

1 **Quantifying the propagation of parametric uncertainty on flux balance analysis**

2 Hoang V. Dinh^{1,2}, Debolina Sarkar¹, Costas D. Maranas^{1,2}

3 ¹Department of Chemical Engineering, Pennsylvania State University, University Park,
4 Pennsylvania, USA

5 ²Center for Advanced Bioenergy and Bioproducts Innovation, The Pennsylvania State University,
6 University Park, PA 16802, USA

7 *Corresponding author: costas@psu.edu (CDM)

8 **Abstract**

9 Flux balance analysis (FBA) and associated techniques operating on stoichiometric genome-scale
10 metabolic models play a central role in quantifying metabolic flows and constraining feasible
11 phenotypes. At the heart of these methods lie two important assumptions: (i) the biomass
12 precursors and energy requirements neither change in response to growth conditions nor
13 environmental/genetic perturbations, and (ii) metabolite production and consumption rates are
14 equal at all times (i.e., steady-state). Despite the stringency of these two assumptions, FBA has
15 been shown to be surprisingly robust at predicting cellular phenotypes. In this paper, we formally
16 assess the impact of these two assumptions on FBA results by quantifying how uncertainty in
17 biomass reaction coefficients, and departures from steady-state due to temporal fluctuations could
18 propagate to FBA results. In the first case, conditional sampling of parameter space is required to
19 re-weight the biomass reaction so as the molecular weight remains equal to 1 g/mmol, and in the
20 second case, metabolite (and elemental) pool conservation must be imposed under temporally
21 varying conditions. Results confirm the importance of enforcing the aforementioned constraints
22 and explain the robustness of FBA biomass yield predictions.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

1. Introduction

Flux balance analysis (FBA) encompasses a wide range of computational analyses that operate on genome-scale metabolic (GSM) models (Lewis et al., 2012; Orth et al., 2010), and provide insights on maximum theoretical yields of biomass or product per mole of substrate or limiting resource (Feist and Palsson, 2010), and metabolic activities in the form of intracellular fluxes under physiological constraints (Long et al., 2017; Mahadevan and Schilling, 2003). Relying on the solution of a linear programming (LP) optimization formulations (Orth et al., 2010), FBA analyses have gained broad acceptance due to the tractability and scalability of the calculations and the wide availability of GSM models for many organisms (Gu et al., 2019). As many as 185 manually curated GSM models have been reported (as of February 2019 (Gu et al., 2019)) with many thousands (Gu et al., 2019; Lieven et al., 2020) automatically generated using workflows such as KBase (Arkin et al., 2018), RAVEN (Wang et al., 2018), and CarveMe (Machado et al., 2018). GSM models aim to encompass all known biochemical conversions occurring in an organism, the participating metabolites, reaction stoichiometry, corresponding enzymes, genes encoding them, and production of biomass constituents in physiologically relevant proportions in cells (Thiele and Palsson, 2010). FBA predictions have been found to recapitulate growth phenotypes reasonably well under nutrient-limited conditions (O'Brien et al., 2013) and after adaptive laboratory evolution (Ibarra et al., 2002). By exploiting the fact that the timescale of organism growth far exceeds that of cellular biochemical reactions, every metabolite can be assumed to be at a pseudo-steady state, whereby the production equals the consumption flux. FBA also requires the establishment of an objective function as a teleological driver of cellular metabolism, which is usually maximizing biomass yield for a given amount of carbon uptake (Feist and Palsson, 2010). This translates computationally to maximizing flux through an imposed biomass drain (i.e.,

1 biomass reaction) which serves to sequester metabolites in their physiological ratios. Thus, at the
2 heart of most FBA studies lie two major assumptions: (i) the biomass precursors and energy
3 requirements are constant and invariant to changes in growth conditions, environmental, and/or
4 genetic perturbations; and (ii) metabolites remain at steady-state with no temporal fluctuations in
5 their concentration or production/consumption rates.

6 Uncertainty or inaccuracies in biomass composition can arise from limitations in measuring
7 techniques (Beck et al., 2018; Lange and Heijnen, 2001), and the metabolically and temporally
8 varying, and/or heterogenous nature of cellular populations (Takhaveev and Heinemann, 2018).
9 Experimental quantification and modeling of metabolic heterogeneity is important for accurate
10 cellular characterization (Takhaveev and Heinemann, 2018) and efficient bioprocessing
11 (González-Cabaleiro et al., 2017). A previous metabolic modeling effort accounting for biomass
12 composition variability considered biomass coefficients to be binomially distributed
13 (MacGillivray et al., 2017). However, this effort did not enforce the requirement that the molecular
14 weight (MW) of the biomass metabolite must always equal 1 gram dry weight (gDW) mmol^{-1}
15 (Thiele and Palsson, 2010), which is needed to correctly map biomass flux to specific growth rate
16 (Chan et al., 2017). In addition to the biomass composition, variance in energy demand parameters
17 can be significant factors controlling predicted maximum theoretical yields. Growth-associated
18 ATP maintenance (GAM) captures the energy demand per unit of produced biomass (Thiele and
19 Palsson, 2010) and is embedded within the biomass reaction. Non-growth associated ATP
20 maintenance (NGAM) accounts for the energy demand associated with cellular processes such as
21 repair and maintenance (Thiele and Palsson, 2010). GAM and NGAM values are estimated using
22 regression on growth data from chemostats (Thiele and Palsson, 2010), and are thus subject to
23 uncertainty due to instrument limitations and population heterogeneity to an even larger extend

1 than biomass composition. Capturing uncertainty in ATP maintenance parameters is essential
2 because fluctuations in energy demand can propagate to other parts of the metabolic network and
3 ultimately affect the growth and product yield.

4 The steady-state condition for all metabolites is a fundamental assumption of FBA. However, this
5 might be too stringent as time-course intracellular metabolomics studies have revealed the
6 transient nature of metabolite pools (Andreas Angermayr et al., 2016; Kultschar et al., 2019; Qiao
7 et al., 2020; Yurkovich et al., 2017) as a cellular response to microenvironmental variations
8 (Schmitz et al., 2017) and over the course of a typical growth cycle (Andreas Angermayr et al.,
9 2016). In addition, periodic variations in metabolism are inherent for many organisms such as
10 phototrophs whose lifestyle is tailored around light availability and diurnal cycle (Andreas
11 Angermayr et al., 2016; Diamond et al., 2015). To account for metabolite level fluctuations within
12 FBA at a given time point, metabolite accumulation/depletion terms (i.e., the right-hand side
13 (RHS) of mass balance constraints) need to be temporally decoupled. Of course, over time all
14 imbalances will have to average out for a stable culture. In a previous study by Reznik and
15 coworkers, allowing for metabolite imbalances was necessary for the FBA model to consistently
16 capture observed changes in pools of growth-limiting metabolites (Reznik et al., 2013). In another
17 study, relaxing metabolite balance constraints yielded a better correlation between FBA
18 predictions and experimentally observed intracellular metabolic fluxes (MacGillivray et al., 2017).
19 In both studies imbalances for different metabolites were assumed to be independent from one
20 another. However, even under unsteady-state conditions conserved metabolite pools (CM-pools)
21 (Famili and Palsson, 2003) must remain invariant. These pools arise because the overall amount
22 of a particular chemical moiety (such as acyl carrier protein, NAD⁺/NADH, or folate species)
23 cannot change (in the absence of active biosynthesis) even under the action of the available

1 (exchange) reactions in the system. The mathematical consequence of a CM-pool is that a weighted
2 sum of the RHS terms for metabolites within the pool must equal zero, consistent with a balance
3 between all accumulation (positive RHS) and depletion (negative RHS) terms (Cornish-Bowden
4 and Hofmeyr, 2002; Nikolaev et al., 2005). Thus, any investigation of temporal variation in
5 metabolite levels and resulting phenotypic effects must directly (or indirectly) account for
6 relationships embedded within CM-pools. In the past, CM-pools have been found by either
7 examining the structure of the stoichiometric matrix directly (Cornish-Bowden and Hofmeyr,
8 2002; De Martino et al., 2014; Famili and Palsson, 2003; Reder, 1988) or examining the feasible
9 region of optimization problems with stoichiometric restrictions included (Nikolaev et al., 2005).
10 As many as 74 aggregated CM-pools, whose subsets of members constitute smaller CM-pools,
11 have been identified for the *Escherichia coli* genome-scale model *iAF1260* (De Martino et al.,
12 2014).

13 In this study, we first quantify how uncertainty in the biomass constituent coefficients propagates
14 to FBA results for maximum theoretical biomass yield calculations using the latest *E. coli* model
15 *iML1515* (Monk et al., 2017). Two cases were considered: (i) naively ignoring and (ii) considering
16 the need to re-normalize coefficients so that the biomass MW is strictly 1 g mmol^{-1} . Failing to
17 rescale biomass reaction coefficients resulted in three-fold higher standard deviation (SD) for the
18 biomass yield upon sampling. Uncertainty in the ATP maintenance parameters had an even greater
19 impact on the predicted biomass yield SD (i.e., six-fold higher than that of biomass composition
20 uncertainty). We found, that in general, any injected random noise (i.e., uncertainty) on biomass
21 coefficients is dampened in the FBA prediction for the biomass yield. This observation is
22 consistent with the surprising robustness of FBA predictions despite the many assumptions
23 inherent in the biomass reaction construction and parameterization. However, predicted

1 intracellular metabolic fluxes can be quite sensitive to the injected random noise. We also explored
2 the effect of allowing for deviations from the metabolite steady-state assumption by temporally
3 relaxing the equality of production and consumption terms (i.e., RHS of balance constraints). We
4 found that FBA problems with randomly sampled non-zero RHS vectors were (almost) always
5 infeasible because CM-pool constraints were violated. Projecting all sampled RHS vectors onto
6 the subspace defined by the CM-pool constraints (thereby ensuring their feasibility) results in a
7 seemingly feasible FBA problem but unless elemental (i.e., C, N, P, etc.) balance constraints are
8 explicitly included, undetected elemental balance violations occur. In steady-state FBA, reaction
9 stoichiometries included in the model are elementally balanced, thus no additional elemental
10 balance constraints are needed. However, whenever the RHS terms are perturbed away from zero,
11 the elemental balances are not automatically satisfied thus requiring their explicit inclusion in any
12 unsteady-state FBA calculations. Additionally, we found that the average FBA predicted biomass
13 yield is significantly decreased compared to the steady-state case. This alludes to a strong selection
14 pressure for strictly maintaining steady-state in metabolite balances, as any departures tend to
15 handicap the cell with a lower (on average) biomass yield potential. This further bolsters the
16 underlying assumption of steady-state in FBA and helps explain its effectiveness in predicting
17 overall cellular phenotypes.

18

19 **2. Methods**

20 *2.1. Sampling protocol of perturbations in the biomass coefficients and ATP maintenance*

21 We used the latest *E. coli* model (*iML1515*) (Monk et al., 2017) in minimal media (M9) with
22 glucose as carbon source to determine the effect of uncertainty in biomass coefficients on flux
23 predictions. The engineered reverse beta-oxidation pathway, which is absent in the wild-type strain

1 but included in the model, was turned off by setting the upper bounds of the reactions ACACT1r-
2 8r to zero. The lower bound of pyruvate:flavodoxin oxidoreductase (POR5) was also fixed to zero,
3 as its activity in the reverse direction is only induced by superoxide generators (Nakayama et al.,
4 2013). Model parameters subjected to uncertainty include biomass precursor coefficients,
5 macromolecular compositions, GAM, and NGAM. Parameters were sampled 10,000 times from a
6 normal distribution with the mean centered at the original value and relative SD of 5%, 10%, 20%,
7 and 30% of the original value (hereafter referred to as uncertainty level). During sampling, we
8 ensured that the following material balances were maintained in the biomass reaction: (i) ATP
9 hydrolysis recapitulating GAM, (ii) pyrophosphate in RNA and DNA synthesis, and (iii) water in
10 protein synthesis. Failure to recognize the ATP hydrolysis balance results in a mismatch in ATP
11 and ADP coefficients resulting in either (i) ATP coefficients being higher and its *de novo* synthesis
12 competing with organism growth, or (ii) ADP coefficients being higher creating a metabolic
13 surplus.

14 Biomass precursor coefficients and the GAM value were first extracted from the biomass reaction
15 with lumped coefficients. Then, random sampling of precursor coefficients or the GAM value was
16 performed. Before reassembling the biomass reaction from sampled parameters, re-normalization
17 of the biomass MW to 1 g mmol⁻¹ was performed as:

$$c_i^{new} = \frac{c_i^{samp}}{\sum_{i \in I_{Biom}} c_i^{samp} MW_i} \quad (\text{Eq. 1})$$

18 where I_{Biom} is the set of biomass constituents in the biomass reaction, c_i^{samp} is the randomly
19 sampled coefficient for metabolite i , MW_i is the MW of metabolite i , and c_i^{new} is the recalibrated
20 biomass coefficient for metabolite i . For macromolecular composition perturbation, the weight
21 fractions of macromolecular classes were sampled and coefficients of constituents in each class

1 uniformly scaled. Classification of 66 biomass constituents is provided in Supplementary
2 Materials 1 for all amino acids, cell wall components, lipids, RNAs, DNAs, inorganic ions, and
3 cofactors and prosthetic groups.

4

5 *2.2. Quantifying uncertainty propagation from biomass coefficients and ATP maintenance to* 6 *biomass yield and predicted metabolic fluxes using the standard deviation ratio (SDR) metric*

7 The sampling workflow and propagation of uncertainty is illustrated in Figure 1. For every sample,
8 a biomass reaction was reconstructed and parsimonious flux balance analysis (pFBA) (Lewis et
9 al., 2010) was performed to obtain a predicted flux distribution. Maximum glucose uptake rate was
10 set at 10 mmol gDW⁻¹ h⁻¹. Mean and SD were calculated for the fluxes of every reaction j in the
11 population, and the propagation of uncertainty was quantified using the SD ratio (SDR) defined
12 as:

$$SDR_{Input}^{Output} = \frac{\Delta(SD/Mean)_{Output}}{\Delta(SD/Mean)_{Input}} = \frac{\Delta(\sigma_v/\mu_v)}{\Delta(\sigma_c/\mu_c)} = SDR_c^v \quad (\text{Eq. 2})$$

13

14 where σ_v and μ_v are the SD and mean for the pFBA output flux or yield v , σ_c and μ_c are the SD
15 and mean for the input parameter c subjected to the injection of normally distributed random noise,
16 and SDR_c^v is the rate of change of the output relative SD with respect to the input relative SD. For
17 example, $SDR_c^{v_{Biom}}$ and $SDR_c^{v_j}$ quantify the amplification (or dampening) of the effect of injecting
18 noise on a parameter c on the pFBA predicted biomass yield v_{Biom} or metabolic flux v_j ,
19 respectively. SDR was calculated with linear regression for the calculated output over the input
20 SD.

1 NADH, NADPH, and ATP cofactor production flux SDRs ($SDR_c^{v_{Cofactor}}$) were calculated by
2 aggregating over all cofactor producing reactions ($v_{Cofactor}$),

$$v_{Cofactor} = \sum_{j \in J_{Cofactor}} S_{Cofactor,j} v_j \quad (\text{Eq. 3})$$

3 where $J_{Cofactor}$ is the set of reactions producing a particular cofactor, $S_{Cofactor,j}$ is the cofactor
4 stoichiometry in reaction j , and v_j is the flux through reaction j . See Supplementary Materials 1
5 for a list of all ATP, NADPH, and NADH producing reactions.

6 SDR for the sum of fluxes through alternative (or parallel) pathways was computed in a similar
7 manner to Eq. 2, by summing all fluxes connecting the same start and end metabolites. This is
8 done to quantify the uncertainty propagating to the net conversion flux, which is typically more
9 dampened than to the individually distributed flux components through alternative pathways. See
10 Supplementary Materials 1 for a list of all “flux sum” quantities and the corresponding reaction
11 sets.

12

13 *2.3. Protocol for sampling departures from the metabolite steady-state and instationary FBA*

14 Departure from metabolite steady-state means one or more RHS terms of the FBA’s mass balance
15 constraints (Orth et al., 2010) (denoted by b_i for metabolite i ’s balance (Nikolaev et al., 2005))
16 may assume non-zero values. A positive (or negative) RHS term indicates metabolite accumulation
17 (or depletion). Randomly sampled RHS terms (almost) always lead to a setup of an infeasible
18 optimization FBA problem because CM-pool constraints are consistently violated. These include
19 CM-pool constraints for a subset of RHS terms corresponding to the conservation of chemical
20 moieties (e.g., ACP or CoA) in metabolites (De Martino et al., 2014; Nikolaev et al., 2005; Reder,

1 1988) that cannot be degraded (and *de novo* synthesized for some moieties) due to a lack of
 2 degradation (and biosynthesis) pathway in the model (Chan et al., 2017). In addition, some nutrient
 3 metabolites and their derivatives form a CM-pool in a specific growth condition (but not in others)
 4 if the corresponding exchange reactions that supply the required nutrients are blocked. Thus, for
 5 metabolites in CM-pools, their accumulation is coupled with the depletion of other metabolites.
 6 As the first step in identifying these coupled metabolites, we introduce an optimization formulation
 7 called conserved metabolite pool check (*CMP-check*) that checks whether a given metabolite i^* is
 8 part of a CM-pool. The key idea here is to introduce an imbalance ε denoting accumulation of
 9 metabolite i^* (or $-\varepsilon$ for depletion), and then determine if (i) it is coupled to steady-state
 10 perturbations to other network metabolites, (ii) can be satisfied through exchange with the growth
 11 medium, or (iii) is always infeasible. Slack variables slv_i^U and slv_i^L denote positive and negative
 12 departures, respectively, from the steady-state condition for metabolite i ($i \neq i^*$). The objective of
 13 *CMP-check* is the minimization of the sum of slack variables across all metabolites, thus
 14 identifying the minimal set of coupled (imbalanced) metabolites. The formulation is as follows:

$$\begin{aligned}
 & \min \sum_{i \in I, i \neq i^*} slv_i^U + slv_i^L \\
 & \text{subject to} \\
 & \sum_{j \in J} S_{i^*j} v_j = \varepsilon \\
 & \sum_{j \in J} S_{ij} v_j = slv_i^U - slv_i^L, \quad \forall i \in I \setminus \{i^*\} \\
 & v_j^L \leq v_j \leq v_j^U, \quad \forall j \in J \setminus \{Biom\} \\
 & v_{Biom} = 0 \\
 & 0 \leq slv_i^U, slv_i^L \leq M, \quad \forall i \in I
 \end{aligned}
 \tag{CMP-check}$$

15
 16 where M is an arbitrarily large scalar. Flux through the biomass reaction (v_{Biom}) was turned off to
 17 prevent metabolite rebalancing through accumulation (or depletion) of the biomass sink flux.

1 *CMP-check* was solved twice for every metabolite i^* for both accumulation (using ε) and depletion
2 (using $-\varepsilon$). We used an arbitrarily small value of 0.01 mmol gDW⁻¹ h⁻¹ for ε , and 1,000 mmol
3 gDW⁻¹ h⁻¹ for M .

4 Because generating of CM-pool specific constraints is computationally inefficient for genome-
5 scale models (see Supplementary Text), we hereby introduce a workflow that injects uncertainty
6 into metabolites' RHS values and quantifies uncertainty propagating to FBA-predicted biomass
7 yield while accounting for CM-pools and elemental balance constraints. Imbalances were
8 introduced to the RHS terms of the metabolite mass balance constraints (i.e., by an amount b_i^{rand})
9 by sampling from a normal distribution centered at zero at the uncertainty levels σ^b of 0.1%, 1%,
10 and 10% of the glucose uptake rate by mass (i.e., in the unit of gram glucose gDW⁻¹ h⁻¹). We start
11 from the mass-basis SD σ^b (i.e., unit of gram gDW⁻¹ h⁻¹) to avoid the injecting of larger deviation
12 to metabolites with larger MW values. For example, molar-basis sampling of the accumulation
13 rate of the metabolite “(O16 antigen) x4 core oligosaccharide lipidA” (o16a4colipa_p), whose
14 MW is 44-fold higher than that of glucose, will introduce a disproportional drain of resource in
15 the system. SD in the molar unit (i.e., mmol gDW⁻¹ h⁻¹) (i.e., denoted as σ_i^b) consistent with the
16 unit of the RHS term can be calculated using the following equation:

$$\sigma_i^b = (\sigma^b) \left[\frac{\text{g}}{\text{gDW.h}} \right] * \frac{1}{MW_i} \left[\frac{\text{mol}}{\text{g}} \right] * 1000 \left[\frac{\text{mmol}}{\text{mol}} \right] = \left(\frac{\sigma^b * 1000}{MW_i} \right) \left[\frac{\text{mmol}}{\text{gDW.h}} \right] \quad (\text{Eq. 5})$$

17 where MW_i is the MW of metabolite i . Using optimization formulation *CMP-check*, metabolites
18 were split into three groups: (i) allowed to only accumulate (i.e., $b_i^{rand} > 0$), (ii) allowed to only
19 go into deficit (i.e., $b_i^{rand} < 0$), and (iii) allowed to either accumulate or go into deficit (i.e.,
20 $b_i^{rand} \in \mathbb{R}$). Steady-state departure variables b_i^{rand} were then sampled from a half-normal
21 distribution for cases (i) and (ii) and from a full normal distribution for case (iii). The normal

1 distribution was centered at zero with a SD of σ_i^b . A half-normal distribution is equivalent to a chi-
 2 square distribution with a single degree-of-freedom and scaled by σ_i^b .
 3 To obtain a vector \mathbf{b}^{feas} composed of individual RHS terms b_i^{feas} for FBA, we introduced and
 4 used the Projection Using Linear Programming (*PULP*) formulation that projects the vector \mathbf{b}^{rand}
 5 composed of individual RHS terms b_i^{rand} to the subspace that satisfies network stoichiometry and
 6 reaction directionality restrictions:

$$\begin{aligned}
 & \min \sum_{i \in I} MW_i (sl_i^U + sl_i^L) \\
 & \text{subject to} \\
 & \quad \sum_{j \in J} S_{ij} v_j = b_i^{feas}, \quad \forall i \in I \\
 & \quad b_i^{feas} = b_i^{rand} + sl_i^U - sl_i^L, \quad \forall i \in I \quad (PULP) \\
 & \quad 0 \leq sl_i^U, sl_i^L \leq M, \quad \forall i \in I \\
 & \quad v_j^L \leq v_j \leq v_j^U, \quad \forall j \in J \setminus \{Biom\} \\
 & \quad v_{Biom} = 0 \\
 & \quad \sum_{i \in I} E_{ie} b_i^{feas} = 0, \quad \forall e \in \mathcal{E}
 \end{aligned}$$

7
 8 where sl_i^U and sl_i^L are positive and negative slack variables, respectively, and E_{ie} denotes the
 9 stoichiometry of element or moiety e in metabolite i 's chemical formula. The set of all elements
 10 and moieties collected from the metabolites' formulae is denoted by \mathcal{E} . By minimizing the sum of
 11 slack variables sl_i^U and sl_i^L , the initial randomly generated RHS vector \mathbf{b}^{rand} is thus projected onto
 12 the feasible RHS solution space. The last constraint in *PULP* denotes elemental balance constraints
 13 dictating the conservation of elements, such as C, N, and P.

14 10,000 RHS vectors (\mathbf{b}^{rand}) were randomly sampled, as described above, for each uncertainty
 15 level σ^b and corresponding feasible RHS vectors (\mathbf{b}^{feas}) generated using *PULP*. In order to

1 quantify the effect of metabolite concentration departures from the steady-state condition, we
 2 introduce the instationary *FBA* (*iFBA*) version of FBA where the RHS terms b_i of mass balance
 3 constraints are set to b_i^{feas} to yield a feasible optimization problem. The optimization formulation
 4 for a standard growth-maximization *iFBA* is formulated as:

$$\begin{aligned}
 & \max v_{Biom} \\
 & \text{subject to} \\
 & \sum_{j \in J} S_{ij} v_j = b_i (= b_i^{feas}), \quad \forall i \in I \\
 & v_j^L \leq v_j \leq v_j^U, \quad \forall j \in J \\
 & b_i \in \mathbb{R}, \quad \forall i \in I
 \end{aligned} \tag{iFBA}$$

5
 6
 7 where I and J are the set of metabolites and reactions, respectively, S_{ij} is the stoichiometric
 8 coefficient of metabolite i in reaction j , v_j is the flux of reaction j , v_{Biom} is the specific growth
 9 rate or biomass dilution flux, and v_j^U and v_j^L are the upper and lower bounds of reaction j ,
 10 respectively. v_j^L assumes a value of zero for irreversible reactions and an arbitrarily large negative
 11 value for reversible reactions. Then, the mean and SD of v_{Biom} normalized by steady-state FBA's
 12 v_{Biom} are calculated to quantify the output variation.

13 Note that in the absence of elemental balance constraints, the RHS vector generated by *PULP*
 14 (almost) always introduces a net deficit or surplus of chemical elements to the system. We also
 15 run the *PULP* without the elemental balance constraints yielding elementally imbalanced \mathbf{b}^{feas}
 16 for which the elemental surplus for an element e normalized by the glucose uptake rate by mass
 17 can be quantified as:

$$\frac{MW_e \sum_{i \in I} E_{ie} b_i^{feas}}{MW_{glc} v_{EX_glc}} \tag{Eq. 6}$$

1
2 where MW_e is the atomic weight for the element e , MW_{glc} is the molecular weight of glucose, and
3 v_{EX_glc} is the glucose uptake rate. A positive elemental surplus value means more moles of an
4 element are consumed than produced leading to a metabolic surplus, and vice-versa for a negative
5 value.

6 Because metabolites' formulae are used to derive the elemental balance constraints, the identity of
7 all generic "R" and "X" groups were disambiguated into 16 specific conserved moieties using the
8 metabolite names for 278 metabolite formulae in the *iML1515* model. For metabolites' MW used
9 throughout the analysis, a zero value was arbitrarily assigned for the MW of the conserved moieties
10 (i.e., formerly "R" and "X" groups) since they are not provided with the model. Thus, the MW of
11 metabolites with a partially defined formulae, such as acyl carrier protein (i.e., formula
12 $C_{11}H_{21}N_2O_7PRS$) is calculated after omitting undefined groups. Updated metabolite formulae and
13 their MW values are provided in the Supplementary Materials 2.

14

15 *2.4. Software implementations*

16 Sampling of biomass coefficients was carried out using the Python 3.6 package NumPy for normal
17 distribution sampling (`numpy.random.normal`) (Virtanen et al., 2020) and COBRAPy for flux
18 balance analysis (Ebrahim et al., 2013). Sampling of metabolite RHS terms was carried out using
19 the Python 3.6 package SciPy for normal and half-normal distribution sampling (`scipy.stats.norm`
20 and `scipy.stats.halfnorm`) (Virtanen et al., 2020). *CMP-check* and *PULP* were implemented in
21 General Algebraic Modeling System (GAMS) programming language (version 33.2.0, GAMS

1 Development Corporation), with IBM ILOG CPLEX (version 12.10.0.0) as the solver. Example
 2 scripts are available at <https://github.com/maranasgroup/uncFBA>.

3

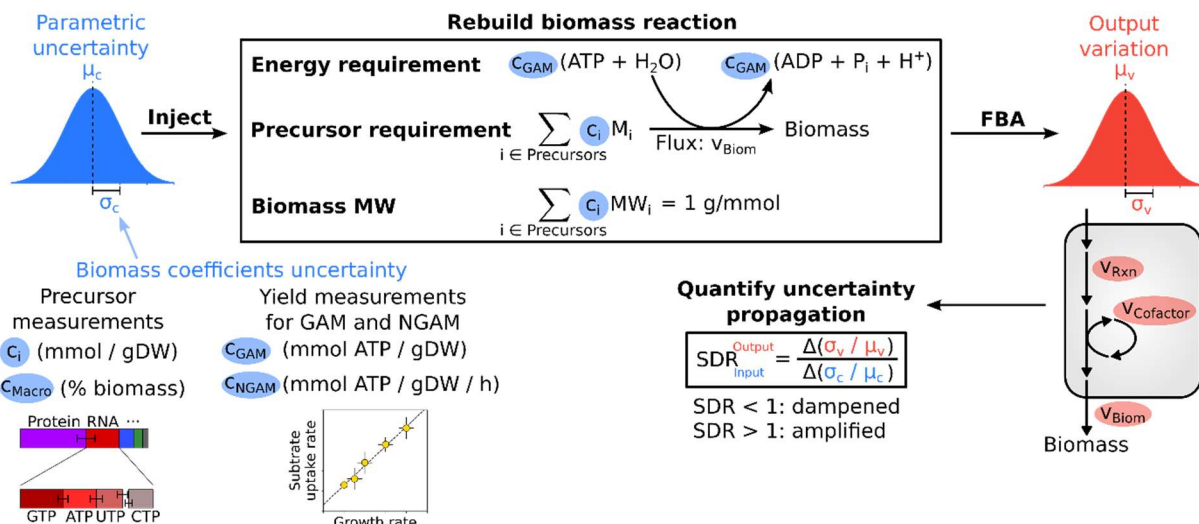
4 3. Results

5 We first describe how the injection of normally distributed random noise into biomass coefficients
 6 and ATP demand propagates in FBA predictions towards biomass and metabolic flux predictions,
 7 and assess the sensitivity (or lack thereof) of predicted flux distributions to input parameters.

8 3.1. Propagation of uncertainty from biomass coefficients and ATP maintenance to FBA 9 predictions of metabolic fluxes and biomass yield

10 The overall workflow which is described in detail in the Methods section (Section 2.1 and 2.2) is
 11 illustrated using Figure 1. We quantify how biomass coefficients and/or ATP maintenance
 12 uncertainty is propagated to FBA predicted biomass yield and metabolic fluxes.

13

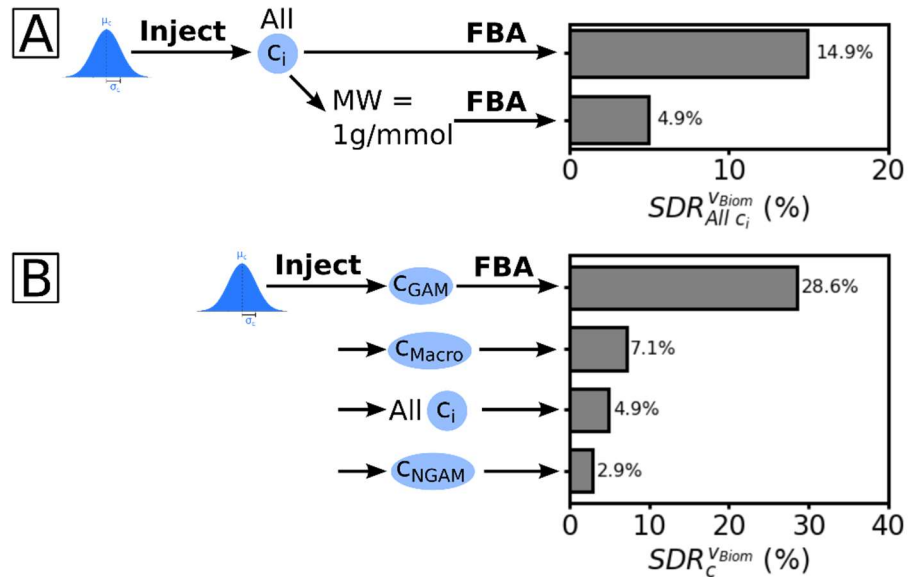


14 **Figure 1.** Overview of the method to inject and quantify propagated uncertainty from biomass
 15 coefficients and ATP maintenance rates to FBA predictions. First, normally distributed uncertainty
 16 is injected to biomass precursor coefficients and ATP maintenance parameters. The biomass
 17 reaction is reassembled while accounting for ATP hydrolysis balance and biomass MW. FBA is
 18

1 next carried out using the new biomass reaction and parameter SDR_{Input}^{Output} quantifies the extent of
2 uncertainty propagation.
3

4 In all cases, the effect of injecting random noise to the biomass coefficients is dampened in the
5 biomass yield predictions by FBA. This is because the random introduction of noise causes some
6 of the coefficients to increase but others to decrease, thus partially counteracting any changes in
7 the overall demand for biomass formation. This is a consequence of the fact that the biomass flux
8 is affected by many constituent coefficients and unless they all change in a coordinated manner,
9 the effect on the biomass flux is attenuated. Even if some random noise injections happen to cause
10 a coordinated increase in the demand for some constituents for the formation of one unit of
11 biomass, perturbations for other constituents are likely to cancel them out. Furthermore, as seen in
12 Figure 2A, by renormalizing the biomass coefficients such that biomass MW is maintained at 1 g
13 mmol⁻¹, the propagation of uncertainty to the biomass yield is revealed to be even smaller than by
14 naively ignoring this renormalization. The relative magnitude of the noise amplitude decreases
15 from 14.9% to 4.9% as quantified by the standard deviation ratio $SDR_{All c_i}^{v_{Biom}}$ metric (Eq. 2).
16 Uncertainty in the macromolecular composition leads to a higher $SDR_c^{v_{Biom}}$ (i.e., 7.1% vs. 4.9%,
17 see Figure 2B). Note that in a previous study examining biomass coefficient uncertainty
18 (MacGillivray et al., 2017), the effect of omitting MW-normalization was not detectable because
19 the imposed uncertainty was much smaller (i.e., between $\sim 10^{-1}$ to $10^{-2} \pm 10^{-6}$ mmol gDW⁻¹ for most
20 metabolites) as compared to this study (i.e., between $\sim 10^{-1}$ to 10^{-2} mmol gDW⁻¹ ± 10 to 30%).

21



1
2 **Figure 2.** Effect of uncertainty in biomass composition and ATP demand on FBA predictions. (A)
3 Neglecting to reweigh biomass MW to 1 g mmol⁻¹ leads to a 3-fold exaggerated effect to the
4 biomass yield output. (B) Contribution of individual biomass coefficients (c_i), macromolecular
5 composition (c_{Macro}), GAM (c_{GAM}), and NGAM (c_{NGAM}) uncertainty on biomass yield output.
6

7 SDR metrics for the biomass yield and cofactor net production fluxes are provided in Table 1. We
8 find that random noise propagation is highly dampened (i.e., $SDR < 10\%$ with few exceptions) for
9 both biomass yield and cofactor fluxes. This may explain why FBA is surprisingly successful at
10 predicting biomass and/or flux yields despite the highly uncertain and/or variable nature of
11 required inputs to describe the carbon and energy tally for biomass formation. Because NADPH
12 is the primary redox cofactor for biosynthesis of biomass constituents, FBA predicted variance for
13 the total NADPH synthesis flux (i.e., $SDR_c^{V_{NADPH}}$) under biomass composition uncertainty is
14 higher than those for biomass, NADH or ATP (see Table 1).

15

16 **Table 1.** Calculated SDR values for biomass yield and cofactor net production due to propagated
17 uncertainty from biomass coefficients and ATP maintenance parameters. Do note that biomass
18 MW is reweighted to 1 g mmol⁻¹ unless mentioned otherwise in the remainder of the manuscript.
19

Sampled parameters	$SDR_c^{v_{Biom}}$	$SDR_c^{v_{NADPH}}$	$SDR_c^{v_{NADH}}$	$SDR_c^{v_{ATP}}$
<i>Species composition</i>				
All coefficients	4.9	10.1	5.5	5.9
All coefficients (unnormalized MW)	14.9	11.2	8.9	10.6
<i>Macromolecular composition*</i>				
All compositions	7.1	16.3	9.6	9.3
RNA fraction (22%)	5.0	8.2	6.4	7.0
Lipid fraction (10%)	4.2	0.3	0.8	3.2
Protein fraction (58%)	1.8	13.9	7.2	5.1
Inorganic ions fraction (1%)	0.7	0.3	0.4	1.3
Cell wall fraction (6%)	0.5	2.5	0.3	1.1
DNA fraction (3%)	0.5	1.2	0.7	0.8
Cofactors fraction (0.3%)	0.01	0.1	0.1	0.7
<i>ATP maintenance</i>				
Growth-associated (GAM)	28.60	28.48	38.73	46.81
Non-growth associated (NGAM)	2.92	2.93	3.94	4.94

*Number in brackets is the mass fraction of the respective macromolecule class in the *iML1515 E. coli* GSM's biomass reaction [10].

1
2
3
4 In contrast to imposing random noise on all biomass constituent coefficients at the same time,
5 change in only the GAM requirement causes more pronounced changes in the biomass yield and
6 cofactor balances (see Table 1 and Figure 2A). This is expected as only a single coefficient for
7 GAM undergoes a fluctuation causing systemic changes in ATP demand. This offers a cautionary
8 tale, as a common practice in model reconstruction is to port unchanged ATP maintenance values
9 from previously built models for related organisms even though GAM values are highly variable
10 (e.g., yeasts: 23 – 167 (Aung et al., 2013; Dinh et al., 2019; Mishra et al., 2018; Tomàs-Gamisans
11 et al., 2018, 2016; Torres et al., 2019) & bacteria: 18 – 75 mmol gDW⁻¹ h⁻¹ (Flahaut et al., 2013;
12 Kavvas et al., 2018; Liao et al., 2011; Monk et al., 2017; Orth et al., 2011; Seif et al., 2019; Thiele
13 et al., 2011)). Growth conditions such as supplied substrate (Tomàs-Gamisans et al., 2018, 2016)
14 and nutrient- and oxygen-limited growth (Dinh et al., 2019; Torres et al., 2019) can also cause
15 significant departures. In contrast, changing only the coefficient for NGAM caused a much smaller

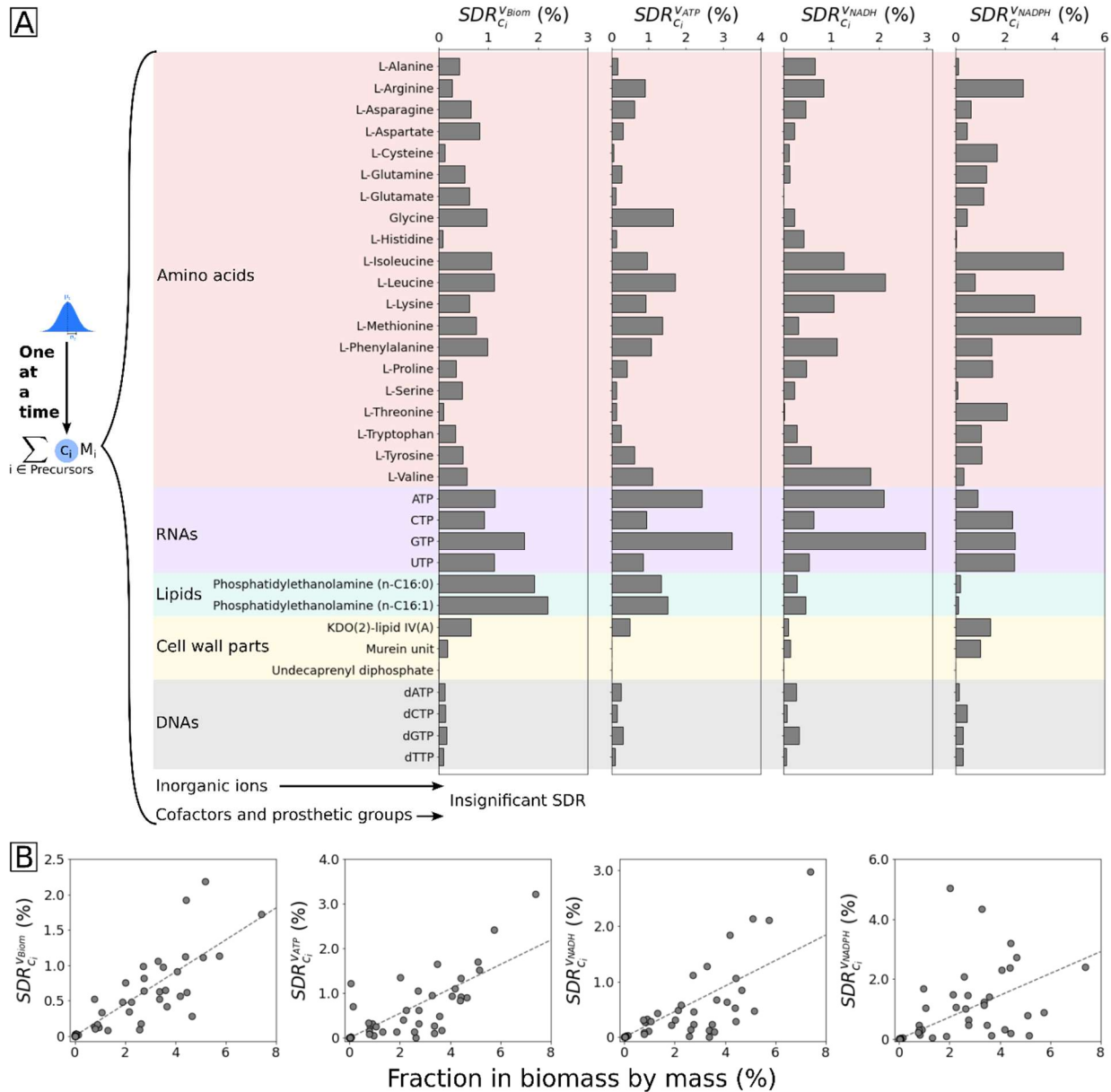
1 change in the predicted biomass yield (i.e., SDR_{cNGAM}^{vBiom} of 2.9%). This is a reflection of the relative
2 demands imposed by GAM and NGAM, respectively, on the consumed glucose (i.e., 27% and
3 3%). The effect of injecting noise on precursor demands for biomass formation (i.e., coefficients
4 or macromolecular compositions) was significantly smaller even though a much larger fraction of
5 glucose (i.e., ~71%) is allocated to their synthesis. Again, the reason for this insensitivity is that
6 the injection of random noise tends to alter their relative proportions but not the total mass (i.e., 1
7 gram of precursors to make 1 gram of biomass). In contrast, random noise on GAM or NGAM
8 requirements directly increases or decreases the net ATP demand without a compensating
9 mechanism. This implies that random errors for coefficients of multiple biomass constituents in a
10 genome-scale model are likely to lead to an attenuated response of FBA biomass yield predictions,
11 whereas systematic errors that directly affect the overall carbon and energy demand will likely
12 cause changes in FBA prediction in proportion to the magnitude of the error.

13 Zooming in closer we also examined the sensitivity of the biomass yield upon injecting noise one-
14 at-a-time to macromolecular groups (see Table 1 and Figure 2C). As expected, uncertainty in
15 macromolecules accounting for a larger fraction of biomass (by mass) gave rise to a higher
16 SDR_{cMacro}^{vBiom} value. These include proteins (58% by weight), RNAs (22% by weight), and lipids
17 (10% by weight). The relative order is determined not only by the net weight fraction, but also by
18 the amount of carbon, ATP, and NAD(P)H consumed during biosynthesis. Even though RNAs
19 and lipids rank after proteins in terms of contributing biomass weight fractions, their uncertainty
20 propagation has a larger effect in biomass yield (Table 1). This is due to higher cofactor costs when
21 forming more covalent bonds for larger RNA ($C_9 - C_{10}$) and lipid (C_{37}) molecules compared to
22 amino acids ($C_2 - C_9$). Interestingly the effect on NAD(P)H production upon the introduction of
23 random noise in the lipid fraction is much less than the effect on biomass yield ($SDR_{c_i}^{vBiom} = 4.2\%$

1 vs. $SDR_{c_i}^{vNADPH} = 0.3\%$ and $SDR_{c_i}^{vNADH} = 0.8\%$). This is due to cofactor promiscuity in enoyl-ACP
2 reductase (catalyzing the last step in a fatty acid elongation cycle) (Bergler et al., 1996) allowing
3 the use of NADH in place of NADPH as the reducing equivalent. As a result, equimolar
4 consumption of ATP, NADH, and NADPH occurs per fatty acid elongation cycle (out of 16 cycles
5 in total). The large but almost equimolar total requirement of 18 ATP, 17 NADPH, and 16 NADH
6 in mol/mol for lipid synthesis could explain the reason for its effect on biomass yield but not
7 cofactor availability as the equimolar cofactor requirement allows the cofactor load to be
8 distributed across different cofactor producing parts of the network. In contrast, cofactor demand
9 of protein synthesis is skewed towards NADPH (e.g., L-methionine requires 8 NADPH, 5 ATP,
10 and 1 NADH in mol/mol) whereas cofactor demand of RNA synthesis is skewed towards more
11 ATP (e.g., *de novo* synthesized ATP requires 6 NADPH and 11 ATP in mol/mol), creating
12 disproportionally heavier loads to NADPH and ATP, respectively which is reflected in the
13 calculated SDR (Table 1). This observation is consistent with the experimentally observed
14 importance of controlling NADPH metabolism in amino acid producing strains of *E. coli* (Xu et
15 al., 2018). Lastly, while minor fraction components such as cofactors and inorganic ions are
16 important to include in expanding the resolution of FBA predictions (Feist and Palsson, 2010), the
17 effect of introducing noise in their respective biomass coefficients on biomass yield is negligible
18 as they constitute only a small fraction (i.e., ~1% of biomass by weight, all $SDR_{c_i}^{vBiom} < 1\%$).

19 Next, we examined the effect of injecting random noise (one-at-a-time) to biomass coefficients
20 and quantified the output SDR (see Figure 3A and Supplementary Material 1) to determine
21 biomass precursors that have a disproportionally large effect on FBA predictions. As seen
22 previously, uncertainty propagation partially correlates with how much each precursor contributes
23 to biomass by mass (i.e., gram per gram biomass, %) (see Figure 3B). Exceptions to this general

1 trend include the biomass constituents of two phosphoethanolamines (i.e., lipid species) affecting
2 biomass yield, GTP and ATP (i.e., RNA species) affecting ATP production, and L-methionine and
3 L-isoleucine (i.e., amino acids) affecting NADPH production due to their differences in cofactor
4 costs, as explained for macromolecular fraction uncertainty. GTP and ATP biomass coefficient
5 values also affect NADH production, as NADH fuels oxidative phosphorylation to generate ATP.
6 Interestingly, NADH production is affected by the L-leucine biomass coefficient because it is
7 produced in the leucine biosynthesis pathway: one from 3-isopropylmalate dehydrogenase and one
8 from producing the precursor acetyl-CoA (from pyruvate via pyruvate dehydrogenase). Based on
9 the reported SDR values (see Supplementary Material 1) this analysis can be used to prioritize
10 measurements of biomass components that could improve the accuracy of flux and yield
11 quantification.

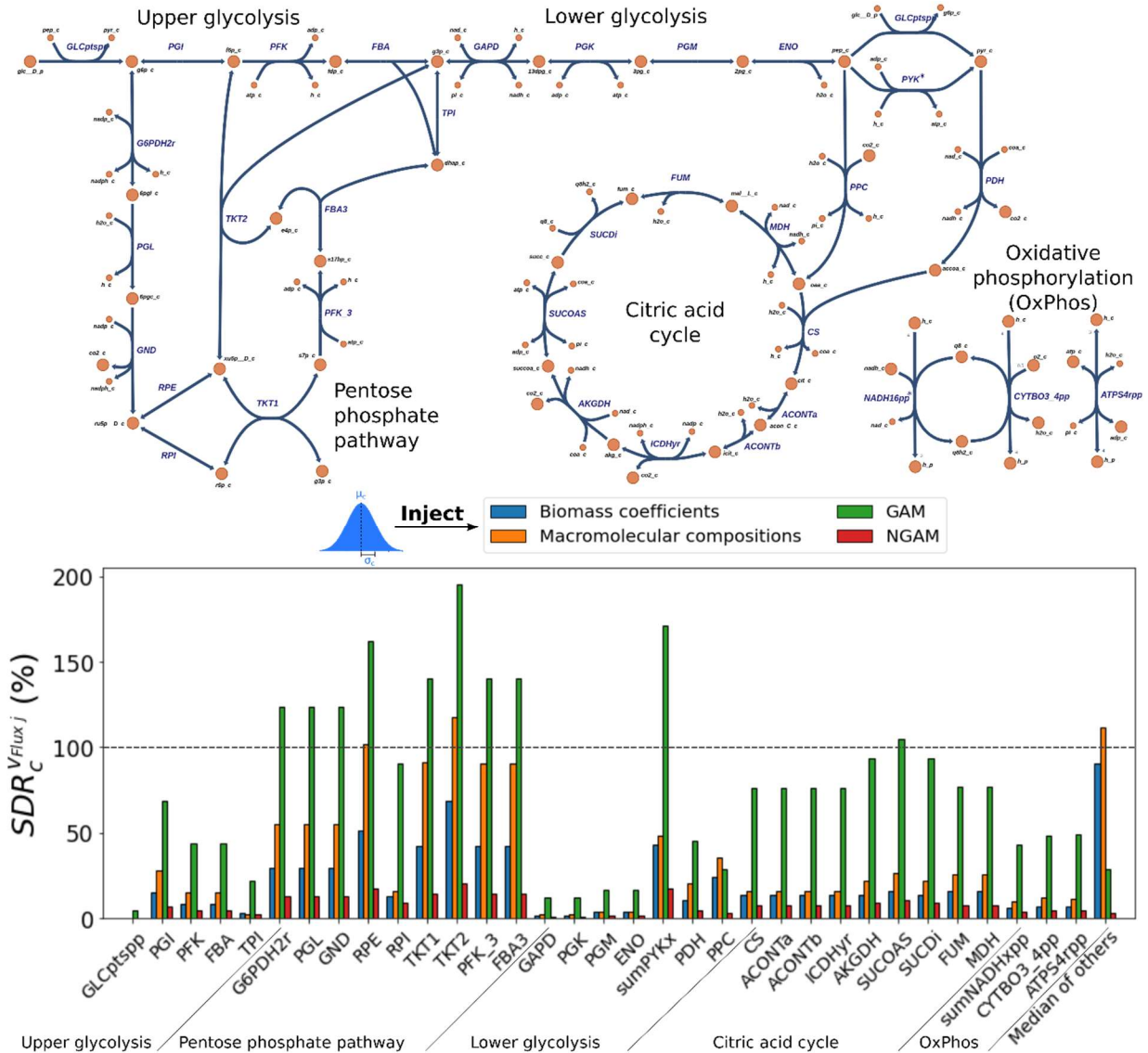


1
 2 **Figure 3.** (A) Responses in output standard deviation for biomass yield and cofactor fluxes due to
 3 “one-at-a-time” noise injection to biomass coefficients. From left to right are responses in the
 4 output standard deviation ratios of biomass yield, ATP, NADH, and NADPH fluxes. (B) Plots of
 5 the aforementioned SDR values vs. coefficient mass fraction by mass in biomass composition. The
 6 dotted lines are the linear regression lines.

7
 8 Unlike biomass yield or aggregated demands for cofactors, injecting random noise in biomass
 9 coefficients can cause varied, and in some cases, an amplified effect for some metabolic fluxes
 10 and pathways (see Figure 4). Central metabolism fluxes exhibited higher SDR values under GAM

1 uncertainty whereas downstream biosynthesis fluxes had larger SDR under biomass composition
2 uncertainty. Notably the SDR for the pentose phosphate pathway was consistently high under
3 either GAM or biomass coefficients uncertainty. Under NGAM uncertainty, as expected, all
4 reaction $SDR_{c_{NGAM}}^{v_j}$ values were below 25% (median = 2.9%). Even though on average the
5 uncertainty propagation from biomass composition random noise injection is still less than
6 proportional for most of the fluxes, it is significantly larger (see Figure 4) than the one for biomass
7 yield (see Table 1). This implies that biomass composition accuracy is more important for
8 predicting individual metabolic fluxes using FBA as opposed to the overall biomass yield.

9



1 Upper glycolysis Pentose phosphate pathway Lower glycolysis Citric acid cycle OxPhos Median of others
 2 **Figure 4.** SDR of individual fluxes in central metabolism. Reaction IDs used in the metabolic map
 3 and the bar chart follow the BiGG notation used by the model *iML1515* [31]. The last entry,
 4 “median of others”, corresponds to the median SDR for all reactions that are not part of central
 5 metabolism (i.e., mostly biosynthesis reactions).
 6

7 Starting from the extracellular glucose node, the phosphotransferase system (PTS) exhibited the
 8 least variation among all reactions upon random noise injection to the parameters (see Figure 4,
 9 reaction *GLCptspp* in “upper glycolysis”). The maximization of biomass yield in FBA indirectly
 10 couples it with maximal glucose uptake. Further downstream, glucose 6-phosphate (G6P) is
 11 funneled towards either G6P isomerase (*PGI*, in glycolysis) or dehydrogenase (*G6PDH2r*, in PPP).

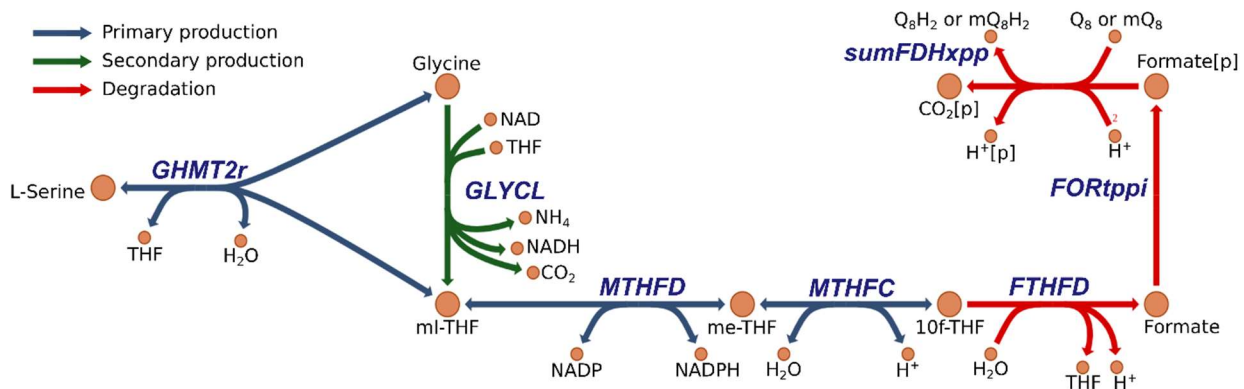
1 We found that the flux towards G6PDH2r exhibited about twice the sensitivity to the injected
2 uncertainty than towards PGI. This is because PGI funnels twice as much flux as G6PDH2r (i.e.,
3 65% vs. 35% of glucose uptake). The split portions rejoin in lower glycolysis, which showed
4 relatively low sensitivity to parametric uncertainty except for the last pyruvate kinase step (Figure
5 3). This is because pyruvate is a precursor for many biomass constituents as well as acetyl-CoA
6 used in TCA cycle to produce additional energy. After pyruvate, fluxes in energy-producing
7 pathways of TCA cycle and oxidative phosphorylation are relatively sensitive to GAM uncertainty
8 as the net ATP demand on the system is affected. The pentose phosphate pathway (PPP, producing
9 NADPH and various biomass precursors) was also affected by GAM uncertainty, as an increase
10 in energy demand requires more glucose to be utilized to produce ATP, in turn releasing more CO₂
11 as the respiratory byproduct. In particular, transketolase-2 reaction (TKT2) exhibited the highest
12 sensitivity possibly because it is the only reaction in the pFBA solution connecting PPP to the
13 fructose-6P (F6P) node, which is a precursor to the murein unit (i.e., a cell wall component) in
14 biomass. Thus, a coordinated increase or decrease in the demand of PPP metabolites and F6P
15 affects TKT2 more than other PPP reactions that are not connected to the F6P node. Beyond
16 glycolysis, a low sensitivity to biomass composition perturbation was observed in the TCA cycle
17 and oxidative phosphorylation, indicating a relatively stable cellular demand for ATP demand
18 despite fluctuations in composition that tend to counteract one another.

19 Biomass coefficient variability can also affect metabolism by activating pathways that did not
20 carry flux in the pFBA solution using the original biomass coefficients, where a total of 434
21 reactions carry flux at maximal growth under glucose utilization. Upon injecting noise to biomass
22 coefficients or GAM uncertainty, this number increases to 471 indicating that alternative pathways
23 may become active in response to a perturbed biomass composition or ATP demand (see

1 Supplementary Text for the list of activated alternative pathways). This implies that changes in the
2 value of the coefficients in the biomass objective function in pFBA can update the optimality order
3 of vertices in the polytope describing the phenotypic space. In some cases, activations are trivial
4 such as switching between quinone and menaquinone but in other cases, physiologically relevant
5 pathway activations are observed involving the use of an alternative redox cofactor (e.g., NADH
6 or NADPH) or energy molecule (e.g., ATP or GTP). Other reaction activation cases involve short
7 linear pathways such as the ones in glycine cleavage (GLYCL), some reactions in one carbon
8 metabolism (FTHFD, FORtppi, sumFDHxpp; Figure 5), GTP biosynthesis (sumGTPsyn), and
9 hydrogen peroxide production (SPODM, sumQMOx).

10 Variability in the coefficients of biomass precursors (such as RNAs) that require one-carbon
11 moiety in the form of 5,10-methylene-tetrahydrofolate (m1-THF) or 10-formyl-tetrahydrofolate
12 (10f-THF) (see Figure 5) can have systemic effects through one carbon metabolism. In *E. coli*,
13 one-carbon moiety is co-produced from glycine via the glycine hydroxymethyltransferase (see
14 Figure 5, “primary production”). In most samples, biomass coefficients for precursors requiring
15 one-carbon moiety are high enough that additional contribution of the moiety from GLYCL is
16 required (see Figure 5, “secondary production”) and only in limited samples (~6% cases for 10%
17 and ~30% cases for 30% uncertainty) the moiety was produced in excess. The surplus amount was
18 degraded to formate (via 10f-THF deformylase (FTHFD)) and subsequently to CO₂ (FORtppi,
19 sumFDHxpp). Interestingly, the activation of this degradation pathway is consistent with the
20 physiological role of FTHFD in controlling the one-carbon pool (Nagy et al., 1995). Another
21 pathway activated under biomass coefficient noise injection was in GTP production (sumGTPsyn),
22 which exhibited a bimodal distribution of flux values across samples. This was due to the presence
23 of two alternative adenylate kinase reactions phosphorylating AMP which can use either ATP

1 (reaction ADK1) or GTP (reaction ADK3) (Rogne et al., 2018). Finally, hydrogen peroxide (H_2O_2)
 2 production, via quinol monooxidase (sumQMOx) and then superoxide dismutase (SPODM),
 3 followed a normal distribution with a relatively wide spread ($\sim 240\% SDR_{All c_i}^{vj}$). This is because
 4 this pathway is not the sole H_2O_2 source in the network and is only used in a subset of samples to
 5 supplement production of the biomass constituent Fe^{3+} (via reaction FESD1s). Overall, analyzing
 6 the activation of pathways upon injection of noise in biomass composition can offer clues as to the
 7 optimality criteria that govern the utilization of specific pathways over others and pinpoint which
 8 alternatives may become activated when the overall carbon and energy demands are altered.



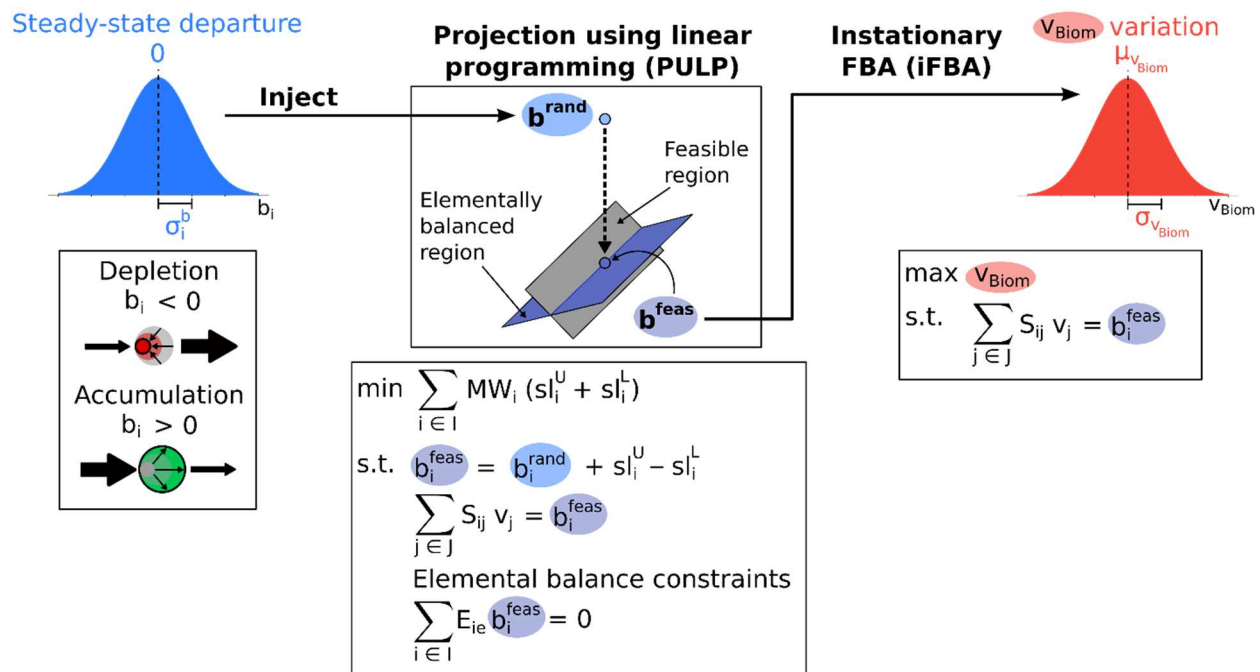
9
 10 **Figure 5.** Overview of the one carbon metabolism pathway. GLYCL (highlighted in green)
 11 complements 10f-THF production whereas its degradation pathway (highlighted in red) is used to
 12 degrade excessive 10f-THF. [p]: in periplasm compartment; THF: tetrahydrofolate; mI-THF: 5,10-
 13 methylene-THF; me-THF: 5,10-methenyl-THF; 10f-THF: 10-formyl-THF; Q₈: ubiquinone-8;
 14 Q₈H₂: ubiquinol-8; mQ₈: menaquinone-8; mQ₈H₂: menaquinol-8; GHMT2r: glycine
 15 hydroxymethyltransferase, GLYCL: glycine cleavage system, MTHFD: methylene-THF
 16 dehydrogenase; MTHFC: methenyl-THF cyclohydrolase; FTHFD: formyl-THF deformylase;
 17 FORtpi: formate transport via diffusion; sumFDHxpp: formate dehydrogenase (Q₈ or mQ₈).
 18

19 *3.2. Propagation of temporally decoupled metabolite balances on FBA biomass yield predictions*
 20 Equality of production with consumption for all metabolites is a key assumption that imposes
 21 restrictions on FBA outputs. However, metabolism need not obey this stringent criterion at all
 22 time-points of the growth cycle, thereby allowing for some surpluses or deficits that can later be

1 evened out. These temporary departures from the strict steady-state assumption may reflect
2 inherent noise in the rate of reactions (Levine and Hwa, 2007), variability among cellular
3 subpopulations in a culture (Kumar et al., 2020), coordinated responses to changes in the
4 microenvironment (Schmitz et al., 2017), or cellular programming (Andreas Angermayr et al.,
5 2016). Sampling unsteady-state regimes may offer the cell an opportunity to jump start metabolism
6 by depleting a stored metabolite pool (i.e., glycogen storage (Shinde et al., 2020)) or deplete
7 oxygen before an oxygen-sensitive enzyme is produced (Milligan et al., 2007; Rabouille et al.,
8 2014). Herein we aim to explore the allowed magnitude of these departures and their overall effect
9 on fitness as quantified by biomass yield. By randomly perturbing the RHS terms of the metabolite
10 balances we (almost) always end up with an infeasible *iFBA* problem (see Section 2.3). This
11 implies that any departure from metabolite steady-state must satisfy the constraints associated with
12 the conservation of conserved metabolite pools (CM-pools, see Section 2.3).

13 Ignoring these constraints overestimates the permitted variability in the network and results in
14 infeasible solutions. One can attempt to classify all metabolites as to their participation CM-pools
15 by examining one of them at a time (see *CMP-check* formulation in Section 2.3). Three outcomes
16 are possible. A metabolite is part of a CM-pool if any accumulation (or depletion) is exactly
17 balanced by another metabolite's pool depletion (or accumulation), respectively. Some metabolites
18 are not a part of any CM-pools as both their accumulation and depletion can be compensated via
19 exchange with the extracellular medium. Finally, the accumulation or depletion of a metabolite
20 cannot be re-balanced by perturbing the pool of other metabolites or via exchange with the medium
21 (see Supplementary Materials 2 for the detailed *CMP-check* results). We found that the generation
22 of CM-pool specific constraints scales poorly and becomes computationally inefficient for
23 genome-scale models. Instead, we chose to indirectly satisfy all such CM-pools by simply

1 projecting all RHS vector samples back to stoichiometric feasibility by minimizing the sum of
 2 violations (see *PULP* formulation in Section 2.3) while ensuring moiety and elemental balances.
 3 Figure 6 provides a pictorial representation of the *PULP* optimization formulation. Randomly
 4 sampled RHS terms are projected onto feasible assignments for the RHS vector as dictated by
 5 network stoichiometry and reaction directionalities. This was implemented by adding slack
 6 variables to the randomly sampled RHS terms and employing as the objective function the
 7 minimization of the MW weighted sum of these slack variables. This weighting converts the
 8 projection to a mass-basis (i.e., gram gDW⁻¹ h⁻¹) from a molar-basis (i.e., mmol gDW⁻¹ h⁻¹), thereby
 9 preventing the RHS vector adjustment to bias the accumulation or depletion of metabolites with
 10 higher MWs. The same mass-basis was used to generate RHS random samples (see Section 2.3
 11 for details on the sampling procedure).

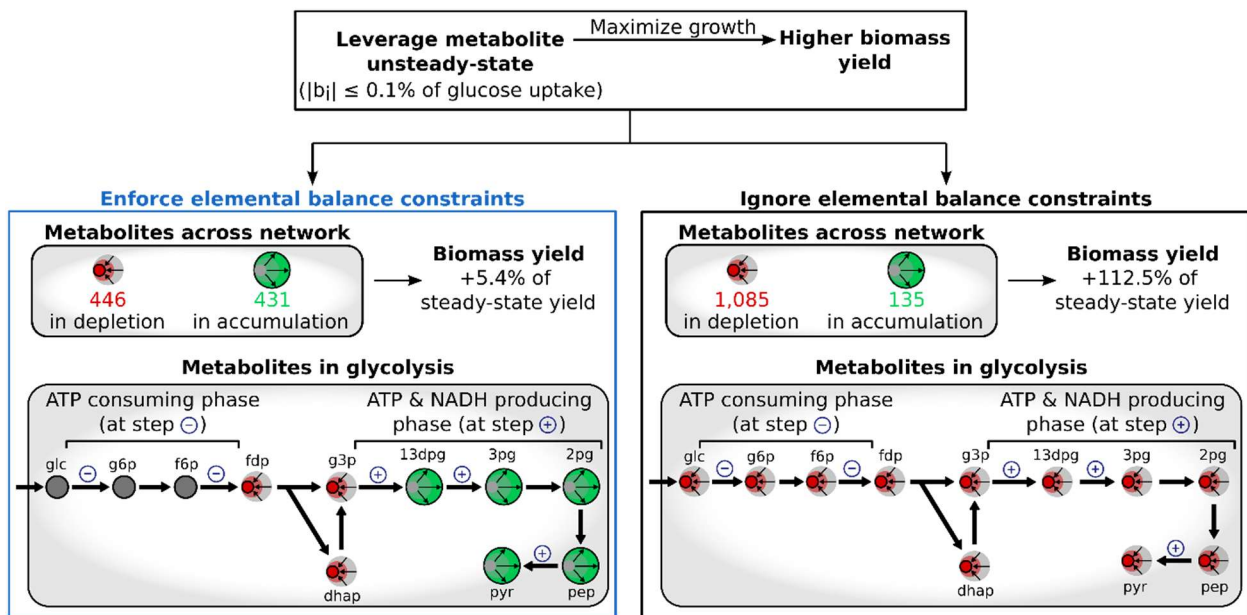


12
 13 **Figure 6.** Overview of the method to sample and propagate uncertainty from metabolites unsteady-
 14 state. First, an ensemble of \mathbf{b}^{rand} vectors is generated by sampling from a normal distribution
 15 $\mathcal{N}(0, \sigma^b)$. Next, the optimization formulation *PULP* is used to project (one-at-a-time) the vector
 16 \mathbf{b}^{rand} onto the closest RHS vector \mathbf{b}^{feas} that satisfies both the conserved metabolite pools and

1 elemental balances. These vectors \mathbf{b}^{feas} are used as inputs for *iFBA* to generate biomass yield
2 predictions under RHS uncertainty.
3

4 Feasible RHS vectors \mathbf{b}^{feas} obtained from *PULP* are then used as inputs to the *iFBA* optimization
5 (see Section 2.3) to calculate flux distributions and biomass yields and perform statistical analysis.
6 Upon plotting the probability distributions of b_i^{feas} , we find that they remain approximately
7 normal for most metabolites (see Supplementary Text for a few example histograms). An
8 important observation is that the \mathbf{b}^{feas} from *PULP* without the direct inclusion of the elemental
9 balance constraints (see Section 2.3) introduces elemental imbalances to the *iFBA* output. This is
10 because under steady-state FBA, all reactions carry out metabolite conversions that are
11 automatically elementally balanced, assuming that the model stoichiometry is elementally
12 balanced (Chan et al., 2017). However, in *iFBA* the departure vectors \mathbf{b}^{feas} introduce metabolite
13 surpluses and/or deficits in a manner unconstrained by the model stoichiometric coefficients. The
14 indirect incorporation of CM-pool constraints based on the *PULP* projection formulation ensures
15 that the conservation of moieties (e.g., ACP, CoA, etc.) is maintained but elemental balances are
16 not necessarily satisfied. This means that elemental balances must be directly imposed in the *PULP*
17 formulation (see Section 2.3). To demonstrate what happens when the elemental balance
18 constraints are not satisfied, we performed calculations using a modified version of *PULP* where
19 we maximized growth under unsteady-state, both with and without the elemental balance
20 constraints (see Figure 7A for an illustration). The upper bound of metabolite accumulation and
21 depletion rates is set to 0.1% of glucose uptake rate, by mass. As expected, the cell can deplete
22 some metabolite pools (i.e., 446) to (temporarily) grow faster by 5.4% while accumulating other
23 metabolites (i.e., 431) (Figure 7B). This is achieved by inducing local imbalances around ATP and
24 NAD(P)H producing reactions. For example, by depleting fructose-1,6P (fdp), glyceraldehyde-3P

1 (g3p) and dihydroxyacetone phosphate (dhap) and simultaneously accumulating downstream
 2 metabolites, the cell can temporarily drive more flux through the payoff phase of glycolysis and
 3 yield up to 0.7% more ATP and 1.7% more NADH (see Figure 7C). However, lack of inclusion
 4 of elemental balances leads to a systematic depletion of metabolite pools (i.e., 1,085 depleted vs.
 5 135 in surplus, respectively). Pool depletion is used as an alternative substrate source thus
 6 bypassing ATP dependent exchange reactions and reaching an erroneously predicted increase in
 7 biomass yield of 112.5% (Figure 7B). Therefore, care must be taken when interpreting unsteady-
 8 state FBA results because when casting the RHS vector as an unconstrained optimization variable,
 9 violations of the conserved metabolite pools and elemental balances can occur, leading to
 10 unrealizable flux distributions and an overprediction of the maximum possible biomass yield.



11 **Figure 7.** Maximization of biomass yield under metabolite unsteady-state for the two cases that
 12 enforce (left) or ignore (right) elemental balance constraints. Shown for each case are the numbers
 13 of metabolites depleting or accumulating, biomass yield change, and a closer look at the glycolysis
 14 pathway.
 15
 16

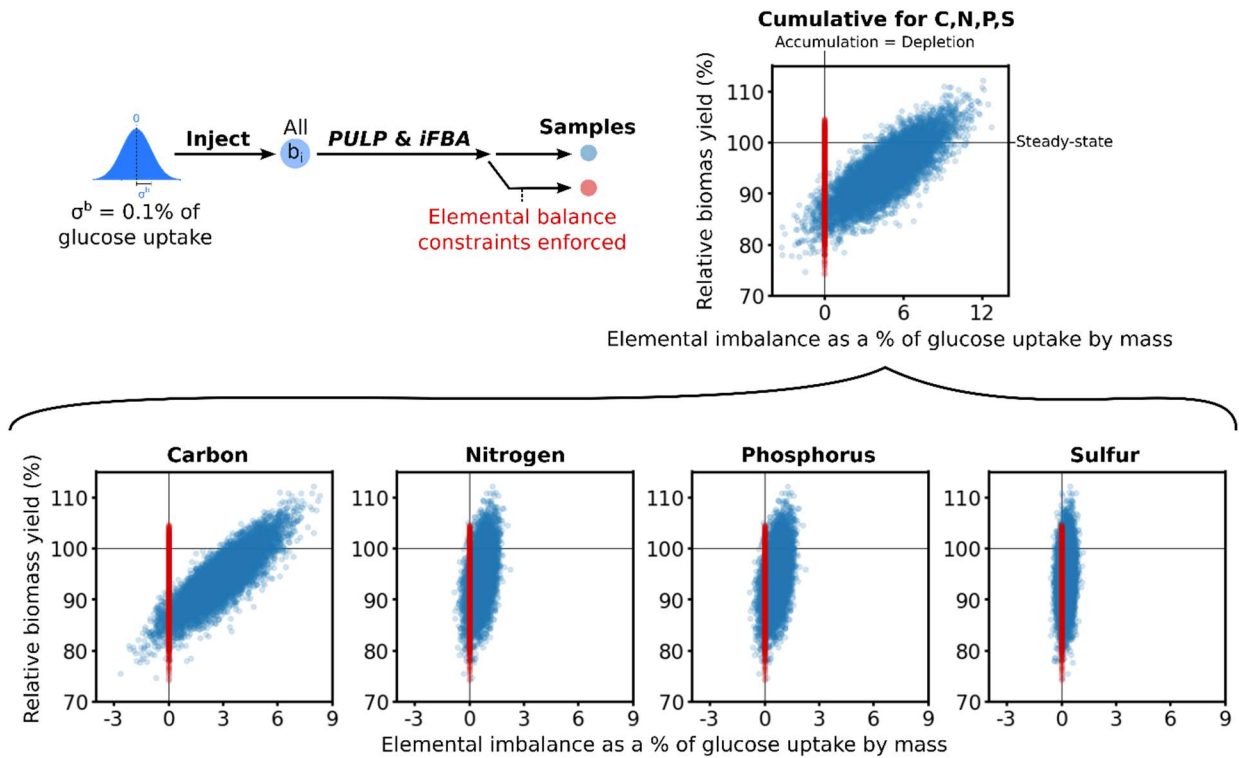
17 Neglecting the explicit introduction of elemental balance constraints during sampling using *PULP*
 18 introduces systematic elemental surplus which artificially boosts biomass yield. Figure 8 illustrates

1 biomass yield predictions (i.e., cloud of blue dots) as a function of the percent elemental imbalance
2 (C, N, P, and S) caused by the sampling without considering elemental balance constraints. Lack
3 of consideration of carbon elemental balance causes most instances to be in overall carbon deficit
4 as *iFBA* maximally drains carbon substrates towards biomass. Despite this negative balance, only
5 a fraction of *iFBA* simulations lead to results that exceed steady-state FBA biomass yield. This is
6 because metabolite imbalances cause biomass constituents to not be produced in the required
7 ratios. The red vertical streak of points depicts the simulation when the elemental balance
8 constraints are imposed. As expected, they always satisfy carbon balance and in the great majority
9 of cases have a significantly decreased biomass yield, alluding to the largely detrimental effect of
10 departing (even temporarily) from steady-state. Even though N and P are not limiting, as
11 ammonium and phosphate uptake rates were left unconstrained, some enhancement of biomass
12 yield as a consequence of their elemental imbalances were still observed (Figure 8). This is because
13 N and P supplied from the growth medium require energy (i.e., NADPH or ATP) to be assimilated,
14 which can be bypassed by violating elemental balance constraints and leveraging imbalances in
15 intracellular metabolites. Note that for nitrogen incorporation through the carrier L-glutamate an
16 NADPH molecule is required whereas for phosphate incorporation to a molecule an ATP molecule
17 is typically required. Notably even though sulfur has a similar assimilation energy requirement via
18 sulphate adenylyltransferase requiring 2 ATPs and a GTP, no significant correlation between
19 biomass yield and sulfur surplus was observed (see Figure 8). This is likely because sulfur accounts
20 for only 0.8% of the biomass by weight compared to 49.3% for carbon, 15.1% for nitrogen and
21 9.2% for phosphorous.

22

1

2



3

4 **Figure 8.** *iFBA* calculated biomass yields vs. incurred elemental imbalance as a % of the glucose
 5 uptake by mass upon ignoring (samples in blue) and enforcing (samples in red) elemental balance
 6 constraints.

7

8 Interestingly, we found that average biomass yield is much more sensitive to noise injection to the

9 RHS vectors of the steady-state constraints than to the biomass composition coefficients. Even for

10 an injected random noise with a standard deviation of just 0.1% of the glucose uptake rate (by

11 mass) to the RHS, the average biomass yield for the population decreases by 7.1% from the steady-

12 state yield (Figure 9). More substantial departures from steady-state attained by injecting noise

13 that is 1% or 10% of the glucose uptake rate, leads to dramatic reductions of the average biomass

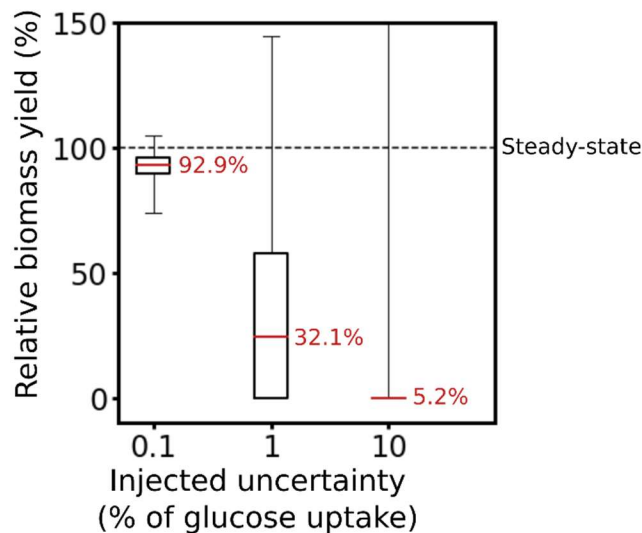
14 yield by 67.9% or 94.8%, respectively. While it is possible for the cell to achieve faster growth by

15 temporarily leveraging metabolite imbalances (see Figure 7 and 9), the effect is rather small and

16 temporary since the depleting metabolite pools will likely be exhausted relatively quickly. Because

1 there is a significant penalty in the biomass yield whenever even temporary excursions from
2 steady-state are carried out in a random manner, a strong selection pressure is implied for cells to
3 maintain exponential growth under steady-state. This further bolsters the steady-state assumption
4 in FBA in predicting cellular phenotypes under growth pressure over a time span.

5



6

7 **Figure 9.** *iFBA*-calculated relative biomass yield when subjected to 0.1%, 1%, and 10%
8 uncertainty in the RHS vector (as a mass % of glucose uptake rate). Biomass yield distributions
9 are visualized with box plots and the numbers in red are the distribution means.

10

11 **4. Discussion**

12 In this work, we systematically quantified the effect of injecting random noise to biomass
13 coefficient values and to the RHS of the metabolite balance equalities on FBA-predicted individual
14 metabolic fluxes and biomass yield. Consistent with the robustness of biomass yield predictions
15 by FBA, we found that random biomass coefficient uncertainty propagates weakly to the predicted
16 biomass yield. This effect is overestimated by about 3-fold if the biomass coefficients are not
17 property rescaled so that they correctly map to the MW of the biomass metabolite being 1 g mmol^{-1} .

18 In contrast, the effect on predicted individual metabolic fluxes can be much more pronounced, and

1 in some cases amplified in comparison to the magnitude of the injected noise. Noise injected in
2 the energy requirement component of the biomass reaction GAM has a proportionate effect on
3 both FBA predicted biomass yield and metabolic fluxes, alluding to the importance of deriving an
4 accurate estimate for GAM in metabolic model reconstruction and avoiding whenever possible
5 direct porting from related organisms. Injection of noise on the RHS of balance constraints
6 revealed the difficulty of recovering feasible flux distributions as metabolite pool and elemental
7 balance constraints maintain strict requirements even under unsteady-state conditions. Metabolite
8 unsteady-state causes dramatic reductions in the predicted average biomass yield even for very
9 small temporary random departures from steady-state, pointing at a strong selection pressure to
10 maintain steady-state during exponential growth. Conceptually, this implies that it is best for all
11 biomass constituents to be available on a “just-in-time” basis to avoid growth arrests because a
12 single required (core) biomass constituent is missing. Any potential benefit of temporarily boosting
13 growth by creating a metabolite imbalance through depletion seems to cause a disproportionate
14 reduction in the long run. The results obtained for both types of uncertainty argue in favor of the
15 core assumptions of constant biomass composition and steady-state underpinning FBA.

16 This study revealed a number of best practices (and things to avoid) whenever assessing the impact
17 of parametric uncertainty on FBA results. Any modification in the biomass reaction should be
18 accompanied by checking for material balances in the lumped reactions (i.e., ATP hydrolysis for
19 GAM), processes (i.e., DNA, RNA, and protein synthesis), and re-normalizing biomass MW to 1
20 g mmol⁻¹. Similarly, updating biomass reaction formulations to account for new processes such as
21 tRNA-charging of amino acids (Xavier et al., 2017) must be carried out so as mass and charge
22 balances are conserved. An important observation from this study is that injected noise in biomass
23 coefficients does not propagate equally to the biomass yield output, but it is dependent on the

1 biosynthetic, energy, and cofactor costs for the corresponding biomass constituent. This
2 establishes a prioritization sequence for experimental determination of various components of the
3 biomass reaction. For example, amino acid composition and protein fraction should be prioritized
4 in order to obtain accurate NADPH yield prediction whereas lipid species composition and fraction
5 are important for accurate biomass yield prediction. Priority should also be given to
6 macromolecular composition as their uncertainty propagation affects FBA-predictions to a larger
7 extent than individual biomass coefficients. To this end, a pipeline was recently put forth
8 (Simensen et al., 2021) to robustly measure macromolecular composition in *E. coli*.

9 Special considerations are also required when exploring departures from metabolite steady-states,
10 as violating CM-pool equality constraints will cause the optimization problems to become
11 infeasible. This can be a problem when incorporating experimental measurements of concentration
12 changes to *iFBA* to quantify the RHS imbalance for metabolites that are parts of CM-pools.
13 Missing or imprecise measurements could render the experimentally derived RHS vector unusable,
14 and may require steady-state departures of unmeasured metabolite(s) to achieve a feasible solution
15 (Bordbar et al., 2017). To this end, *PULP* can be a useful tool for projecting experimental
16 measurements back to the feasible RHS solution space. Explicit inclusion of elemental balances
17 within *PULP* is necessary to avoid elemental imbalances in the predicted flux distribution as under
18 unsteady-state conditions the model stoichiometry cannot solely maintain elemental balancing
19 leading to systematic depletions.

20 The methods developed here enable the prediction of flux phenotypes subject to an input of
21 parametric uncertainty. They are also applicable to the study of systems that exhibit periodic
22 variations in their metabolism (Andreas Angermayr et al., 2016) or the study of non-growing cells
23 that only exhibit metabolite concentration fluctuations (Yurkovich et al., 2017). The assumption

1 of normality for the injected uncertainty can be easily removed by generating samples conforming
2 to any probability distribution. For example, one can assess what input probability distribution
3 recovers the observed bimodal distributions of phenotypes (Elsafadi et al., 2016; Schmitz et al.,
4 2017). Overall, the methods presented in this work could be used to further studies investigating
5 metabolic flux and yield variation under parametric uncertainty in FBA.

6

7 **CRedit authorship contribution statement**

8 **Hoang V. Dinh:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data
9 Curation, Writing – Original Draft. **Debolina Sarkar:** Methodology, Software, Formal analysis,
10 Writing – Review & Editing. **Costas D. Maranas:** Conceptualization, Methodology, Resources,
11 Supervision, Writing – Review & Editing, Funding acquisition.

12

13 **Acknowledgment**

14 We would like to thank Charles Foster (from The Pennsylvania State University) for a critical
15 review of the manuscript. Computations for this research were performed on the Pennsylvania
16 State University's Institute for Computational and Data Sciences' Roar supercomputer. This work
17 was partially funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation
18 (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research
19 under Award Number DE-SC0018420). Any opinions, findings, and conclusions or
20 recommendations expressed in this publication are those of the author(s) and do not necessarily
21 reflect the views of the U.S. Department of Energy. Funding also provided by the DOE Office of
22 Science, Office of Biological and Environmental Research (Award Number DE-SC0018260).

1 Funding also provided by The Center for Bioenergy Innovation a U.S. Department of Energy
2 Research Center supported by the Office of Biological and Environmental Research in the DOE
3 Office of Science.

4

5 **References**

- 6 Andreas Angermayr, S., van Alphen, P., Hasdemir, D., Kramer, G., Iqbal, M., van Grondelle, W.,
7 Hoefsloot, H.C., Choi, Y.H., Hellingwerf, K.J., 2016. Culturing *Synechocystis* sp. Strain PCC
8 6803 with N₂ and CO₂ in a diel regime reveals multiphase glycogen dynamics with low
9 maintenance costs. *Appl. Environ. Microbiol.* 82, 4180–4189.
10 <https://doi.org/10.1128/AEM.00256-16>
- 11 Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P.,
12 Ware, D., Perez, F., Canon, S., Sneddon, M.W., Henderson, M.L., Riehl, W.J., Murphy-
13 Olson, D., Chan, S.Y., Kamimura, R.T., Kumari, S., Drake, M.M., Brettin, T.S., Glass, E.M.,
14 Chivian, D., Gunter, D., Weston, D.J., Allen, B.H., Baumohl, J., Best, A.A., Bowen, B.,
15 Brenner, S.E., Bun, C.C., Chandonia, J.-M., Chia, J.-M., Colasanti, R., Conrad, N., Davis,
16 J.J., Davison, B.H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J.N.,
17 Fang, G., Faria, J.P., Frybarger, P.M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J.,
18 Haun, H.L., He, F., Jain, R., Joachimiak, M.P., Keegan, K.P., Kondo, S., Kumar, V., Land,
19 M.L., Meyer, F., Mills, M., Novichkov, P.S., Oh, T., Olsen, G.J., Olson, R., Parrello, B.,
20 Pasternak, S., Pearson, E., Poon, S.S., Price, G.A., Ramakrishnan, S., Ranjan, P., Ronald,
21 P.C., Schatz, M.C., Seaver, S.M.D., Shukla, M., Sutormin, R.A., Syed, M.H., Thomason, J.,
22 Tintle, N.L., Wang, D., Xia, F., Yoo, H., Yoo, S., Yu, D., 2018. KBase: The United States
23 Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* 36, 566–569.
24 <https://doi.org/10.1038/nbt.4163>
- 25 Aung, H.W., Henry, S.A., Walker, L.P., 2013. Revising the representation of fatty acid,
26 glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast
27 metabolism. *Ind. Biotechnol.* 9, 215–228. <https://doi.org/10.1089/ind.2013.0013>
- 28 Beck, A., Hunt, K., Carlson, R., 2018. Measuring Cellular Biomass Composition for
29 Computational Biology Applications. *Processes* 6, 38. <https://doi.org/10.3390/pr6050038>
- 30 Bergler, H., Fuchsbichler, S., Högenauer, G., Turnowsky, F., 1996. The enoyl-[acyl-carrier-
31 protein] reductase (FabI) of *Escherichia coli*, which catalyzes a key regulatory step in fatty
32 acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-
33 CoA. *Eur. J. Biochem.* 242, 689–694. <https://doi.org/10.1111/j.1432-1033.1996.0689r.x>
- 34 Bordbar, A., Yurkovich, J.T., Paglia, G., Rolfsson, O., Sigurjónsson, Ó.E., Palsson, B.O., 2017.
35 Elucidating dynamic metabolic physiology through network integration of quantitative time-
36 course metabolomics. *Sci. Rep.* 7, 1–12. <https://doi.org/10.1038/srep46249>
- 37 Chan, S.H.J., Cai, J., Wang, L., Simons-Senftle, M.N., Maranas, C.D., 2017. Standardizing
38 biomass reactions and ensuring complete mass balance in genome-scale metabolic models.
39 *Bioinformatics* 33, 3603–3609. <https://doi.org/10.1093/bioinformatics/btx453>
- 40 Cornish-Bowden, A., Hofmeyr, J.H.S., 2002. The role of stoichiometric analysis in studies of

- 1 metabolism: An example. *J. Theor. Biol.* 216, 179–191.
2 <https://doi.org/10.1006/jtbi.2002.2547>
- 3 De Martino, A., De Martino, D., Mulet, R., Pagnani, A., 2014. Identifying all moiety conservation
4 laws in genome-scale metabolic networks. *PLoS One* 9, 100750.
5 <https://doi.org/10.1371/journal.pone.0100750>
- 6 Diamond, S., Jun, D., Rubin, B.E., Golden, S.S., 2015. The circadian oscillator in *Synechococcus*
7 *elongatus* controls metabolite partitioning during diurnal growth. *Proc. Natl. Acad. Sci. U. S.*
8 *A.* 112, E1916–E1925. <https://doi.org/10.1073/pnas.1504576112>
- 9 Dinh, H. V., Suthers, P.F., Chan, S.H.J., Shen, Y., Xiao, T., Deewan, A., Jagtap, S.S., Zhao, H.,
10 Rao, C. V., Rabinowitz, J.D., Maranas, C.D., 2019. A comprehensive genome-scale model
11 for *Rhodospiridium toruloides* IFO0880 accounting for functional genomics and phenotypic
12 data. *Metab. Eng. Commun.* 9, e00101. <https://doi.org/10.1016/j.mec.2019.e00101>
- 13 Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. COBRAPy: CONstraints-Based
14 Reconstruction and Analysis for Python. *BMC Syst. Biol.* 7, 74.
15 <https://doi.org/10.1186/1752-0509-7-74>
- 16 Elsafadi, M., Manikandan, M., Atteya, M., Hashmi, J.A., Iqbal, Z., Aldahmash, A., Alfayez, M.,
17 Kassem, M., Mahmood, A., 2016. Characterization of Cellular and Molecular Heterogeneity
18 of Bone Marrow Stromal Cells. *Stem Cells Int.* 2016. <https://doi.org/10.1155/2016/9378081>
- 19 Famili, I., Palsson, B.O., 2003. The convex basis of the left null space of the stoichiometric matrix
20 leads to the definition of metabolically meaningful pools. *Biophys. J.* 85, 16–26.
21 [https://doi.org/10.1016/S0006-3495\(03\)74450-6](https://doi.org/10.1016/S0006-3495(03)74450-6)
- 22 Feist, A.M., Palsson, B.O., 2010. The biomass objective function. *Curr. Opin. Microbiol.* 13, 344–
23 349. <https://doi.org/10.1016/J.MIB.2010.03.003>
- 24 Flahaut, N.A.L., Wiersma, A., Van De Bunt, B., Martens, D.E., Schaap, P.J., Sijtsma, L., Dos
25 Santos, V.A.M., De Vos, W.M., 2013. Genome-scale metabolic model for *Lactococcus lactis*
26 MG1363 and its application to the analysis of flavor formation. *Appl. Microbiol. Biotechnol.*
27 97, 8729–8739. <https://doi.org/10.1007/s00253-013-5140-2>
- 28 González-Cabaleiro, R., Mitchell, A.M., Smith, W., Wipat, A., Ofiteru, I.D., 2017. Heterogeneity
29 in pure microbial systems: Experimental measurements and modeling. *Front. Microbiol.*
30 <https://doi.org/10.3389/fmicb.2017.01813>
- 31 Gu, C., Kim, G.B., Kim, W.J., Kim, H.U., Lee, S.Y., 2019. Current status and applications of
32 genome-scale metabolic models. *Genome Biol.* <https://doi.org/10.1186/s13059-019-1730-3>
- 33 Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *Escherichia coli* K-12 undergoes adaptive
34 evolution to achieve in silico predicted optimal growth. *Nature* 420, 186–189.
35 <https://doi.org/10.1038/nature01149>
- 36 Kavvas, E.S., Seif, Y., Yurkovich, J.T., Norsigian, C., Poudel, S., Greenwald, W.W., Ghatak, S.,
37 Palsson, B.O., Monk, J.M., 2018. Updated and standardized genome-scale reconstruction of
38 *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of
39 physiological conditions. *BMC Syst. Biol.* 12. <https://doi.org/10.1186/s12918-018-0557-y>
- 40 Kultschar, B., Dudley, E., Wilson, S., Llewellyn, C.A., 2019. Intracellular and Extracellular
41 Metabolites from the Cyanobacterium *Chlorogloeopsis fritschii*, PCC 6912, During 48 Hours
42 of UV-B Exposure. *Metabolites* 9, 74. <https://doi.org/10.3390/metabo9040074>
- 43 Kumar, R., Ghosh, M., Kumar, S., Prasad, M., 2020. Single Cell Metabolomics: A Future Tool to
44 Unmask Cellular Heterogeneity and Virus-Host Interaction in Context of Emerging Viral
45 Diseases. *Front. Microbiol.* 0, 1152. <https://doi.org/10.3389/FMICB.2020.01152>
- 46 Lange, H.C., Heijnen, J.J., 2001. Statistical reconciliation of the elemental and molecular biomass

- 1 composition of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 75, 334–44.
- 2 Levine, E., Hwa, T., 2007. Stochastic fluctuations in metabolic pathways. *Proc. Natl. Acad. Sci.*
3 104, 9224–9229. <https://doi.org/10.1073/PNAS.0610987104>
- 4 Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins,
5 J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith,
6 R.D., Palsson, B.Ø., 2010. Omic data from evolved *E. coli* are consistent with computed
7 optimal growth from genome-scale models. *Mol. Syst. Biol.* 6, 390.
8 <https://doi.org/10.1038/msb.2010.47>
- 9 Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype
10 relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10, 291–305.
11 <https://doi.org/10.1038/nrmicro2737>
- 12 Liao, Y.C., Huang, T.W., Chen, F.C., Charusanti, P., Hong, J.S.J., Chang, H.Y., Tsai, S.F.,
13 Palsson, B.O., Hsiung, C.A., 2011. An experimentally validated genome-scale metabolic
14 reconstruction of *Klebsiella pneumoniae* MGH 78578, iYL1228. *J. Bacteriol.* 193, 1710–
15 1717. <https://doi.org/10.1128/JB.01218-10>
- 16 Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A.,
17 Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B.E., Edirisinghe, J.N.,
18 Faria, J.P., Feist, A.M., Fengos, G., Fleming, R.M.T., García-Jiménez, B., Hatzimanikatis,
19 V., van Helvoirt, W., Henry, C.S., Hermjakob, H., Herrgård, M.J., Kaafarani, A., Kim, H.U.,
20 King, Z., Klamt, S., Klipp, E., Koehorst, J.J., König, M., Lakshmanan, M., Lee, D.Y., Lee,
21 S.Y., Lee, S., Lewis, N.E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P.,
22 Mardinoglu, A., Medlock, G.L., Monk, J.M., Nielsen, J., Nielsen, L.K., Nogales, J.,
23 Nookaew, I., Palsson, B.O., Papin, J.A., Patil, K.R., Poolman, M., Price, N.D., Resendis-
24 Antonio, O., Richelle, A., Rocha, I., Sánchez, B.J., Schaap, P.J., Malik Sheriff, R.S., Shoaie,
25 S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J.O., Wodke, J.A.H., Xavier, J.C., Yuan,
26 Q., Zakhartsev, M., Zhang, C., 2020. MEMOTE for standardized genome-scale metabolic
27 model testing. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0446-y>
- 28 Long, M.R., Reed, J.L., Wren, J., 2017. Improving flux predictions by integrating data from
29 multiple strains. *Bioinformatics* 33, 893–900. <https://doi.org/10.1093/bioinformatics/btw706>
- 30 MacGillivray, M., Ko, A., Gruber, E., Sawyer, M., Almaas, E., Holder, A., 2017. Robust Analysis
31 of Fluxes in Genome-Scale Metabolic Pathways. *Sci. Rep.* 7, 268.
32 <https://doi.org/10.1038/s41598-017-00170-3>
- 33 Machado, D., Andrejev, S., Tramontano, M., Patil, K.R., 2018. Fast automated reconstruction of
34 genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.*
35 46, 7542–7553. <https://doi.org/10.1093/nar/gky537>
- 36 Mahadevan, R., Schilling, C.H.H., 2003. The effects of alternate optimal solutions in constraint-
37 based genome-scale metabolic models. *Metab. Eng.* 5, 264–276.
38 <https://doi.org/10.1016/J.YMBEN.2003.09.002>
- 39 Milligan, A.J., Berman-Frank, I., Gerchman, Y., Dismukes, G.C., Falkowski, P.G., 2007. Light-
40 dependent oxygen consumption in nitrogen-fixing cyanobacteria plays a key role in
41 nitrogenase protection. *J. Phycol.* 43, 845–852. <https://doi.org/10.1111/J.1529-8817.2007.00395.X>
- 42
- 43 Mishra, P., Lee, N.R., Lakshmanan, M., Kim, M., Kim, B.G., Lee, D.Y., 2018. Genome-scale
44 model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. *BMC*
45 *Syst. Biol.* 12, 12. <https://doi.org/10.1186/s12918-018-0542-5>
- 46 Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W.,

- 1 Zhang, Z., Mori, H., Feist, A.M., Palsson, B.O., 2017. iML1515, a knowledgebase that
2 computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904–908.
3 <https://doi.org/10.1038/nbt.3956>
- 4 Nagy, P.L., Marolewski, A., Benkovic, S.J., Zalkin, H., 1995. Formyltetrahydrofolate hydrolase,
5 a regulatory enzyme that functions to balance pools of tetrahydrofolate and one-carbon
6 tetrahydrofolate adducts in *Escherichia coli*. *J. Bacteriol.* 177, 1292–1298.
7 <https://doi.org/10.1128/jb.177.5.1292-1298.1995>
- 8 Nakayama, T., Yonekura, S.-I., Yonei, S., Zhang-Akiyama, Q.-M., 2013. *Escherichia coli*
9 pyruvate:flavodoxin oxidoreductase, YdbK - regulation of expression and biological roles in
10 protection against oxidative stress. *Genes Genet. Syst.* 88, 175–88.
11 <https://doi.org/10.1266/ggs.88.175>
- 12 Nikolaev, E. V., Burgard, A.P., Maranas, C.D., 2005. Elucidation and Structural Analysis of
13 Conserved Pools for Genome-Scale Metabolic Reconstructions. *Biophys. J.* 88, 37–49.
14 <https://doi.org/10.1529/BIOPHYSJ.104.043489>
- 15 O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B., 2013. Genome-scale models
16 of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst.*
17 *Biol.* 9, 693. <https://doi.org/10.1038/msb.2013.52>
- 18 Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Palsson, B., 2011. A
19 comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol.*
20 *Syst. Biol.* 7, 535. <https://doi.org/10.1038/msb.2011.65>
- 21 Orth, J.D., Thiele, I., Palsson, B.Ø., 2010. What is flux balance analysis? *Nat. Biotechnol.* 28, 245–
22 248. <https://doi.org/10.1038/nbt.1614>
- 23 Qiao, Y., Liu, G., Lv, X., Fan, X., Zhang, Y., Meng, L., Ai, M., Feng, Z., 2020. Metabolic Pathway
24 Profiling in Intracellular and Extracellular Environments of *Streptococcus thermophilus*
25 During pH-Controlled Batch Fermentations. *Front. Microbiol.* 10.
26 <https://doi.org/10.3389/fmicb.2019.03144>
- 27 Rabouille, S., Van de Waal, D.B., Matthijs, H.C.P., Huisman, J., 2014. Nitrogen fixation and
28 respiratory electron transport in the cyanobacterium *Cyanothece* under different light/dark
29 cycles. *FEMS Microbiol. Ecol.* 87, 630–638. <https://doi.org/10.1111/1574-6941.12251>
- 30 Reder, C., 1988. Metabolic control theory: A structural approach. *J. Theor. Biol.* 135, 175–201.
31 [https://doi.org/10.1016/S0022-5193\(88\)80073-0](https://doi.org/10.1016/S0022-5193(88)80073-0)
- 32 Reznik, E., Mehta, P., Segrè, D., 2013. Flux Imbalance Analysis and the Sensitivity of Cellular
33 Growth to Changes in Metabolite Pools. *PLoS Comput. Biol.* 9, e1003195.
34 <https://doi.org/10.1371/journal.pcbi.1003195>
- 35 Rogne, P., Rosselin, M., Grundström, C., Hedberg, C., Sauer, U.H., Wolf-Watz, M., 2018.
36 Molecular mechanism of ATP versus GTP selectivity of adenylate kinase. *Proc. Natl. Acad.*
37 *Sci. U. S. A.* 115, 3012–3017. <https://doi.org/10.1073/pnas.1721508115>
- 38 Schmitz, A.C., Hartline, C.J., Zhang, F., 2017. Engineering Microbial Metabolite Dynamics and
39 Heterogeneity. *Biotechnol. J.* 12, 1700422. <https://doi.org/10.1002/biot.201700422>
- 40 Seif, Y., Monk, J.M., Mih, N., Tsunemoto, H., Poudel, S., Zuniga, C., Broddrick, J., Zengler, K.,
41 Palsson, B.O., 2019. A computational knowledge-base elucidates the response of
42 *Staphylococcus aureus* to different media types. *PLoS Comput. Biol.* 15.
43 <https://doi.org/10.1371/journal.pcbi.1006644>
- 44 Shinde, S., Zhang, X., Singapuri, S.P., Kalra, I., Liu, X., Morgan-Kiss, R.M., Wang, X., 2020.
45 Glycogen Metabolism Supports Photosynthesis Start through the Oxidative Pentose
46 Phosphate Pathway in Cyanobacteria. *Plant Physiol.* 182, 507–517.

- 1 <https://doi.org/10.1104/PP.19.01184>
- 2 Simensen, V., Schulz, C., Karlsen, E., Bråtelund, S., Burgos, I., Thorfinnsdottir, L.B., García-
3 Calvo, L., Bruheim, P., Almaas, E., 2021. Quantification of macromolecular biomass
4 composition for constraint-based metabolic modeling. *bioRxiv* 2021.08.20.457062.
5 <https://doi.org/10.1101/2021.08.20.457062>
- 6 Takhaveev, V., Heinemann, M., 2018. Metabolic heterogeneity in clonal microbial populations.
7 *Curr. Opin. Microbiol.* <https://doi.org/10.1016/j.mib.2018.02.004>
- 8 Thiele, I., Hyduke, D.R., Steeb, B., Fankam, G., Allen, D.K., Bazzani, S., Charusanti, P., Chen,
9 F.C., Fleming, R.M.T., Hsiung, C.A., De Keersmaecker, S.C.J., Liao, Y.C., Marchal, K., Mo,
10 M.L., Özdemir, E., Raghunathan, A., Reed, J.L., Shin, S. II, Sigurbjörnsdóttir, S., Steinmann,
11 J., Sudarsan, S., Swainston, N., Thijs, I.M., Zengler, K., Palsson, B.O., Adkins, J.N., Bumann,
12 D., 2011. A community effort towards a knowledge-base and mathematical model of the
13 human pathogen *Salmonella Typhimurium* LT2. *BMC Syst. Biol.* 5, 8.
14 <https://doi.org/10.1186/1752-0509-5-8>
- 15 Thiele, I., Palsson, B., 2010. A protocol for generating a high quality genome-scale metabolic
16 reconstruction. *Nat. Protoc.* 5, 93–121.
- 17 Tomàs-Gamisans, M., Ferrer, P., Albiol, J., 2018. Fine-tuning the *P. pastoris* iMT1026 genome-
18 scale metabolic model for improved prediction of growth on methanol or glycerol as sole
19 carbon sources. *Microb. Biotechnol.* 11, 224–237. <https://doi.org/10.1111/1751-7915.12871>
- 20 Tomàs-Gamisans, M., Ferrer, P., Albiol, J., Tomas-Gamisans, M., Ferrer, P., Albiol, J., 2016.
21 Integration and Validation of the Genome-Scale Metabolic Models of *Pichia pastoris*: A
22 Comprehensive Update of Protein Glycosylation Pathways, Lipid and Energy Metabolism.
23 *PLoS One* 11, e0148031. <https://doi.org/10.1371/journal.pone.0148031>
- 24 Torres, P., Saa, P.A., Albiol, J., Ferrer, P., Agosin, E., 2019. Contextualized genome-scale model
25 unveils high-order metabolic effects of the specific growth rate and oxygenation level in
26 recombinant *Pichia pastoris*. *Metab. Eng. Commun.* e00103.
27 <https://doi.org/10.1016/j.mec.2019.e00103>
- 28 Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski,
29 E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman,
30 K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng,
31 Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I.,
32 Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt,
33 P., Vijaykumar, A., Bardelli, A. Pietro, Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A.,
34 Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C.,
35 Nicholson, D.A., Hagen, D.R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva,
36 F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R.,
37 Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J.,
38 Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington,
39 J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M.,
40 Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov,
41 N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier,
42 S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J.,
43 Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay,
44 U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for
45 scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- 46

- 1 Wang, H., Marcišauskas, S., Sánchez, B.J., Domenzain, I., Hermansson, D., Agren, R., Nielsen,
2 J., Kerkhoven, E.J., 2018. RAVEN 2.0: A versatile toolbox for metabolic network
3 reconstruction and a case study on *Streptomyces coelicolor*. PLOS Comput. Biol. 14,
4 e1006541. <https://doi.org/10.1371/journal.pcbi.1006541>
- 5 Xavier, J.C., Patil, K.R., Rocha, I., 2017. Integration of Biomass Formulations of Genome-Scale
6 Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in
7 Prokaryotes. Metab. Eng. 39, 200–208. <https://doi.org/10.1016/J.YMBEN.2016.12.002>
- 8 Xu, J.Z., Yang, H.K., Zhang, W.G., 2018. NADPH metabolism: a survey of its theoretical
9 characteristics and manipulation strategies in amino acid biosynthesis. Crit. Rev. Biotechnol.
10 <https://doi.org/10.1080/07388551.2018.1437387>
- 11 Yurkovich, J.T., Zielinski, D.C., Yang, L., Paglia, G., Rolfsson, O., Sigurjónsson, Ó.E., Broddrick,
12 J.T., Bordbar, A., Wichuk, X.K., Brynjólfsson, S., Palsson, S., Gudmundsson, S., Palsson,
13 B.O., 2017. Quantitative time-course metabolomics in human red blood cells reveal the
14 temperature dependence of human metabolic networks. J. Biol. Chem. 292, 19556–19564.
15 <https://doi.org/10.1074/jbc.M117.804914>
16