Computational and Structural Biotechnology Journal 19 (2021) 4626–4640

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Review

Advances in flux balance analysis by integrating machine learning and mechanism-based models

Ankur Sahu, Mary-Ann Blätke, Jędrzej Jakub Szymański, Nadine Töpfer *

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstraße 3, 06466 Gatersleben, Germany

ARTICLE INFO

Article history:
Received 15 May 2021
Received in revised form 3 August 2021
Accepted 3 August 2021
Available online 5 August 2021

Keywords:
Flux balance analysis
Genome-scale modeling
Machine learning
Kinetic models
Petri-nets
Multi-scale modeling

ABSTRACT

The availability of multi-omics data sets and genome-scale metabolic models for various organisms provide a platform for modeling and analyzing genotype-to-phenotype relationships. Flux balance analysis is the main tool for predicting flux distributions in genome-scale metabolic models and various data-integrative approaches enable modeling context-specific network behavior. Due to its linear nature, this optimization framework is readily scalable to multi-tissue or -organ and even multi-organism models. However, both data and model size can hamper a straightforward biological interpretation of the estimated fluxes. Moreover, flux balance analysis simulates metabolism at steady-state and thus, in its most basic form, does not consider kinetics or regulatory events. The integration of flux balance analysis with complementary data analysis and modeling techniques offers the potential to overcome these challenges. In particular machine learning approaches have emerged as the tool of choice for data reduction and selection of most important variables in big data sets. Kinetic models and formal languages can be used to simulate dynamic behavior. This review article provides an overview of integrative studies that combine flux balance analysis with machine learning approaches, kinetic models, such as physiology-based pharmacokinetic models, and formal graphical modeling languages, such as Petri nets. We discuss the mathematical aspects and biological applications of these integrated approaches and outline challenges and future perspectives.

Contents

Abbreviations: Asn, asparagine; C, carbon; Gln, glutamine; N, nitrogen; Suc, sucrose.

* Corresponding author.
E-mail address: toepfer@ipk-gatersleben.de (N. Töpfer).

https://doi.org/10.1016/j.csbj.2021.08.004

1. Introduction

In the past two decades, genome-scale metabolic modeling has rapidly advanced in terms of number and quality of available network reconstructions as well as modeling approaches based on flux balance analysis (FBA). FBA approaches have been applied to study microbial [1–3], human [4,5], and plant metabolic networks [6–10] and aided, for instance, designing microbial strains for optimized compound production [11,12], understanding human diseases [13–15], and elucidating constraints in plant metabolism [16–18]. Recent developments enable the study of interacting systems such as the human gut microbiome [19–22] or tissues and cell types in humans [4,13,23,24] and plants [25–31].

FBA is a constraint-based optimization framework that relies on a stoichiometric representation of a large- to genome-scale metabolic network (*i.e.*, core models vs. full models), a set of known input and output constraints, such as measured uptake or excretion rates, and an optimality assumption [32]. If available, kinetic information can be considered in the optimization problem formulation. For instance, v_{max} values can serve as lower and upper flux boundaries. A schematic overview of the approach is shown in (Fig. 1). For microbial systems, the assumed optimality criterion is usually the maximization of growth, *i.e.,* biomass production [33]. The solution to a FBA problem is a set of metabolic flux distributions that are optimal with respect to the assumed objective. It is worth noticing that the objective value to an optimization problem is unique, however the flux distributions seldomly are. This problem of non-unique flux distributions can be tackled by applying a second optimization criterion such as the minimization of the sum of fluxes (also termed parsimonious FBA or pFBA), performing flux sampling or flux variability analysis (FVA). Due to its linear nature, and thus low computational cost, FBA can be readily scaled to metabolic networks with several thousands of reactions. It is therefore particularly amenable to modeling genome-scale metabolic networks, interacting metabolic networks, such as microbial communities or tissues and organs within one organism, and performing parameter scans across a range of external conditions [18]. Thus, FBA complements laborious and computationally expensive Metabolic Flux Analysis which determines metabolic flux distributions based on measured metabolite labeling patterns [34–36]. Aspects like temporal resolution, regulatory constraints, or experimentally determined transcript, protein or metabolite abundances as well as kinetic parameters are *per se* not included in FBA formulations. However, some of these aspects have been tackled by dynamic extensions of FBA [37,38], approaches for integrating regulatory networks [39,40], transcript and protein data [41–43], and metabolite data [44,45] or by developing enzyme-constraint FBA techniques [46–52].

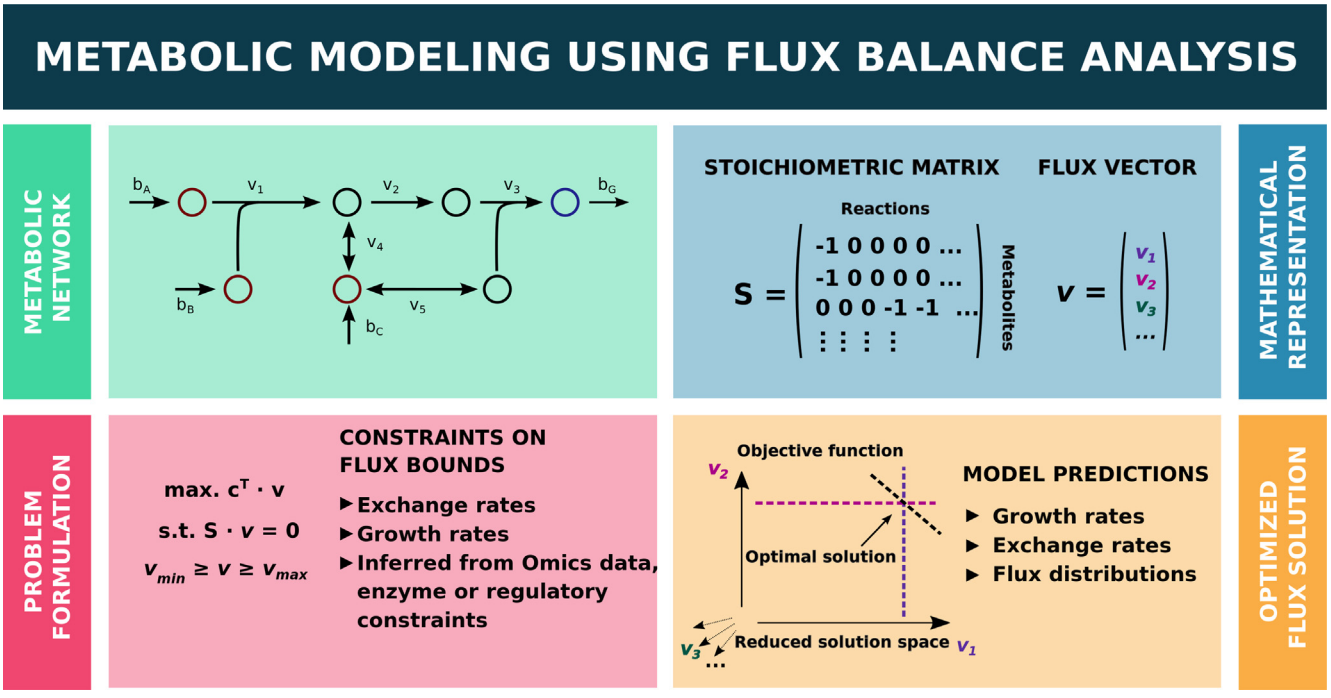

Fig. 1. Schematic representation of FBA. The top part illustrates a toy metabolic network and its mathematical representation in the form of a stoichiometric matrix and a flux vector. (Left) The system comprises metabolites A to G, exchange fluxes b_A, b_B, b_C, and b_G and reactions r_1 to r_5 with the respective reaction fluxes v_1 to v_5. (Right) The stoichiometric matrix S represents the reaction stoichiometries of the network metabolites (rows) in the respective reactions (columns) and *v* is the vector of fluxes. The bottom part illustrates the FBA problem formulation. (Left) Find a flux distribution that maximizes (minimizes) the objective function, where c^T is the transposed vector of weights, indicating how much each reaction contributes to the objective function (first line) subject to the system is at steady-state (second line) and fluxes are within their lower and upper bound (third line). (Right) The solution to the optimization problem is a flux distribution that satisfies the applied constraints (colored lines) and optimizes the objective function (black line), *e.g.* growth rate (biomass production). Note, that depending on the available experimental data, exchange and growth rates can be both, an input to the model or a model prediction.

Table 1
Overview of studies integrating flux balance analysis with machine learning, kinetic models, and petri nets.

Article	Integrative Component	Organism	Purpose of integration
Machine Learning and Flux Balance Analysis			
Sánchez *et al.*, 2019 [66]	Principal component analysis	*Saccharomyces cerevisiae*	Random sampling of lipid specific flux distributions
Bhadra *et al.*, 2018 [59]	Principal Component Analysis	*Pichia pastoris* and *Saccharomyces cerevisiae*	Determining variability in flux data
Dai *et al.*, 2018 [67]	Singular Value Decomposition	*Escherichia coli*	Selecting important metabolite constraint for flux estimation in cell-free protein synthesis system
Patané *et al.*, 2019 [68]	Density-Based Spatial Clustering of Applications with Noise	*Escherichia coli*	Finding clusters of flux solutions for improved ethanol production
Kim *et al.*, 2016 [60]	Recurrent Neural Networks, Lasso Regression, Ensemble Learning	*Escherichia coli*	Extracting features from multi-omics data for setting FBA constraint
Occhipinti *et al.* 2018 [80]	Elastic Net Regression	*Pseudomonas putida*	Grouping flux distributions for maximal rhamnolipids production
Shaked *et al.*, 2016 [81]	Support Vector Machines, Random Forest	Human	Classifying metabolic fluxes based on drug side effects
DiMucci et al., 2018 [82]	Random Forest	*Escherichia coli* and gut associated bacteria	Classifying metabolic fluxes for finding interspecies interaction
Bordbar *et al.*, 2017 [61]	Principal Component Analysis and Linear Regression	Human, *Saccharomyces cerevisiae*	Determining the rate of change of metabolites
Jalili *et al.*, 2021 [83]	Principal Component Analysis, Random Forest	Human	Extracting FBA-based metabolic signatures for cancer cell
Yaneske *et al.*, 2018 [84]	Hierarchical Clustering, K-means Clustering, Elastic Net Regression	Human	Finding highly correlated fluxes for multi-omics age prediction
Vijayakumar *et al.* 2020 [85]	Principal Component Analysis, K-means Clustering, and Lasso Regression	*Synechococcus* sp.	Extracting growth promoting and limiting features from flux distributions
Culley *et al.*, 2020 [86]	Multimodal Artificial Neural Networks	*Saccharomyces cerevisiae*	Employing flux distributions as features for model training
Magazzù *et al.*, 2021 [87]	Lasso Regression, Multimodal Artificial Neural Networks	*Saccharomyces cerevisiae*	Employing flux distributions as features for model training
Available modeling tools and web interfaces	PMFA [59], GEESE [95], SWIFTCORE [163]		
Kinetic Models and Flux Balance Analysis			
Krauss *et al.*, 2012 [106]	PBPK model	Human liver	Investigating hyperuricemia therapy, ammonia detoxification and paracetamol-induced toxication
Guebila & Thiele, 2015, 2016 [112,113]	PBPK model	Human small intestine enterocyte	Investigating dietary strategies for treating Parkinson's disease symptoms
Shepelyuk *et al.*, 2016 [114]	PBPK model	Human platelet	Determining mutual influence of platelet metabolism and blood glucose levels
Grafahrend-Belau *et al.*, 2013 [28]	Whole-plant functional model	Barley (*Hordeum vulgare*)	Studying source-sink relationships
Mallmann *et al.* 2014 [133]	Kinetic model of photosynthesis	C3-C4 intermediates in the genus *Flaveria*	Studying the evolution of C4 photosynthesis
Shaw & Cheung, 2018 [31]	Equations for balanced growth theory	*Arabidopsis thaliana*	Studying optimal resource partitioning
Available modeling tools and web interfaces	MUFINS [105], COMETS [125], PKSim® [164], MoBi® [165]		
Petri Nets and Flux Balance Analysis			
Fisher *et al.* 2013 [143]	Quasi-Steady Petri Net	Human liver	Studying feedback mechanism in bile acid homeostasis
Ptak *et al.*, 2016 [146]	Agent-based Quasi-Steady Petri Net	Human liver	Studying the cooperative effect of cell communication during bile acid homeostasis
Simone *et al.*, 2020 [147]	Extended Stochastic Petri Net	Human pancreas	Studying effect of oxidative phosphorylation in pancreatic cancer
Self *et al.*, 2018 [152]	Stochastic Petri net	*Escherichia coli*	Deriving a biomass proxy for the dynamic analysis of different growth conditions
Available modeling tools and web interfaces	MUFINS [105], Snoopy [166], SurreyFBA [145], GreatSPN [148], Charlie [167], Marcie [168]		

Here we focus on the integration and application of FBA with other, complementary modeling techniques. This review covers the main areas of development - on the one hand, the integration with "black-box" machine learning (ML) approaches and on the other hand, the integration with mechanistic approaches, such as kinetic models and graph-theoretical Petri nets (see Table 1). Each section briefly introduces the modeling technique, presents examples for their integration with FBA, highlights benefits of the integrated analysis, and if available lists dedicated modeling tools. In this review, we cover seminal earlier studies and emphasize recent developments in the respective field. We conclude by highlighting current challenges and anticipated developments.

2. Machine learning and flux balance analysis

In recent years ML has become a key element in biological research. In contrast to statistical models which are applied to infer relationships between specific classes of variables, ML algorithms put emphasis on model performance and can be trained on highly heterogeneous data. The latter aspect becomes more important with the growing complexity of genome-scale metabolic models and the increasing availability of various biological datasets (*e.g.* gene expression, metabolite levels, categorical phenotypes, *etc.*) which can be integrated with these models [53,54]. Thus, ML algorithms provide a bridge between the knowledge-driven predictive models of FBA and heterogeneous biological data sets. While FBA is a predictive approach, it can be combined with both predictive and descriptive ML models [55]. Predictive ML models are trained by supervised learning to create a mapping between input data and defined output variables. The term supervised learning relates to the fact that data are labeled and the algorithms are trained to perform predictions according to these labels. In contrast, descriptive ML models use unsupervised learning to identify emergent patterns in the input data without the need for providing any labels.

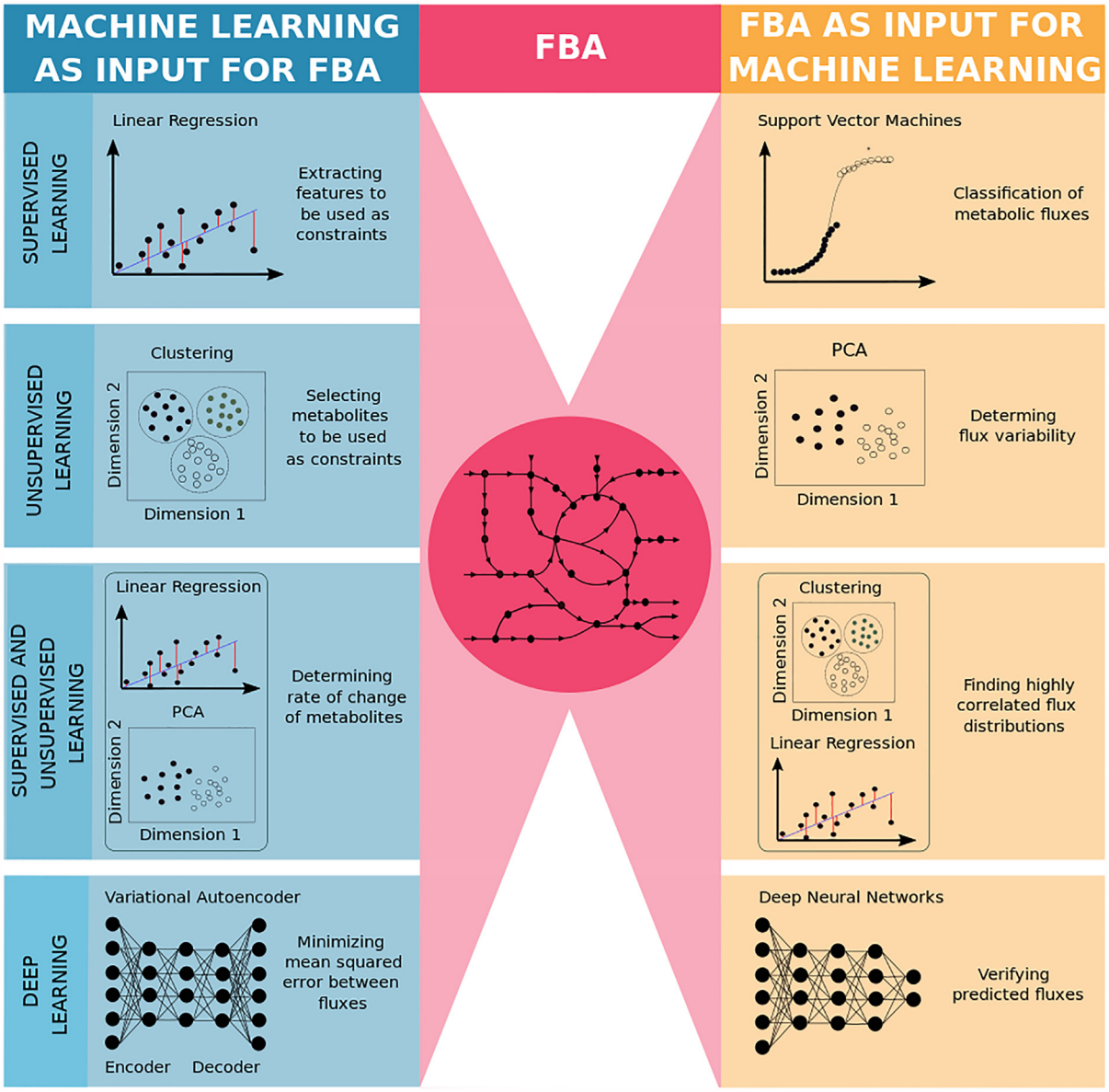

Fig. 2. Overview of different approaches for integrating ML and FBA. The column on the left illustrates ML approaches used for extracting inputs for FBA simulations. The column on the right illustrates applications where FBA simulations serve as input for ML methods. We classified the depicted integration procedures based on the types of ML approaches and follow the structure of the main text, *i.e.*, supervised, unsupervised, supervised and unsupervised, and Deep Learning. Note that the direct integration of ML and FBA has not been illustrated here.

In unsupervised approaches, patterns are identified according to defined mathematical criterion (*e.g.* number of clusters or variance independence). Both learning methods have been widely used to analyze large-scale biological datasets [56,57] and have also been integrated with FBA. The ML-FBA integration includes three major cases: (1) ML can be used to select molecular features (*e.g.* genes, transcripts, metabolites) that are passed as inputs to the metabolic models [53,58]; (2) some ML methods, such as Principal Component Analysis (PCA), can be integrated directly with the FBA algorithm [59]; (3) and finally, results of FBA simulations can be analyzed by ML approaches (Fig. 2). Hence, ML complements FBA by extracting relevant information from interacting datasets (genomic, proteomic, and metabolomic) [60] and improving the interpretation of the results [61]. All but one of the presented studies fall either into category (1) or (3). In the following sections we highlight selected studies that provide a balanced overview of the field.

2.1. Integrating unsupervised learning and flux balance analysis

Some of the widely used unsupervised learning techniques are PCA [62], Hierarchical Clustering, K-means Clustering [63], Singular Value Decomposition [64] and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [65]. In combination with FBA these methods have been mainly applied to explore patterns in flux distributions and in the following we highlight some of these integrative studies.

PCA is a data dimensionality reduction approach for identifying the most dominating sources of variance in complex data. The approach identifies orthogonal linear combinations of variables, called the principal components, which maximally explain the variation in the data. PCA has been used to investigate variations in flux distributions as well as directly integrated with the FBA formulation.

Sánchez *et al.* [66] developed a method termed 'Split Lipids Into Measurable Entities reactions' (SLIMEr) for improved representation of lipid requirements in genome-scale metabolic models. SLIMEr uses a mathematical construct to split lipid components into lipid classes and acyl chain distributions and applies constraints on both. The method was tested on a model of *Saccharomyces cerevisiae* (budding yeast) and PCA was used on randomly sampled flux distributions for both the extended and basic model and compared to experimental values. The authors demonstrated that the extended model could accurately represent amounts of lipid species, analyze network flexibility, and compute the energetic requirements for transitions between metabolic states.

Identification of principal components can be constrained using known relationships between variables, such as molecular interactions between genes and metabolites or similar functional annotations. Bhadra *et al.* [59] presented an approach in which they modified the PCA algorithm to capture sample variance under consideration of the network structure in terms of metabolic pathways and flux modes. The approach, named Principal Metabolic Flux Mode Analysis (PMFA), was implemented as a PCA optimization problem with structural constraints derived from FBA. As a result, PMFA highlights sets of reactions, the so-called principal metabolic flux modes, which are approximately in steady-state and explain most of the data variance. The authors benchmarked their approach by comparing results from PCA, FBA, elementary flux mode analysis, and PMFA using data sets and metabolic models for the two yeast strains *Pichia pastoris* and *S. cerevisiae*. They found PMFA to be more efficient in capturing variance in sets of experiments than the other approaches tested. Furthermore, they used PMFA to analyze transcriptomics data for *S. cerevisiae* under different oxygen conditions. The analysis identified six mitochondrial pathways responsive to changes in oxygen availability, thus underlining the power of pathway-centric variation analysis.

Singular Value Decomposition is another dimensionality reduction method. It performs matrix factorization to identify independent sources of variance in the data and the respective contribution of each variable to this variance. Integrated with FBA, Singular Value Decomposition can be used to select variables (*e.g.* reactions or metabolites) whose variation contributes the most to the emergence of specific flux distributions across various conditions or genotypes.

Dai *et al.* [67] integrated dynamic FBA with Singular Value Decomposition in a cell-free protein synthesis system based on an *Escherichia coli* metabolic network (removing reactions that are not present in the cell-free system).The main objective of the study was to increase the system's protein synthesis capacity by understanding which metabolites were performance limiting. Thus, Singular Value Decomposition was used to select metabolite measurements to be set as constraints for flux estimations. To this end, the stoichiometric matrix was decomposed into 105 modes and the top 36 metabolites with the highest singular value-weighted sum were selected as constraints. The authors showed that direct integration of metabolite measurements into the flux estimation problem improved the prediction accuracy of the protein synthesis rates and thus enabled optimization of the cell-free protein synthesis system.

Besides dimensionality reduction, unsupervised learning approaches can be used for cluster analysis and clustering methods in combination with FBA have been used to group flux distributions.

Patané *et al.* [68] used DBSCAN with FBA and Pareto optimality to predict engineering strategies for maximized ethanol production in eight different microorganisms. In microbial engineering approaches the production rate of specific compounds and biomass production often represent competing objectives. Thus, the authors developed a multi-objective metabolic engineering approach to investigate how genetic manipulations could affect the production rates of one or more metabolites of interest. In their framework, DBSCAN was used to cluster multiple alternative flux solutions (where ethanol and biomass production act as competing objective functions). For each metabolically engineered organism, the approach identified clusters of solutions, of which one provided the optimal combination of ethanol and biomass production. In other clusters, a slight increase in ethanol production led to a decrease in biomass production. Thus, the integrative approach was able to predict the best engineering strategies for improved ethanol production without penalizing biomass production.

2.2. Integrating supervised learning and flux balance analysis

Supervised learning [69] approaches can be used to generate data-derived flux boundaries for FBA simulations and thus reduce the flux solution space [53]. Most common applications of supervised learning integrated with FBA include Linear Regression [70], Regularized Regressions such as Lasso Regression [71] and Elastic Net Regression [71], Decision Trees [72], Artificial Neural Networks [73], and Support Vector Machines [74]. In the following, we will highlight some of these integrative studies.

Kim *et al.* [60] presented a Multi-omics Model and Analytics (MOMA) framework for omics-driven predictive modeling of multilayer interacting data (transcriptional regulatory, protein–protein interaction, and metabolic reaction network). MOMA uses an ensemble of ML techniques, including Recurrent Neural Networks [75], Lasso Regression, and Ensemble Learning [76], in combination with FBA to predict and analyze the growth of *E. coli*. The authors implemented ML approaches on the aggregated multi-omics data to select features for constraining FBA simulations. Recurrent Neural Networks were used to extract features from transcriptomics data, while Lasso Regression and Ensemble Learning were used to extract features from protein and metabolomics data. When compared with two other approaches, namely Metabolism and Gene Expression Model [77] and Expression Balance Analysis [78,79], MOMA predicted genome-wide expression and growth with the highest performance.

Elastic Net is, besides Lasso, another regularized Regression method. Occhipinti *et al.* [80] integrated Elastic Net Regression with FBA to identify genes correlating with the overproduction of rhamnolipids in the soil bacterium *Pseudomonas putida*. The authors reconstructed a generic metabolic model by incorporating reactions related to rhamnosyltransferase enzymes and integrated it with 40 samples of gene expression data of *P. putida*. For each condition-specific sample, FBA was used to estimate optimal flux distributions for maximal rhamnolipids production. The predicted flux rates were grouped into wild and mutant and Elastic Net Regression was used to identify reactions and pathways which differentiated these two groups. The integrative approach revealed two pathways involved in purine metabolism and fatty acid biosynthesis to be significant for rhamnolipids production.

Support-Vector Machines are supervised learning models able to address both linear and non-linear classification problems. These models separate data into classes finding possible hyperplanes (decision boundaries) based on the maximum distance between classes of data points.

Shaked *et al.* [81] integrated Support Vector Machines and FBA to develop 'model-based phenotype predictors' as indicators for drug side effects. In their analysis, the authors used a comprehensive data set of human disease and drugs responsive phenotypic data. In their workflow, Flux Variability Analysis was applied in drug-targeted gene knockout genome-scale metabolic models of human cells (generic). The resulting flux boundaries were used to

create a feature matrix in which each drug was represented as columns of upper and lower flux boundaries. Support Vector Machines were then trained on a known drug side effect matrix for the classification of potential drug side-effects. By examining the perturbation effects on the overall flux distributions, this analysis enabled the classification of drugs as either side effect-causing or not side-effect causing.

DiMucci *et al.* [82] used Random Forests, an ensemble ML classification method based on Decision Trees and combined it with FBA to study latent interspecies interactions in microbial communities. For this purpose, the author used dynamic FBA which considered total biomass as a time-dependent variable. The difference between final biomass accumulation values for each microbe in co– and mono-culture was used to calculate the relative yield. On the basis of those relative yield values Random Forest learning was used to classify metabolites showing competitive or facilitative interactions. The study identified fructose exchange to be crucial for the competitive interaction between different microbes and thus contributed to understanding hidden interactions in microbial communities.

2.3. Integrating supervised and unsupervised learning with flux balance analysis

Different ML techniques can be used in combination and in this section we review studies integrating FBA with both supervised and unsupervised learning.

Bordbar *et al.* [61] developed a method termed 'unsteady-state FBA' and combined it with PCA and Linear Regression. The approach enables integrating time-course metabolomics data with genome-scale metabolic models to predict metabolic fluxes under dynamic conditions. The method was tested on three time-course metabolomics datasets for red blood cells, platelets, and *S. cerevisiae*, and for steady-state data from *E. coli*. First PCA was applied to each metabolomics dataset to divide the intracellular and extracellular metabolite profiles into discrete time profiles. For each of these discrete time profiles, Linear Regression was applied to determine the rate of change of each metabolite. Metabolites with a significant rate of change were used as constraints for unsteady-state FBA while unmeasured metabolites were considered to be at steady-state. The approach successfully explained the utilization of extracellular citrate by red blood cells. The inclusion of intracellular metabolite profiles with unsteady-state FBA provided more accurate flux predictions than classical FBA.

Jalili *et al.* [83] used Random Forest classification together with PCA and FBA to identify cancer-specific metabolic signatures in an omics-data integrative genome-scale model of human metabolism (Recon3D). FBA was employed to calculate flux distributions for each cancer model. FBA-based features were then extracted using PCA and Random Forest approaches. PCA indicated the highest variation (response variables) in the flux distributions of cancer models. Random Forest then used these response variables to classify key fluxes (indicating which sub-cellular systems were affected by the corresponding processes). Using this approach, the authors discovered different metabolic patterns (extracellular transport, mitochondrial transporters, fatty acid synthesis, and the pentose phosphate pathway) which differentiate normal and cancer cell metabolism.

Yaneske *et al.* [84] employed a combination of K-means Clustering, PCA, Elastic Net Regression, and FBA as a metabolic age predictor from age-associated gene expression data. The authors used gene expression data from individual humans to determine the correlation between age and metabolic flux distributions. To this end, gene expression data was used to construct individual-based genome-scale metabolic models of CD4 T-helper cells. K-means Clustering analysis of the predicted flux distributions was used to group individuals and PCA was used to determine variation in the flux distributions of each individual group. Those grouped flux distributions were then analyzed by Elastic Net Regression to select strongly correlated variables. The authors found that the combination of FBA and ML could act as a multi-omics-based age predictor providing more accurate predictions of chronological age than transcriptome-based approaches alone.

Vijayakumar *et al.* [85] presented a pipeline which integrates genome-scale metabolic modeling with multi-omics data and applies FBA, PCA, K-means Clustering, and Lasso Regression for improved phenotypic prediction (*i.e.* phototrophic growth) in the renewable-biofuels-producing cyanobacterium *Synechococcus* sp.. In this pipeline, gene expression data were integrated with a genome-scale metabolic model of *Synechococcus* sp. to generate condition-specific models (light intensity, temperature, salinity, and oxygen or carbon dioxide). Multi-objective FBA with quadratic optimization was implemented to obtain unique flux distributions (where biomass was considered as the primary and ATP maintenance, photosystem I or II as the secondary objective). ML approaches were then applied to identify key genes and reactions for different conditions. PCA was used to determine variance in the flux distribution for each objective pair. Further, K-means Clustering was used to identify clusters of different growth conditions and Lasso Regression was applied to select reactions related to growth rates. The combination of FBA and ML methods revealed genes and reactions related to growth-promoting or growth-limiting conditions and thus aided understanding the mechanisms related to the alternation of light intensity and salinity.

Culley *et al.* [86] proposed a framework in which they integrated FBA with Multimodal Artificial Neural Networks to analyze growth-related mechanisms of several *S. cerevisiae* strains. Aim of the study was to overcome the black-box limitation of data-driven ML approaches by integrating it with mechanistic knowledge (flux distributions). The framework was tested in a *S. cerevisiae* genome-scale metabolic model and gene expression data of 1,484 single gene knock-out strains of *S. cerevisiae* combined with their relative growth rates. Growth rates were used as constraints for pFBA. The resulting flux distributions and gene expression data were used as input for Multimodal Artificial Neural Networks. To this end, Neural Networks were stacked with different layers, where gene expression data and metabolic fluxes were used as an individual layer, the concatenated gene expression and metabolic fluxes were used as an additional hidden layer. Also, a feed-forward hidden layer was used for better understanding cross-modal relationships. The results of the study suggest that the integrated analysis of flux distributions with the gene expression data improves Multimodal Artificial Neural Networks prediction in comparison to training with individual data sets.

In a follow-up study, Magazzù *et al.* [87] compared the performance of Multimodal Artificial Neural Networks and Lasso Regression using the same test case of *S. cerevisiae*. The study was conducted in two ways - single-view fashion, where gene expression data and metabolic fluxes distributions were separately used and in multi-view fashion, where both data sets were concatenated and then used as input for Lasso Regression and Multimodal Artificial Neural Networks. The authors demonstrated that Lasso Regression can outperform Multimodal Artificial Neural Networks and the multi-view approach improved output predictions, the interpretability of the input features as well as biological understanding.

2.4. Integrating Deep learning with flux balance analysis

Deep Learning is a branch of ML which is based on Artificial Neural Networks. In recent years, Deep Learning has emerged as a powerful method for image classification, speech recognition,

and biomedical signal processing, as well as integrating multi-omics biological data [88,89]. Given a large training set, Deep Learning methods can be used for classification, regression, and dimensionality reduction and can deal with highly non-linear problems, as well as unstructured data [89,90].

Artificial Neural Networks are composed of connected layers of artificial neurons called units. Reminiscent to real neurons, each unit represents an activation function that translates the strength of the unit inputs into an integrated output signal. The number of units in each layer, the number of layers, the pattern of connectivity, and the type of individual activation functions describe the Artificial Neural Networks 'architecture' and determine their applicability and performance. Artificial Neural Network architectures with more than one "hidden" layer between the input and the output layers are called 'deep neural networks'. Applications of Deep Neural Networks in FBA studies utilize their ability to address highly non-linear classification problems as well as the capability to be trained on very heterogeneous data sets [91].

Two notable, preprint studies combine FBA with a special type of Deep Neural Networks called Autoencoders. Autoencoders are Deep Neural Networks, whose architecture contains a 'bottleneck' termed 'code' layer of a relatively low number of units. During training, the network encodes complex data in a low-dimensional 'code' layer (encoding step) and then reconstructs the input data from that code (decoding step). The consequence of such operation is noise reduction and extraction of significant data features in the code layer. Thus, Autoencoders are widely used to *e.g.* process noisy image data [92] or sound recordings [93]. In systems biology, Autoencoders are successfully used to link layers of omics data and extract features from *e.g.* transcriptomic profiles related to specific metabolic functions and macro-phenotypes.

In a representative study, Guo *et al.* [94] combined FBA with Deep Neural Networks. The authors used a five-layer Autoencoder model, with two encoder layers representing gene expression and protein abundance of *E. coli*, respectively and the code layer representing quantitative phenotypes. In this study, the connectivity between the layers was based on known biological interactions, *e.g.* between transcripts and their protein products. To define connectivity between the protein and phenotype layer FBA of a genome-scale model of *E. coli* was used to select proteins essential for a given phenotype. In the pre-training step, the Autoencoder reduced the dimensionality of the data to the variance components that could be used to link gene expression with phenotypes. In the consecutive training step, the decoder layer was removed from the network and the pre-trained network was 'fine-tuned' using experimentally measured phenotype data. The approach, termed 'Deep-Metabolism' significantly improved genotype-phenotype predictions in comparison to fully connected Artificial Neural Networks.

Variational Autoencoders are probabilistic generative models, which once trained, sample from trained distributions to generate input-like data in the decoding step. Barsacchi *et al.* [95] developed a framework termed Gene Expression latEnt Space Encoder (GEESE), which uses Variational Autoencoders for learning the inherent structure of gene expression data related to the regulation of metabolic fluxes in *E. coli*. In the study, the reference flux distributions were estimated by FBA using gene expression across a wide range of treatments as constraints. In parallel, the fluxes were estimated in an FBA-independent method by training a Variational Autoencoder coupled to a 'flux approximator' (a fully connected Artificial Neural Networks). The study showed that Autoencoders can be used to identify gene expression patterns related to the regulation of metabolic fluxes and generate synthetic gene expression data which can translate to FBA-based flux distributions.

3. Kinetic models and flux balance analysis

Due to the underlying steady-state assumption, FBA is usually restricted to model non-dynamic processes. To overcome this drawback several dynamic or pseudo-dynamic extensions to FBA have been developed [18,37,96,97]. Of these, the static dynamic FBA (dFBA) formalism is computationally feasible even for large-scale metabolic networks and has thus been widely used to suggest metabolic engineering strategies for the overproduction of desired compounds in microbial strains and the design and evolution of microbial communities [98–101]. In brief, the static dFBA implementation divides the modeled time frame into several intervals and iteratively solves an instantaneous FBA optimization problem at the beginning of each time interval followed by dynamic integration over the interval. This way, exchange fluxes predicted by FBA can be translated into updated metabolite concentrations which in turn serve as an input for the next iteration step. This static dFBA approach has further been used to couple FBA and ordinary differential equations (ODEs) models and in particular physiology-based pharmacokinetic (PBPK) models. Thus, here we treat PBPK models as an extension to dFBA. A general scheme for the model integration is shown in (Fig. 3). In this section, we focus on selected studies that combine PBPK or other kinetic models with FBA to and we demonstrate how these models can be applied to tackle a wide range of biological questions such as treatments for diseases of human metabolism, microbial community dynamics, or evolutionary or developmental questions in plant biology.

3.1. Integrating models of Physiology-based pharmacokinetics and flux balance for human tissues

PBPK models are routinely used by the pharmaceutical industry to determine the distribution and effects of drugs on a whole-body scale [102–104]. These models are ODE systems with hundreds of equations which describe the absorption, distribution, metabolization, and excretion of compounds and commonly include kinetics for blood flow, exchange rates for organs or tissues, and knowledge of anatomy, physiology, and compound properties. Thus, these models are heavily parameterized and require extensive sets of kinetic information. Several toolboxes for modeling PBPK models which already contain collections of the necessary physiological parameters exist such as MUFINS [105] and the commercial tools PKSim® and MoBi® (part of the Computational Systems Biology Software Suite of Bayer Technology Services GmbH (Leverkusen, Germany)). On the contrary, FBA models rely only on the reaction stoichiometries of the underlying metabolic network and allow simulating metabolism at a large scale. Thus, the integration of PBPK and FBA models allows linking cellular processes to whole-body models. The integration procedure can be rationalized by assuming a separation of time constants. Metabolic reactions are typically much faster than processes such as absorption, distributions between compartments, and clearance in the cell. Thus, one can assume that the metabolic network reaches steady-state within the PBPK model's integration time-step.

A formalism for integrating PBPK and FBA models was first introduced by Krauss *et al.* [106]. The authors defined feedforward or indirect coupling as when concentrations of a compound in the PBPK model constrain enzyme activities and thus flux boundaries in the FBA model. Feedback or indirect coupling is defined as when FBA flux solutions are used to set reaction rates in the PBPK model. It is important to notice that unlike in dFBA where FBA simulations and dynamic integration form an iterative cycle the coupling of PBPK and FBA models can involve feedforward, feedback integration or both. As a case study, the authors integrated a genome-scale model of human liver cells with a PBPK model of an adult

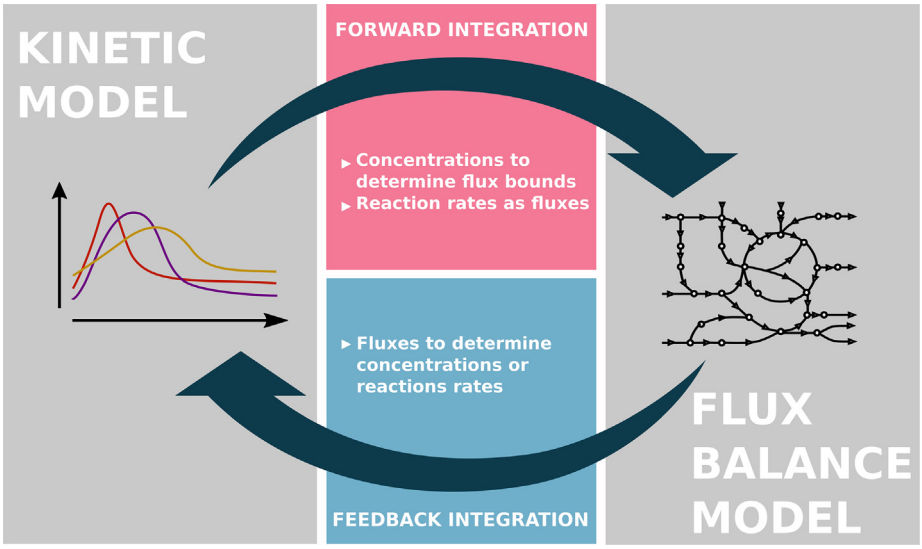

Fig. 3. Schematic workflow of forward and backward integration of kinetic and flux balance models. Initial concentrations can be either fed into the kinetic model to calculate an initial response and thus new concentrations and reaction rates or they can be used to determine maximum uptake rates in the flux balance model by dividing the available concentration of metabolite X by the simulated time interval. Based on the imposed rate constraints the flux balance model can be used to simulate metabolism at a large-scale. Model outputs, such as growth rates or flux values can be used either directly or to calculate updated metabolites concentrations which then feedback into the kinetic model. Note that we distinguish between reaction rates from the kinetic model and steady-state reaction rates, *i.e.*, fluxes in the stoichiometric model.

human and studied three distinct scenarios. Firstly, they modeled the distribution and action of allopurinol for treating hyperuricemia and found that the predicted uric acid concentration after multiple dosing was in line with *in vivo* observations. Secondly, they demonstrated the approach's capability for identifying biomarkers specific for the effect of impaired ammonia metabolism on blood plasma levels. Thirdly, the model was applied to study paracetamol-induced toxication on liver function. The presented case studies highlighted the applicability of the modeling framework in scientific research, clinical applications, and drug development.

In the following years, several other studies integrated models of liver metabolism with PBPK models to study various metabolic response processes. This included modeling alcohol metabolism in the human body [107], simulating the regulation of blood glucose in type I diabetes [108], predicting amino acid biomarkers for a set of inborn errors of metabolism [109], understanding the metabolic effects of the female sex hormone estradiol [110], and studying the effect of isoniazid (an antibacterial drug against *Mycobacterium tuberculosis*) on human liver metabolism [111].

Guebila and Thiele [112,113] combined a whole-body PBPK model with a FBA model of the small intestine enterocyte to study the interplay between dietary amino acids and the adsorption of the Parkinson's disease prodrug levodopa. In the brain levodopa is biotransformed into dopamine, allowing the reversal of Parkinson's disease symptoms. While there were indications that dietary amino acids could improve the absorption of levodopa, a systemic analysis of dietary amino acids uptake and levodopa absorption was still lacking. The model integration was achieved by setting the levodopa absorption rate as an upper bound for the FBA model and subsequently, the obtained flux values were set as rates for the PBPK model. By applying their model to study different nutritional strategies the authors found that several dietary scenarios could have an effect on levodopa's bioavailability. Altogether, their studies contributed to gaining further understanding of diet-levodopa interactions in Parkinson's disease patients.

Shepelyuk *et al.* [114] analyzed the energetics of blood cells by combining a PBPK and a FBA model. More precisely, the authors investigated the metabolism of quiescent platelets in the context of a human glucose/insulin/glucagon PBPK model and systematically investigated the mutual influence of platelet metabolism and blood glucose levels. The authors found that platelet metabolism had a minor effect on overall blood glucose levels, the influence of blood glucose on platelet metabolism strongly depended on the fraction of the platelet's glucose transporter, and that platelets stored sufficient glycogen for a case of prolonged fasting.

Some of the presented studies were also previously reviewed here [115–117]. Noteworthy, Thiele *et al.* [118] outlined a computational pipeline for personalizing these multi-scale models by including gut-microbial metabolism and dietary information into the model formulation and thus highlighted the potential applications in personalized medicine.

It is worth noting here that some studies applied other dynamic variants of FBA to study human metabolism. For instance, Nilsson *et al.* developed a simplified version of dFBA which they referred to as piecewise flux-balance analysis or pwFBA [119]. This approach uses a set of ODEs to describe how cell dry weight and external metabolite concentrations change over time and uses forward integration to fit uptake fluxes and feedback integration to identify time intervals suitable for steady-state analysis by finding metabolic depletion events. The approach was applied in conjunction with experimental measurements to study liver cancer cell metabolism and put forward hypotheses about the involvement of glutamate in the cytosol and the mitochondria and pointed at potential drug targets that could reduce growth of liver cancer cells.

3.2. Combined kinetic and flux balance models to study microbes and microbial communities

Dynamic FBA has been extensively used to study growth of microbes and microbial communities in natural and industrial systems (reviewed here [98,100,120,121]) and several toolboxes and

strain optimization procedures are available [122,123]. Several modeling approaches extended the temporal nature of dFBA by accounting for resource allocation and diffusion processes or regulatory processes [124].

Harcombe *et al.* presented 'Computation of Microbial Ecosystems in Time and Space (COMETS)' - a modeling framework which integrates dFBA with diffusion on a lattice and applied it to design microbial communities [125]. To achieve the model integration, the authors coupled a hybrid kinetic-dFBA model and a diffusion model which employs a standard 2D diffusion equation on a lattice. In the integrated model the simulation of diffusion steps alternates with growth steps. The authors modeled synthetic two- and three-species consortia. Due to the spatial nature of the modeling framework, they were able to answer questions related to colony arrangements and cross-feeding interdependencies. They found that the assumption that individual species locally optimize intracellular resource allocation determined interspecies interactions and microbial community dynamics. Together with experimental testing of their findings, they demonstrated the predictability of complex microbial communities.

Phalak *et al.* modeled a community of *Pseudomonas aeruginosa* and *Staphylococcus aureus*, two bacteria found in chronic wound film [126]. To achieve temporal and spatial resolution they solved a series of FBA problems using genome-scale models of the two bacteria with different experimentally motivated objectives. To describe convective and diffusional processes in the biofilm layer, the authors used partial differential equations which were spatially discretized to yield large-sets of ODEs which were then coupled with the FBA problem. The study enabled quantifying the impact of nutrient competition, cross-feeding, and inhibition of *S. aureus* by a small molecule secreted by *P. aeruginosa*. Overall, the study demonstrated the relevance of such integrative modeling studied for biomedical applications.

Additionally, several multi-layer integrative approaches exist, which couple signaling, regulatory and metabolic networks by extending the dFBA formalism. For instance, Lee *et al.* presented integrated dynamic FBA [127] which integrates 'slow' reactions in a time-delayed manner with the 'fast' steady-state reactions of the stoichiometric metabolic model. The authors exemplified their modeling framework by simulating signaling, metabolic, and regulatory processes using a genome-scale model of the yeast strain *S. cerevisiae*. In another study, Covert *et al.* presented a method called 'integrated FBA', which integrates regulatory FBA (an FBA extension which uses Boolean logic to consider reactions active or inactive) and ODEs [128]. Their modeling employs forward and feedback integration and was applied to study *E. coli* wild-type and single gene perturbation phenotypes for diauxic growth on glucose/lactose and glucose/glucose-6-phosphate. The authors found their approach to have higher predictive power than regulatory FBA and ODE modeling individually.

3.3. Combined kinetic and flux balance models to study developmental and evolutionary dynamics in plants

In planta, Grafahrend-Belau *et al.* presented a multi-model dFBA approach by coupling a previously published dynamic whole-plant functional model [129,130] and a large-scale stoichiometric model of barley metabolism with 890 metabolites and 971 reactions to study carbon and nitrogen balance under different environments [28]. The functional plant model comprises a photosynthesis model and a multiorgan model describing the dynamics of carbon and nitrogen allocation between intracellular compartments and organs across the entire life cycle of the barley plant. This model was coupled to the metabolic model via leaf-specific exchange rates of Suc, Asn, and Gln with the phloem, starch synthesis, and Suc transport into the vacuole. Additionally, a time-dependent seed biomass composition was included in the optimization procedure. This framework enabled the authors to systematically analyze metabolic interactions during barley development. The analysis revealed a sink-to-source shift of the barley stem to meet the nutrient requirements of the growing seed caused by senescence-induced declining leaf source capacity.

Mallmann *et al.* studied the evolution of C4 photosynthesis, a heat-adapted form of photosynthesis in which initial C-fixation and re-fixation by the main photosynthetic enzyme RuBisCO are spatially separated into two distinct cell types. This separation and subsequent C enrichment around RuBisCO suppress the RuBisCO's oxygenase activity and thus prevents the need for cost intensive recycling reactions. To study the transition from the ancestral C3 to the specialized C4 state in the genus *Flaveria*, the authors combined a kinetic model of photosynthesis and a stoichiometric model of leaf metabolism [131–133]. The kinetic model yielded key parameters of photosynthesis of C3-C4 intermediates and was used to constrain a genome-scale stoichiometric model of C4 photosynthesis. Amongst other things, the authors identified that nitrogen mis-balance in intermediate states was a driving force for the evolution from C3 to C4 photosynthesis.

Both studies used a mechanistic model to constrain a flux balance model, *i.e.*, performed a feedforward integration only. Recently, Shaw and Cheung presented a multi-tissue dFBA model of plant metabolism which employed both a feedforward and feedback integration and was used to analyze carbon and nitrogen metabolism and resource partitioning during *Arabidopsis thaliana* growth [31]. To this end, they employed their previously presented diel FBA framework [37] and extended it to represent a leaf-root system in which leaf and root biomass and thus uptake and metabolic capacities of the subsystems were dynamically updated. N uptake dynamics were simulated by calculating the maximum nitrate uptake rate by a Michaelis-Menten equation (feedforward) and in dependence of the present root biomass (feedback). Analogously, the maximum photon uptake was updated in each simulation step by determining the current projected leaf area (feedback). The authors used their model to study the effect of different nutrient availability as well as herbivory, the addition of nutrients in the soil, and shades on plant metabolism. Their approach demonstrated the predictive power of combining dynamically updated external conditions and adjusted metabolic capacities of different plant organs to study their overall contribution to plant growth.

4. Petri nets and flux balance analysis

Metabolic networks do not operate in isolation but are tightly intertwined with regulatory mechanisms such as signaling or gene-regulatory networks. Petri nets (PNs) offer a modeling paradigm for integrating these different biological networks into concise models. Carl Adam introduced PNs in the early 60 s as a graphical modeling formalism allowing for a general representation of interacting components in a network [134]. PNs are a multi-purpose modeling solution and have been applied to a variety of molecular networks, including gene regulation [135], signaling [136], and metabolic pathways [137]. They have also been widely used to model concurrency (events occurring in parallel) and conflicts (events using the same resources) in various systems [138].

In mathematical terms, PNs are bi-partite directed graphs comprising two non-overlapping sets of nodes - transitions and places - allowing for a general representation of interacting components in a network. A graphical representation of the PN formalism is given in Fig. 4. While places represent the components in a network, transitions represent their interactions. In a molecular biological context, places may represent single ions or atoms, simple chemi-

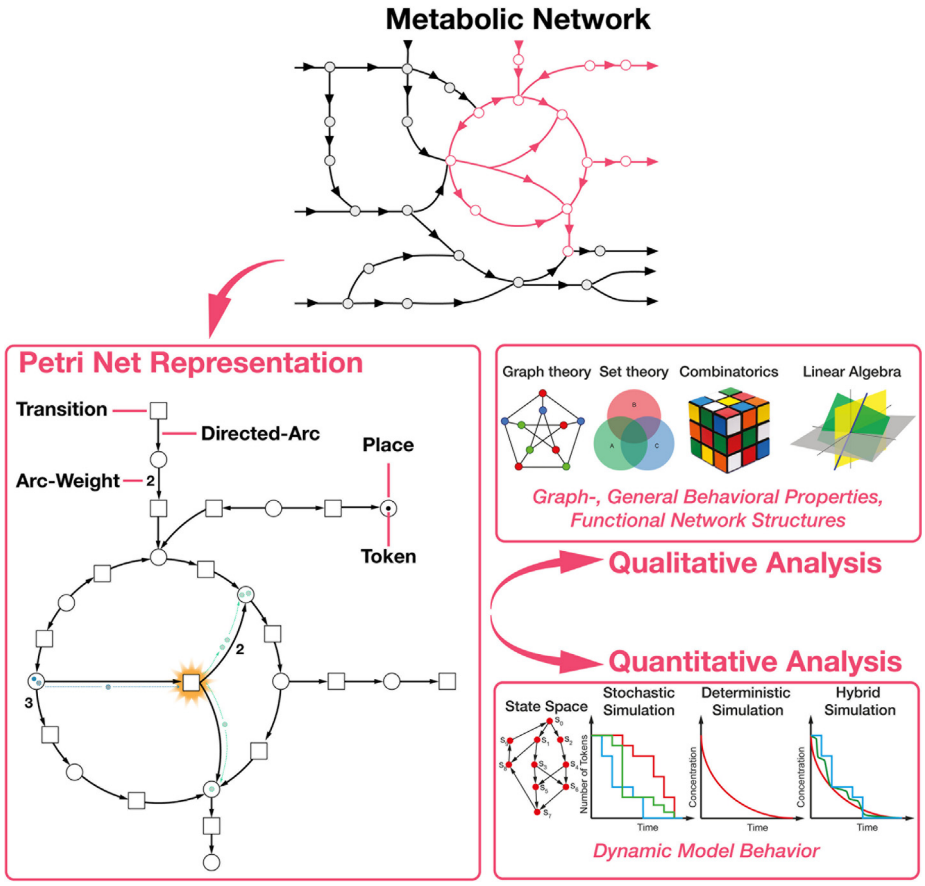

Fig. 4. Petri Nets Basics. Metabolic and other molecular networks can easily be translated and displayed by a PN graph. Metabolites, proteins, transcripts or genes are defined by places (circles), while reactions and interactions are encoded by transitions (squares). Directed arcs relate inputs (reactants) and outputs (products) of each transition. Arc-weights indicate stoichiometric coefficients. Tokens specify the value of a place *e.g.* the number of molecules or a discretized concentration level. A transition is enabled if its pre-places (reactants) hold at least as many tokens as defined by the arc-weight. An enabled transition may fire and thus remove or add tokens from its pre- and post-places. PNs allow for the application of various qualitative and quantitative analysis approaches [169,170]. Qualitative analysis approaches employ i.a. graph and set theory, combinatorics and linear algebra to determine graph and structurally-defined behavioral properties, as well as functional network structures. Quantitative analysis approaches focus on exploring the dynamic behavior using different simulation paradigms (stochastic, continuous, hybrid) or analytical state-space construction.

cal compounds or macromolecules such as lipids, proteins, or nucleic acids. Transitions describe, i.a. all kinds of reactions, structural modifications, transport, and binding processes. Directed arcs connect places with transitions and vice versa. Thus, arcs define places as substrates or products of a reaction encoded by the respective transitions. Tokens specify the value of a place. In the context of molecular networks, tokens refer to the number of molecules or a discrete concentration or activity levels. The distribution of tokens over all places in a PN represents the current state of the system. A transition is called enabled if its pre-places hold at least as many tokens as required by the corresponding arc-weight (stoichiometry). At each time step, only one of the enabled transitions is allowed to fire. While firing, the transition removes tokens from its substrate places and adds new tokens to its product places according to the stoichiometry indicated by the arc-weights. The introduction of timed Petri nets, particularly stochastic, continuous, and hybrid Petri nets, extends standard PNs by associating firing rates to each transition which in turn allow incorporating reaction kinetics, stochastic rate equations or both. The PN formalism offers several advantages over other modeling approaches such as ordinary or stochastic differential equations. Firstly, PN classes are interconvertible. Thus, network behavior can be explored using different simulation paradigms while maintaining the qualitative PN structure among the different classes. Secondly, the flexibility of the PN language allows modeling various biological systems, such as metabolic, signaling, and gene-regulatory networks, and thus creating multi-scale models of biological systems within one framework. Finally, compared to ODE models, the graphical notation using PNs is more accessible to biologists without in-depth mathematical training.

4.1. Metabolic network analysis using Petri nets

The application of PNs to metabolic networks allows for analyses of the underlying network structure, the identification of metabolic routes, and exploration of the state-space [139,140]. The state space is itself a directed graph, which encompasses all possible behaviors in the form of single connected states. Analytical approaches can be used to compute all possible states, based on the defined PN model. However, an increase in model size results in a broader range of possible behaviors reflected by a growing state space. In the worst-case scenario, the state space grows infinitely. This phenomenon is also called state-space explosion and cannot be handled with analytical approaches which makes the analysis of genome-scale metabolic networks using such approaches infeasible. Koch *et al.* [141] tackled the problem of state-space explosion by applying decomposition and network reduction techniques based on structural PN graph properties to extract steady-state pathways in a PN representation of the *A. thaliana* metabolic network. In the case of state-space explosion or infinite state spaces, simulative approaches are beneficial but often hampered by the lack of kinetic data. This data is needed to

define firing rates of the transitions based on *e.g.* mass action, Michaelis-Menten, or Hill kinetics. Thus, the modeler often has to rely on parameter estimates. Furthermore, the simulation efficiency of large-scale systems is low, either due to numerical problems of the ODE-solvers or long runtimes for stochastic simulation algorithms [142]. Thus, PNs have been integrated with FBA to tackle some of these shortcomings. In the following, we summarize studies combining both modeling frameworks.

4.2. Quasi steady state Petri nets

Quasi Steady State Petri Nets (QSSPNs) are the first approach for combining PNs and FBA [143]. QSSPNs facilitate simulating the effects of signaling or gene-regulatory mechanisms on metabolic models and allows incorporating rate constants which is an advantage over standard FBA techniques. Both types of networks, regulatory and metabolic networks, are connected by two special sets of nodes which are part of look-up tables. In analogy to the 'feed-forward' and 'feed-back' formalism introduced for coupling PBPK and FBA models, PNs employ constraint and objective places. Constraint places connect the regulatory mechanisms with the metabolic network. The look-up table of a constraint place defines how the flux boundaries of the connected reactions are updated based on the number of tokens on this constraint place. Vice versa, objective places connect metabolism with regulatory mechanisms. The look-up table of an objective place defines how the number of tokens of the objective has to be updated based on the value of the objective function of the metabolic network. Instead of obtaining a steady-state flux distribution using FBA, the QSSPN algorithm employs Monte Carlo sampling of flux trajectories which can then be further examined to detect dynamic behavior of interest.

Fisher *et al.* [143] employed QSSPNs to bile acid homeostasis in the human liver by combining a model of human hepatocyte metabolism [144] with gene regulation and signaling processes which are known to be involved in bile acid homeostasis. In this process, the liver reacts to cholesterol by clearing and converting it to bile acid. However, increased bile acid levels are toxic and thus cholesterol clearance must be slowed down, resulting in a negative feedback loop. The QSSPN analysis could reproduce most of the experimentally observed behavior. *In silico* gene knock-out studies identified genes and signaling molecules responsible for the physiological imbalance in cholesterol clearance and disturbance of the bile acid pool. Furthermore, examining the simulated trajectories identified the coupling of the chenodiol and cholate branch of the bile acid synthesis pathway through the regulatory network. The QSSPN algorithm is part of the SurreyFBA software [145], which has been integrated with MUFINS, a multi-formalism interaction network simulator [105].

While QSSPNs allows simulating only single-cell models, its agent-based extension AB-QSSPN [146] enables simulations of cell populations with millions of cells even on small clusters of PCs. Applying AB-QSSPNs to the bile acid homeostasis model demonstrated that the cooperative effect of cellular communication resulted in faster mean response times to the increased burst of bile acid. As a result, the negative feedback loop, which slows down cholesterol clearance and prevents toxic bile acid levels, responded more swiftly than in the previous study. This insight on the cooperative effect of cellular communication is even more relevant for modeling and simulating tumor growth or immune responses where cell–cell communication might be more prominent.

4.3. Extended stochastic Petri nets

Extended Stochastic Petri Nets (ESPNs) are another approach for integrating PNs with FBA in a general-purpose modeling framework that combines different solution techniques within a single graphical formalism [147]. Regulatory and metabolic processes are integrated into one coherent PN using the tool GreatSPN [148]. ESPNs distinguish between two types of transitions: (1) standard transitions which apply mass action law as firing rate, and (2) general transitions which use more specialized functions, such as Michaelis-Menten kinetic or Hill kinetic as firing rates. General transitions encode reactions that are considered to be at steady-state, including reactions of the metabolic network. An ODE system is derived from the ESPN model and a file storing the kinetics of the general transitions. For a biochemical reaction system, a PN is a structured description of the corresponding ODE system [149]. The chemical reaction scheme and its PN representation refer to the same stoichiometric matrix. Transitions in the PN also describe reaction rate kinetics. Therefore, the structurally defined stoichiometric matrix of a PN and the provided reaction rate kinetics allow modelers to derive the ODE system and in turn describe the change of concentrations over time. Simone *et al.* applied ESPNs to a metabolic model of pancreatic cancer [147] which they extended with a sub-network for oxidative phosphorylation [150]. The transition speeds of the oxidative phosphorylation sub-network were computed using FBA (assuming the maximization of ATP and mitochondrial fumarate production). The authors showed that in the presence of oxygen, the oxidative phosphorylation provided the main source for ATP production. Furthermore, including the oxidative phosphorylation had a global impact on the system's behavior. In their example, FBA acted as a global source and sink system for specific metabolites such as ATP and mitochondrial fumarate.

4.4. Simulating Genome-Scale metabolic models using Petri nets

FBA's biomass function is a complex agglomeration of substrates with non-integer stoichiometries reflecting the ratios of various biomass components [151]. As such the biomass function cannot directly be translated into the PN formalism as the arc-weights which represent these stoichiometries in a PN must be integers. Thus, in order to include this information, it is necessary to derive an approximation. Self *et al.* [152] suggested a routine for deriving such an approximation which can then be used to simulate the dynamics of a metabolic network given as a PN. In the first step, the authors converted a model of *E. coli*'s core metabolism [153] into a PN. They stochastically simulated the behavior of the PN for a range of growth conditions using a discrete-time leap method for stochastic simulations [154]. In parallel, they obtained the corresponding values of the objective function by using FBA. To derive a proxy for the biomass function suitable for a PN formalism, the authors employed stepwise regression and a Random Forest-based variable selection algorithm [155]. The final regression model contained twelve predictors: seven variables for metabolites (*e.g.* fructose, D-glucose, oxygen, hydrogen), three variables for reactions (*e.g.* glutaminase, ribose-5-phosphate isomerase, succinyl-CoA synthetase), one variable for a combination of two growth conditions (ethanol, fructose, glucose, glutamine, acetaldehyde, glutamate, fumarate, malate, lactate, and succinate), and one variable for the aerobic/anaerobic condition. The authors argued that the observed regression function could mimic FBA's biomass function with high accuracy. However, while the derived proxy biomass function is a predictive measure for the biomass at steady-state, its composition of terms related to metabolites and reactions does not provide a biological interpretation of the model's behavior and currently does not capture time-dependent biomass compositions [156]. Furthermore, the dynamic analysis of different growth conditions showed that the biomass value of two paired growth conditions under aerobic conditions was always between 1 and 7% larger than the sum of biomass values for the single growth conditions. Thus, aerobic growth on two substrates

had a slightly super-additive effect, and thus biomass production was more effective. For a follow-up study, the authors suggested investigating the proxy biomass function with respect to transient states and the effect of dynamic changes. Furthermore, the authors suggested applying their analysis framework to an unreduced genome-scale metabolic network to further evaluate the capabilities of metabolic models based on PN for analyzing dynamic behavior.

4.5. Flexible nets

Júlvez *et al.* [157] introduced Flexible Nets (FNs) as a framework for modeling, analyzing and controlling dynamic biological systems with uncertainties. While FNs are inspired by PN formalism, their syntax and execution semantics differ significantly. FNs are tripartite graphs which, in addition to places and transitions, use a third vertex named handlers. FN consists of two connected subnets, an event net and an intensity net, separating the stoichiometry of the system (event net) from the reaction kinetics (intensity net). Handlers represent an intermediate layer between places (which model metabolites) and transitions (which model reactions) and allow incorporating parameter uncertainties, modeling partially observable systems, constrained control actions, and resource allocation and transient-states [158]. Therefore, FNs combine the modeling capabilities of both constraint-based models and differential equations. As in FBA approaches, the modeling of metabolic networks using FNs allows considering flux boundaries of input, output, and metabolic reactions and optimization problems defined by the objective functions in FBA. As a proof of principle, Júlvez *et al.* [157], demonstrate how well FNs perform in modeling, analyzing, and controlling by applying the formalisms to the simple system of glucose consumption and utilization in yeast. Furthermore, the authors used FN to build a quantitative and predictive model of Wilson disease (a heritable defect in copper utilization) and to evaluate the relative efficacy of different therapeutic options. In a follow-up study, the authors used the FN approach to model a continuous cell culture of HeLa cells in a bioreactor [159]. Here, a method that resembles dynamic FBA was used to analyze potential steady-states of the system with the objective to maximize the cell density in dependence on the system's dilution rate. The authors state that FNs can efficiently handle genome-scale metabolic networks for which kinetic information is mainly given by flux boundaries and linear dependencies on the concentration of metabolites [159].

5. Perspectives

Over the past two decades, FBA approaches have developed into a standard tool for modeling and analyzing large-scale metabolic networks in microorganisms, humans, and plant systems. Standard FBA's limitations to modeling steady-state metabolism and the lack of consideration of regulatory and kinetic information have been addressed by a plethora of various integrative approaches [39–41,43–45] some of which we reviewed here. Recently, omics data and the advent of machine learning techniques to analyze these data have opened up opportunities for combining selections of the most important variables in omics data sets with the explanatory power of mechanistic models. Deep Learning in combination with FBA enables integrating various data types as inputs for genome-scale metabolic models as well as developing metabolic engineering strategies. The major limitation of Deep Learning in systems biology is a still relatively small size of training sets with respect to the number of variables that can be quantified in a single sample using omics approaches. Thanks to technological progress in data acquisition and infrastructure this picture is quickly changing in biotechnology and medicine. However, applications in less data-rich fields are still in their infancy.

Both kinetic and PN models of regulation and metabolism enable the representation of complex processes and mapping their effect on metabolism. In this context, PNs are an appealing formalism due to its intuitive and executable modeling style and its capability to model different bio-molecular network types in a mathematically concise manner. However, kinetic models and especially PBPK models and some PN approaches require extensive knowledge of kinetic parameters such as maximum reaction rates v_{max} and affinity constants K_m of the modeled kinetic reactions. While these parameters might be available for humans, they are typically sparse for microbial and plant systems. Additionally, solving these models at a large-scale is computationally expensive and thus one of the critical issues which are hampering the application of kinetics- and PN-based approaches to selected model systems. Adapting existing tools to the recent advances in GPU technology will reduce computation time and permit the analysis of large multiscale networks. Together, these approaches pave the way for an integrated interpretation of omics data in their physiological context.

Data reproducibility and the development of community and cross-community standards remain big challenges in systems biology. FAIR (Findable, Accessible, Useable, and Interoperable) guiding principles [160] and the Systems Biology Markup Language (SBML) [161,162] set a standard for data and model management. However, the usage of non-standardized metabolite and reaction nomenclatures in metabolic models, the need for individualized model curation and modification steps as well as case-specific model integration procedures challenge easy reproducibility. These issues can only be solved by community-driven standards. With these challenges addressed, we envision data- and model-integrative approaches to become part of the standard portfolio in microbial strain design, plant breeding programs, and systems medicine applications.

CRediT authorship contribution statement

Ankur Sahu: Writing – original draft, Visualization. **Mary-Ann Blätke:** Writing - original draft, Visualization. **Jędrzej Jakub Szymański:** Writing - review & editing. **Nadine Töpfer:** Conceptualization, Writing - original draft, Writing - review & editing, Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kim WJ, Kim HU, Lee SY. Current state and applications of microbial genome-scale metabolic models. Curr Opin Syst Biol 2017;2:10–8.
[2] Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2:886–97.
[3] Fang X, Lloyd CJ, Palsson BO. Reconstructing organisms in silico: genome-scale models and their emerging applications. Nat Rev Microbiol 2020;18:731–43.
[4] Geng J, Nielsen J. In silico analysis of human metabolism: Reconstruction, contextualization and application of genome-scale models. Curr Opin Syst Biol 2017;2:29–38.
[5] Ryu JY, Kim HU, Lee SY. Reconstruction of genome-scale human metabolic models using omics data. Integr Biol 2015;7:859–68.
[6] de Oliveira Dal'Molin CG, Nielsen LK. Plant genome-scale metabolic reconstruction and modelling. Curr Opin Biotechnol 2013;24:271–7.
[7] Collakova E, Yen JY, Senger RS. Are we ready for genome-scale modeling in plants? Plant Sci 2012;191:53–70.

[8] de Oliveira Dal'Molin CG, Nielsen LK. Plant genome-scale reconstruction: from single cell to multi-tissue modelling and omics analyses. Curr Opin Biotechnol 2018;49:42–8.
[9] Sweetlove LJ, Ratcliffe RG. Flux-balance modeling of plant metabolism. Front Plant Sci 2011;2:38–38.
[10] Töpfer N. Environment-coupled models of leaf metabolism. Biochem Soc Trans 2021;49:119–29.
[11] Jensen MK, Keasling J. Synthetic Metabolic Pathways. Springer; 2018.
[12] Xu N, Ye C, Liu L. Genome-scale biological models for industrial microbial systems. Appl Microbiol Biotechnol 2018;102:3439–51.
[13] Cook DJ, Nielsen J. Genome-scale metabolic models applied to human health and disease. Wiley Interdiscip Rev Syst Biol Med 2017;9:e1393–e1393.
[14] Nilsson A, Nielsen J. Genome scale metabolic modeling of cancer. Metab Eng 2017;43:103–12.
[15] Li H, Yang Y, Hong W, Huang M, Wu M, Zhao X. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. Signal Transduct Target Ther 2020;5:1–23.
[16] Shameer S, Baghalian K, Cheung CYM, Ratcliffe RG, Sweetlove LJ. Computational analysis of the productivity potential of CAM. Nat Plants 2018;4:165–71.
[17] Shameer S, Vallarino JG, Fernie AR, Ratcliffe RG, Sweetlove LJ. Flux balance analysis of metabolism during growth by osmotic cell expansion and its application to tomato fruits. Plant J 2020;103.
[18] Töpfer N, Braam T, Shameer S, Ratcliffe RG, Sweetlove LJ. Alternative Crassulacean Acid Metabolism Modes Provide Environment-Specific Water-Saving Benefits in a Leaf Metabolic Model. Plant Cell 2020;32:3689–705.
[19] Bauer E, Thiele I. From network analysis to functional metabolic modeling of the human gut microbiota. MSystems 2018;3.
[20] Sen P, Orešič M. Metabolic modeling of human gut microbiota on a genome scale: an overview. Metabolites 2019;9:22–22.
[21] van der Ark KCH, van Heck RGA, Dos Santos VAPM, Belzer C, de Vos WM. More than just a gut feeling: constraint-based genome-scale metabolic models for predicting functions of human intestinal microbes. Microbiome 2017;5:1–13.
[22] Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 2017;35:81–9.
[23] Cho JS, Gu C, Han TH, Ryu JY, Lee SY. Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring. Curr Opin Syst Biol 2019;15:1–11.
[24] Dunphy LJ, Papin JA. Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. Curr Opin Biotechnol 2018;51:70–9.
[25] Moreira TB, Shaw R, Luo X, Ganguly O, Kim H-S, Coelho LGF, et al. A genome-scale metabolic model of soybean (Glycine max) highlights metabolic fluxes in seedlings. Plant Physiol 2019;180:1912–29.
[26] Shaw R, Cheung CYM. A mass and charge balanced metabolic model of Setaria viridis revealed mechanisms of proton balancing in C4 plants. BMC Bioinf 2019;20:1–11.
[27] Scheunemann M, Brady SM, Nikoloski Z. Integration of large-scale data for extraction of integrated Arabidopsis root cell-type specific models. Sci Rep 2018;8:1–15.
[28] Grafahrend-Belau E, Junker A, Eschenröder A, Müller J, Schreiber F, Junker BH. Multiscale metabolic modeling: dynamic flux balance analysis on a whole-plant scale. Plant Physiol 2013;163:637–47.
[29] Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T. Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. Proc Natl Acad Sci 2012;109:339–44.
[30] Gomes de Oliveira Dal'Molin C, Quek L-E, Saa PA, Nielsen LK. A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. Front Plant Sci 2015;6:4–4.
[31] Shaw R, Cheung CY. A dynamic multi-tissue flux balance model captures carbon and nitrogen metabolism and optimal resource partitioning during Arabidopsis growth. Front Plant Sci 2018;9:884–884.
[32] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28:245–8.
[33] Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol 2010;13:344–9.
[34] Antoniewicz MR. Methods and advances in metabolic flux analysis: a mini-review. J Ind Microbiol Biotechnol 2015;42:317–25.
[35] Nikoloski Z, Perez-Storey R, Sweetlove LJ. Inference and prediction of metabolic network fluxes. Plant Physiol 2015;169:1443–55.
[36] Dai Z, Locasale JW. Understanding metabolism with flux analysis: From theory to application. Metab Eng 2017;43:94–102.
[37] Cheung CYM, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C3 and crassulacean acid metabolism leaves. Plant Physiol 2014;165:917–29.
[38] Mahadevan R, Edwards JS, Doyle III FJ. Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 2002;83:1331–40.
[39] Imam S, Schäuble S, Brooks AN, Baliga NS, Price ND. Data-driven integration of genome-scale regulatory and metabolic network models. Front Microbiol 2015;6:409–409.
[40] Cruz F, Faria JP, Rocha M, Rocha I, Dias O. A review of methods for the reconstruction and analysis of integrated genome-scale models of metabolism and regulation. Biochem Soc Trans 2020;48:1889–903.
[41] Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 2012;3:299–299.
[42] Kim MK, Lun DS. Methods for integration of transcriptomic data in genome-scale metabolic models. Comput Struct Biotechnol J 2014;11:59–65.
[43] Robaina Estévez S, Nikoloski Z. Generalized framework for context-specific metabolic model extraction methods. Front Plant Sci 2014;5:491–491.
[44] Töpfer N, Kleessen S, Nikoloski Z. Integration of metabolomics data into metabolic networks. Front Plant Sci 2015;6:49–49.
[45] Töpfer N, Seaver SMD, Aharoni A. Integration of plant metabolomics data with metabolic networks: Progresses and challenges. Plant Metabolomics: Springer; 2018. p. 297–310.
[46] Sánchez BJ, Zhang C, Nilsson A, Lahtvee P, 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.
[47] Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási A-L, et al. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. Proc Natl Acad Sci 2007;104:12663–8.
[48] Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. PLoS Comput Biol 2012;8:e1002575.
[49] Goelzer A, Fromion V, Scorletti G. Cell design in bacteria as a convex optimization problem. Automatica 2011;47:1210–8.
[50] Goelzer A, Fromion V. Resource allocation in living organisms. Biochem Soc Trans 2017;45:945–52.
[51] O'brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 2013;9:693–693.
[52] Nilsson A, Nielsen J, Palsson BO. Metabolic models of protein allocation call for the kinetome. Cell Syst 2017;5:538–41.
[53] Rana P, Berry C, Ghosh P, Fong SS. Recent advances on constraint-based models by integrating machine learning. Curr Opin Biotechnol 2020;64:85–91.
[54] Xu C, Jackson SA. Machine learning and complex biological data 2019.
[55] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015;349:255–60.
[56] Gilpin W, Huang Y, Forger DB. Learning dynamics from large biological datasets: machine learning meets systems biology. Curr Opin. Syst Biol 2020.
[57] Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS Comput Biol 2016;12:e1004977–e1004977.
[58] Zampieri G, Vijayakumar S, Yaneske E, Angione C. Machine and deep learning meet genome-scale metabolic modeling. PLoS Comput Biol 2019;15: e1007084–e1007084.
[59] Bhadra S, Blomberg P, Castillo S, Rousu J. Principal metabolic flux mode analysis. Bioinformatics 2018;34:2409–17.
[60] Kim M, Rai N, Zorraquino V, Tagkopoulos I. Multi-omics integration accurately predicts cellular state in unexplored conditions for Escherichia coli. Nat Commun 2016;7:1–12.
[61] Bordbar A, Yurkovich JT, Paglia G, Rolfsson O, Sigurjónsson ÓE, Palsson BO. Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. Sci Rep 2017;7:46249–46249.
[62] Lever J, Krzywinski M, Altman N. Points of significance: Principal component analysis. Nat Methods 2017;14:641–3.
[63] Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. Neurocomputing 2017;267:664–81.
[64] Wang Y, Zhu L. Research and implementation of SVD in machine learning. IEEE 2017:471–5.
[65] Garcia-Dias R, Vieira S, Pinaya WHL, Mechelli A. Clustering analysis. Mach. Learn., Elsevier; 2020, p. 227–47.
[66] Sánchez BJ, Li F, Kerkhoven EJ, Nielsen J. SLIMEr: probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. BMC Syst Biol 2019;13:1–9.
[67] Dai D, Horvath N, Varner J. Dynamic sequence specific constraint-based modeling of cell-free protein synthesis. Processes 2018;6:132–132.
[68] Patané A, Jansen G, Conca P, Carapezza G, Costanza J, Nicosia G. Multi-objective optimization of genome-scale metabolic models: the case of ethanol production. Ann Oper Res 2019;276:211–27.
[69] Singh A, Thakur N. Sharma A. A review of supervised machine learning algorithms: Ieee; 2016. p. 1310–5.
[70] Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. John Wiley & Sons; 2021.
[71] Ogutu JO, Schulz-Streeck T, Piepho H-P. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions 2012;vol. 6:1–6.
[72] Navada A, Ansari AN, Patil S, Sonkamble BA. Overview of use of decision tree algorithms in machine learning. IEEE 2011:37–42.
[73] Van Gerven M, Bohte S. Artificial neural networks as models of neural information processing. Front Comput Neurosci 2017;11:114.
[74] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics 2018;15:41–51.

[75] Grossberg S. Recurrent neural networks. Scholarpedia 2013;8:1888.
[76] Dong X, Yu Z, Cao W, Shi Y, Ma Q. A survey on ensemble learning. Front Comput Sci 2020;14:241–58.
[77] Carrera J, Estrela R, Luo J, Rai N, Tsoukalas A, Tagkopoulos I. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of E scherichia coli. Mol Syst Biol 2014;10:735–735.
[78] Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, et al. In silico method for modelling metabolism and gene product expression at genome scale. Nat Commun 2012;3:1–10.
[79] O'brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 2013;9:693.
[80] Occhipinti A, Eyassu F, Rahman TJ, Rahman PKSM, Angione C. In silico engineering of Pseudomonas metabolism reveals new biomarkers for increased biosurfactant production. PeerJ 2018;6:e6046–e6046.
[81] Shaked I, Oberhardt MA, Atias N, Sharan R, Ruppin E. Metabolic network prediction of drug side effects. Cell Syst 2016;2:209–13.
[82] DiMucci D, Kon M, Segrè D. Machine learning reveals missing edges and putative interaction mechanisms in microbial ecosystem networks. Msystems 2018;3.
[83] Jalili M, Scharm M, Wolkenhauer O, Damaghi M, Salehzadeh-Yazdi A. Exploring the Metabolic Heterogeneity of Cancers: A Benchmark Study of Context-Specific Models. J Pers Med 2021;11:496.
[84] Yaneske E, Angione C. The poly-omics of ageing through individual-based metabolic modelling. BMC Bioinf 2018;19:83–96.
[85] Vijayakumar S, Rahman PK, Angione C. A hybrid flux balance analysis and machine learning pipeline elucidates metabolic adaptation in cyanobacteria. Iscience 2020;23:101818.
[86] Culley C, Vijayakumar S, Zampieri G, Angione C. A mechanism-aware and multiomic machine-learning pipeline characterizes yeast cell growth. Proc Natl Acad Sci 2020;117:18869–79.
[87] Magazzù G, Zampieri G, Angione C. Multimodal regularised linear models with flux balance analysis for mechanistic integration of omics data. Bioinformatics 2021.
[88] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[89] Sen P, Lamichhane S, Mathema VB, McGlinchey A, Dickens AM, Khoomrung S, et al. Deep learning meets metabolomics: A methodological perspective. Brief Bioinform 2020.
[90] Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol 2016;12:878–878.
[91] Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. Neurocomputing 2017;234:11–26.
[92] Yasenko L, Klyatchenko Y, Tarasenko-Klyatchenko O. In: Image noise reduction by denoising autoencoder. IEEE; 2020. p. 351–5.
[93] Banerjee R, Ghose A. In: A semi-supervised approach for identifying abnormal heart sounds using variational autoencoder. IEEE; 2020. p. 1249–53.
[94] Guo W, Xu Y, Feng X. DeepMetabolism: a deep learning system to predict phenotype from genome sequencing. ArXiv Prepr ArXiv170503094 2017.
[95] Barsacchi M, Terre HA, Lió P. GEESE: Metabolically driven latent space learning for gene expression data. BioRxiv 2018:365643–365643.
[96] Radhakrishnan M, Edwards S, Doyle FJ. Dynamic flux balance analysis of diauxic growth in Eschericha coli. Biophys J 2002;83:3–3.
[97] Rügen M, Bockmayr A, Steuer R. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. Sci Rep 2015;5:1–16.
[98] Kim OD, Rocha M, Maia P. A review of dynamic modeling approaches and their application in computational strain optimization for metabolic engineering. Front Microbiol 2018;9:1690–1690.
[99] Popp D, Centler F. µbialSim: constraint-based dynamic simulation of complex microbiomes. Front Bioeng. Biotechnol 2020;8.
[100] Perez-Garcia O, Lear G, Singhal N. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. Front Microbiol 2016;7:673–673.
[101] Bosi E, Bacci G, Mengoni A, Fondi M. Perspectives and challenges in microbial communities metabolic modeling. Front Genet 2017;8:88–88.
[102] Kostewicz ES, Aarons L, Bergstrand M, Bolger MB, Galetin A, Hatley O, et al. PBPK models for the prediction of in vivo performance of oral dosage forms. Eur J Pharm Sci 2014;57:300–21.
[103] Sager JE, Yu J, Ragueneau-Majlessi I, Isoherranen N. Physiologically based pharmacokinetic (PBPK) modeling and simulation approaches: a systematic review of published models, applications, and model verification. Drug Metab Dispos 2015;43:1823–37.
[104] Zhao P, Zhang L, Grillo JA, Liu Q, Bullock JM, Moon YJ, et al. Applications of physiologically based pharmacokinetic (PBPK) modeling and simulation during regulatory review. Clin Pharmacol Ther 2011;89:259–67.
[105] Wu H, Von Kamp A, Leoncikas V, Mori W, Sahin N, Gevorgyan A, et al. MUFINS: multi-formalism interaction network simulator. NPJ Syst Biol Appl 2016;2:1–10.
[106] Krauss M, Schaller S, Borchers S, Findeisen R, Lippert J, Kuepfer L. Integrating cellular metabolism into a multiscale whole-body model. PLoS Comput Biol 2012;8:e1002750–e1002750.
[107] Toroghi MK, Cluett WR, Mahadevan R. Multiscale metabolic modeling approach for predicting blood alcohol concentration. IEEE Life Sci Lett 2016;2:59–62.
[108] Wadehn F, Schaller S, Eissing T, Krauss M, Küpfer L. A multiscale, model-based analysis of the multi-tissue interplay underlying blood glucose regulation in type I diabetes. 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, IEEE; 2016, p. 1417–21
[109] Toroghi MK, Cluett WR, Mahadevan R. A multi-scale model of the whole human body based on dynamic parsimonious flux balance analysis. IFAC-Pap 2016;49:937–42.
[110] Sier JH, Thumser AE, Plant NJ. Linking physiologically-based pharmacokinetic and genome-scale metabolic networks to understand estradiol biology. BMC Syst Biol 2017;11:1–16.
[111] Cordes H, Thiel C, Baier V, Blank LM, Kuepfer L. Integration of genome-scale metabolic networks into whole-body PBPK models shows phenotype-specific cases of drug-induced metabolic perturbation. NPJ Syst Biol Appl 2018;4:1–11.
[112] Guebila MB, Thiele I. Systems pharmacology of levodopa absorption. Adv Syst Synth Biol 2015.
[113] Guebila MB, Thiele I. Model-based dietary optimization for late-stage, levodopa-treated, Parkinson's disease patients. NPJ Syst Biol Appl 2016;2:1–8.
[114] Shepelyuk TO, Panteleev MA, Sveshnikova AN. Computational modeling of quiescent platelet energy metabolism in the context of whole-body glucose turnover. Math Model Nat Phenom 2016;11:91–101.
[115] Maldonado EM, Leoncikas V, Fisher CP, Moore JB, Plant NJ, Kierzek AM. Integration of genome scale metabolic networks and gene regulation of metabolic enzymes with physiologically based pharmacokinetics. CPT Pharmacomet Syst Pharmacol 2017;6:732–46.
[116] Øyås O, Stelling J. Genome-scale metabolic networks in time and space. Curr Opin Syst Biol 2018;8:51–8.
[117] Martins Conde P do R, Sauter T, Pfau T. Constraint based modeling going multicellular. Front Mol Biosci 2016;3:3–3.
[118] Thiele I, Clancy CM, Heinken A, Fleming RMT. Quantitative systems pharmacology and the personalized drug–microbiota–diet axis. Curr Opin Syst Biol 2017;4:43–52.
[119] Nilsson A, Haanstra JR, Engqvist M, Gerding A, Bakker BM, Klingmüller U, et al. Quantitative analysis of amino acid metabolism in liver cancer links glutamate excretion to nucleotide synthesis. Proc Natl Acad Sci 2020;117:10294–304.
[120] Henson MA, Hanly TJ. Dynamic flux balance analysis for synthetic microbial communities. IET Syst Biol 2014;8:214–29. https://doi.org/10.1049/iet-syb.2013.0021.
[121] Gottstein W, Olivier BG, Bruggeman FJ, Teusink B. Constraint-based stoichiometric modelling from single organisms to microbial communities. J R Soc Interface 2016;13. https://doi.org/10.1098/rsif.2016.0627.
[122] Zomorrodi AR, Islam MM, Maranas CD. d-OptCom: Dynamic multi-level and multi-objective metabolic modeling of microbial communities. ACS Synth Biol 2014;3:247–57. https://doi.org/10.1021/sb4001307.
[123] Zhuang K, Yang L, Cluett WR, Mahadevan R. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. DySScO strategy for strain design. BMC Biotechnol 2013;13:8–8. 10.1186/1472-6750-13-8.
[124] Chen J, Gomez JA, Höffner K, Phalak P, Barton PI, Henson MA. Spatiotemporal modeling of microbial metabolism. BMC Syst Biol 2016;10:21–21. 10.1186/s12918-016-0259-2.
[125] Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. Cell Rep 2014;7:1104–15.
[126] Phalak P, Chen J, Carlson RP, Henson MA. Metabolic modeling of a chronic wound biofilm consortium predicts spatial partitioning of bacterial species. BMC Syst Biol 2016;10:1–20.
[127] Lee JM, Gianchandani EP, Eddy JA, Papin JA. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. PLoS Comput Biol 2008;4:e1000086–e1000086.
[128] Covert MW, Xiao N, Chen TJ, Karr JR. Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. Bioinformatics 2008;24:2044–50.
[129] Mueller J, Eschenroeder A, Christen O, Junker B, Schreiber F. ProNet-CN model: a dynamic and multi-scale process network combining photosynthesis, primary carbon metabolism and effects of leaf nitrogen status. 2012 IEEE 4th Int. Symp. Plant Growth Model. Simul. Vis. Appl., IEEE; 2012, p. 289–96.
[130] Kang MZ, Dumont Y, Guo Y. Plant growth modeling, simulation, visualization and applications. Proceedings PMA12: The Fourth International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications, Shanghai, China, 31 October-3 November 2012 2012.
[131] Von Caemmerer S. Biochemical models of leaf photosynthesis. Csiro publishing 2000.
[132] de Oliveira Dal'Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK. C4GEM, a genome-scale metabolic model to study C4 plant metabolism. Plant Physiol 2010;154:1871–85.
[133] Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber APM, Westhoff P, et al. The role of photorespiration during the evolution of C4 photosynthesis in the genus Flaveria. Elife 2014;3:e02478–e02478.
[134] Petri Carl A. kommunikation mit automaten. PhD Univ Bonn West Ger 1962.
[135] Matsuno H, Doi A, Nagasaki M, Miyano S. Hybrid Petri net representation of gene regulatory network. Biocomput. 2000, World Scientific; 1999, p. 341–52.

[136] Sackmann A, Heiner M, Koch I. Application of Petri net based analysis techniques to signal transduction pathways. BMC Bioinformatics 2006;7:482–482.
[137] Koch I, Junker BH, Heiner M. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. Bioinformatics 2005;21:1219–26.
[138] Murata T. Petri nets: properties, analysis and applications. Proceed IEEE 1989;77(4):541–80.
[139] Heiner M, Koch I. Petri net based model validation in systems biology. Int. Conf. Appl. Theory Petri Nets: Springer; 2004. p. 216–37.
[140] Baldan P, Cocco N, Marin A, Simeoni M. Petri nets for modelling metabolic pathways: a survey. Nat Comput 2010;9:955–89.
[141] Koch I, Nöthen J, Schleiff E. Modeling the metabolism of Arabidopsis thaliana: Application of network decomposition and network reduction in the context of Petri nets. Front Genet 2017;8:85–85.
[142] Smallbone K, Mendes P. Large-scale metabolic models: From reconstruction to differential equations. Ind Biotechnol 2013;9:179–84.
[143] Fisher CP, Plant NJ, Moore JB, Kierzek AM. QSSPN: dynamic simulation of molecular interaction networks describing gene regulation, signalling and whole-cell metabolism in human cells. Bioinformatics 2013;29:3181–90.
[144] Gille C, Bölling C, Hoppe A, Bulik S, Hoffmann S, Hübner K, et al. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. Mol Syst Biol 2010;6:411–411.
[145] Gevorgyan A, Bushell ME, Avignone-Rossa C, Kierzek AM. SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. Bioinformatics 2011;27:433–4.
[146] Ptak W, Kierzek AM, Sroka J. AB-QSSPN: Integration of Agent-Based Simulation of Cellular Populations with Quasi-Steady State Simulation of Genome Scale Intracellular Networks. Int. Conf. Appl. Theory Petri Nets Concurr.: Springer; 2016. p. 113–22.
[147] Simone P, Laura F, Gianfranco B, Luciano M, Giulia S, Niccoló T, et al. Integrating Petri Nets and Flux Balance Methods in Computational Biology Models: a Methodological and Computational Practice. Fundam Informaticae 2020;171:367–92.
[148] Amparore EG, Beccuti M, Donatelli S. (Stochastic) model checking in GreatSPN. Int. Conf. Appl. Theory Petri Nets Concurr.: Springer; 2014. p. 354–63.
[149] Gilbert D, Heiner M. From Petri Nets to Differential Equations – An Integrative Approach for Biochemical Network Analysis BT - Petri Nets and Other Models of Concurrency - ICATPN 2006. In: Donatelli S, Thiagarajan PS, editors., Berlin, Heidelberg: Springer Berlin Heidelberg; 2006, p. 181–200.
[150] Roy M, Finley SD. Computational model predicts the effects of targeting cellular metabolism in pancreatic cancer. Front Physiol 2017;8:217–217.
[151] Palsson B. Systems biology. Cambridge University Press; 2015.
[152] Self T, Gilbert D, Heiner M. Derivation of a biomass proxy for dynamic analysis of whole genome metabolic models. Int. Conf. Comput. Methods Syst. Biol. 2018:39–58.
[153] Orth JD, Fleming RMT, Palsson BO. Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. EcoSal Plus 2010.
[154] Rohr C. Discrete-time leap method for stochastic simulation. Fundam Informaticae 2018;160:181–98.
[155] Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw 2010;36:1–13.
[156] Heinken A, Basile A, Thiele I. Advances in constraint-based modelling of microbial communities. Curr Opin. Syst Biol 2021.
[157] Júlvez J, Dikicioglu D, Oliver SG. Handling variability and incompleteness of biological data by flexible nets: a case study for Wilson disease. NPJ Syst Biol Appl 2018;4:1–12.
[158] Júlvez J, Oliver SG. Flexible Nets: a modeling formalism for dynamic systems with uncertain parameters. Discrete Event Dyn Syst 2019;29:367–92.
[159] Júlvez J, Oliver SG. A unifying modelling formalism for the integration of stoichiometric and kinetic models. J R Soc Interface 2020;17:20200341.
[160] Wilkinson MD, Dumontier M, IjJ Aalbersberg, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:1–9.
[161] Zhang F, Smith LP, Blinov ML, Faeder J, Hlavacek WS, Tapia JJ, et al. Systems biology markup language (SBML) level 3 package: multistate, multicomponent and multicompartment species, version 1, release 2. J Integr Bioinforma 2020;17.
[162] Keating SM, Waltemath D, König M, Zhang F, Dräger A, Chaouiya C, et al. SBML Level 3: an extensible format for the exchange and reuse of biological models. Mol Syst Biol 2020;16:e9110–e9110.
[163] Tefagh M, Boyd SP. SWIFTCORE: a tool for the context-specific reconstruction of genome-scale metabolic networks. BMC Bioinf 2020;21:1–14.
[164] Willmann S, Thelen K, Lippert J. Integration of dissolution into physiologically-based pharmacokinetic models III: PK-Sim®. J Pharm Pharmacol 2012;64:997–1007.
[165] Eissing T, Kuepfer L, Becker C, Block M, Coboeken K, Gaub T, et al. A computational systems biology software platform for multiscale modeling and simulation: integrating whole-body physiology, disease biology, and molecular reaction networks. Front Physiol 2011;2:4.
[166] Heiner M, Herajy M, Liu F, Rohr C, Schwarick M. Snoopy–a unifying Petri net tool. Int. Conf. Appl. Theory Petri Nets Concurr.: Springer; 2012. p. 398–407.
[167] Heiner M, Schwarick M, Wegener J-T. Charlie–an extensible Petri net analysis tool. Springer 2015:200–11.
[168] Heiner M, Rohr C, Schwarick M. MARCIE–model checking and reachability analysis done efficiently. Springer 2013:389–99.
[169] Blätke MA, Heiner M, Marwan W. Chapter 7 - BioModel Engineering with Petri Nets. In: Robeva RS, editor. Algebr. Discrete Math. Methods Mod. Biol., Boston: Academic Press; 2015, p. 141–92. 10.1016/B978-0-12-801213-0.00007-1.
[170] Blätke MA, Rohr C, Heiner M, Marwan W. A Petri-Net-Based Framework for Biomodel Engineering. In: Benner P, Findeisen R, Flockerzi D, Reichl U, Sundmacher K, editors. Large-Scale Netw. Eng. Life Sci., Cham: Springer International Publishing; 2014, p. 317–66. 10.1007/978-3-319-08437-4_6.