

## REVIEW

# Flux balance analysis-based metabolic modeling of microbial secondary metabolism: Current status and outlook

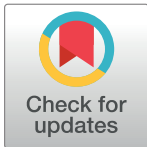
Sizhe Qiu<sup>1,2</sup>, Aidong Yang<sup>2</sup>, Hong Zeng<sup>1\*</sup>

**1** School of Food and Health, Beijing Technology and Business University, Beijing, China, **2** Department of Engineering Science, University of Oxford, Oxford, United Kingdom

\* [zenghong@btbu.edu.cn](mailto:zenghong@btbu.edu.cn)

## Abstract

In microorganisms, different from primary metabolism for cellular growth, secondary metabolism is for ecological interactions and stress responses and an important source of natural products widely used in various areas such as pharmaceuticals and food additives. With advancements of sequencing technologies and bioinformatics tools, a large number of biosynthetic gene clusters of secondary metabolites have been discovered from microbial genomes. However, due to challenges from the difficulty of genome-scale pathway reconstruction and the limitation of conventional flux balance analysis (FBA) on secondary metabolism, the quantitative modeling of secondary metabolism is poorly established, in contrast to that of primary metabolism. This review first discusses current efforts on the reconstruction of secondary metabolic pathways in genome-scale metabolic models (GSMMs), as well as related FBA-based modeling techniques. Additionally, potential extensions of FBA are suggested to improve the prediction accuracy of secondary metabolite production. As this review posits, biosynthetic pathway reconstruction for various secondary metabolites will become automated and a modeling framework capturing secondary metabolism onset will enhance the predictive power. Expectedly, an improved FBA-based modeling workflow will facilitate quantitative study of secondary metabolism and in silico design of engineering strategies for natural product production.



## OPEN ACCESS

**Citation:** Qiu S, Yang A, Zeng H (2023) Flux balance analysis-based metabolic modeling of microbial secondary metabolism: Current status and outlook. *PLoS Comput Biol* 19(8): e1011391. <https://doi.org/10.1371/journal.pcbi.1011391>

**Editor:** Christos A. Ouzounis, CPERI, GREECE

**Published:** August 24, 2023

**Copyright:** © 2023 Qiu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors declare that there is no conflict of interests.

## 1. Introduction

The cellular metabolism in microorganisms is the set of biochemical reactions to maintain life, and it can be divided into 2 branches, primary and secondary metabolisms: primary metabolism generates energy and synthesizes cellular biomass, while secondary metabolism mediates microorganisms' adaptation to the living environment [1]. Secondary metabolites have been defined as metabolites that are unessential for growth and reproduction [2–5], but they usually have specialized functions and the biosynthesis of them is sometimes species specific. Though secondary metabolites have little contribution to cell growth from the simple mass balance point of view, they are ecologically and evolutionarily important as they often contribute to inter-species antagonistic or mutualistic interactions [6] and responses to environmental stresses [7–9].

As a rich source of valuable natural products, microbial secondary metabolites have been contributing to pharmaceutical, cosmetic, food, and agricultural industries since the discovery of penicillin [10], for example, cyclic lipopeptides derived from *Bacillus subtilis* used as anti-bacterial agents in both medicine and agriculture [11], lactic acid bacteria-derived exopolysaccharides (EPSs) used as food additives [12], and sacran produced by cyanobacteria for skincare [13]. Compared to plant-based processes, microbial bioproduction processes are more efficient and implementable, as the doubling time of a microbial cell is usually 20 to 60 min [14], much shorter than that of a plant cell, which typically takes days [15]. In addition, a microbial system is more controllable than a plant system, as the latter is more subject to fluctuating environmental factors [16]. There have been many successful applications to optimize the microbial production of secondary metabolites, e.g., through eliminating competing pathways, using genome shuffling to obtain a high-yield strain [17], or fine-tuning gene expression by promoter engineering [18]. Overall, microorganisms are ideal platforms to produce secondary metabolites, due to the richness of valuable natural products and inherent advantages as cell factories.

Data-driven computational modeling has become widely used in the quantitative study of the synthetic biology of secondary metabolism [19] due to the fast growing amount of multiomics data and computational resources [20,21]. Two most commonly used modeling methods of cellular metabolism are differential equation-based models [22,23] and flux balance analysis (FBA)-based models [24]. Admittedly, differential equation-based models can be easily connected with process control, but most of them treat the whole metabolic network as a “black box” [25,26]. On the other hand, fine-grained “white-box” differential equation-based models [23] are typically costly to construct due to limited data availability [19,27] and complex enzyme kinetic mechanisms [28]. Alternatively, the reconstruction of a genome-scale metabolic model (GSMM) is fast and mostly automated [29], and the major analytical approach of a GSMM, i.e., FBA using linear programming, is less computationally expensive than solving a large-scale differential equation set [30,31]. Therefore, this review focuses on FBA-based modeling techniques and GSMMs integrated with secondary metabolic pathways (smGSMMs).

In recent years, utilizing FBA to model and engineer secondary metabolism has gained significant interest in synthetic biology, especially in the production of antibiotics [32]. Consequently, review articles on this topic have emerged: Breitling and colleagues focused on modeling challenges caused by incomplete and uncertain information and suggested quantitative consideration of parameter uncertainties [19]; Mohite and colleagues provided an overview of various smGSMMs for actinomycetes’ and modeling techniques applied for the production of antibiotics [32]; Weber and Kim summarized tools for searching secondary metabolites’ biosynthetic gene clusters (BGCs) and smGSMM reconstruction [21]. Nevertheless, existing reviews cover either BGC identification which is only part of the modeling process or modeling techniques for a specific class of secondary metabolites; therefore, this review intends to elucidate this topic more holistically and generically via presenting the current progress of the complete modeling process, i.e., from building an smGSMM for target secondary metabolites to predicting secondary metabolite production flux within the scope of FBA-based modeling techniques. We begin with compiling recent developments on the reconstruction of secondary metabolic pathways, which are necessary for building smGSMMs. Then, we move on to discuss FBA-based modeling techniques for predicting secondary metabolite production, where various modeling techniques that manage to capture secondary metabolism are critically reviewed. Finally, suggestions are given to potential extensions of FBA to improve the prediction accuracy of secondary metabolite production. Unlike an early relevant work that discusses how to incorporate parameter uncertainties that result from incomplete information

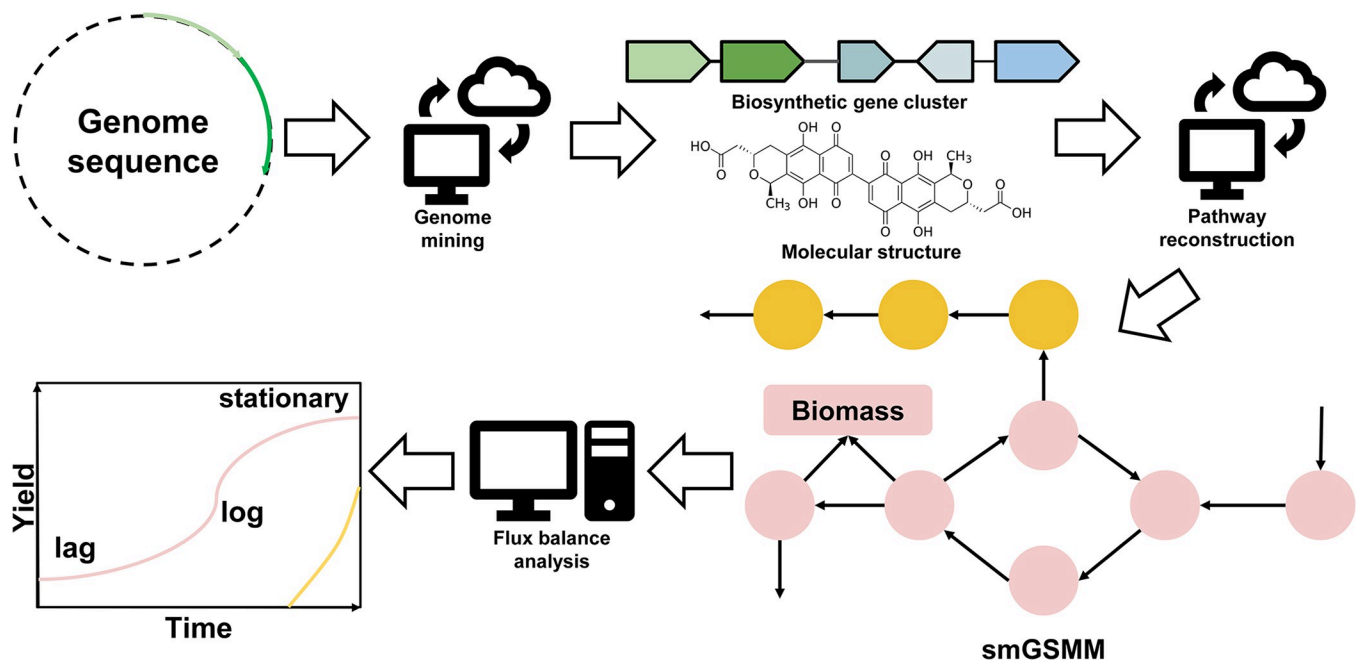
of secondary metabolism [19], this review directs attention to approaches that manage to quantitatively capture secondary metabolite production with pathway reconstruction and FBA-based modeling techniques developed in recent years.

## 2. Pathway reconstruction for microbial-derived secondary metabolites

To model secondary metabolism, 3 key steps are needed: (1) the identification of secondary metabolites and associated BGCs through genome mining; (2) reconstruction of secondary metabolic pathways; and (3) the simulation of metabolic fluxes (Fig 1). The first 2 steps are the basis of the third step. For BGC identification, a number of well-developed genome mining tools that can find BGCs for the biosynthesis of secondary metabolites are available, such as antiSMASH [33], PRISM [34], or BAGEL [35]. The development of genome mining tools in recent years has been reviewed by many well-written articles [21,36–41], from which interested readers can find in-depth information. Compared to genome mining tools applicable to finding BGCs of secondary metabolites, commonly used metabolic network reconstruction tools show limitations in assembling the biosynthetic pathways of secondary metabolites, yet there exists a few tools capable of addressing such a challenge [42,43].

### 2.1. Limitations of current genome-scale metabolic network reconstruction tools on assembling secondary metabolic pathways

The computational modeling of secondary metabolism requires in silico biosynthetic pathway reconstruction of secondary metabolites. Commonly used automated GSM reconstruction tools are CarveMe [44], ModelSEED [45], RAVEN [46], Merlin [47], AutoKEGGRec [48], AuReMe [49], MetaDraft [50], Pathway Tools [51], FAME [52], and GEMSiRV [53]. They rely on metabolic reaction databases: BiGG [54], SEED [45], MetaCyC [55], and KEGG [56].



**Fig 1. Graphical abstract of the workflow of FBA-based metabolic modeling for secondary metabolism.** Orange circle: secondary metabolite; pink circle: primary metabolite. FBA, flux balance analysis; smGSMM, genome-scale metabolic model with secondary metabolic pathway.

<https://doi.org/10.1371/journal.pcbi.1011391.g001>

Among them, BiGG and SEED both integrate and standardize genome annotations, reactions, and GSMMs of different microorganisms, but leave many gaps in peripheral pathways associated with secondary metabolites [44]. MetaCyc contains 747 secondary metabolic pathways [57], but most of them are plant specific [58,59]. Reconstruction of 24 penicillium GSMMs by Prigent and colleagues reveals that MetaCyc allows automatic reconstruction of 34 biosynthetic pathways of terpenoids, polyketides (PKs), phenylpropanoids, alkaloids, and nonribosomal peptides (NRPs) [57]. Compared to MetaCyc, KEGG has significantly depleted “Metabolism Of Terpenoids And Polyketides” and enriched “Alkaloid And Other Secondary Metabolite Biosynthesis” pathway classes [60]. Despite advances in smGSMMs of penicillium [57] and actinomycetes [32], the incomplete description of species-specific secondary metabolism in databases still makes it hard to reconstruct secondary metabolic pathways based on genome annotation alone without supplementary experimental information [19,61]. In short, among the currently available genome-scale metabolic network reconstruction tools, it is difficult to recommend a state of the art one for reconstructing complete biosynthetic pathways of secondary metabolites.

Given the issues presented in automated pathway reconstruction tools, manual curation is sometimes used as an expedient solution, e.g., the paulomycin pathway in the smGSMM of *Streptomyces albus* J1074 [62], the lumped biosynthetic reaction of EPS from monosaccharides in *Lactobacillus casei* LC2W [63]. However, manual curation is laborious, inefficient, and subject to human error. Most manual approaches are not fine-grained enough, which means intermediary metabolites are left out. As a result, smGSMMs are incapable of accounting for bottlenecks that affect secondary metabolite production, such as precursor depletion, poor enzyme capacity of intermediate step, or toxic intermediate accumulation [19,64]. The limited performance calls for pathway reconstruction tools specialized for secondary metabolism. Fortunately, there are still some tools for automated pathway reconstruction of certain secondary metabolites, which were developed via bottom-up (BGC-based) or top-down (retrosynthesis-based) approaches (Table 1).

## 2.2. BGC-based pathway reconstruction

Firstly, the bottom-up approach, namely the BGC-based approach, reconstructs secondary metabolic pathways from identified BGCs. Genome mining tools such as antiSMASH identify BGCs of secondary metabolites and output genbank files containing annotated functional domains. Subsequently, reconstruction tools such as BiGMeC [42] and DDAP [65] take identified BGCs as inputs and assemble reactions from template smGSMMs (e.g., Sco-GEM for *Streptomyces coelicolor* [66]) or pre-curated pathway databases (e.g., DDPA database for type I PKS pathways [65]) based on annotated genes in BGCs. The BGC-based method reconstructs

**Table 1. Automated pathway reconstruction tools for microbial secondary metabolism.**

Tool	Scope	Input	Output	Reference
BiGMeC	PKs, NRPs	Genbank files of BGCs from antiSMASH	Json files containing reconstructed pathways	[42]
DDAP	Type I PK synthase	Fasta/genbank/csv files of polyketide synthase sequences	A list of possible pathways and the associated product SMILES strings	[65]
RetroPath 2.0	All classes of secondary metabolites	Source and sink compounds' SMILES strings and reaction rules	A reaction network linking the source set to the sink set	[67]
BioNavi-NP	All classes of secondary metabolites	Products' SMILES strings and reaction rules	Possible precursors and biosynthetic pathways	[68]

BGC, biosynthetic gene cluster; NRP, nonribosomal peptide; PK, polyketide.

<https://doi.org/10.1371/journal.pcbi.1011391.t001>

genome-scale biosynthetic pathways of secondary metabolites with sufficient accuracy; however, its scope is limited, i.e., the reconstruction usually works for PKs and NRPs only [42], because genome mining tools like antiSMASH or PRISM sometimes cannot detect BGCs of other secondary metabolites, such as EPSs or fatty acid derivatives [40]. Also, the pre-curated databases and template smGSMMs have low coverage of biosynthetic pathways of secondary metabolites other than PKs or NRPs.

### 2.3. Retrosynthesis-based pathway reconstruction

Unlike reconstructing pathways from BGCs, a “top-down” approach predicts pathways based on retrosynthesis of end products. Tools like **RetroPath 2.0** [67] or **BioNavi-NP** [68] can predict reactions to produce target secondary metabolites using the retrosynthesis algorithm. This algorithm uses a dataset of curated reactions from databases, such as SimPheny [69] or BNICE [70], as the input to generate reaction rules, which are used to decompose the end product’s molecular structure into possible precursors [71]. The final output is a route linking precursors to the end product, in which all reactions are essential for their viability [67]. For example, BioNavi-NP decomposes Sterhirsutin J into potential candidate precursors, such as colletorin D acid, and those candidates’ biosynthesis are further traced back to simple building blocks (e.g., malonyl CoA), finally outputting most possible biosynthetic pathways [68]. The advantage of retrosynthesis is that the pathway reconstruction is no longer restricted to certain classes of secondary metabolites, as its pathway prediction is based on reaction rules and the molecular structure of the end product. However, retrosynthesis might give multiple solutions and the suggested solutions are not always consistent with the actual genome of the studied microbe, since reaction rules from databases are not species specific. Therefore, for better reconstruction accuracy, users need to adjust either the list of reaction rules or the reconstructed pathway based on the organism’s protein sequences, before making predictions.

In short, to build smGSMMs, the practical approach seems to be first reconstructing secondary metabolic pathways with automated methods, and then integrating them to the primary metabolic network. For PKs and NRPs, BGC-based tools (BiGMeC and DDAP) are advantageous, while retrosynthesis can predict the biosynthetic pathways for other secondary metabolites.

## 3. Towards predicting microbial production of secondary metabolites: FBA-based modeling techniques

### 3.1. Challenges faced by flux balance analysis to predict secondary metabolite production

FBA models cellular metabolism by utilizing linear programming to compute metabolic fluxes, optimizing an objective function, usually biomass formation ( $v_{growth}$ ) or a tailored objective ( $c_{1 \times R} * [v_1, v_2, \dots, v_R]^T$ ,  $c_{1 \times R}$  is a vector of coefficients, T stands for matrix transpose) (Eq 1),  $v$  stands for the reaction flux. The linear programming is solved in a constrained solution space of mass conservation (Eq 2) and upper/lower bounds of reaction fluxes,  $v_{i,max}$  and  $v_{i,min}$  (Eq 3) [24]. Eq 2 is based on the pseudo-steady state assumption, and hence, FBA is often applicable only for modeling different stabilized steady states but not transient phases where intracellular metabolite concentrations will change [72].

$$\text{Maximize objective} = v_{growth} \text{ or } c_{1 \times R} * [v_1, v_2, \dots, v_R]^T \quad (1)$$

$$\text{Mass conservation} : S_{M \times R} * [v_1, v_2, \dots, v_R]^T = 0 \quad (2)$$

$$\text{Upper/lower bounds : } v_{i,\min} \leq v_i \leq v_{i,\max} \quad (3)$$

$v_{growth}$  is the biomass formation rate normalized to 1 gram dry weight, also considered as the growth rate.  $S_{M \times R}$  is the stoichiometric matrix of the metabolic network with  $M$  metabolites and  $R$  reactions. A lot of toolboxes have been developed for conducting FBA, such as COBRA Toolbox [73], MetaFlux [74], and FBA-SimVis [75], and it has become increasingly common to use FBA to quantitatively study cellular metabolism, e.g., elucidation of metabolic responses to different culture conditions [76,77], identification of gene knockout to improve metabolite production [78], prediction of metabolite cross-feeding in microbial communities [79].

Although FBA has been successful in predicting microbial growth and primary metabolism, its application in predicting secondary metabolite production faces challenges from several different aspects. The conventional FBA is tailored to simulate the growth and metabolism of microbes in the exponential phase, which is not ideal for simulating secondary metabolites' biosynthesis that typically occurs in the stationary phase or late growth phase [32,72]. Ergo, the first apparent challenge is the unsuitable objective function. Generally, the maximization of biomass synthesis is used as the objective function, based on the assumption that cells use the available nutrients most efficiently for biomass synthesis [80]. Consequently, the metabolic fluxes involved in the biosynthesis of secondary metabolites, which are not essential components of biomass, are often neglected. Therefore, selecting a suitable objective function is crucial when simulating the metabolic switch from primary metabolism to secondary metabolism using FBA [32]. Another inherent shortcoming of FBA is its lack of characterization of regulation. The metabolic switch is controlled by gene expression regulation, and therefore, the prediction of secondary metabolite production will become more accurate if FBA incorporates the gene regulatory network (GRN) [81]. Besides, the overlook of environmental stresses (such as temperature [7,82], pH [83], nutrient availability [84,85], etc.), which impact the expression of enzymes in secondary metabolic pathways, also limits the accuracy of predictions. The optimality principle used to predict microbial metabolism should keep the balance between biomass formation and minimization of death under stress conditions, but the latter is ignored in conventional FBA [72]. Without incorporating important cellular activities such as stress response, FBA would not be able to predict nongrowth associated metabolic fluxes.

### 3.2. Existing FBA-based modeling techniques for secondary metabolism

To address the limitations of conventional FBA on secondary metabolism discussed in Section 3.1, several techniques have been developed, including: (1) adding the targeted secondary metabolite into the growth function (biomass formation); (2) switching from the classical biomass formation objective to a secondary metabolism-associated objective; (3) estimating metabolic fluxes through sampling in a strictly constrained space; and (4) integrating gene expression data in FBA. The first 3 techniques aim to overcome the limitation caused by the unsuitable objective function, while the last one manages to incorporate gene expression regulation. The existing modeling techniques and their applications are summarized in [Table 2](#), and their predictive power is also discussed.

Firstly, the flux distribution obtained from conventional FBA highly depends on the specific objective function used. If maximizing biomass formation is selected as the objective function, as commonly done, FBA would predict zero flux through the secondary metabolic pathways as it does not contribute to cell growth [86]. To resolve this type of limitation, it has been proposed to add the secondary metabolite into the biomass formation reaction, so that FBA will optimize both the synthesis of biomass precursors and the target secondary metabolite. This

Table 2. Summary of different modeling techniques to predict secondary metabolite production.

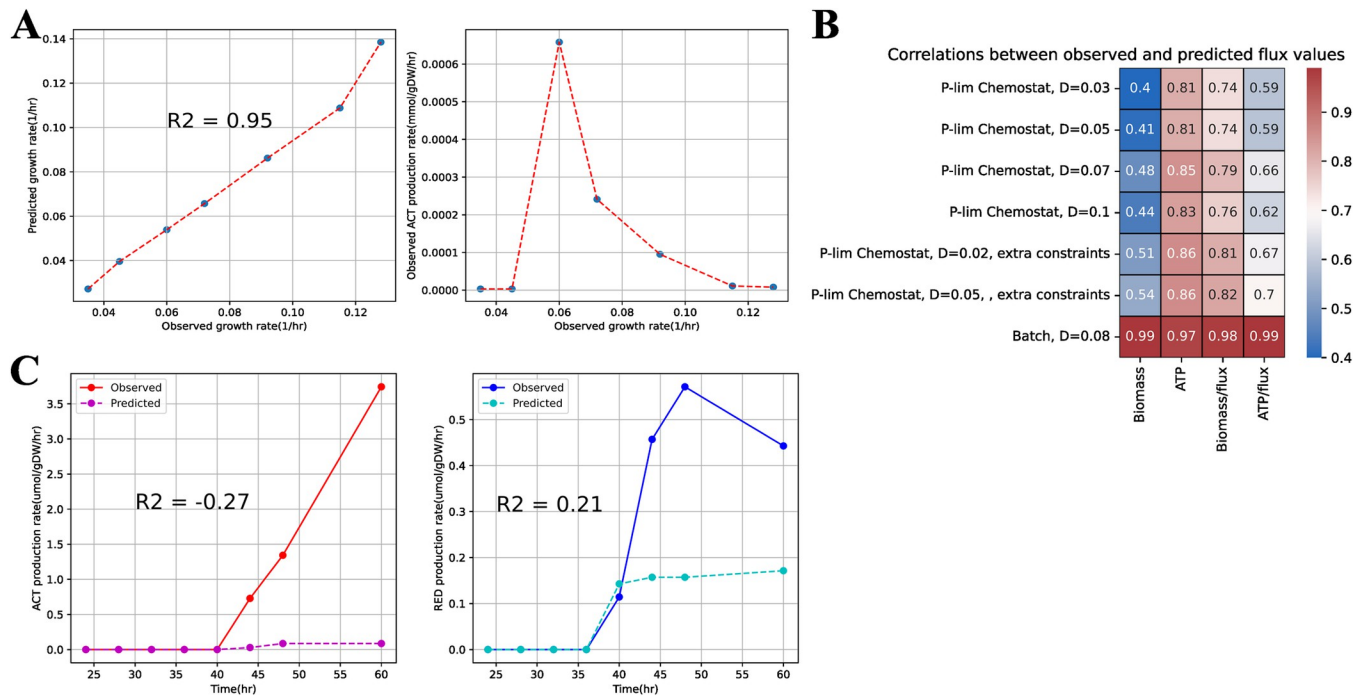
Secondary metabolite	Organism	Modeling technique	Predictive power	Reference
ACT, RED	<i>Streptomyces coelicolor</i>	Add the secondary metabolite into biomass and maximize biomass formation	Good accuracy but restricted by settings	[87,88]
Clavulanic acid	<i>Streptomyces clavuligerus</i>	Use ATP yield as the objective function	Good accuracy and unrestricted by settings	[89]
Riboflavin	<i>Ashbya gossypii</i>	Use secondary metabolite yield as the objective function	Good accuracy and unrestricted by settings	[90]
Acetoin, diacetyl, acetaldehyde, benzaldehyde, 2,3-butanediol, and amino acid-derived flavor metabolites, 2-methylbutanal, 2-methylpropanoic acid, etc.	<i>Lactococcus lactis</i> subsp. <i>cremoris</i> , <i>Lactococcus lactis</i> subsp. <i>lactis</i> , <i>Streptococcus thermophilus</i> , <i>Leuconostoc mesenteroides</i>	Maximize secondary metabolite yield in the solution space of FVA	Not quantitatively examined	[92]
ACT, RED, CDA, Cpk	<i>Streptomyces coelicolor</i>	Random sampling with enzyme constraints	Qualitatively accurate but not quantitatively examined	[66]
ACT, RED	<i>Streptomyces coelicolor</i>	Integration of transcriptomics data into FBA	Low quantitative accuracy	[97]

ACT, actinorhodin; CDA, calcium dependent antibiotics; Cpk, coelimycin P1; FBA, flux balance analysis; FVA, flux variability analysis; RED, undecylprodigiosin.

<https://doi.org/10.1371/journal.pcbi.1011391.t002>

modeling technique was used to simulate the antibiotic production in *Streptomyces coelicolor*: based on the observed antibiotics production rate, the amount of antibiotics was increased dynamically when the cell transitioned from primary metabolism to secondary metabolism [87,88]. Because stoichiometric coefficients of undecylprodigiosin (RED) and actinorhodin (ACT) in biomass formation objective function were set based on experimental measurement, the accuracy in predicting the growth rate, R-squared value = 0.95, could reflect the accuracy of predicting secondary metabolite production (Fig 2A). This technique could account for the metabolic switch, as the studies by Alam and colleagues [87] and Amara and colleagues [88] demonstrated significant correlations between gene expression levels and predicted metabolic fluxes for most genes in secondary metabolic pathways. However, anti-correlations of predicted fluxes and gene expression levels for some genes in calcium-dependent antibiotics (CDAs) and RED biosynthesis indicated that the failure of including regulatory constraints was an important source of error [87,88]. From the perspective of modeling technique, in this case, FBA still optimizes the growth rate and makes the production of secondary metabolites growth-associated, which contradicts with the biological fact that the biosynthesis of most secondary metabolites is growth-unassociated. Also, this modeling technique manually fixes the ratio between secondary metabolite production and biomass formation (i.e., product yield per unit biomass) with experimental measurements, which restricts its applicability for cases where the yield of the secondary metabolite is varied or unknown. Due to the lack of a solid biological basis, this approach can only model flux distributions if the microbial cell behaves according to the artificially adjusted objective function. Though such an approach can, to some extent, reflect statuses of primary and secondary metabolisms, it can hardly be used directly for design purposes, because it cannot predict secondary metabolite production for different design settings of interest (e.g., different growth media, different strains).

As an alternative to the manipulation of the biomass formation while still adopting the classic objective function that maximizes biomass growth, switching to a new objective function can avoid introducing inappropriate growth association. The commonly used candidates of alternative objective functions are ATP yield, ATP yield per unit flux, biomass formation per unit flux, and target secondary metabolite production. Toro and colleagues compared 4 different objective functions' performances on both primary metabolism and clavulanic acid



**Fig 2. Quantitative assessment of existing FBA-based modeling techniques for predicting secondary metabolite production.** (A) Comparison of predicted and observed growth rates and ACT production fluxes at different observed growth rates in Alam and colleagues [87]. (B) Comparison of 4 different objective functions used in FBA to predict both primary metabolism and clavulanic acid production flux [89]. Correlation scores are computed for predicted and observed fluxes. P-lim: limited phosphorus content. (C) Comparison of predicted and observed production fluxes of ACT and RED in Kim and colleagues [97]. ACT, actinorhodin; FBA, flux balance analysis.

<https://doi.org/10.1371/journal.pcbi.1011391.g002>

production of *Streptomyces clavuligerus* [89], and the study found that maximization of ATP yield could predict primary metabolism and clavulanic acid production flux with best accuracy (Fig 2B). ATP yield maximization can be assumed to be the objective for microorganisms in both exponential growth and stationary phases, as the cell is always in need of energy to maintain its activity. A more direct objective function is to maximize the target secondary metabolite production. In the simulation of riboflavin overproduction by *Ashbya gossypii*, Ledesma-Amaro and colleagues assumed that *Ashbya gossypii* in the stationary phase switched from maximizing biomass formation entirely to maximizing riboflavin production [90]. The result showed that the simulated value of riboflavin production rate, 0.0156 mmol/gDW/h [90], was close to the experimental value, 0.0126 mmol/gDW/h [91]. Using flux variability analysis (FVA) to maximize the target secondary metabolite production from the flux solutions that maximize biomass growth is another approach, which has been used in modeling the production of flavor metabolites, such as acetoin, in lactic acid bacteria [92], but this approach has not been quantitatively examined with experimental data. The mathematical formulation of secondary metabolism-associated objective function in place of biomass formation has shown its ability to predict the production of some secondary metabolites, independently of condition-specific experimental data. However, changing the objective function may be too simplistic to account for metabolic activities that keep the balance of growth and minimization of death (stress resistance) as discussed in Section 3.1.

Since the FBA solution can be inaccurate with an unsuitable objective function, a modeling technique that is independent of objective functions has been adopted to resolve the issue. In the simulation using the Sco-GEM model, a consensus smGSMM for *Streptomyces coelicolor*,



the metabolic fluxes were strictly constrained with the enzyme capacity,  $v \leq k_{cat}[E]$ .  $k_{cat}$  is the turnover rate and  $[E]$  is the concentration of the enzyme. Fluxes were approximated using unbiased random sampling in the constrained solution space instead of optimizing a defined objective function [66]. The metabolic switch from glycolysis, fatty acid, and nucleotide biosynthesis to ACT, RED, CDA, and coelimycin P1 (Cpk) biosynthesis was captured by this technique [66], but no comparison of predicted and observed reaction fluxes was performed for quantitative validation. Because random sampling without a highly constrained space will output multiple flux states with wide ranges, the accuracy of the predicted fluxes relies on the strict constraints imposed by quantified enzyme protein concentrations and enzyme kinetic parameters, which are strain specific and condition specific. Therefore, this modeling technique cannot compute metabolic fluxes through secondary metabolic pathways without proteomic data measured under a given setting. The cost of proteomics quantification makes it hard to predict secondary metabolite production with different growth conditions.

Modeling techniques discussed above do not take into account the inherent lack of characterizing regulation in conventional FBA. Reactions in a GSMM, if not gap filled, have associated gene-protein-reaction (GPR) rules, allowing the integration of quantitative gene expression data, such as transcriptomics or proteomics. If gene expression levels are integrated into FBA, the predicted fluxes can reflect metabolic enzyme activities and thus characterize regulations. The generic algorithm is to minimize the utilization of reactions with low gene expression levels and allow reactions with high gene expression levels to have higher absolute flux values.

There have been various developed techniques for integrating gene expression data into FBA, e.g., GIMME [93], E-flux [94], iMAT [95], but most were applied only on primary metabolism [96]. However, there exists a transcriptomics-based strain optimization tool for secondary metabolite production that uses iMAT to predict the secondary metabolism of *Streptomyces coelicolor* [97]. The comparison of different gene expression integrated FBA modeling techniques, iMAT was found to be the only technique that can capture the onset of secondary metabolism in *Streptomyces coelicolor* [97]. Nonetheless, the discrepancy between predicted and observed production fluxes of ACT and RED indicates that iMAT only performs well in a qualitative manner (Fig 2C). Also, this technique relies on condition-specific gene expression data, limiting the scope of extrapolation. It can predict metabolism for various design settings of interest only if different gene expression data corresponding to the different settings are provided. Otherwise, this technique, to a large extent, can merely compute the flux distribution corresponding to a specific set of gene expression levels. In short, to our knowledge, so far there has not been a gene expression integrated modeling technique that predicts secondary metabolite production with quantitative accuracy.

In a nutshell, existing FBA-based modeling techniques that manipulate the objective function cannot capture 2 distinct life objectives of a microbial cell concurrently, i.e., maximizing growth and minimizing death [72]. Besides, how to formulate a mathematical model that keeps the balance between the 2 objectives still remains an open question. Furthermore, no mechanistic representation of gene expression regulation upstream to metabolism has been built and integrated in FBA to make better use of omics datasets than simply determining “on” and “off” of reactions for fixed settings. In summary, for existing FBA-based modeling techniques that predict secondary metabolite production, those without integration of gene expression data are either limited to specific cases or too simplistic that only work under certain ideal assumptions, while gene expression integrated techniques are scarce and unable to make predictions for different settings without costly data generation.

### 3.3. Potential extensions of FBA towards predicting secondary metabolite production

The associated GPR rules, already targeted by some of the existing approaches to predict the secondary metabolite production as summarized in **Section 3.2**, suggests that the ability to determine the activity of enzymes in secondary metabolic pathways could be the key to the accuracy of modeling the secondary metabolite production. Here, this review proposes 2 potential extensions of FBA that may resolve issues of existing modeling techniques discussed in Section 3.2.

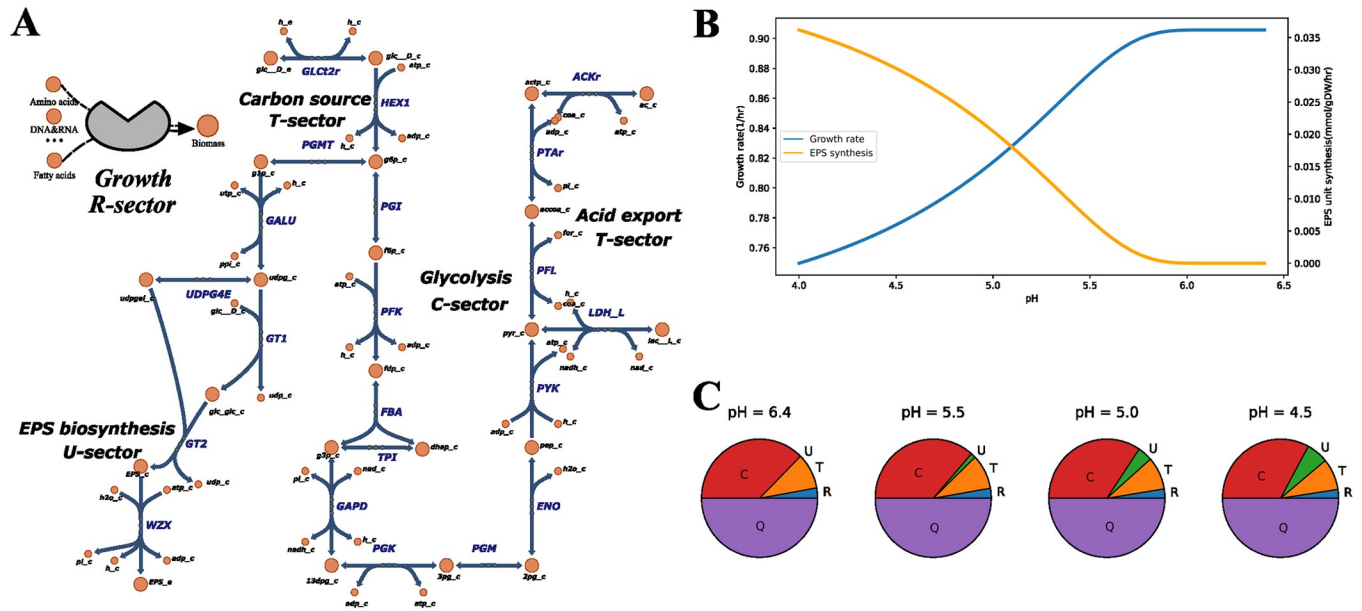
#### 3.3.1. Constrained proteome allocation for secondary metabolism

Proteome allocation, namely the distribution of proteome resources in different pathways, governs cellular metabolism by controlling maximum reaction fluxes [98]. The constrained proteome allocation theory has been proposed to model the competition for proteome resources among different functional sectors such as catabolism, anabolism, transportation, etc., characterizing the coordination of proteome partitions and cellular metabolism under different growth conditions [99–101]. The theoretical model has been applied in various FBA-based models, such as CAFBA [102] or ME-model [103], via converting it into proteomic constraints on reaction fluxes. There are different ways to divide the proteome based on the modeling requirement. For example, the proteome was divided into fermentation, respiration, and biomass formation sectors in the prediction of overflow metabolism in *Escherichia coli* [104]. Usually, glycolytic enzymes are clustered as the sector of catabolism or energy, membrane transporter proteins are clustered as the sector of transportation, and a lumped proteome resource for the growth function (biomass formation) is considered the sector of anabolism [102,105,106]. The neglect of other enzymes' proteome costs normally will not significantly affect the simulation, as fluxes through those pathways are far smaller than that through central carbon metabolism [102,105,106].

Though constrained proteome allocation-embedded FBA models have achieved good accuracy in case studies such as predicting overflow metabolism in *E. coli* [105] or explaining lactic acid production in lactic acid bacteria [106], to our knowledge it has not been used to predict the secondary metabolite production yet. The proteome partitioning has not included a sector defined for secondary metabolism, as we found in various proteome allocation models. To apply proteome allocation to predict secondary metabolite production, one potential approach is integrating synthetic chemostat model (SCM) into FBA [72]. SCM is a differential equation-based model of microbial growth kinetics that divides cellular metabolism into a P-component for growth and a U-component for stress resistance [107]. Being a macroscopic bioreactor-level model, SCM is too coarse to characterize different functional proteome sectors, but the embedded concept can be used to modify the original constrained proteome allocation framework. If a U-sector containing enzymes in biosynthetic pathways of secondary metabolites is included and stress response is introduced into proteome allocation, then the modified model might be able to characterize both growth and stress response via proteome allocation at the metabolic switch that leads to secondary metabolite production. Below, an illustrative example of pH-induced EPS production in lactic acid bacteria [108,109] is presented for the proposed constrained proteome allocation model for both branches of metabolism (**Fig 3**).

$$\phi_Q(50\%) + \phi_U + \phi_C + \phi_R + \phi_T \leq 1 \quad (4)$$

$$\frac{\phi_U}{\phi_U + \phi_C + \phi_R + \phi_T} \geq k_1 e^{\frac{-k_2}{(6.5-pH)^2}}, k_1, k_2 > 0 \quad (5)$$



**Fig 3. An illustrative example of pH-induced EPS production in lactic acid bacteria, used to explain the constrained proteome allocation model for both primary and secondary metabolism.** (A) The metabolic network of lactic acid bacteria for both primary metabolism (C, R, T sectors) and secondary metabolism (U sector). (B) Simulated metabolic response to pH: the increase of acidity inhibits the growth rate and induces EPS production. (C) Simulated proteome allocation in response to pH: the increase of acidity activates secondary metabolism, and more proteome resources get allocated to the U sector. Note: this “toy” model is for illustration only. EPS, exopolysaccharide.

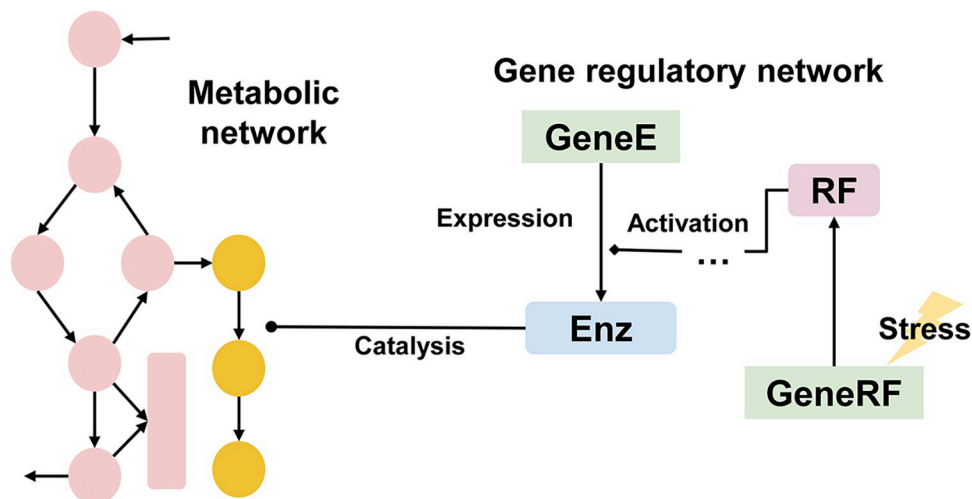
<https://doi.org/10.1371/journal.pcbi.1011391.g003>

$$v_i \leq kcat_i[E_i], \sum_i [E_i] \leq \phi_x [P_{TOT}], x = U, C, R, T \tag{6}$$

The cellular proteome is divided into inflexible housekeeping sector (Q-sector), catabolic sector (C-sector), ribosomal sector for protein translation (R-sector), transportation sector (T-sector), and secondary metabolism sector (U-sector).  $\phi_x$  is the mass fraction of the sector  $x$ , and  $\phi_Q$  is assumed to 50% [102,105,106] (Eqs 4, 6). The total amount of proteome resources is conserved (Eq 4). In response to environmental stress, which is acidity in this case, the fraction of proteome resources allocated to the U-sector, which is EPS biosynthesis, rises up (Eq 5).  $k_1$  and  $k_2$  in Eq 5 are empirical coefficients. The metabolic fluxes are constrained by the total amount of proteome resources allocated to the sector (Eq 6).  $[E_i]$ ,  $kcat_i$  are the concentration and turnover rate of the enzyme  $i$ , respectively, and  $[P_{TOT}]$  is the concentration of total cellular proteins. The proteome allocation of secondary metabolism (Eq 5) can materialize in different forms in actual implementation, e.g., modeling the consequence of differing proteomic costs caused for different stress factors, such as temperature, inhibitors, or nutrient limitation [102,110,111].

### 3.3.2. Combine gene regulatory and metabolic networks to capture secondary metabolism onset

Constrained proteome allocation-embedded FBA models can capture the shift in metabolic states via modeling resource distribution among functional proteome sectors [102]. However, they do not characterize regulatory interactions between regulatory factors (RFs) and genes at transcriptional or translational level. On the other hand, existing gene expression integrated



**Fig 4. Schematic diagram of the combination of metabolic and GRNs to predict the secondary metabolite production.** The stress signal stimulates the expression of RF, resulting in the activation of enzymes catalyzing reactions for secondary metabolite biosynthesis. Orange circle: secondary metabolite; pink circle: primary metabolite. Enz, enzyme; GRN, gene regulatory network; RF, regulatory factor.

<https://doi.org/10.1371/journal.pcbi.1011391.g004>

FBA modeling techniques, such as iMAT, do not have a mechanistic representation of gene expression regulation, and hence, their applications on the prediction of secondary metabolite production are restricted by condition-specific data availability. Therefore, a regulatory network model needs to be combined with FBA, as illustrated in Fig 4, to resolve the limitation of direct integration of gene expression into FBA. Though not for predicting secondary metabolite production, several integrated regulatory-metabolic models, also called regulatory FBA (rFBA), have been reported for a while [112–115]. The generic framework of dynamic rFBA is a combination of FBA for intracellular metabolic fluxes, a Boolean logic network modeling regulatory interactions, and a set of differential equations modeling concentrations of RFs, biomass, and extracellular metabolites [116]. The rFBA model constrains metabolic fluxes by predicting the expression status of the gene associated with the reaction, which means that metabolic fluxes only flow through active reactions.

Enlightened by the idea of dynamic rFBA, if regulatory interactions related to secondary metabolism are elucidated in a Boolean logic network, dynamic simulation of gene expression regulation at metabolic switch can be performed, and thus quantitatively capture the onset of secondary metabolite production (Fig 4). The stress response can be appended into rFBA, as the process from stress sensing to regulation can be represented by signal transduction in a network [117], though such approach has not been applied to the modeling of secondary metabolism yet. In essence, rFBA has the potential to overcome the limitation of conventional FBA by including regulation, an important cellular activity apart from biomass formation, which will enhance the predictive power of FBA on secondary metabolite production.

#### 4. Conclusion and outlook

This review summarized the current key challenges that limit the modeling of secondary metabolism, which are mainly derived from 2 aspects: (1) the difficulty in reconstructing the complete biosynthetic route of a secondary metabolite in an enzymatically detailed manner; and (2) the inability of FBA-based modeling techniques to make quantitatively accurate predictions for secondary metabolism under different conditions for engineering and practical

purposes. As for the first challenge, except for PKs/NRPs, the pathway reconstruction for secondary metabolites generally lacks template models and databases containing information of biochemical reactions associated with secondary metabolism. Consequently, this process is much less automated. For example, although databases related with EPS biosynthesis are available, e.g., EPS-DB [118], CAZy [119], the information on enzymes and molecular structures is not assembled together, which makes automated reconstruction of EPS biosynthetic pathways currently infeasible.

The existing FBA-based modeling techniques that are intended to tackle the second challenge have been evaluated in this review. It shows that they are either too simplistic or have restricted application scopes. None of them are predictive enough to capture the production of secondary metabolites. Nevertheless, those modeling techniques show usefulness in certain cases, such as manipulating the objective function and incorporating transcriptomics data. With the aim of building a modeling technique that captures the mechanistic understanding of secondary metabolism, we have suggested the application of 2 potential extensions of FBA-based modeling techniques. They respectively incorporate constrained proteome allocation that explicitly account for the enzyme capacity for secondary metabolism and connect with the regulatory network model for gene expression states that control reaction activities with high resolution. Detailed implementation approaches in those directions remain to be explored, which we envisage will benefit from further and more mechanistic understanding of cellular responses leading to the production of secondary metabolites. Additionally, future curation works of proteomic and gene regulatory information, both in connection with typical or specific secondary metabolic pathways, will contribute to the advancement of those approaches.

When a high coverage of biosynthetic pathways for secondary metabolites is achieved in metabolic reaction databases for smGSMM reconstruction, and a modeling technique that accurately captures the mechanism of secondary metabolism is developed, the complex bio-process of secondary metabolite production will then become more “white-boxed.” Researchers, particularly those in the field of synthetic biology, will be able to conduct more reliable *in silico* analysis of secondary metabolite production. This advancement will contribute to the development of model-based design and pathway engineering approaches aimed at controlling secondary metabolism, such as enhancing the productivity of high-value secondary metabolites.

## Author Contributions

**Supervision:** Hong Zeng.

**Visualization:** Sizhe Qiu.

**Writing – original draft:** Sizhe Qiu.

**Writing – review & editing:** Aidong Yang, Hong Zeng.

## References

1. Seyedsayamdost MR. Toward a global picture of bacterial secondary metabolism. *J Ind Microbiol Biotechnol*. 2019; 46:301–311. <https://doi.org/10.1007/s10295-019-02136-y> PMID: 30684124
2. Mann J. *Secondary Metabolism*. Oxford University Press; 1987. Available from: [https://books.google.com/books/about/Secondary\\_Metabolism.html?hl=&id=VHjwAAAAMAAJ](https://books.google.com/books/about/Secondary_Metabolism.html?hl=&id=VHjwAAAAMAAJ).
3. Mosunova O, Navarro-Muñoz JC, Collemare J. The Biosynthesis of Fungal Secondary Metabolites: From Fundamentals to Biotechnological Applications. *Encyclopedia of Mycology*. 2021;458–476.
4. Thirumurugan D, Cholarajan A, Raja SSS, Vijayakumar R. An Introductory Chapter: Secondary Metabolites. *Secondary Metabolites—Sources and Applications*. 2018. <https://doi.org/10.5772/intechopen.79766>

5. Abdel-Aziz SM, Abo Elsoud MM, Anise AAH. Microbial Biosynthesis: A Repertory of Vital Natural Products. *Food Biosynthesis*. 2017;25–54.
6. O'Brien J, Wright GD. An ecological perspective of microbial secondary metabolism. *Curr Opin Biotechnol*. 2011; 22:552–558. <https://doi.org/10.1016/j.copbio.2011.03.010> PMID: 21498065
7. Yoon V, Nodwell JR. Activating secondary metabolism with stress and chemicals. *J Ind Microbiol Biotechnol*. 2014; 41: 415–424. <https://doi.org/10.1007/s10295-013-1387-y> PMID: 24326978
8. Then Bergh K, Brakhage AA. Regulation of the *Aspergillus nidulans* penicillin biosynthesis gene *acvA* (*pcbAB*) by amino acids: implication for involvement of transcription factor PACC. *Appl Environ Microbiol*. 1998; 64:843–849. <https://doi.org/10.1128/AEM.64.3.843-849.1998> PMID: 9501424
9. Oslizlo A, Stefanic P, Dogsa I, Mandic-Mulec I. Private link between signal and response in *Bacillus subtilis* quorum sensing. *Proc Natl Acad Sci U S A*. 2014; 111:1586–1591. <https://doi.org/10.1073/pnas.1316283111> PMID: 24425772
10. Sanchez S, Guzmán-Trampe S, Ávalos M, Ruiz B, Rodríguez-Sanoja R, Jiménez-Estrada M. Microbial Natural Products. *Natural Products*. *Chem Biol*. 2012:65–108. <https://doi.org/10.1371/journal.pone.0035398>
11. Kaspar F, Neubauer P, Gimpel M. Bioactive Secondary Metabolites from *Bacillus subtilis*: A Comprehensive Review. *J Nat Prod*. 2019; 82:2038–2053. <https://doi.org/10.1021/acs.jnatprod.9b00110> PMID: 31287310
12. Korcz E, Varga L. Exopolysaccharides from lactic acid bacteria: Techno-functional application in the food industry. *Trends Food Sci Technol*. 2021; 110:375–384.
13. Derikvand P, Llewellyn CA, Purton S. Cyanobacterial metabolites as a source of sunscreens and moisturizers: a comparison with current synthetic compounds. *Eur J Phycol*. 2017; 52:43–56.
14. Gibson B, Wilson DJ, Feil E, Eyre-Walker A. The distribution of bacterial doubling times in the wild. *Proc R Soc B Biol Sci*. 2018; 285:20180789. <https://doi.org/10.1098/rspb.2018.0789> PMID: 29899074
15. Brownleader D. Plant cell biotechnology. *Plant Biochem*.
16. Barcelos MCS, Vespermann KAC, Pelissari FM, Molina G. Current status of biotechnological production and applications of microbial exopolysaccharides. *Crit Rev Food Sci Nutr*. 2020; 60:1475–1495. <https://doi.org/10.1080/10408398.2019.1575791> PMID: 30740985
17. Pham JV, Yilma MA, Feliz A, Majid MT, Maffetone N, Walker JR, et al. A Review of the Microbial Production of Bioactive Natural Products and Biologics. *Front Microbiol*. 2019; 10. <https://doi.org/10.3389/fmicb.2019.01404> PMID: 31281299
18. Del Carratore F, Hanco EK, Breitling R, Takano E. Biotechnological application of *Streptomyces* for the production of clinical drugs and other bioactive molecules. *Curr Opin Biotechnol*. 2022; 77:102762. <https://doi.org/10.1016/j.copbio.2022.102762> PMID: 35908316
19. Breitling R, Achcar F, Takano E. Modeling challenges in the synthetic biology of secondary metabolism. *ACS Synth Biol*. 2013; 2:373–378. <https://doi.org/10.1021/sb4000228> PMID: 23659212
20. Palazzotto E, Weber T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Curr Opin Microbiol*. 2018; 45:109–116. <https://doi.org/10.1016/j.mib.2018.03.004> PMID: 29656009
21. Weber T, Kim HU. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol*. 2016; 1:69–79. <https://doi.org/10.1016/j.synbio.2015.12.002> PMID: 29062930
22. Costa RS, Hartmann A, Vinga S. Kinetic modeling of cell metabolism for microbial production. *J Biotechnol*. 2016; 219:126–141. <https://doi.org/10.1016/j.jbiotec.2015.12.023> PMID: 26724578
23. Foster CJ, Wang L, Dinh HV, Suthers PF, Maranas CD. Building kinetic models for metabolic engineering. *Curr Opin Biotechnol*. 2021; 67:35–41. <https://doi.org/10.1016/j.copbio.2020.11.010> PMID: 33360621
24. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010; 28:245–248. <https://doi.org/10.1038/nbt.1614> PMID: 20212490
25. Dochain D, Bastin G. ADAPTIVE CONTROL OF FEDBATCH BIOREACTORS. *Chem Eng Commun*. 1990; 87:67–85.
26. Bouguettoucha A, Balannec B. Unstructured models for lactic acid fermentation—a review. *Food Technol*. 2011. Available from: [https://www.researchgate.net/profile/Abdallah-Bouguettoucha/publication/228355166\\_Unstructured\\_Models\\_for\\_Lactic\\_Acid\\_Fermentation-A\\_Review/links/02e7e53bedb3eb05fe000000/Unstructured-Models-for-Lactic-Acid-Fermentation-A-Review.pdf](https://www.researchgate.net/profile/Abdallah-Bouguettoucha/publication/228355166_Unstructured_Models_for_Lactic_Acid_Fermentation-A_Review/links/02e7e53bedb3eb05fe000000/Unstructured-Models-for-Lactic-Acid-Fermentation-A-Review.pdf).
27. Nilsson A, Nielsen J, Palsson BO. Metabolic Models of Protein Allocation Call for the Kinetome. *Cell Syst*. 2017; 5:538–541. <https://doi.org/10.1016/j.cels.2017.11.013> PMID: 29284126

28. Ulusu NN. Evolution of Enzyme Kinetic Mechanisms. *J Mol Evol*. 2015; 80:251–257. <https://doi.org/10.1007/s00239-015-9681-0> PMID: 25987355
29. Faria JP, Rocha M, Rocha I, Henry CS. Methods for automated genome-scale metabolic model reconstruction. *Biochem Soc Trans*. 2018; 46:931–936. <https://doi.org/10.1042/BST20170246> PMID: 30065105
30. Damiani C, Di Filippo M, Pescini D, Maspero D, Colombo R, Mauri G. popFBA: tackling intratumour heterogeneity with Flux Balance Analysis. *Bioinformatics*. 2017; 33:i311–i318. <https://doi.org/10.1093/bioinformatics/btx251> PMID: 28881985
31. Martelli C, De Martino A, Marinari E, Marsili M, Pérez Castillo I. Identifying essential genes in *Escherichia coli* from a metabolic optimization principle. *Proc Natl Acad Sci U S A*. 2009; 106:2607–2611. <https://doi.org/10.1073/pnas.0813229106> PMID: 19196991
32. Mohite OS, Weber T, Kim HU, Lee SY. Genome-Scale Metabolic Reconstruction of Actinomycetes for Antibiotics Production. *Biotechnol J*. 2019; 14:e1800377. <https://doi.org/10.1002/biot.201800377> PMID: 30117679
33. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*. 2011; 39:W339–46. <https://doi.org/10.1093/nar/gkr466> PMID: 21672958
34. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res*. 2017; 45:W49–W54. <https://doi.org/10.1093/nar/gkx320> PMID: 28460067
35. de Jong A, van Hijum SAFT, Bijlsma JJE, Kok J, Kuipers OP. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res*. 2006; 34:W273–9. <https://doi.org/10.1093/nar/gkl237> PMID: 16845009
36. Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol*. 2015; 11:639–648. <https://doi.org/10.1038/nchembio.1884> PMID: 26284671
37. Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes—a review. *Nat Prod Rep*. 2016; 33:988–1005. <https://doi.org/10.1039/c6np00025h> PMID: 27272205
38. Tania, Goyal M, Baranwal M. In Silico Approach to Analyze the Biochemical Pathways of Bacterial Metabolite Synthesis. In: Choudhary DK, Kumar M, Prasad R, Kumar V, editors. *In Silico Approach for Sustainable Agriculture*. Springer Singapore: Singapore, 2018. p 141–159.
39. Chavali AK, Rhee SY. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief Bioinform*. 2018; 19:1022–1034. <https://doi.org/10.1093/bib/bbx020> PMID: 28398567
40. Blin K, Kim HU, Medema MH, Weber T. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief Bioinform*. 2019; 20:1103–1113. <https://doi.org/10.1093/bib/bbx146> PMID: 29112695
41. Ren H, Shi C, Zhao H. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. *iScience*. 2020; 23:100795. <https://doi.org/10.1016/j.isci.2019.100795> PMID: 31926431
42. Sulheim S, Fossheim FA, Wentzel A, Almaas E. Automatic reconstruction of metabolic pathways from identified biosynthetic gene clusters. *BMC Bioinformatics*. 2021; 22:81. <https://doi.org/10.1186/s12859-021-03985-0> PMID: 33622234
43. Wang L, Dash S, Ng CY, Maranas CD. A review of computational tools for design and reconstruction of metabolic pathways. *Synth Syst Biotechnol*. 2017; 2:243–252. <https://doi.org/10.1016/j.synbio.2017.11.002> PMID: 29552648
44. Machado D, Andrejev S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res*. 2018; 46:7542–7553. <https://doi.org/10.1093/nar/gky537> PMID: 30192979
45. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res*. 2021; 49:D575–D588. <https://doi.org/10.1093/nar/gkaa746> PMID: 32986834
46. Wang H, Marcišauskas S, Sánchez BJ, Domenzain I, Hermansson D, Agren R, et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput Biol*. 2018; 14:e1006541. <https://doi.org/10.1371/journal.pcbi.1006541> PMID: 30335785

47. Capela J, Lagoa D, Rodrigues R, Cunha E, Cruz F, Barbosa A, et al. merlin v4.0: an updated platform for the reconstruction of high-quality genome-scale metabolic models. *bioRxiv*. 2021; 2021.02.24.432752.
48. Karlsen E, Schulz C, Almaas E. Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics*. 2018; 19:467. <https://doi.org/10.1186/s12859-018-2472-z> PMID: 30514205
49. Aite M, Chevallier M, Frioux C, Trottier C, Got J, Cortés MP, et al. Traceability, reproducibility and wiki-exploration for 'à-la-carte' reconstructions of genome-scale metabolic models. *PLoS Comput Biol*. 2018; 14:e1006146.
50. Olivier B. *SystemsBioinformatics/cbmpy-metadraft: MetaDraft is now available*. 2018. <https://doi.org/10.5281/zenodo.2398336>
51. Karp PD, Midford PE, Billington R, Kothari A, Krummenacker M, Latendresse M, et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform*. 2019; 22:109–126.
52. Boele J, Olivier BG, Teusink B. FAME, the Flux Analysis and Modeling Environment. *BMC Syst Biol*. 2012; 6:8. <https://doi.org/10.1186/1752-0509-6-8> PMID: 22289213
53. Liao Y-C, Tsai M-H, Chen F-C, Hsiung CA. GEMSiRV: a software platform for GENome-scale metabolic model simulation, reconstruction and visualization. *Bioinformatics*. 2012; 28:1752–1758. <https://doi.org/10.1093/bioinformatics/bts267> PMID: 22563070
54. Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res*. 2020; 48:D402–D406. <https://doi.org/10.1093/nar/gkz1054> PMID: 31696234
55. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res*. 2018; 46:D633–D639. <https://doi.org/10.1093/nar/gkx935> PMID: 29059334
56. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017; 45:D353–D361. <https://doi.org/10.1093/nar/gkw1092> PMID: 27899662
57. Prigent S, Nielsen JC, Frisvad JC, Nielsen J. Reconstruction of 24 Penicillium genome-scale metabolic models shows diversity based on their secondary metabolism. *Biotechnol Bioeng*. 2018; 115:2604–2612. <https://doi.org/10.1002/bit.26739> PMID: 29873086
58. Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, et al. MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol*. 2005; 138:27–37. <https://doi.org/10.1104/pp.105.060376> PMID: 15888675
59. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res*. 2020; 48:D445–D453. <https://doi.org/10.1093/nar/gkz862> PMID: 31586394
60. Altman T, Travers M, Kothari A, Caspi R, Karp PD. A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics*. 2013; 14:112. <https://doi.org/10.1186/1471-2105-14-112> PMID: 23530693
61. Monk J, Nogales J, Palsson BO. Optimizing genome-scale network reconstructions. *Nat Biotechnol*. 2014; 32:447–452. <https://doi.org/10.1038/nbt.2870> PMID: 24811519
62. Kittikunapong C, Ye S, Magadán-Corpas P, Pérez-Valero Á, Villar CJ, Lombó F, et al. Reconstruction of a Genome-Scale Metabolic Model of *Streptomyces albus* J1074: Improved Engineering Strategies in Natural Product Synthesis. *Metabolites*. 2021; 11. <https://doi.org/10.3390/metabo11050304> PMID: 34064751
63. Xu N, Liu J, Ai L, Liu L. Reconstruction and analysis of the genome-scale metabolic model of *Lactobacillus casei* LC2W. *Gene*. 2015; 554:140–147. <https://doi.org/10.1016/j.gene.2014.10.034> PMID: 25452194
64. Nguyen Q-T, Merlo ME, Medema MH, Jankevics A, Breitling R, Takano E. Metabolomics methods for the synthetic biology of secondary metabolism. *FEBS Lett*. 2012; 586:2177–2183. <https://doi.org/10.1016/j.febslet.2012.02.008> PMID: 22710183
65. Li T, Tripathi A, Yu F, Sherman DH, Rao A. DDAP: docking domain affinity and biosynthetic pathway prediction tool for type I polyketide synthases. *Bioinformatics*. 2020; 36:942–944. <https://doi.org/10.1093/bioinformatics/btz677> PMID: 31504190
66. Sulheim S, Kumelj T, van Dissel D, Salehzadeh-Yazdi A, Du C, van Wezel GP, et al. Enzyme-Constrained Models and Omics Analysis of *Streptomyces coelicolor* Reveal Metabolic Changes that Enhance Heterologous Production. *iScience*. 2020; 23:101525. <https://doi.org/10.1016/j.isci.2020.101525> PMID: 32942174



67. Delépine B, Duigou T, Carbonell P, Faulon J-L. RetroPath2.0: A retrosynthesis workflow for metabolic engineers. *Metab Eng*. 2018; 45:158–170. <https://doi.org/10.1016/j.ymben.2017.12.002> PMID: 29233745
68. Zheng S, Zeng T, Li C, Chen B, Coley CW, Yang Y, et al. BioNavi-NP: Biosynthesis Navigator for Natural Products. *arXiv [q-bio.QM]*. 2021. Available from: <http://arxiv.org/abs/2105.13121>.
69. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol*. 2011; 7:445–452. <https://doi.org/10.1038/nchembio.580> PMID: 21602812
70. Hadadi N, Hafner J, Shajkofci A, Zisaki A, Hatzimanikatis V. ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies. *ACS Synth Biol*. 2016; 5:1155–1166. <https://doi.org/10.1021/acssynbio.6b00054> PMID: 27404214
71. Watson IA, Wang J, Nicolaou CA. A retrosynthetic analysis algorithm implementation. *J Chem*. 2019; 11:1. <https://doi.org/10.1186/s13321-018-0323-6> PMID: 30604073
72. Panikov NS. Genome-Scale Reconstruction of Microbial Dynamic Phenotype: Successes and Challenges. *Microorganisms*. 2021; 9. <https://doi.org/10.3390/microorganisms9112352> PMID: 34835477
73. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc*. 2007; 2:727–738. <https://doi.org/10.1038/nprot.2007.99> PMID: 17406635
74. Latendresse M, Ong WK, Karp PD. Metabolic Modeling with MetaFlux. *Methods Mol Biol*. 2022; 2349:259–289. [https://doi.org/10.1007/978-1-0716-1585-0\\_12](https://doi.org/10.1007/978-1-0716-1585-0_12) PMID: 34718999
75. Grafahrend-Belau E, Klukas C, Junker BH, Schreiber F. FBA-SimVis: interactive visualization of constraint-based metabolic models. *Bioinformatics*. 2009; 25:2755–2757. <https://doi.org/10.1093/bioinformatics/btp408> PMID: 19578041
76. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. Synergy between <sup>13</sup>C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*. *Metab Eng*. 2011; 13:38–48.
77. Vijayakumar S, Rahman PKS, Angione C. A Hybrid Flux Balance Analysis and Machine Learning Pipeline Elucidates Metabolic Adaptation in Cyanobacteria. *iScience*. 2020; 23:101818. <https://doi.org/10.1016/j.isci.2020.101818> PMID: 33354660
78. Burgard AP, Pharkya P, Maranas CD. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003; 84:647–657. <https://doi.org/10.1002/bit.10803> PMID: 14595777
79. Zorrilla F, Buric F, Patil KR, Zelezniak A. metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res*. 2021; 49:e126. <https://doi.org/10.1093/nar/gkab815> PMID: 34614189
80. Toya Y, Shimizu H. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol Adv*. 2013; 31:818–826. <https://doi.org/10.1016/j.biotechadv.2013.05.002> PMID: 23680193
81. Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen ØM, et al. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics*. 2010; 11:10. <https://doi.org/10.1186/1471-2164-11-10> PMID: 20053288
82. Lind AL, Smith TD, Saterlee T, Calvo AM, Rokas A. Regulation of Secondary Metabolism by the Velvet Complex Is Temperature-Responsive in *Aspergillus*. *G3*. 2016; 6:4023–4033. <https://doi.org/10.1534/g3.116.033084> PMID: 27694115
83. Brakhage AA. Regulation of fungal secondary metabolism. *Nat Rev Microbiol*. 2013; 11:21–32. <https://doi.org/10.1038/nrmicro2916> PMID: 23178386
84. Martín JF, Sola-Landa A, Santos-Beneit F, Fernández-Martínez LT, Prieto C, Rodríguez-García A. Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. *J Microbial Biotechnol*. 2011; 4:165–174. <https://doi.org/10.1111/j.1751-7915.2010.00235.x> PMID: 21342462
85. Demain AL. Carbon source regulation of idiolite biosynthesis in actinomycetes. *Regulation of secondary metabolism in Actinomycetes*. <https://doi.org/10.1201/9781003068600-4/carbon-source-regulation-idiolite-biosynthesis-actinomycetes-arnold-demain>
86. García Sánchez CE, Torres Sáez RG. Comparison and analysis of objective functions in flux balance analysis. *Biotechnol Prog*. 2014; 30:985–991. <https://doi.org/10.1002/btpr.1949> PMID: 25044958
87. Alam MT, Merlo ME, STREAM Consortium, Hodgson DA, Wellington EMH, Takano E, et al. Metabolic modeling and analysis of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics*. 2010; 11:202. <https://doi.org/10.1186/1471-2164-11-202> PMID: 20338070

88. Amara A, Takano E, Breitling R. Development and validation of an updated computational model of *Streptomyces coelicolor* primary and secondary metabolism. *BMC Genomics*. 2018; 19:519. <https://doi.org/10.1186/s12864-018-4905-5> PMID: 29973148
89. Toro L, Pinilla L, Avignone-Rossa C, Ríos-Esteva R. An enhanced genome-scale metabolic reconstruction of *Streptomyces clavuligerus* identifies novel strain improvement strategies. *Bioprocess Biosyst Eng*. 2018; 41:657–669. <https://doi.org/10.1007/s00449-018-1900-9> PMID: 29404683
90. Ledesma-Amaro R, Kerkhoven EJ, Revuelta JL, Nielsen J. Genome scale metabolic modeling of the riboflavin overproducer *Ashbya gossypii*. *Biotechnol Bioeng*. 2014; 111:1191–1199. <https://doi.org/10.1002/bit.25167> PMID: 24374726
91. Stahmann KP, Arst HN Jr, Althöfer H, Revuelta JL, Monschau N, Schlüpen C, et al. Riboflavin, overproduced during sporulation of *Ashbya gossypii*, protects its hyaline spores against ultraviolet light. *Environ Microbiol*. 2001; 3:545–550. <https://doi.org/10.1046/j.1462-2920.2001.00225.x> PMID: 11683864
92. Özcan E, Seven M, Şirin B, Çakır T, Nikerel E, Teusink B, et al. Dynamic co-culture metabolic models reveal the fermentation dynamics, metabolic capacities and interplays of cheese starter cultures. *Biotechnol Bioeng*. 2021; 118:223–237. <https://doi.org/10.1002/bit.27565> PMID: 32926401
93. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*. 2008; 4:e1000082. <https://doi.org/10.1371/journal.pcbi.1000082> PMID: 18483554
94. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol*. 2009; 5:e1000489. <https://doi.org/10.1371/journal.pcbi.1000489> PMID: 19714220
95. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. *Bioinformatics*. 2010; 26:3140–3142. <https://doi.org/10.1093/bioinformatics/btq602> PMID: 21081510
96. Kim MK, Lun DS. Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput Struct Biotechnol J*. 2014; 11:59–65. <https://doi.org/10.1016/j.csbj.2014.08.009> PMID: 25379144
97. Kim M, Yi JS, Lakshmanan M, Lee D-Y, Kim B-G. Transcriptomics-based strain optimization tool for designing secondary metabolite overproducing strains of *Streptomyces coelicolor*. *Biotechnol Bioeng*. 2016; 113:651–660. <https://doi.org/10.1002/bit.25830> PMID: 26369755
98. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P-J, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol*. 2017; 13:935. <https://doi.org/10.15252/msb.20167411> PMID: 28779005
99. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of cell growth and gene expression: origins and consequences. *Science*. 2010; 330:1099–1102. <https://doi.org/10.1126/science.1192588> PMID: 21097934
100. Hui S, Silverman JM, Chen SS, Erickson DW, Basan M, Wang J, et al. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol Syst Biol*. 2015; 11:784. <https://doi.org/10.15252/msb.20145697> PMID: 25678603
101. Zeng H, Rohani R, Huang WE, Yang A. Understanding and mathematical modelling of cellular resource allocation in microorganisms: a comparative synthesis. *BMC Bioinformatics*. 2021; 22:467. <https://doi.org/10.1186/s12859-021-04382-3> PMID: 34583645
102. Mori M, Hwa T, Martin OC, De Martino A, Marinari E. Constrained Allocation Flux Balance Analysis. *PLoS Comput Biol*. 2016; 12:e1004913. <https://doi.org/10.1371/journal.pcbi.1004913> PMID: 27355325
103. Yang L, Yurkovich JT, King ZA, Palsson BO. Modeling the multi-scale mechanisms of macromolecular resource allocation. *Curr Opin Microbiol*. 2018; 45:8–15. <https://doi.org/10.1016/j.mib.2018.01.002> PMID: 29367175
104. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*. 2015; 528:99–104. <https://doi.org/10.1038/nature15765> PMID: 26632588
105. Zeng H, Yang A. Modelling overflow metabolism in *Escherichia coli* with flux balance analysis incorporating differential proteomic efficiencies of energy pathways. *BMC Syst Biol*. 2019; 13:3. <https://doi.org/10.1186/s12918-018-0677-4> PMID: 30630470
106. Regueira A, Rombouts JL, Aljoscha Wahl S, Mauricio-Iglesias M, Lema JM, Kleerebezem R. Resource allocation explains lactic acid production in mixed-culture anaerobic fermentations. *Biotechnol Bioeng*. 2021; 118:745–758. <https://doi.org/10.1002/bit.27605> PMID: 33073364
107. Panikov NS. Mechanistic mathematical models of microbial growth in bioreactors and in natural soils: Explanation of complex phenomena. *Math Comput Simul*. 1996; 42:179–186.

108. Torino MI, Taranto MP, Sesma F, de Valdez GF. Heterofermentative pattern and exopolysaccharide production by *Lactobacillus helveticus* ATCC 15807 in response to environmental pH. *J Appl Microbiol.* 2001; 91:846–852. <https://doi.org/10.1046/j.1365-2672.2001.01450.x> PMID: 11722662
109. Nguyen P-T, Nguyen T-T, Bui D-C, Hong P-T, Hoang Q-K, Nguyen H-T. Exopolysaccharide production by lactic acid bacteria: the manipulation of environmental stresses for industrial applications. *AIMS Microbiol.* 2020; 6:451–469. <https://doi.org/10.3934/microbiol.2020027> PMID: 33364538
110. Páez-Watson T, van Loosdrecht MCM, Wahl SA. Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: Application to polyphosphate accumulating organisms. *Water Res.* 2023; 228:119365. <https://doi.org/10.1016/j.watres.2022.119365> PMID: 36413834
111. Qiu S, Zeng H, Yang Z, Hung W-L, Wang B, Yang A. Dynamic metagenome-scale metabolic modeling of a yogurt bacterial community. *Biotechnol Bioeng.* 2023. <https://doi.org/10.1002/bit.28492> PMID: 37428554
112. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature.* 2004; 429:92–96. <https://doi.org/10.1038/nature02456> PMID: 15129285
113. Herrgård MJ, Lee B-S, Portnoy V, Palsson BØ. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* 2006; 16:627–635. <https://doi.org/10.1101/gr.4083206> PMID: 16606697
114. Goelzer A, Bekkal Brikci F, Martin-Verstraete I, Noirot P, Bessières P, Aymerich S, et al. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol.* 2008; 2:20. <https://doi.org/10.1186/1752-0509-2-20> PMID: 18302748
115. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A.* 2010; 107:17845–17850. <https://doi.org/10.1073/pnas.1005139107> PMID: 20876091
116. Liu L, Bockmayr A. Regulatory dynamic enzyme-cost flux balance analysis: A unifying framework for constraint-based modeling. *J Theor Biol.* 2020; 501:110317. <https://doi.org/10.1016/j.jtbi.2020.110317> PMID: 32446743
117. Gonçalves E, Bucher J, Ryll A, Niklas J, Mauch K, Klamt S, et al. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol Biosyst.* 2013; 9:1576–1583. <https://doi.org/10.1039/c3mb25489e> PMID: 23525368
118. Birch J, Van Calsteren M-R, Pérez S, Svensson B. The exopolysaccharide properties and structures database: EPS-DB. Application to bacterial exopolysaccharides. *Carbohydr Polym.* 2019; 205:565–570. <https://doi.org/10.1016/j.carbpol.2018.10.063> PMID: 30446142
119. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 2009; 37:D233–8. <https://doi.org/10.1093/nar/gkn663> PMID: 18838391