three-dimensional view of gene variation in human metabolism," *Nat Biotechnol,* vol. 36, no. 3, pp. 272-281, Mar 2018, doi: 10.1038/nbt.4072.
[229] H. M. Berman *et al.*, "The Protein Data Bank," *Nucleic Acids Res,* vol. 28, no. 1, pp. 235-242, 2000.
[230] J. R. Karr *et al.*, "A whole-cell computational model predicts phenotype from genotype," *Cell,* vol. 150, no. 2, pp. 389-401, Jul 20 2012, doi: 10.1016/j.cell.2012.05.044.
[231] "IBM ILOG CPLEX.
V13. 1: User's Manual for CPLEX," *International Business Machines Corporation,* vol. 46, no. 53, p. 157, 2009.
[232] L. Gurobi Optimization, "Gurobi Optimizer Reference Manual," 2020. [Online]. Available: "http://www.gurobi.com.
[233] T. Achterberg, "SCIP: solving constraint integer programs," *Mathematical Programming Computation,* vol. 1, no. 1, pp. 1-41, 2009, doi: 10.1007/s12532-008-0001-1.
[234] A. M. Gleixner, D. E. Steffy, and K. Wolter, "Iterative Refinement for Linear Programming," *ZIB Report,* pp. 1-98, 2015.

[235] A. Bar-Even *et al.*, "The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters," *Biochemistry,* vol. 50, no. 21, pp. 4402-10, May 31 2011, doi: 10.1021/bi2002289.
[236] N. Le Novere *et al.*, "Minimum information requested in the annotation of biochemical models (MIRIAM)," *Nat Biotechnol,* vol. 23, no. 12, pp. 1509-15, Dec 2005, doi: 10.1038/nbt1156.
[237] M. V. Schneider, Ed. *In Silico Systems Biology* (Methods in Molecular Biology). Humana Press, 2013.
[238] E. Demir *et al.*, "The BioPAX community standard for pathway data sharing," *Nat Biotechnol,* vol. 28, no. 9, pp. 935-42, Sep 2010, doi: 10.1038/nbt.1666.
[239] S. M. Wimalaratne *et al.*, "Uniform resolution of compact identifiers for biomedical data," *Sci Data,* vol. 5, p. 180029, May 8 2018, doi: 10.1038/sdata.2018.29.
[240] N. M. Gizzatkulov, I. I. Groyanin, E. A. Metelkin, E. A. Mogilevskaya, K. V. Peskov, and O. V. Demin, "DBSolve Optimum: a software package for kinetic modeling which allows dynamic visualization of simulation results," *BMC Syst Biol,* vol. 4, no. 109, pp. 1-11, 2010.
[241] F. T. Bergmann *et al.*, "COPASI and its applications in biotechnology," *J Biotechnol,* vol. 261, pp. 215-220, Nov 10 2017, doi: 10.1016/j.jbiotec.2017.06.1200.