# PLOS ONE

RESEARCH ARTICLE

# Defining the nutritional input for genome-scale metabolic models: A roadmap

**Georgios Marinos**[1], **Christoph Kaleta**[1]*, **Silvio Waschina**[2]*

**1** Research Group Medical Systems Biology, Institute of Experimental Medicine, Kiel University, University Medical Center Schleswig-Holstein, Kiel, Schleswig-Holstein, Germany, **2** Division of Nutriinformatics, Institute for Human Nutrition and Food Sciences, Kiel University, Kiel, Schleswig-Holstein, Germany

* s.waschina@nutrinf.uni-kiel.de (SW); c.kaleta@iem.uni-kiel.de (CK)

## Abstract

The reconstruction and application of genome-scale metabolic network models is a central topic in the field of systems biology with numerous applications in biotechnology, ecology, and medicine. However, there is no agreed upon standard for the definition of the nutritional environment for these models. The objective of this article is to provide a guideline and a clear paradigm on how to translate nutritional information into an *in-silico* representation of the chemical environment. Step-by-step procedures explain how to characterise and categorise the nutritional input and to successfully apply it to constraint-based metabolic models. In parallel, we illustrate the proposed procedure with a case study of the growth of *Escherichia coli* in a complex nutritional medium and show that an accurate representation of the medium is crucial for physiological predictions. The proposed framework will assist researchers to expand their existing metabolic models of their microbial systems of interest with detailed representations of the nutritional environment, which allows more accurate and reproducible predictions of microbial metabolic processes.

OPEN ACCESS

## Introduction

A key approach in Systems Biology is the elucidation of the metabolic potential of an organism and the transformation of this information into networks of metabolic reactions. Such networks are called genome-scale metabolic models and can be used in computer simulations of metabolism of cells and cellular communities. For instance, models for various eukaryotic and prokaryotic organisms and cell types have been reconstructed (e.g. as reviewed in [1]) and applied in microbial ecology, biotechnology, and medical research [2–7].

By mathematical approaches, such as the application of Flux Balance Analysis (FBA), we are able to optimise a predefined objective function, which is based on nutritional constraints that represent the chemical environment of cells [4]. A widely used objective function is the production of biomass, which constitutes building block metabolites (i.e. amino acids, lipids, carbohydrates, nucleotides), inorganic ions, vitamins, and co-factors [8]. Flux Balance Analysis (FBA) enables the estimation of maximal biomass production, i.e. yield or rate, depending on how the model is constrained [9]. For example, biomass production rate can be predicted by Flux Balance Analysis (FBA) if the inflow of nutrients to the metabolic network is constrained

by the maximum nutrient uptake rate [10]. However, these uptake rates are context-, organism-, and compound-specific, which render the experimental measurement of the values for many FBA applications infeasible especially in chemically complex environments (i.e. broad range of potential nutrients). In contrast, if FBA models are directly constrained by the concentrations of nutrients, the biomass production yield can be predicted in units of grams biomass per gram nutrient [9]. The accuracy of the FBA model predictions are dependent on three factors: (1) A properly defined linear objective function that accurately represents the cell's biochemical objective, (2) an accurate reconstruction of the metabolic network architecture that represents the cell's metabolic capabilities, and (3) a quantitative representation of the chemical environment with all nutrients that are available for uptake. Several approaches and methods have been developed to construct realistic biomass objective functions [11,12] and to increase the accuracy of the metabolic network structure [13–16]. A realistic representation of the nutritional environment has received far less attention, although it is of increasing importance, as flux balance analysis is progressively applied in contexts, such as environmental and clinical microbiology. For example, it is challenging to study the nutritional microenvironment of bacterial cells in the mammal's intestine. In such cases, the exact chemical composition of the environment can only be vaguely defined, as the individual's diet is a source of variation [17] and the bioavailability of nutrients (especially important for the bioactive ones) is dependent on various processes inside the intestine [18].

Thus, it is a common issue that there is no unanimous approach on how to model the nutritional input in constraint-based modelling. For instance, it is usually not well-defined how to proceed in representing various compounds (e.g. macromolecules such as proteins, isomeric forms of compounds such as L- and D-Lactate, volatile compounds, gases, and trace elements) in the models and how to proceed in calculating their inflow values (i.e. constraints) starting from their concentrations. Researchers often unintentionally fail to provide information on their exact methodology how nutritional constraints were calculated and in numerous instances arbitrary numbers are used for nutritional constraints [19]. As a result, the validity of phenotypic predictions is potentially impaired and the applicability of FBA models therefore limited. Thus, new methods and guidelines are required to increase the predictive potential of FBA models especially in chemically complex environments and to facilitate reproducibility of results obtained from computational models of metabolism.

Here, we provide clear guidelines on how to design a realistic computational representation of the nutritional input for metabolic models for environments, which are initially only defined on the level of chemically complex components such as yeast extract, tryptone or milk powder. Therefore, a step-by-step procedure is proposed that incorporates publicly available databases and resources and facilitates reproducible calculations from chemically complex components to the quantities of specific molecules that are presumably available for uptake by the cell types of interest. The procedure is illustrated on the example of Lysogeny Broth (LB) medium, a commonly used microbial growth medium (experimentally and *in-silico*) that contains, besides water and NaCl, the chemically complex components tryptone and yeast extract which provide a broad range of nutrients for microbial growth. Based on the example we emphasise the importance for rationally designed nutritional constraints in FBA models in order to obtain meaningful physiological predictions.

## Material and methods

### Software for data analysis and metabolic modelling

Nutritional data was stored and processed using R (v. 3.6.3, [20]) and the R-package 'readODS' (v. 1.6.7, [21]). FBA and further model analyses were performed using the R-packages

'glpkAPI' (v. 1.3.2, [22]), 'sybil' (v. 2.1.5, [23]), 'sybilSBML' (v. 3.0.1, [23]), and the software libSBML (v. 5.17.2 [24]) and GLPK (v. 4.65 [25]). For simulating the metabolic network of *Escherichia coli* str. K-12 substr. MG1655, we employed the model 'iJO1366' [26], which can be downloaded from the BiGG database (http://bigg.ucsd.edu/) [27,28]. The model's default, "core" biomass reaction with an ATP-consumption estimate for growth associated mainte-nance (GAM) of 53.95 ATP gDW$^{-1}$ was defined as an objective function. For data processing and visualisation the R-packages 'dplyr' (v. 0.8.5, [29]) and 'ggplot2' (v. 3.3.0, [30]), 'egg' (v.0.4.5, [31]) were utilised.

## Stepwise procedures

The methodology is graphically depicted in Fig 1. In the following section, we are going to describe the modelling procedure, which is subdivided into 6 distinct steps. To illustrate the proposed procedure, this protocol is applied to computationally model the media conditions of Lysogeny Broth (LB) for cultures of *Escherichia coli* str. K-12 substr. MG1655 (S1 File: Step-wise procedure of designing nutritional input (LB medium)).

## Step 1: Understanding the chemical composition of the growth medium and the physiology of the model organisms

The first task of nutritional modelling is the collection of information regarding the growth medium of the organism. One of the aims of this initial step is to gather information about the
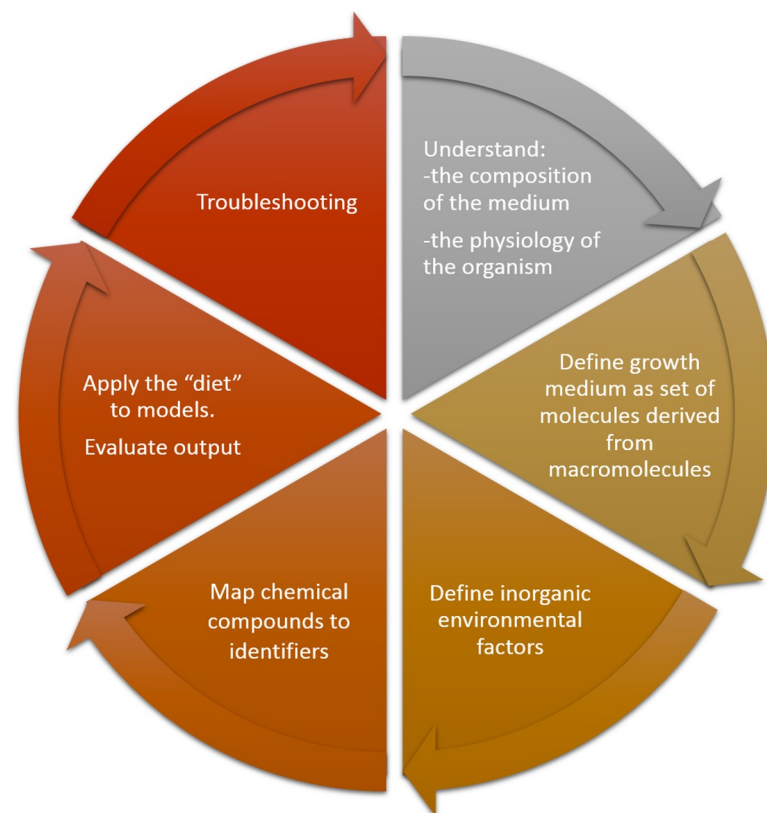


**Fig 1. Graphical summary of stepwise procedures.** Roadmap to definition of the nutritional input for genome-scale metabolic models.

https://doi.org/10.1371/journal.pone.0236890.g001

**Table 1. Examples of nutritional model for bacteria, humans and worms.**

| Nutritional model or medium | Source |
|---|---|
| Defined culture medium for *Escherichia coli* | *Brock Biology of Microorganisms*, p. 100, ISBN: 978-1-292-01831-7 [33] |
| Various human diets (e.g. Vegetarian, Mediterranean) | www.vmh.life [34,35] |
| *Caenorhabditis elegans* Maintenance Medium | Szewczyk NJ *et al.* [36] |

https://doi.org/10.1371/journal.pone.0236890.t001

background of the organism of interest (e.g. metabolic functions, auxotrophies). Additionally, a further aim is the collection of data that will enable us to enumerate and quantify those components that can potentially serve as nutrients for the model of this organism.

In laboratory settings, this information can usually be extracted from the definition of the culture medium. For that purpose, the modelling of qualitatively and quantitatively detailed defined growth media can be straightforward, as direct measurement/definitions of specific molecules (e.g. measurement of amino acids in a liquid [32]) might be provided. On the other hand, in environmental settings (e.g. soil, sea water, mammalian intestine), data on the chemical composition of the growth environment might be less accessible but could potentially be gathered from online databases and available literature (see Tables 1 and 2 for example resources).

Furthermore, modelling the metabolism of a cell requires detailed knowledge of the underlying (micro)environment, because cells are dependent on nutrients provided either through the medium or from other neighbouring cells. Therefore, they cannot be regarded as isolated organisms *per se*.

For instance, the simulation of cells, which are part of a multicellular system (e.g. host cells and bacteria in the vicinity [47]), requires more elaborate equations that link the molecular

**Table 2. Collection of the most useful resources for modelling nutrition.**

| Database | Website / ISBN of the book | Short Description |
|---|---|---|
| Metabolic modelling databases | | |
| VMH | www.vmh.life [34,35] | Metabolic models of bacteria, human cells, and collection of human diets |
| BiGG | https://bigg.ucsd.edu [27,28] | Metabolic models of various organisms |
| ModelSeed | www.modelseed.org [37] | Construction of metabolic models with collection of reactions/metabolites |
| MetaNetX | www.metanetx.org [38–41] | Integration of different metabolic modelling databases |
| Chemical and nutritional databases | | |
| PubChem | https://pubchem.ncbi.nlm.nih.gov [42] | Chemical compounds and their properties |
| FoodData Central | https://fdc.nal.usda.gov [43] | Chemical composition of ready and raw foods. |
| | | Until 2019 the database was available through USDA Food Composition Databases (https://ndb.nal.usda.gov/ndb/) |
| FooDB | www.foodb.ca/ [44] | Chemical and biological database specialised on foods and nutrients. |
| Fundamental nutritional literature | | |
| Ernährung des Menschen | ISBN 9783825287481 [45] | Detailed overview of human nutrition (in German) |
| Advanced Nutrition and Human Metabolism | ISBN 9781305627857 [46] | Detailed overview of human nutrition and metabolism (in English) |

https://doi.org/10.1371/journal.pone.0236890.t002

quantities of compounds with what is available for a model as input. For example, to simulate the nutritional environment of bacterial cells in the lower gastroenterological tract, we should account for the chemical and biological processes that take place in the upper tract, as host processes contribute to the bacterial microenvironment. Such processes may include absorption [for details see [48]], enzymatic and spontaneous degradation or modification of chemical compounds, and other time-dependent ones (e.g. peristalsis).

On the other hand, bacterial cell cultures involve well-designed experiments and many parameters are known and/or adjusted according to the scientific question (e.g. chemically defined media, defined environmental conditions). In such cases, this information can be directly incorporated into *in-silico* experiments. In conclusion, modelling should be firstly focused on clarifying the nutritional environment of an organism.

Here, we will consider the Lysogeny Broth (LB) medium as a sample case [49]. This medium is considered complex and nutrient-rich and is widely used for the cultivation of a range of different bacterial species [50]. It is an aqueous mixture composed of tryptone, yeast extract, and sodium chloride, while sodium hydroxide is added to adjust the pH level to 7.0 [49]. Although production of yeast extract is not unanimous and may differ across different products, in this study, we assumed that the yeast extract (*Saccharomyces cerevisiae*) is autolysed (i.e. lysis of yeast contents by the cell's own enzymes) [51].

Tryptone and yeast extract are the major constituents of this growth medium. The chemical composition of those ingredients can often be obtained directly from the manufacturer's product specifications. Such manuals ideally contain detailed nutritional information about the major elements and amino acids, which are available in yeast extract and tryptone. However, these manuals typically lack information about the content of minor minerals, vitamins, nucleic acids, fatty acids, and fibres. These constituents were for example inferred from the studies by Grant *et al*. [52], by Sarwar *et al*. [53] and as well as by the *Bionumbers* [54,55]— along with the bionutrients technical manual published by the company BD [56].

## Step 2: Define growth medium as a set of molecules derived from macromolecules

At this stage, the objective is the establishment of a list of molecules and their quantities based on the known ingredients of the medium (see step 1). Such a representation of the composition of the growth medium has to be derived, because only a small set of compounds of the medium are represented in the models.

Nutrients in organismal metabolic models are usually depicted as unique molecules with known chemical formula, structure, and mass such as vitamins, ions, amino acids, fatty acids, mono-/oligosaccharides, and specific fibres. By using primary literature, textbook information, and databases, (see Table 2 for example resources), it is possible to translate macromolecules or mixture of molecules into their constituent molecular components (e.g. albumin into its amino acid composition or mineral mixes to individual vitamins and minerals, see Table 2 for examples).

After collecting all the compounds and their quantities, unit transformations to the reference unit of the study (e.g. mmol/L) should be performed. Usually, chemical compounds are dissociated or hydrated in aqueous solutions (e.g. salts) or are dependent on the pH of the solution (e.g. conjugate bases and acids). Such cases ought to ideally be represented by one molecule or ion (e.g. butyrate not both butyrate and butyric acid), since models can often only handle one form of molecules or ions. A special case are natural polymers (e.g. fibres), which can occur in many different configurations and lengths. However, metabolic models often comprise only specific forms of the polymers. In such cases, one solution is to assume that the

focal polymer occurs only in the forms that the model supports. Another possibility to deal with the multiplicity of polymer variants is to translate the total polymer mass within the growth medium to a concentration of the respective monomers. However, in this case, the model would not require the enzymatic hydrolysis of the polymer.

Commercial mixture of compounds (e.g. yeast extract, casitone) can be translated into their molecular components (e.g. ions, amino acids) according to the manufacturer's manuals. To sum up, the focus of this step is the enumeration of compounds which constitute the growth medium of interest. In our example, we followed these principles by adding ions of salts (e.g. NaCl recorded as ion Na+ and ion Cl-) or simple compounds (e.g. polymer cellulose recorded as its dimer cellobiose).

## Step 3: Define the inorganic environmental factors

The objective of this step is to define the environmental compounds needed for cell growth. This step is important, because there are environmental constraints of the actual experimental or physiological setting that should be taken into consideration in addition to the nutritional compounds. For example, culturing conditions are usually well defined with respect to the oxygen regime as well as pH. Thus, the concentrations of the substances oxygen and protons need to be adapted accordingly (i.e. [oxygen] = 0 mM for anaerobic cultures, [protons] = $10^{-7}$ M for medium pH 7).

In our illustrative example (LB medium), the original experiments are conducted in aerobic conditions in an aqueous solution. Assuming that water's density is 1 g/cm$^3$, concentration of water in pure water is $5.55^*10^4$ mmol/L and the concentration of oxygen was set to 18.2 mmol/L, based on data from *E. coli* cultures [57]. Concentration of protons in the computational diet should reflect the pH of the actual experiments (pH = 7 which is equal to $10^{-7}$ M or $10^{-4}$ mM of protons) [49].

## Step 4: Map chemical compounds to metabolite identifiers of the metabolic model

The goal of this step is to map molecules to their representations in the model. It is a crucial step of metabolic modelling, as the more precisely the list of compounds is matched to the metabolites of the model, the more accurate will be the outcome of model simulations. In general, metabolites are linked to specific exchange reactions, i.e. pseudo-reactions that enable the exchange of metabolites with the external compartment of the model, which represents the theoretical chemical environment of the model.

In the general convention for metabolic modelling, exchange reactions that represent the chemical environment have identifiers with prefix "EX_" followed by the identifier of the focal metabolite/nutrient and the suffix for the external compartment (i.e. "(e)" or "_e"). However, the exact nomenclature for exchange reactions depends on the underlying reaction database and modelling framework. Although databases are not always compatible, approaches for mapping identifiers are available (e.g. MetaNetX [38–41]). Nonetheless, some databases such as VMH [34,35] and BIGG [27,28] share some but not all metabolite and reaction identifiers. The most commonly used biochemistry databases for metabolic modelling can be found in Table 2.

Furthermore, compounds, which are part of the growth medium, are frequently not available in a model (e.g. fibres). There are three options available to address this issue: (1) to remove these compounds from the setting (e.g. dyes and indicators of *in-vitro* cultures like phenolphthalein can be excluded if we can assume that they do influence the organism's metabolic processes), (2) choice of another molecule which is related to the missing compound (e.g.

nicotinate instead of nicotinamide), or (3) if feasible, to curate the models, i.e. to extend the models by adding the exchange and internal metabolic reactions needed based on the available genetic or physiological evidence. All in all, the objective of this step is the most precise representation of the molecular composition of the chemical environment in the model, as incorrect assignment of metabolite(s) to model compounds frequently leads to errors in subsequent model applications (see troubleshooting step).

Our LB practical example is based on the iJO1366 [26] model of *E. coli*, which can be found in the BIGG [27] database of models, reactions, and compounds. In total, 53 compounds were found in the metabolite list of the model. For 13 cases we could not identify appropriate exchange reactions in the model (e.g. metals, folate), so they were not added in the setting. Moreover, some media compounds are matched with more than one model metabolite. For instance, concentration of ions in media definitions are frequently provided without specifications of their oxidation states (e.g. Fe2+ / Fe3+), while metabolic models require this information as the redox potential is crucial for most reactions that involve these ions. A quantitative estimation of the oxidation states ratio of ions in rich media such as LB will be in most cases not possible without experimental measurements, as such ratios are highly variable over time e.g. by reacting with dissolved oxygen. In our LB medium example and in the case of Fe ions, we have chosen a 50% Fe2+ / 50% Fe3+, which reflects the measured intracellular ratio of the oxidation states in *E. coli* W3110 [58]. For chemically well-defined media formulations, the concentration of the different oxidation states can be inferred from the amount of the source salt (e.g. FeSO4 provides Fe2+, FeCl3 provides Fe3+ when dissolved).

## Step 5: Apply the "diet" to the model and evaluate the output

This step focuses on adjusting the exchange reactions linked to the metabolite identifiers, which were selected in step 5, and on performing growth simulations. This adjustment is essential for the accuracy of the growth simulation, as the outcome (i.e. growth) depends on the availability of each metabolite according to the actual concentration of the diet.

Notably and in the framework of FBA, the flux of the exchange reactions should be constrained to represent the availability of the nutrients. For this purpose, exchange reactions are constrained by upper and lower bounds. Usually, the uptake of a compound is represented by a negative value and excretion by a positive value. In case the FBA modelling approach aims to predict the organism's growth rate ($hr^{-1}$), the concentrations defined after step 1 to 4 need to be transformed to the maximum uptake fluxes in the unit $mmol^*(gDW^*hr)^{-1}$. This conversion is not trivial as it relies on kinetics (e.g. following the Monod equation), which are dependent on the compound and the organism of interest. In case the aim is to predict maximum growth/biomass yield (gDW per litre of growth medium), one can assign the concentration of the nutrients (e.g. in mmol per litre of growth medium) to the uptake bound as explained in the introduction.

Having defined constraints on the exchange reactions, the models can be optimised typically by maximising flux through the biomass reactions. Usually software like sybil [23], which utilises FBA [9], or the various COBRA suites/tools [59–62] are useful for this purpose. If the input of the exchange reactions reflects the concentration of compounds, the result of the optimisation corresponds to the biomass yield based on the growth medium. Otherwise, if the input reflects fluxes, the result corresponds to the actual growth rate of the organism.

In our LB example, the models followed the nomenclature scheme "EX_" (for exchange) + metabolite identifier +"(e)" (for external) to represent exchange reactions. In our case, by utilising sybil [23], the exchange reactions are constrained based on the concentration of the growth medium and used to conduct FBA. For compounds not present in the diet we set the lower bound to zero.

## Step 6: Troubleshooting

The aim of this step is to reveal and solve issues that occur after applying the computational diet and optimisation methods (i.e. FBA). Troubleshooting and fine-tuning should ideally guarantee that the outcome is biologically meaningful.

For instance, although it might be known that the cells can grow on specific medium, the simulation may lead to an unexpected outcome, where the model does not grow after following all the above steps. Except for the case of an erroneously defined diet, this issue may be associated with low quality of the model, which may miss reactions or the model may have gaps and/or errors. If the latter is the case, then either the model needs to be refined (model curation) and gap-filling algorithms can be applied [63,64].

In case of an erroneously defined nutritional growth environment, reduced cost calculations of uptake reactions can provide insights on nutrients that are absent or have too low concentrations in the modelled medium but which are essential for cell growth [65]. Reduced costs are sensitivity parameters for each reaction that are part of an FBA solution. Briefly, these values can be interpreted as the impact of an increase of the flux through the respective reaction on the final objective value (i.e. growth yield) [66]. Thus, when applied to the nutrient uptake reactions, the reduced costs allow the identification of compounds that could be added to the media in order to obtain increased or enabled growth [67].

Therefore, reduced costs can indicate compounds, which are part of the actual nutritional environment, but are not included in the modelled computational growth medium due to missing chemical data and knowledge of the media composition. In such cases, steps 1–5 have to be re-evaluated potentially by including additional metabolomic/chemical information on the medium components from the literature or even performing targeted chemical analytics to quantify the metabolite in question. In addition, the reduced cost analysis could also point to compounds, which may already be part of the modelled nutritional environment, but whose quantity may not be sufficient to facilitate model growth. This could be for instance the case if the lower bound for ATP hydrolysis, which is often part of cellular models to represent energetic costs for cell maintenance, cannot be reached due to too low availability of potential energy sources.

In this context, it also needs to be noted that a zero or infeasible growth solution returned by the FBA model under the reconstructed nutritional environment does not necessarily indicate missing or limiting compounds in the defined diet. Instead, a *nil* growth solution could also be caused by inconsistencies in the model itself, i.e. due to missing reactions, incorrect reaction bounds, or erroneous definition of the biomass reaction. Thus, in case of *nil* growth it is recommended to consider also options for model curation and algorithmic gapfilling approaches (e.g. [13,63,64]). Alterations to the modelled nutritional environment in order to enable growth should be primarily considered in cases where the model itself underwent already substantial curation efforts and where uncertainty of the exact chemical composition of the growth environment exists. The latter is especially relevant for growth media, which are generally chemically undefined such as LB medium or other complex growth media that are frequently used for animal/human cell line cultures [68].

## Results

### *E. coli* growth in modelled LB-medium

As an exemplary case, we applied the proposed procedure to the chemically complex growth medium Lysogeny Broth (LB) [49]. Following step 5, the optimal growth yield of the model of interest based on the user-defined nutritional constraints can be calculated. The modelled LB

medium was applied to constrain the genome-scale metabolic model of Escherichia coli K-12 (model designation: iJO1366). At first, the *E. coli* model predicted that the organism is not able to produce biomass in LB medium (objective value of zero). To troubleshoot this issue, we searched for the reduced costs. The missing compound with the highest absolute reduced cost was molybdate, EX_mobd(e), a molecule comprising molybdenum and oxygen. Ions of molybdate are the only biologically relevant source of molybdenum and play an important role in redox enzymes [69]. Molybdenum occurs in a variety of food categories [70]. The concentration of molybdate can be inferred by the value of molybdenum in this case. Adding molybdate (lower bound $-3.07^*10^{-4}$; value as molybdenum in STEP 3, based on the initial value from [52]), the optimisation was possible and the value of objective function was 1.54 g per Litre of LB medium for iJO1366. This case illustrates that calculating the reduced costs can indicate essential compounds that are missing in the nutritional constraints of the model and to adjust die diet composition accordingly.

In addition, we applied the modelled LB medium (incl. molybdate) to the core model of *E. coli* metabolism, which is a sub-network of the *Escherichia coli* K-12 genome-scale metabolic network [71] and which can be downloaded from the BiGG database (http://bigg.ucsd.edu/) [27,28]. The predicted biomass yield using this core model reached 0.59 g/L, which is only 38% of the growth yield predicted using the genome-scale model. This decrease is explained by the fact that the core model does not include the reactions and transporters that are required to utilise a number of nutrients from the LB medium, which the genome-scale model is able to use for growth.

*E. coli* is known to produce acetate as metabolic by-product at high growth rates, even in the presence of oxygen as potential electron acceptor [72]. Using our *in silico* representation of the LB medium, we also predicted acetate production using the core model (4.5 mmol/L yield) and the genome-scale model (46.72 mmol/L).

## Impact of diet modifications

Based on the rationale that an accurately modelled growth medium or diet allows realistic prediction of the cellular metabolism and cell growth, those models should also be able to accurately predict the metabolic- and growth response to medium modifications, such as nutrient supplementations. To test this, we searched for dietary compounds that increase the predicted objective function value (i.e. *E. coli* growth yield) if added as a supplement to the computational LB medium. To this end, we repeated the optimisation by adding +1mM (case A) or 30% (case B) of all compounds contained in the LB medium. We used these two different supplementation approaches to obtain substantial (cut-off of relative difference = $10^{-6}$) effects on growth both when supplementing compounds which are initially low in abundance or absent or whose molecular weight is high (case A) and when supplementing initially highly abundant compounds or low-molecular weight compounds (case B). Additionally, we have extended the list of supplemented compounds (case A) to the carbohydrates D-glucose and L-arabinose, as glucose is very often used in growth experiments and arabinose has been identified in yeast extracts in a previous publication of Sarwar *et al.* (53). The analysis revealed that most compounds, which increased the growth of the model, mainly belonged to the group of amino acids, and secondarily to nucleobases, carbohydrates, and fatty acids (Fig 2A). Oxygen had the strongest effect on biomass yield (more than 17% increase of yield while supplementing +30%; see Fig 2B).

Moreover, supplementation of carbohydrates had the strongest effect on the predicted yield compared to the other nutrient groups (Fig 2A). This suggests that the available amount of carbohydrates in LB medium is growth yield limiting. Interestingly, the effect of supplementing
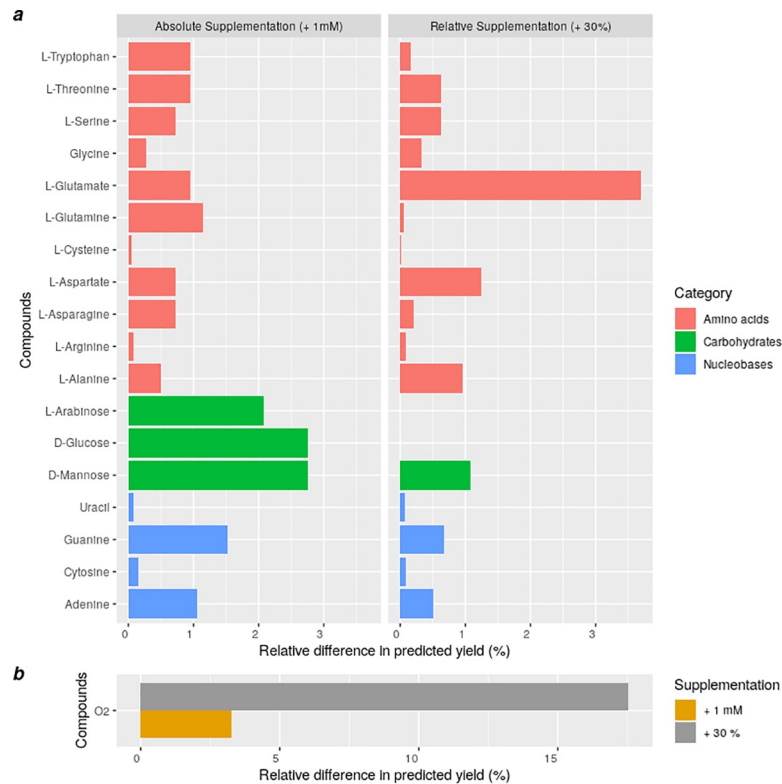
**Fig 2. Impact of diet modifications.** (a) Supplementation of amino acids and nucleobases to the computational LB medium (+30% of the original amount or +1 mM for each compound) and relative differences to the optimisation results for growth yield of the model iJO1366. (b) Similarly to (a) for oxygen. Cut-off of relative difference = $10^{-6}$.

the purines guanine and adenine is substantially bigger in comparison to the effect for uracil and cytosine, possibly due to the difference in the initial concentrations (mean concentration for G/A ~0.32mM, for U/C ~0.24mM). In contrast, the supplementation of arginine, cysteine, and glycine is predicted to have only limited effects on the growth yield (Fig 2A).

## Discussion

### Impact of defining precise nutrient levels for metabolic models

The aim of this work is to provide guidance for estimating the molecular concentrations of nutrients, which can be used as nutritional input in metabolic models (Fig 2). The utilisation of generic and arbitrary values can have a strong impact on downstream consequences on model predictions. For instance, recently published western and high-fibre diets, used for modelling the nutritional environment of the gut microbiome [19], contain values of 1 mmol/(gDW*h) for a big set of exchange reactions in the section "Minerals, vitamins, other", suggesting that there is uncertainty regarding the values that should be used. Another similar example is the LB medium, which is stored in the KBase database which is a popular tool for automated model reconstructions [73]. In this case, all fluxes are 100 mol/(gDW*h) for all compounds, which is physiologically unrealistic and, thus, could result in unlikely growth and metabolic flux predictions. Additionally, automated gap-filling on inaccurate media as part of the model construction could not only cause wrong simulation outcomes (fluxes / growth), but also can cause errors in the network structure itself by making models reliant on an unrealistically high

inflow of specific compounds. As a result, such models may also require unrealistic nutrient input for growth.

As a result, a crucial question arises: Is it worth spending effort and resources in designing a computational medium with accurate quantification? To answer this question, we constrained a publicly available and manually curated *Escherichia coli* K-12 model [26] with the LB medium, which was modelled using the above described 6-step procedure and scrutinised the FBA model prediction with experimental data (growth yield, nutrient utilisation, and the effect of nutrient supplementation) of the physiology of *E. coli* cultures in LB medium.

The growth patterns of *E. coli* in general have been extensively studied [74]. For instance, it is known that the optical density of the cultures in LB Medium can achieve a value of $OD600_{nm}$ 5 [75] and up to $OD600_{nm}$ 6.49 [49]. Additionally, it has been empirically estimated that an optical density of $OD600_{nm}$ 1 corresponds to ~0.3 g/L of bacterial dry weight [55,76]. By combining the information, a yield of (5*0.3) 1.5 g/L can be calculated for the phase when growth curves reach a plateau, which is in line with our optimisation result (1.54 g/L). Interestingly, we observed acetate as metabolic by-product alongside the formation of biomass, which is in line with experimental studies that report acetate production under aerobic conditions in LB medium [72]. Taking into account these results and based on the modelled diet for LB medium with rationally defined quantities for all included compounds, it can be suggested that an accurate representation of the chemical composition of the environment is crucial for precise predictions of biomass formation and metabolic functions using FBA models.

Such realistic results can be particularly helpful for making context-specific predictions. For instance, it is possible to investigate the effect of manipulating the nutritional composition on the predicted phenotypes. In Fig 2A we presented the relative difference of the predicted yields of the *E. coli* model iJO1366 after supplementing the computational medium with each compound contained in LB medium. The analysis highlighted that glucose can increase the yield of the model, which is in line with the fact that glucose can boost the growth of *E. coli* cultures [78]. It is worth mentioning that by adding 20 mM of glucose to the simulations of the *E. coli* model, the predicted yield increased by 54%, which is an overestimation compared to the published experimental value of 39% [77]. Nevertheless, the simulations *in-silico* reproduced the strong growth-supporting effects of glucose on *E. coli* growth. Additionally, supplementing the computational diet by adding 1mM of amino acids significantly supports the growth of the model following this order: tryptophan, threonine, glutamate, serine, aspartate, alanine, glycine, while supplementing the diet by adding +30% by this order: glutamate, aspartate, alanine, threonine, serine, glycine, tryptophan. Similarly, *E. coli* has been reported to sequentially consume these amino acids in the order of serine, aspartate, tryptophan, glutamate, glycine, threonine, and alanine in a tryptone medium [49,78]. This suggests the order of consumption might be associated with the growth benefits for the cell. Taking everything into account, such predictions may be helpful for optimising cultures *in-vitro* by targeted modifications of the medium composition [79].

## Quality of FBA model prediction depend on the accuracy of the modelled environment

Although Flux Balance Analysis can be used to make useful physiological predictions, these are not limitation-free. FBA-based predictions are especially influenced by the exact values for the nutrient uptake reaction constraints. For instance, in the above-mentioned example of *E. coli* growth in LB medium, we showed that increasing the bound of oxygen uptake by 30% results in more than 17% increased growth yield. The amount of oxygen in our simulations is based on the uptake rate observed in experiments [57] and not on the environmental amount

of oxygen, because various phenomena including the diffusion from air to the medium [80] and the uptake limits by the cell [81] restrict the available amount of oxygen. Moreover it is sometimes laborious to identify such phenomena and to include them in the static models, but also it can be challenging to simply quantify some environmental constraints (e.g. gases) or variability of the nutritional compounds in the media, as discussed above, although the output of the models severely depend on them (e.g. oxygen, yeast amino acids). As a result, it is worth noting that as long as an accurate modelled environment cannot be designed, the FBA predictions will be prone to shortcomings.

## Relevance of nutritional modelling for human microbiome research

Computational predictions based on constraint-based models of microbial metabolism have emerged as a powerful tool in microbiome research and broadened its horizons [5,19,82,83]. The microbiota, especially within the human gut, has been acknowledged to be a substantial contributing factor in health and disease of the host [47]. Since the models are utilised in exploring health/disease-related bacteria-bacteria and host-bacteria interactions, the importance of diets has been recognised [84]. For instance, the predictions derived from testing various "diets" in metabolic models led to successful dietary interventions in humans [85].

By providing the community with a roadmap and a detailed example we set the basis for the development of standard operating procedures for reconstructing the nutritional environment in constraint-based metabolic modelling. The procedure for *in-silico* modelling of the nutritional environment of cells described in this study may allow efficient recognition, tracking, and handling of inaccuracies and artefacts in constraint-based metabolic modelling approaches.

## Limitations and outlook

We introduced and applied the proposed procedure to reconstruct the nutritional environment of metabolic models in the context of Flux Balance Analysis. Furthermore, the 6-step procedure can be analogously applied to other metabolic modelling frameworks such as Ordinary Differential Equation (ODE)—based kinetic models or Elementary Mode (EM) analysis; with the only exception of the proposed reduced costs analysis in step 6 ('troubleshooting'). This analysis is limited to FBA applications, as reduced costs are sensitivity parameters that are directly associated with an FBA solution [85].

A precise definition of chemical growth environment is especially valuable for kinetic modelling, as such models typically allow more quantitative predictions than FBA simulations [86]. These predictions include for example the temporal-dynamic changes of individual metabolite levels, including metabolic by-products. However, kinetic models are usually limited to small or medium-scale networks that comprise fewer reactions than genome-scale metabolic network reconstructions. Thus, kinetic models may not contain all pathways in which exogenous compounds from a complex nutritional environment enter the cell's metabolism. For the example of *E. coli*, kinetic models are often derived from a so-called *core model*, which is a sub-network of the *E. coli* genome-scale model that comprises glycolysis, the pentose phosphate pathway, the TCA-cycle, and the oxidative respiration chain [87]. This sub-network does not include the uptake and utilisation of a majority of compounds that are part of complex growth media used for the cultivation of *E. coli* such as LB medium and that are known to be utilised by *E. coli* (e.g. most proteinogenic amino acids). Hence, kinetic models on the basis of this core network cannot reproduce the metabolism of *E. coli* in complex nutritional environments. We therefore emphasise that future research should focus on the extension of kinetic models that include the uptake and degradation of several nutrients found in complex

media. In combination with the incorporation of regulatory circuits such as catabolite repression [88], kinetic modelling will highly benefit from the here proposed procedure to model complex growth environments.

Growth media for eukaryotic cell cultures are also often molecularly complex due to components such as bovine serum or meat extract, which provide a wide range of potential nutrients for cell metabolism [68]. Thus, computational approaches to simulate metabolic processes in eukaryotic cell cultures (e.g. human/animal cell lines) also depend on an accurate representation of the chemical composition of the environment. In this context, the here proposed procedure to model complex nutritional environments will also promote the predictive potential of *in silico* models of eukaryotic cell metabolism.

## Supporting information

**S1 File. Diet table.** Stepwise procedure of designing nutritional input (LB medium).
(ODS)

## Acknowledgments

The authors thank Johannes Zimmermann for helpful discussions.

## Author Contributions

**Conceptualization:** Georgios Marinos, Silvio Waschina.

**Data curation:** Georgios Marinos.

**Formal analysis:** Georgios Marinos.

**Funding acquisition:** Christoph Kaleta.

**Investigation:** Georgios Marinos, Silvio Waschina.

**Methodology:** Georgios Marinos, Silvio Waschina.

**Project administration:** Georgios Marinos, Silvio Waschina.

**Resources:** Christoph Kaleta.

**Software:** Georgios Marinos.

**Supervision:** Christoph Kaleta, Silvio Waschina.

**Validation:** Georgios Marinos.

**Visualization:** Georgios Marinos.

**Writing – original draft:** Georgios Marinos.

**Writing – review & editing:** Christoph Kaleta, Silvio Waschina.

## References

1. Reed JL. Genome-Scale Metabolic Modeling and Its Application to Microbial Communities. In: The Chemistry of Microbiomes: Proceedings of a Seminar Series. National Academies Press (US); 2017.

2. Bauer E, Thiele I. From metagenomic data to personalized in silico microbiotas: predicting dietary supplements for Crohn's disease. NPJ Syst Biol Appl. 2018 Aug 1; 4:27. https://doi.org/10.1038/s41540-018-0063-2 PMID: 30083388

3. Zhang C, Hua Q. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. Front Physiol. 2015; 6:413. https://doi.org/10.3389/fphys.2015.00413 PMID: 26779040

4. Muller EEL, Faust K, Widder S, Herold M, Martínez Arbas S, Wilmes P. Using metabolic networks to resolve ecological properties of microbiomes. Current Opinion in Systems Biology. 2018 Apr; 8:73–80.

5. Graspeuntner S, Waschina S, Künzel S, Twisselmann N, Rausch TK, Cloppenborg-Schmidt K, et al. Gut Dysbiosis With Bacilli Dominance and Accumulation of Fermentation Products Precedes Late-onset Sepsis in Preterm Infants. Clin Infect Dis. 2019 Jul 2; 69(2):268–77. https://doi.org/10.1093/cid/ciy882 PMID: 30329017

6. Aden K, Rehman A, Waschina S, Pan W-H, Walker A, Lucio M, et al. Metabolic Functions of Gut Microbes Associate With Efficacy of Tumor Necrosis Factor Antagonists in Patients With Inflammatory Bowel Diseases. Gastroenterology. 2019 Nov; 157(5):1279–92.e11. https://doi.org/10.1053/j.gastro.2019.07.025 PMID: 31326413

7. Pryor R, Norvaisas P, Marinos G, Best L, Thingholm LB, Quintaneiro LM, et al. Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy. Cell. 2019 Sep 5; 178(6):1299–312.e29. https://doi.org/10.1016/j.cell.2019.08.003 PMID: 31474368

8. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010 Mar; 28(3):245–8. https://doi.org/10.1038/nbt.1614 PMID: 20212490

9. 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 Jul 5; 8(7): e1002575. https://doi.org/10.1371/journal.pcbi.1002575 PMID: 22792053

10. Varma A, Palsson BO. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Nat Biotechnol. 1994 Oct; 12(10):994–8.

11. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol. 2007 Jul 10; 3:119. https://doi.org/10.1038/msb4100162 PMID: 17625511

12. Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol. 2010 Jun; 13(3):344–9. https://doi.org/10.1016/j.mib.2010.03.003 PMID: 20430689

13. Schmidt R, Waschina S, Boettger-Schmidt D, Kost C, Kaleta C. Computing autocatalytic sets to unravel inconsistencies in metabolic network reconstructions. Bioinformatics. 2015 Feb 1; 31(3):373–81. https://doi.org/10.1093/bioinformatics/btu658 PMID: 25286919

14. Gevorgyan A, Poolman MG, Fell DA. Detection of stoichiometric inconsistencies in biomolecular models. Bioinformatics. 2008 Oct 1; 24(19):2245–51. https://doi.org/10.1093/bioinformatics/btn425 PMID: 18697772

15. Pitkänen E, Rousu J, Ukkonen E. Computational methods for metabolic reconstruction. Curr Opin Biotechnol. 2010 Feb; 21(1):70–7. https://doi.org/10.1016/j.copbio.2010.01.010 PMID: 20171871

16. Orth JD, Palsson BØ. Systematizing the generation of missing metabolic knowledge. Biotechnol Bioeng. 2010 Oct 15; 107(3):403–12. https://doi.org/10.1002/bit.22844 PMID: 20589842

17. Salonen A, Lahti L, Salojärvi J, Holtrop G, Korpela K, Duncan SH, et al. Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. ISME J. 2014 Nov; 8(11):2218–30. https://doi.org/10.1038/ismej.2014.63 PMID: 24763370

18. Rein MJ, Renouf M, Cruz-Hernandez C, Actis-Goretta L, Thakkar SK, da Silva Pinto M. Bioavailability of bioactive food compounds: a challenging journey to bioefficacy. Br J Clin Pharmacol. 2013 Mar; 75 (3):588–602. https://doi.org/10.1111/j.1365-2125.2012.04425.x PMID: 22897361

19. 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 Jan; 35(1):81–9. https://doi.org/10.1038/nbt.3703 PMID: 27893703

20. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

21. Schutten Gerrit-Jan, Chan Chung-hong, Leeper Thomas J. et al. readODS: Read and Write ODS Files [Internet]. 2018. Available from: https://CRAN.R-project.org/package=readODS

22. Gelius-Dietrich G. glpkAPI: R Interface to C API of GLPK version [Internet]. 2020 [cited 2019 Jul 17]. Available from: https://CRAN.R-project.org/package=glpkAPI

23. Gelius-Dietrich G, Fritzemeier CJ, Desouki AA, Lercher MJ. sybil—Efficient constraint-based modelling in R. BMC Systems Biology. 2013; 7(1):125.

24. Bornstein BJ, Keating SM, Jouraku A, Hucka M. LibSBML: an API library for SBML. Bioinformatics. 2008 Mar 15; 24(6):880–1. https://doi.org/10.1093/bioinformatics/btn051 PMID: 18252737

25. Makhorin A, Moscow A. GLPK-GNU Project-Free Software Foundation (FSF) [Internet]. 2012. Available from: https://www.gnu.org/software/glpk/

26. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism. Mol Syst Biol. 2011 Oct 11; 7:535. https://doi.org/10.1038/msb.2011.65 PMID: 21988831

27. BiGG [Internet]. [cited 2019 Jul 11]. Available from: http://bigg.ucsd.edu/

28. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 2016 Jan 4; 44(D1):D515–22. https://doi.org/10.1093/nar/gkv1049 PMID: 26476456

29. Wickham Hadley, Romain François, Henry Lionel, Kirill Müller. dplyr: A Grammar of Data Manipulation [Internet]. 2020. Available from: https://CRAN.R-project.org/package=dplyr

30. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.

31. Auguie B. egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size [Internet]. 2019. Available from: https://CRAN.R-project.org/packag=egge

32. Godel H, Graser T, Földi P, Pfaender P, Fürst P. Measurement of free amino acids in human biological fluids by high-performance liquid chromatography. J Chromatogr. 1984 Aug 3; 297:49–61. https://doi.org/10.1016/s0021-9673(01)89028-2 PMID: 6490773

33. Madigan MT, Martinko JM, Bender KS, Buckley DH, Stahl DA. Brock Biology of Microorganisms. Pearson; 2014.

34. Virtual Metabolic Human [Internet]. [cited 2019 Jul 11]. Available from: https://www.vmh.life/

35. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol. 2018 Mar; 36(3):272–81. https://doi.org/10.1038/nbt.4072 PMID: 29457794

36. Szewczyk NJ, Kozak E, Conley CA. Chemically defined medium and Caenorhabditis elegans. BMC Biotechnol. 2003 Oct 27; 3:19. https://doi.org/10.1186/1472-6750-3-19 PMID: 14580264

37. ModelSEED [Internet]. [cited 2019 Jul 11]. Available from: http://modelseed.org/

38. MetaNetX [Internet]. 2011 [cited 2019 Jul 11]. Available from: https://www.metanetx.org

39. Moretti S, Martin O, Van Du Tran T, Bridge A, Morgat A, Pagni M. MetaNetX/MNXref–reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. Nucleic Acids Research. 2016; 44(D1):D523–6. https://doi.org/10.1093/nar/gkv1117 PMID: 26527720

40. Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I, Pagni M. Reconciliation of metabolites and biochemical reactions for metabolic networks. Brief Bioinform. 2014 Jan; 15(1):123–35. https://doi.org/10.1093/bib/bbs058 PMID: 23172809

41. Ganter M, Bernard T, Moretti S, Stelling J, Pagni M. MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. Bioinformatics. 2013 Mar 15; 29(6):815–6. https://doi.org/10.1093/bioinformatics/btt036 PMID: 23357920

42. Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019; 47(D1):D1102–9. https://doi.org/10.1093/nar/gky1033 PMID: 30371825

43. FoodData Central [Internet]. [cited 2020 Mar 11]. Available from: https://fdc.nal.usda.gov/

44. FooDB [Internet]. [cited 2019 Dec 10]. Available from: http://foodb.ca/

45. Elmadfa I, Leitzmann C. Ernährung des Menschen. UTB; 2019.

46. Gropper SS, Smith JL, Carr TP. Advanced Nutrition and Human Metabolism. Cengage Learning; 2016. 640 p.

47. Esser D, Lange J, Marinos G, Sieber M, Best L, Prasse D, et al. Functions of the Microbiota for the Physiology of Animal Metaorganisms. J Innate Immun. 2019; 11(5):393–404. https://doi.org/10.1159/000495115 PMID: 30566939

48. Kiela PR, Ghishan FK. Physiology of Intestinal Absorption and Secretion. Best Pract Res Clin Gastroenterol. 2016 Apr; 30(2):145–59. https://doi.org/10.1016/j.bpg.2016.02.007 PMID: 27086882

49. Sezonov G, Joseleau-Petit D, D'Ari R. Escherichia coli physiology in Luria-Bertani broth. J Bacteriol. 2007 Dec; 189(23):8746–9. https://doi.org/10.1128/JB.01368-07 PMID: 17905994

50. MacWilliams MP, Liao M-K. Luria Broth (LB) and Luria Agar (LA) Media and Their Uses Protocol. 2006 Oct 9 [cited 2019 Jul 11]; Available from: http://www.asmscience.org/content/education/protocol/protocol.3031

51. Zarei O, Dastmalchi S, Hamzeh-Mivehroud M. A Simple and Rapid Protocol for Producing Yeast Extract from Saccharomyces cerevisiae Suitable for Preparing Bacterial Culture Media. Iranian Journal of Pharmaceutical Research: IJPR. 2016; 15(4):907. PMID: 28243289

52. Grant CL, Pramer D. Minor element composition of yeast extract. J Bacteriol. 1962 Oct; 84:869–70. https://doi.org/10.1128/JB.84.4.869-870.1962 PMID: 13949853

53. Sarwar G, Shah BG, Mongeau R, Hoppner K. Nucleic Acid, Fiber and Nutrient Composition of Inactive Dried Food Yeast Products. Journal of Food Science. 1985; 50(2):353–7.

54. Concentrations of metals in components of DGC—Generic—BNID 114108 [Internet]. [cited 2019 Jul 11]. Available from: https://bionumbers.hms.harvard.edu/bionumber.aspx?id=114108&ver=3

55. Milo R, Jorgensen P, Moran U, Weber G, Springer M. BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Res. 2010 Jan; 38(Database issue):D750–3. https://doi.org/10.1093/nar/gkp889 PMID: 19854939

56. BD [Internet]. [cited 2019 Jul 11]. Available from: http://www.bd.com/documents/guides/user-guides/DS_CM_Bionutrients-technical-manual_UG_EN.pdf

57. Fischer E, Zamboni N, Sauer U. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Anal Biochem. 2004 Feb 15; 325(2):308–16.

58. Matzanke BF, Bill E, Trautwein AX. Main components of iron metabolism in microbial systems—Analyzed by in vivo Mössbauer spectroscopy. Hyperfine Interact. 1992 Apr; 71(1):1259–62.

59. Heirendt L, Thiele I, Fleming RMT. DistributedFBA.jl: high-level, high-performance flux balance analysis in Julia. Bioinformatics. 2017 May 1; 33(9):1421–3. https://doi.org/10.1093/bioinformatics/btw838 PMID: 28453682

60. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol. 2013 Aug 8; 7:74. https://doi.org/10.1186/1752-0509-7-74 PMID: 23927696

61. Jamshidi N, Palsson BØ. Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models. Biophys J. 2010 Jan 20; 98(2):175–85. https://doi.org/10.1016/j.bpj.2009.09.064 PMID: 20338839

62. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc. 2019 Feb 20; 14 (3):639–702. https://doi.org/10.1038/s41596-018-0098-2 PMID: 30787451

63. Thiele I, Vlassis N, Fleming RMT. fastGapFill: efficient gap filling in metabolic networks. Bioinformatics. 2014 Sep 1; 30(17):2529–31. https://doi.org/10.1093/bioinformatics/btu321 PMID: 24812336

64. Karp PD, Weaver D, Latendresse M. How accurate is automated gap filling of metabolic models? BMC Syst Biol. 2018 Jun 19; 12(1):73. https://doi.org/10.1186/s12918-018-0593-7 PMID: 29914471

65. Maarleveld TR, Khandelwal RA, Olivier BG, Teusink B, Bruggeman FJ. Basic concepts and principles of stoichiometric modeling of metabolic networks. Biotechnol J. 2013 Sep; 8(9):997–1008. https://doi.org/10.1002/biot.201200291 PMID: 23893965

66. Teusink B, Wiersma A, Molenaar D, Francke C, de Vos WM, Siezen RJ, et al. Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model. J Biol Chem. 2006 Dec 29; 281(52):40041–8. https://doi.org/10.1074/jbc.M606263200 PMID: 17062565

67. Goffin P, van de Bunt B, Giovane M, Leveau JHJ, Höppener-Ogawa S, Teusink B, et al. Understanding the physiology of Lactobacillus plantarum at zero growth. Mol Syst Biol. 2010 Sep 21; 6:413. https://doi.org/10.1038/msb.2010.67 PMID: 20865006

68. Yao T, Asayama Y. Animal-cell Culture Media: History, Characteristics, and Current Issues. Reprod Med Biol. 2017 Mar 21; 16(2):99–117. https://doi.org/10.1002/rmb2.12024 PMID: 29259457

69. Cell biology of molybdenum in plants and humans. Biochimica et Biophysica Acta (BBA)—Molecular Cell Research. 2012 Sep 1; 1823(9):1568–79.

70. Hunt CD, Meacham SL. Aluminum, boron, calcium, copper, iron, magnesium, manganese, molybdenum, phosphorus, potassium, sodium, and zinc: concentrations in common western foods and estimated daily intakes by infants; toddlers; and male and female adolescents, adults, and seniors in the United States. J Am Diet Assoc. 2001 Sep; 101(9):1058–60. https://doi.org/10.1016/S0002-8223(01)00260-7 PMID: 11573761

71. Orth JD, Fleming RMT, Palsson BØ. Reconstruction and Use of Microbial Metabolic Networks: The Core Escherichia Coli Metabolic Model as an Educational Guide. EcoSal Plus. 2010 Sep; 4(1).

72. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, et al. Overflow Metabolism in Escherichia Coli Results From Efficient Proteome Allocation. Nature. 2015 Dec 3; 528(7580):99–104. https://doi.org/10.1038/nature15765 PMID: 26632588

73. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018 Jul 6; 36(7):566–9. https://doi.org/10.1038/nbt.4163 PMID: 29979655

74. Blount ZD. The unexhausted potential of E. coli. Elife. 2015 Mar 25; 4.

75. Baev MV, Baev D, Radek AJ, Campbell JW. Growth of Escherichia coli MG1655 on LB medium: determining metabolic strategy with transcriptional microarrays. Appl Microbiol Biotechnol. 2006 Jul; 71 (3):323–8. https://doi.org/10.1007/s00253-006-0392-8 PMID: 16645822

76. One unit of OD600 corresponds to a cell dry w—Bacteria Escherichia coli—BNID 109835 [Internet]. [cited 2019 Jul 12]. Available from: https://bionumbers.hms.harvard.edu/bionumber.aspx?s=n&v= 4&id=109835

77. Medina MB, Uknalis J, Tu S-I. EFFECTS OF SUGAR ADDITION IN LURIA BERTANI (LB) MEDIA ON ESCHERICHIA COLI O157:H7. J Food Saf. 2011 Aug 1; 31(3):386–94.

78. Prüss BM, Nelms JM, Park C, Wolfe AJ. Mutations in NADH:ubiquinone oxidoreductase of Escherichia coli affect growth on mixed amino acids. J Bacteriol. 1994 Apr; 176(8):2143–50. https://doi.org/10.1128/ jb.176.8.2143-2150.1994 PMID: 8157582

79. Chen Y, Dong F, Wang Y. Systematic development and optimization of chemically defined medium supporting high cell density growth of Bacillus coagulans. Appl Microbiol Biotechnol. 2016 Sep; 100 (18):8121–34. https://doi.org/10.1007/s00253-016-7644-z PMID: 27262567

80. Somerville GA, Proctor RA. Cultivation conditions and the diffusion of oxygen into culture media: The rationale for the flask-to-medium ratio in microbiology. BMC Microbiol. 2013 Jan 16; 13(1):1–2.

81. Varma A, Boesch BW, Palsson BO. Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl Environ Microbiol. 1993 Aug; 59(8):2465–73. https://doi.org/10. 1128/AEM.59.8.2465-2473.1993 PMID: 8368835

82. Bauer E, Zimmermann J, Baldini F, Thiele I, Kaleta C. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. PLoS Comput Biol. 2017 May 22; 13(5): e1005544. https://doi.org/10.1371/journal.pcbi.1005544 PMID: 28531184

83. Magnúsdóttir S, Thiele I. Modeling metabolism of the human gut microbiome. Curr Opin Biotechnol. 2018 Jun; 51:90–6. https://doi.org/10.1016/j.copbio.2017.12.005 PMID: 29258014

84. Sen P, Orešič M. Metabolic Modeling of Human Gut Microbiota on a Genome Scale: An Overview. Metabolites. 2019 Jan 28; 9(2).

85. Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. Cell Metab. 2015 Aug 4; 22(2):320–31. https://doi.org/10.1016/j.cmet.2015.07.001 PMID: 26244934

86. Stalidzans E, Seiman A, Peebo K, Komasilovs V, Pentjuss A. Model-based Metabolism Design: Constraints for Kinetic and Stoichiometric Models. Biochem Soc Trans. 2018 Apr 17; 46(2):261–267. https://doi.org/10.1042/BST20170263 PMID: 29472367

87. Chakrabarti A, Miskovic L, Soh KC, Hatzimanikatis V. Towards Kinetic Modeling of Genome-Scale Metabolic Networks Without Sacrificing Stoichiometric, Thermodynamic and Physiological Constraints. Biotechnol J. 2013 Sep; 8(9):1043–57. https://doi.org/10.1002/biot.201300091 PMID: 23868566

88. Kremling A, Geiselmann J, Ropers D, de Jong H. Understanding Carbon Catabolite Repression in Escherichia Coli Using Quantitative Models. Trends Microbiol. 2015 Feb; 23(2):99–109. https://doi.org/ 10.1016/j.tim.2014.11.002 PMID: 25475882