Heinonen, Markus; Osmala, Maria; Mannerström, Henrik; Wallenius, Janne; Kaski, Samuel; Rousu, Juho; Lähdesmäki, Harri

Bayesian metabolic flux analysis reveals intracellular flux couplings

Published in:
Bioinformatics

DOI:
10.1093/bioinformatics/btz315

Published: 15/07/2019

Document Version
Publisher's PDF, also known as Version of record

Bioinformatics, 35, 2019, i548–i557
doi: 10.1093/bioinformatics/btz315
ISMB/ECCB 2019

Bayesian metabolic flux analysis reveals intracellular flux couplings

Markus Heinonen[1,2,*,†], Maria Osmala[1,†], Henrik Mannerström[1], Janne Wallenius[3], Samuel Kaski[1,2], Juho Rousu[1,2] and Harri Lähdesmäki[1]

[1]Department of Computer Science, Aalto University, Espoo 02150, Finland, [2]Helsinki Institute for Information Technology, Espoo 02150, Finland and [3]Institute for Molecular Medicine Finland, Helsinki 00290, Finland

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Motivation: Metabolic flux balance analysis (FBA) is a standard tool in analyzing metabolic reaction rates compatible with measurements, steady-state and the metabolic reaction network stoichiometry. Flux analysis methods commonly place model assumptions on fluxes due to the convenience of formulating the problem as a linear programing model, while many methods do not consider the inherent uncertainty in flux estimates.

Results: We introduce a novel paradigm of Bayesian metabolic flux analysis that models the reactions of the whole genome-scale cellular system in probabilistic terms, and can infer the full flux vector distribution of genome-scale metabolic systems based on exchange and intracellular (e.g. 13C) flux measurements, steady-state assumptions, and objective function assumptions. The Bayesian model couples all fluxes jointly together in a simple truncated multivariate posterior distribution, which reveals informative flux couplings. Our model is a plug-in replacement to conventional metabolic balance methods, such as FBA. Our experiments indicate that we can characterize the genome-scale flux covariances, reveal flux couplings, and determine more intracellular unobserved fluxes in *Clostridium acetobutylicum* from 13C data than flux variability analysis.

Availability and implementation: The COBRA compatible software is available at github.com/markusheinonen/bamfa.

Contact: markus.o.heinonen@aalto.fi

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Metabolic modeling considers networks of up to thousands of chemical reactions that transform metabolite molecules within cellular organisms (Palsson, 2015). The key problem of metabolism is estimation of the reaction rates, or fluxes, of the system of the highly interdependent intracellular fluxes from measurements of few exchange fluxes that transfer nutrients or products between the external medium and the cell.

The dominant approach to flux estimation is the celebrated flux balance analysis (FBA) framework that finds reaction rates that maximize pre-specified cellular growth or other target objective function (Feist and Palsson, 2010), while assuming the cell is in a steady-state, where concentrations of intracellular metabolites do not change (Almaas *et al.*, 2004). Such FBA problem can be casted as a convenient and computationally efficient linear programing problem of solving a system of linear steady-state constraints while maximizing a linear target objective (Orth *et al.*, 2010), and where flux measurements can be encoded as constraints to the fluxes (Carreira *et al.*, 2014). FBA is commonly used to characterize intracellular fluxes in various simulated objective conditions (Mo *et al.*, 2010). In metabolic flux analysis (MFA) values of unknown fluxes are directly estimated based on measurements of some determined fluxes without explicit maximal growth or other objective assumption (Kim *et al.*, 2008). In both approaches a point estimate for up to thousands of highly interdependent fluxes are determined (Bordbar *et al.*, 2014).

The standard metabolic analyses contain three major model assumptions that warrant careful methodological protocol to achieve biologically meaningful results. First, the exact steady-state constraint can be an unrealistic assumption since metabolites can accumulate or deplete to adapt to dynamic situations, such as during

responses to changing nutrient conditions (MacGillivray *et al.*, 2017; Pakula *et al.*, 2016). Second, in FBA maximal growth (or other target objective) is often assumed as a constraint, while it only holds at the highest growth phase in practise. Third, due to a large number of metabolic reactions and limited number of experimental data, flux point estimates commonly used in the field ignore the notable uncertainty involved in FBA and MFA solutions. The flux variances are key in characterizing metabolic systems and uncertainties emerging from the use of insufficient and noisy data.

Numerous separate extensions to flux analysis have been introduced to alleviate these three assumptions. The robust FBA framework considers the effect of measurement uncertainties to the maximal growth (Zavlanos and Julius, 2011). The steady-state assumption was recently relaxed by the robust analysis of metabolic pathways (RAMP) model (MacGillivray *et al.*, 2017). In contrast to point flux estimates of FBA and MFA, the flux variability analysis (FVA) characterizes the sensitivity of the objective function to independent flux perturbations, resulting in upper and lower bounds around the FBA solution (Gudmundsson and Thiele, 2010; Mahadevan and Schilling, 2003). In principal flux mode analysis, the eigenvectors of steady-state flux cone characterize the flux variability (Bhadra *et al.*, 2018). Alternatively the solution space of the fluxes can be sampled (Schellenberger *et al.*, 2010) by considering only optimal fluxes from border of the flux hypercone (Bordel *et al.*, 2010) or by sampling also inoptimal fluxes from the inside the hypercone (Mo *et al.*, 2010; Saa and Nielsen, 2016a). The sampling methods use the hit-and-run formalism (Smith, 1984) as either the artificially centered hit-and-run (ACHR) (Kaufman and Smith, 1998; Megchelenbrink *et al.*, 2014) or the coordinate hit-and-run with rounding (CCHR) (Haraldsdóttir *et al.*, 2017) sampling algorithms to cope with the large flux space. A related approach uses possibility calculus (Dubois *et al.*, 1996) to iteratively refine the estimate of possible and impossible flux states (Llaneras *et al.*, 2009).

Bayesian methods have been scarcely applied in flux analysis. Small-scale Bayesian construction of kinetics was proposed by Saa and Nielsen (2016b). The Metabolica method proposed modeling distributions of fluxes of skeletal muscle metabolism (Heino *et al.*, 2007, 2010), but did not include modeling of target objectives or genome-scale metabolic models. Bayesian methods have also been developed for 13C labeling data (Kadirkamanathan *et al.*, 2006; Theorell *et al.*, 2017) by assuming fixed steady-state and without incorporating any target objectives.

In this article, we treat all three model assumptions simultaneously by introducing a novel paradigm of Bayesian MFA where the genome-scale, interdependent flux vector distributions are estimated. Our model allows probabilistic relaxation of steady-state and target objective constraints. In contrast to earlier uniform sampling approaches, we place priors on flux distributions, and estimate posterior distributions that characterize and quantify the probability of all flux states that are compatible with flux measurements, steady-state assumption and stoichiometry. Our model reveals flux dependencies in explicit form and characterizes the full space of flux states in principled fashion. The Bayesian flux analysis can be used as a drop-in replacement to standard FBA, MFA, FVA and sampling methods. We provide public implementation of the Bayesian flux analysis using the standard COBRA framework (Becker *et al.*, 2007; Schellenberger *et al.*, 2011).

2 Materials and methods

The goal of this article is a probabilistic formulation of static steady-state metabolic systems that can be applied to whole genome MFA, FBA and FVA. We propose the Bayesian method as a direct replacement to these classic FBA, MFA and FVA tools. We start by assuming a metabolic system of M metabolites and N reactions has been characterized by a constant stoichiometric matrix $S \in \mathbb{Z}^{M\times N}$, where the rows denote metabolite participations in all reactions, while the columns denote reactants and products of metabolites by individual reactions. The flux vector $\mathbf{v} = (v_1, \ldots, v_N)^T \in \mathbb{R}^N$ denotes the reaction rates of the system. The steady-state equation can be stated as

$$S\mathbf{v} = \dot{\mathbf{x}} = 0,$$

which encodes that metabolite concentration changes $\dot{\mathbf{x}} \in \mathbb{R}^M$ are zero and hence the metabolite concentrations $\mathbf{x} \in \mathbb{R}^M$ do not change. Throughout the article, we assume a subset of fluxes have been observed or determined (for instance, some of the exchange fluxes), while the remaining fluxes are unknown. Our goal is to infer the distribution of all unknown fluxes given the observed fluxes, the steady-state constraints and the flux lower and upper bounds.

2.1 Bayesian metabolic model

We formulate a Bayesian flux model (see Supplementary Material for a graphical model), which starts by assuming multivariate Gaussian priors for fluxes as

$$\mathbf{v}|\mathbf{m}_v, \boldsymbol{\sigma}_v \sim \mathcal{N}(\mathbf{m}_v, \Sigma_v),$$

with means $\mathbf{m}_v \in \mathbb{R}^N$ and diagonal covariances $\Sigma_v = diag(\boldsymbol{\sigma}_v^2) = diag(\sigma_{v_1}^2, \ldots, \sigma_{v_N}^2)$. The prior means are set to zero, or to the closest value to zero considering the flux upper and lower bounds. The variances σ_{v_i} are hyperparameters that characterize the a priori values the flux can take. The prior distribution converges towards an uninformative uniform prior as the prior variances increase.

We assume a Gaussian prior also for the metabolite changes

$$\dot{\mathbf{x}}\,|\mathbf{m}_{\dot{\mathbf{x}}}, \boldsymbol{\sigma}_{\dot{\mathbf{x}}} \sim \mathcal{N}(\mathbf{m}_{\dot{\mathbf{x}}}, \Sigma_{\dot{\mathbf{x}}}),$$

where $\mathbf{m}_{\dot{\mathbf{x}}} \in \mathbb{R}^M$ are the a priori mean accumulations or depletions of metabolite species. The diagonal covariances $\Sigma_{\dot{\mathbf{x}}} = diag(\boldsymbol{\sigma}_{\dot{\mathbf{x}}}^2) = diag(\sigma_{\dot{\mathbf{x}}_1}^2, \ldots, \sigma_{\dot{\mathbf{x}}_M}^2)$ encode the variances around prior metabolite changes. In strict steady-state, the prior for metabolite change becomes Dirac's delta function at zero. By increasing the variances $\boldsymbol{\sigma}_{\dot{\mathbf{x}}}^2$ we can relax the steady-state assumption on individual metabolites, and encode allowance for accumulations or depletions of them.

The joint distribution of fluxes $\mathbf{v}$ and metabolite changes $\dot{\mathbf{x}}$ can now be stated as a joint multivariate Gaussian distribution

$$\begin{bmatrix}\mathbf{v}\\ \dot{\mathbf{x}}\end{bmatrix} = \begin{bmatrix}\mathbf{v}\\ S\mathbf{v}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{m}_v\\ S\mathbf{m}_v\end{bmatrix}, \begin{bmatrix}\Sigma_v & \Sigma_v S^T\\ S\Sigma_v & S\Sigma_v S^T\end{bmatrix}\right),$$

that encodes the exact (We add small numerical tolerance κI within the inverse to ensure invertibility of the matrix.) relation $S\mathbf{v} = \dot{\mathbf{x}}$. The conditional distribution of fluxes given a specific realization of metabolite changes $\dot{\mathbf{x}}$ (e.g. 0) is then from standard Gaussian identities

$$\begin{aligned}\mathbf{v}|\,\dot{\mathbf{x}} \;&\sim \mathcal{N}(\mathbf{m}_v + \Sigma_v S^T (S\Sigma_v S^T)^{-1}(\dot{\mathbf{x}} - S\mathbf{m}_v), \Sigma_v - \Sigma_v S^T (S\Sigma_v S^T)^{-1} S\Sigma_v)\\ &\sim \mathcal{N}(\mathbf{m}_v + A(\dot{\mathbf{x}} - S\mathbf{m}_v), \Sigma_v - AS\Sigma_v)\end{aligned}$$

where $A = \Sigma_v S^T (S\Sigma_v S^T)^{-1}$. Since we do not in general have access to exact metabolite change values $\dot{\mathbf{x}}$, we marginalize the conditional flux distribution over the change prior distribution $p(\dot{\mathbf{x}})$ resulting in

$$p(\mathbf{v}|\boldsymbol{\sigma}_v, \boldsymbol{\sigma}_{\dot{\mathbf{x}}}) = \int p(\mathbf{v}|\dot{\mathbf{x}}) p(\dot{\mathbf{x}}) d\dot{\mathbf{x}} = \mathcal{N}(\mathbf{v}|\boldsymbol{\mu}, \mathrm{C}), \quad (1)$$

where $\boldsymbol{\mu} = \mathbf{m}_v + \mathrm{A}(\mathbf{m}_{\dot{\mathbf{x}}} - S\mathbf{m}_v)$ and $C = \Sigma_v - AS\Sigma_v + A\Sigma_{\dot{x}}A^T$.

2.2 Conditioning the model with observations

Assume we have access to noisy observations $\mathbf{y}_o = \mathbf{v}_o + \varepsilon$ from a subset of observed fluxes $\mathbf{v}_o \subset \mathbf{v}$. The observations can be empirical measurements, 13C flux estimations, or flux hypotheses determined by the user. We assume independent additive Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, \omega_i^2)$ with variances collected in a matrix $\boldsymbol{\Omega}_o = diag(\omega_1^2, \ldots, \omega_{N_{obs}}^2)$, and hence the likelihood of observed fluxes is

$$p(\mathbf{y}_o|\mathbf{v}_o, \boldsymbol{\omega}_o) = \mathcal{N}(\mathbf{y}_o|\mathbf{v}_o, \boldsymbol{\Omega}_o).$$

The joint distribution of all fluxes $\mathbf{v}$ and noisy flux observations $\mathbf{y}_o$ is now

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{v}_o + \varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{v} \\ \mathbf{y}_o \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}_o \end{bmatrix}, \begin{bmatrix} \mathrm{C}_{NN} & \mathrm{C}_{No} \\ \mathrm{C}_{oN} & \mathrm{C}_{oo} + \boldsymbol{\Omega}_o \end{bmatrix} \right),$$

which gives a conditional distribution of all fluxes given the observations as

$$\mathbf{v}|\mathbf{y}_o \sim \mathcal{N}(\boldsymbol{\mu} + \mathrm{C}_{No}\mathrm{C}_{\mathrm{y}}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o), \mathrm{C} - \mathrm{C}_{No}\mathrm{C}_{\mathrm{y}}^{-1}\mathrm{C}_{oN}). \quad (2)$$

where $C_y = C_{oo} + \Omega_o$ is the noisy covariance, CNN is the full $(N \times N)$ covariance matrix, $C_{No} = C_{oN}^T$ is the $(N_{obs} \times N)$ covariance matrix between observed fluxes and all fluxes, and Coo is the $(N_{obs} \times N_{obs})$ covariance matrix between observed fluxes. Note that observed fluxes are in both $\mathbf{v}$ and in $\mathbf{v}_o$. Note also that the model works with no observations at all as the conditional distribution in Equation (2) reduces back to the prior in Equation (1).

Finally, we add the flux upper and lower bounds by truncating the distribution with the known flux lower lb and upper ub bounds resulting in the final truncated normal flux posterior

$$\mathbf{v}|\mathbf{y}_o \sim \mathcal{TN}(\boldsymbol{\mu} + \mathrm{C}_{No}\mathrm{C}_{\mathrm{y}}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o), \mathrm{C} - \mathrm{C}_{No}\mathrm{C}_{\mathrm{y}}^{-1}\mathrm{C}_{oN}, lb, ub).$$

The posterior encodes the distribution of bounded fluxes that are compatible with the flux observations, flux priors, and where steady-state applies according to the tolerances determined by the steady-state prior means $\mathbf{m}_{\dot{x}}$ and variances $\boldsymbol{\sigma}_{\dot{\mathbf{x}}}^2$.

The derived flux posterior is an unimodal truncated multivariate normal (TMVN) distribution where flux dependencies are represented through the covariance matrix C, which encodes all flux relationships with high rank. The flux posterior as a whole characterizes the distribution of all valid flux vectors. The main characterizations of interest are the individual flux distributions (the marginals) and flux combination distributions (multi-variate marginals). Marginals of TMVN's are not analytically tractable, nor are they TMVN distributions (Horrace, 2005). We resort to Markov Chain Monte Carlo (MCMC) sampling from the TMVN flux distribution to reveal individual flux, flux pair or flux group distributions.

2.3 Gibbs sampling truncated MVN's

A recent review summarizes sampling approaches for TMVNca (Altmann *et al.*, 2014). The conditionals of TMVNhe are still TMVNs (Horrace, 2005), which has led to many Gibbs-based samplers (Emery *et al.*, 2014; Geweke, 1991; Horrace, 2005; Kotecha and Djuric, 1999; Li and Ghosh, 2015). In addition, Hamiltonian Monte Carlo samplers (Pakman and Paninski, 2014) have been proposed, while elliptical slice samplers would also fit well to the problem (Murray *et al.*, 2010). We experimented with the three main approaches, and found out that Gibbs sampling has consistently the best performance in genome-scale metabolic models up to 4000 fluxes (data not shown). In the remainder of the article we apply Gibbs sampling.

To sample the distribution $\mathbf{v} \sim \mathcal{TN}(\boldsymbol{\mu}, \mathrm{C}, lb, ub)$ we begin by transforming it into whitened domain by Cholesky decomposition $C = LL^T$ with transformed fluxes $\tilde{\mathbf{v}} = L^{-1}(\mathbf{v} - \boldsymbol{\mu})$ with white distribution $\tilde{\mathbf{v}} \sim \mathcal{TN}(0, I, \mathbf{a}, \mathbf{b})$, where $\mathbf{a} = lb - L\boldsymbol{\mu}$ and $\mathbf{b} = ub - L\boldsymbol{\mu}$. We sample from the univariate conditional distributions

$$\tilde{v}_i|\tilde{\mathbf{v}}_{-i} \sim \mathcal{TN}(0, 1, a(\tilde{\mathbf{v}}_{-i}), b(\tilde{\mathbf{v}}_{-i})), \quad (3)$$

which is a standard Normal with bounds in the white domain are (We refer to Li and Ghosh, 2015; for detailed explanation):

$$a(\tilde{\mathbf{v}}_{-i}) = \max\left\{ \max_{j:L_{ji}>0} \frac{a_j - L_{j,-i}\tilde{\mathbf{v}}_{-i}}{L_{ji}}, \max_{j:L_{ji}<0} \frac{b_j - \mathbf{l}_{j,-i}\tilde{\mathbf{v}}_{-i}}{L_{ji}} \right\}$$
$$b(\tilde{\mathbf{v}}_{-i}) = \max\left\{ \min_{j:L_{ji}>0} \frac{b_j - \mathbf{l}_{j,-i}\tilde{\mathbf{v}}_{-i}}{L_{ji}}, \min_{j:L_{ji}<0} \frac{a_j - \mathbf{l}_{j,-i}\tilde{\mathbf{v}}_{-i}}{L_{ji}} \right\}.$$

The bounds are functions of the remaining (whitened) fluxes $\tilde{\mathbf{v}}_{-i}$ and need to be updated after each change to flux values. We iteratively update each whitened flux $\tilde{v}_i$ by sampling a new value from the conditional distribution $\tilde{v}_i|\tilde{\mathbf{v}}_{-i}$ with the minimax tilting method (Botev, 2016), which we found out to outperform the alternative Chopin's algorithm (Chopin, 2011). Finally, we transform the whitened variables back into original domain by $\mathbf{v} = L\tilde{\mathbf{v}} + \boldsymbol{\mu}$.

We notice that the method of Li and Ghosh (2015) can be further optimized by running multiple chains in parallel by considering a flux sample matrix $\tilde{V} = (\tilde{\mathbf{v}}^1, \ldots, \tilde{\mathbf{v}}^{N_c})$ containing whitened flux vector chains $\tilde{\mathbf{v}}^c$. The bound function is then represented as

$$\mathbf{a}(\tilde{V}_{-i}) = \max\left\{ \max_{j:L_{ji}>0} \frac{\mathbf{a} - L_{-i}\tilde{V}_{-i}}{L_{ji}}, \max_{j:L_{ji}<0} \frac{\mathbf{b} - L_{-i}\tilde{V}_{-i}}{L_{ji}} \right\}$$
$$\mathbf{b}(\tilde{V}_{-i}) = \max\left\{ \min_{j:L_{ji}>0} \frac{\mathbf{b} - L_{-i}\tilde{V}_{-i}}{L_{ji}}, \min_{j:L_{ji}<0} \frac{\mathbf{a} - L_{-i}\tilde{V}_{-i}}{L_{ji}} \right\},$$

where $\tilde{V}_{-i}$ is the sample matrix $\tilde{V}$ without the ith row, $L_{:,-i}$ is the Cholesky matrix without the ith column. By sampling several chains in parallel with one CPU node we can utilize the full bandwidth of the CPU. This is especially useful for Gibbs sampling.

2.4 Sampling the flux posterior

We set the initial flux vector $\mathbf{v}^{(0)} = \mathbf{v}_{\mathrm{MAP}} = \arg\max_{\mathbf{v}} p(\mathbf{v})$ to the maximum a posteriori of the truncated normal distribution, which we compute using quadratic programing. The truncated normal distribution is unimodal, and hence we begin sampling from the mode of the distribution providing efficient optimization.

The fluxes can be arranged into distinct bounded fluxes $\mathbf{v}_b$ and unbounded fluxes $\mathbf{v}_u$, where $\mathbf{v} = (\mathbf{v}_b, \mathbf{v}_u)^T$. The conditional distribution of a truncated normal is still a truncated normal (Horrace, 2005). We only need to sample the bounded fluxes $\mathbf{v}_b$ with Gibbs MCMC, and afterwards the distribution of unbounded fluxes $\mathbf{v}_u$ conditioned on the bounded flux samples $\mathbf{v}_b$ can be drawn from untruncated normal as

$$\mathbf{v}_u|\mathbf{v}_b \sim \mathcal{N}(\boldsymbol{\mu}_u + C_{ub}C_{bb}^{-1}(\mathbf{v}_b - \boldsymbol{\mu}_b), C_{uu} - C_{ub}C_{bb}^{-1}C_{bu}),$$

where we arrange $\boldsymbol{\mu} = (\boldsymbol{\mu}_b, \boldsymbol{\mu}_u)^T$ and $\mathrm{C} = \begin{pmatrix} C_{bb} & C_{bu} \\ C_{ub} & C_{uu} \end{pmatrix}$.

We implement the MCMC sampling in Matlab. We run multiple independent Markov chains in a vectorized form. By default we run Nc = 10 chains of Ns = 500 flux samples for a total of 5000 flux vector samples. We use thinning and only accept every Nt = 100th

flux sample. The MCMC chains have converged if successive samples are uncorrelated; chains are indistinguishable and have effectively forgotten the initial value. Convergent chains indicate that the MCMC sampler has characterized the whole flux posterior. We use potential scale reduction factor $\hat{R}$ to approximate convergence (Gelman and Rubin, 1992). An optimal value of $\hat{R} = 1$ indicates convergence, while values $\hat{R} < 1.1$ are considered sufficient for convergence. We also compute the effective number of samples N_{eff} per flux (Gelman *et al.*, 2013).

We assume that flux means $\mathbf{m}_v$ are fixed to either zero or the lower bound of each respective flux. The model has then two main hyperparameters $\boldsymbol{\sigma}_{\text{v}}$ and $\boldsymbol{\sigma}_{\dot{\mathbf{x}}}$ that affect the posterior. The $\boldsymbol{\sigma}_{\dot{\mathbf{x}}}$ determines how much the mass balance can be relaxed, and can be set according to the prior knowledge of the modeler. To enforce mass balance, a small value such as $\boldsymbol{\sigma}_{\dot{\mathbf{x}}} = 0.001$ should be chosen. The prior flux variance $\boldsymbol{\sigma}_{\text{v}}^2$ determines how much fluxes are driven towards zero a priori, but also should be set to sufficiently high value not to exclude possibly high fluxes. In practise we set the variance $\sigma_{v_i}^2 = 100^2$ for all fluxes v_i.

2.5 Sampling FBA solutions

The presented Bayesian model is a MFA model designed to characterize the global flux configurations $\mathbf{v} \sim p(\mathbf{v}|\mathbf{y}_{obs})$ compatible with mass balance assumption, observations, and bounds. The method can as well be applied as a FBA method, where an objective function—such as biomass reaction—is maximized. For FBA mode we first find the standard FBA solution objective flux $\mathbf{v}_{\text{obj}}^{\text{FBA}}$ with linear programing, and encode it as a flux observation $\mathbf{y}_{\text{obj}} \pm \omega_c^2$, where the variance determines how closely we sample from the maximal objective. By default, we set the standard deviation to 0.1% of the objective flux. To run maximum growth Bayesian FBA we would set an observation for the biomass pseudoreaction to the classical FBA maximum growth value and condition the model with the pseudo-measurement as in Section 2.2.

3 Results

We first perform *in silico* experiments to highlight the capabilities of the Bayesian FBA and MFA models in Sections 3.1–3.3. Our goal is to compare the computational approach against the conventional FBA and FVA methods, and to showcase the method's *in silico* performance in various metabolic models. We also compare Bayesian FBA solutions with the uniform samples obtained by ACHR and CHRR, two popular tools implemented in Cobra Toolbox to explore and infer properties of metabolic networks. The main experiment of this article is application of the Bayesian flux analysis to the 13C analysis of the *Clostridium acetobutylicum* in Section 3.5, where we can elucidate fluxes on a genome-scale from a small set of intracellular flux measurements.

3.1 *In silico* metabolic models

Table 1 indicates the stoichiometric models that were considered. We considered four organisms, seven genome-scale metabolic models and one core model. All models were downloaded from the BiGG database (http://bigg.ucsd.edu). For all models we run the Bayesian model in FBA mode—by specifying a growth objective—with standard exchange flux measurements included as bounds. We sampled 10 chains of 500 flux vector samples from the full space containing all intracellular and extracellular fluxes. These 5000 flux vectors represent possible flux configurations compatible with the experimental setting. The sampling thinning parameter determines how uncorrelated successive MCMC samples are. We applied thinning values of 100 and 1000, with linear effect on the running time.

The effective number of independent simulation draws for all models from Bayesian flux analysis with different thinning parameter are shown in Figure 1 using the potential scale reduction factor. The x-axis corresponds to individual fluxes sorted based on the effective number of samples. The number of reactions is different for different models. In all cases majority of the fluxes have over 100 effective number of independent samples, which indicates that the samples represent the flux posterior well. A minority of the fluxes have low effective sample sizes. These are usually central branching fluxes that are highly dependent across the genome-scale metabolism, and hence converge slowly. The thinning parameter has a large effect on some models (core, iJR904, iAF1260 and iJO1366) whereas for some other models there are not much change (CORECO, iYO844, iMM904, 7.6).

3.2 Bayesian FBA and MFA

We illustrate the characteristics of the Bayesian model using *Escherichiacoli* central carbon metabolism model (BiGG model e_coli_core). The model contains 95 fluxes and 72 metabolites. It should be noted that the model was not further constrained and do not represent the native *E.coli* strain as such, while it allows, e.g. carbon fixation. The model was rather used to theoretically compare our modeling approach to conventional FBA methodology. The conventional FBA solution achieves a growth flux $v_{\text{growth}}^{\text{FBA}} \approx 0.873$ by limiting the glucose exchange flux with a $lb_{\text{glc}} = -10$. We consider three cases of Bayesian analysis: (i) 50% growth by defining only the biomass flux observation 0.436 ± 0.0043, (ii) maximal growth scenario by defining the biomass flux observation 0.873 ± 0.0087, and (iii) maximal growth with additional observations for nine key exchange fluxes: glucose (GLC), O_2, CO_2, H_2O, H+, HPO_4, SO_4, NH_4 and ethanol that were all set to their conventional FBA solutions with a SDs of 0.01. In all three experiments the remaining fluxes were free with only a prior distribution with a SD of 100 specified. We defined a nearly strict steady-state by defining $\sigma_{\dot{x}} = 0.01$. We sample a total of 5000 flux vectors with the Gibbs sampler using 10 chains and 500 samples each. We use 1D kernel density estimates as proxies of marginal flux posterior distributions. The small jaggedness of the distributions is an artifact from the MCMC sampling. By considering longer chains these would eventually smoothen out.

Figure 2 shows the flux distributions of 30 fluxes. The blue color indicates the 50% growth flux distributions, the green color the maximal growth distributions, the red color maximal growth with exchange fluxes specified, and the conventional FBA is shown with a black line. The Bayesian distributions represent the space of all allowed steady-state flux configurations given the observations and objective function. Figure 2 shows that maximal growth can still support a large variance in many fluxes, with the FBA point estimate misleading by only considering one flux configuration. Similarly to conventional FVA our approach elucidates directly the possible variance in a given flux. For instance, the pentose-phosphate pathway flux G6PDH2r: D-Glucose 6-phosphate + NADP ⟺ 6-Phosphogluconolactone + H^+ + NADPH can vary between −8 and −2 in maximal growth. The conventional FBA yields zero flux for glyoxylate cycle flux malate synhase (MALS): Acetyl-CoA + Glyoxylate + H2O ⇒ CoA + H^+ + Malate, while the flux space indicates that values up to four are possible.

Table 1. Metabolic models analyzed by Bayesian flux analysis by sampling 500 samples from the flux posteriors

Organism	Model	n	M	Runtime thin 100	thin 1000
E.coli	core	95	72	2 min	20 min
E.coli	iJR904	1075	761	2 hr	1 day 9 h
E.coli	iAF1260	2382	1668	7 hr	4 days 12 h
E.coli	iJO1366	2583	1805	9 hr	4 days 16 h
Bacillus subtilis	iYO844	1250	992	3 hr	1 day 11 h
C.acetobutylicum	Wallenius (2013)	592	444	23 min	3 h
Saccharomyces cerevisiae	iMM904	1577	1226	3 hr	2 days
S.cerevisiae	7.6	3493	2220	10 hr	5 days 15 h
Trichoderma reesei	CORECO	4008	3292	8 hr	4 days 15 h

Note: The runtime is shown for thinning rate 100 and 1000. n denotes the number of reactions and M the number of metabolites in the model.

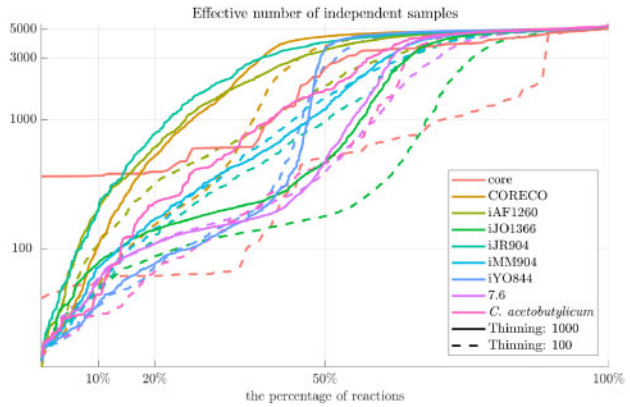

Fig. 1. The effective number of independent simulation draws for individual fluxes for a subset of models from Bayesian flux analysis with different thinning parameters. The x-axis corresponds to individual fluxes sorted based on the effective number of samples

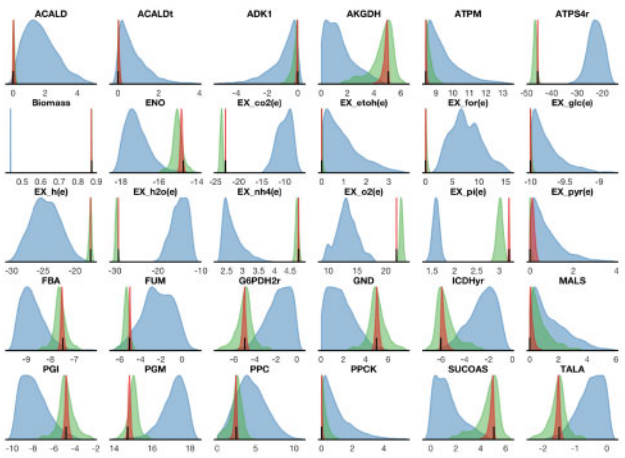

Fig. 2. Posterior flux distributions of *E.coli* core model. The blue color indicates fluxes in 50% growth, the green color maximal growth, the red color maximal growth with nine exchange fluxes specified, and the conventional FBA solution is a black line. Reaction abbreviations and names are listed in Supplementary Table S1

The red distributions indicate how the intracellular fluxes get more and more specified as the model is better specified by inclusion of exchange measurements. Variance of almost all fluxes reduces by more than half. For instance, the flux FUM: Fumarate + H_2O $\Longleftrightarrow$

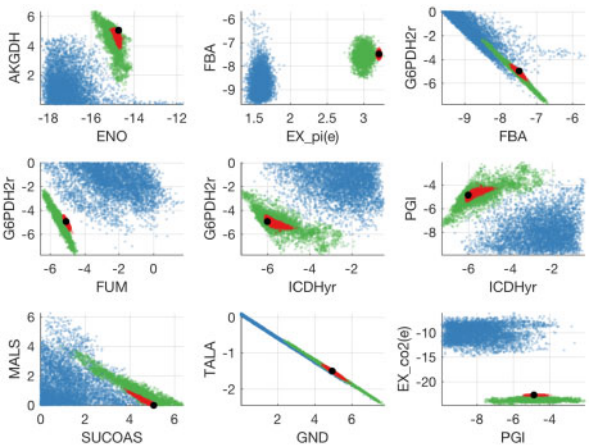

Fig. 3. Examples of flux covariance distributions of E. coli core network. Blue points represent 50% growth, green points maximal growth, red points maximal growth with nine exchange fluxes specified and the conventional FBA solution is a black dot. Scatter plots represent pair-wise (2D) marginal posterior distributions as obtained from the MCMC samples. Reaction abbreviations and names are listed in Supplementary Table S1

Malate is specified to a range of $[-5.1, -4.9]$ from a range of $[-6.5, 4.4]$ without exchange measurements.

The blue color indicates the cellular flux state when the cell is only growing at 50% of the maximum growth rate. Most fluxes have a higher variance in this scenario. Interestingly the glucose intake is still kept at a relatively high rate. Instead of biomass production, the excess carbon from glucose can be diverted to other carbon sinks, such as formate and ethanol production. The ethanol and formate effluxes can grow up to 3 and 15, respectively. The carbon dioxide exchange decreases by over half into a range of $[-15, -3]$ from maximal growth exchange range of $[-25, -23]$.

3.3 Flux couplings

The flux variations are in general not independent from each other. To understand the intracellular flux space, we need to consider higher-order flux dependencies. The flux sample covariances indicate flux couplings, where the variation in one flux is constrained by other fluxes. Figure 3 highlights nine example flux pair patterns out of the total of $\frac{95 \cdot 94}{2} = 4465$ in the core model. Blue points represent 50% growth, green points maximal growth, red points maximal growth with nine exchange fluxes specified, and the conventional FBA solution is a black dot.

The flux covariations become consistently more constrained while traversing from the loose 50% growth model (blue) towards

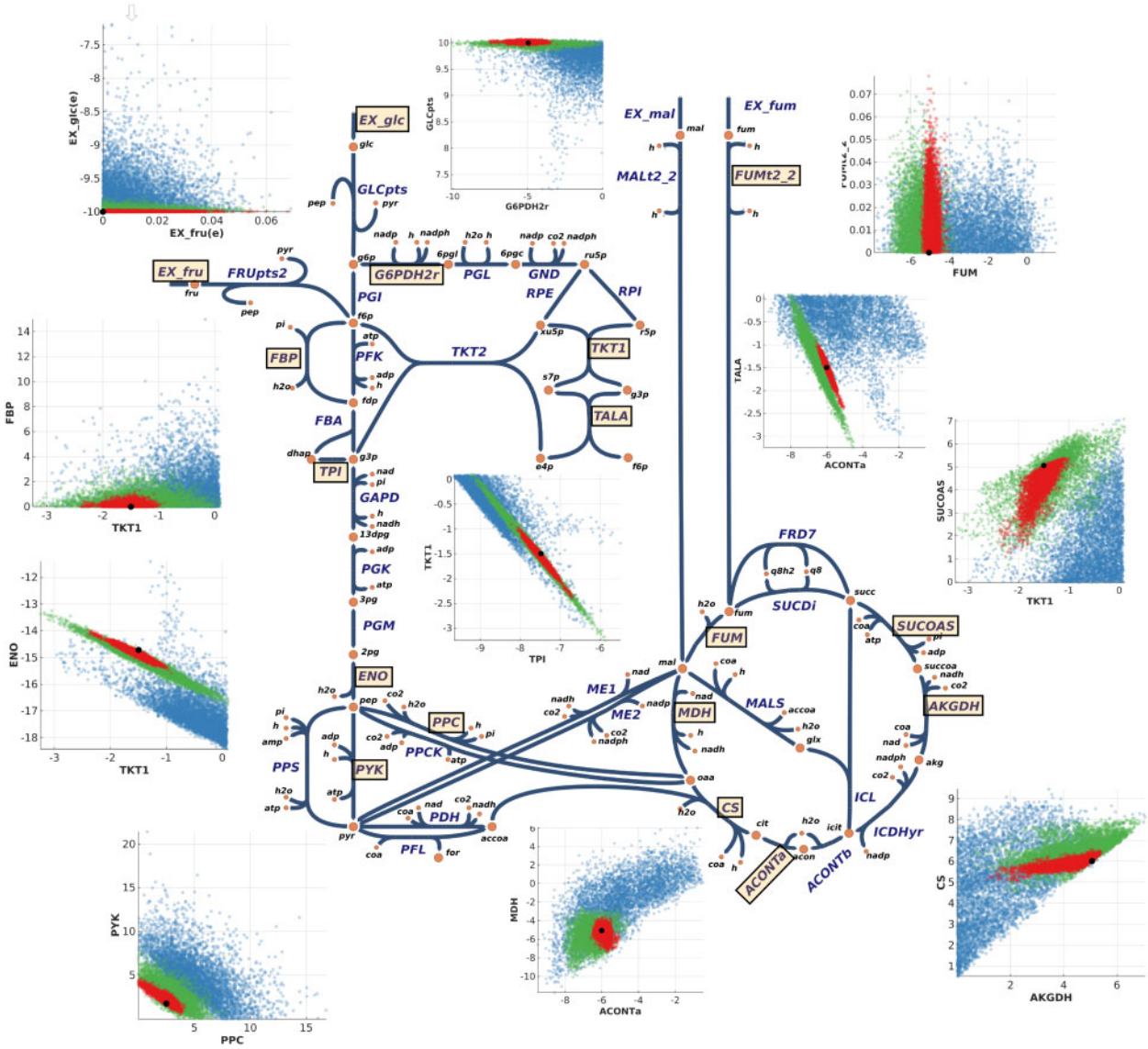

Fig. 4. Pair-wise marginal posterior fluxes presented together in the metabolic network map. The visualized fluxes are highlighted. Blue points represent flux values in 50% growth, green points in maximal growth, red points in maximal growth with nine exchange fluxes specified and the conventional FBA solution is a black dot. Reaction abbreviations and names are listed in Supplementary Table S1

100% maximal growth (green). By measuring the exchange fluxes (red), we can already pinpoint most flux patterns very closely around the theoretically optimal flux pattern as defined by the conventional FBA (black).

Multiple patterns of covariation can be immediately identified. There is an exact coupling between pentose-phosphate pathway reactions GND: 6-Phospho-D-gluconate + NADP ⇒ CO_2 + NADPH + D-Ribulose 5-phosphate and TALA: Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate ⟺ D-Erythrose 4-phosphate + D-Fructose 6-phosphate, as expected from the stoichiometry. Glycolysis related FBA: D-Fructose 1, 6-bisphosphate ⟺ Dihydroxyacetone phosphate + Glyceraldehyde 3-phosphate and pentose-phosphate pathway related G6PDH2r have also a strong, but not exact, negative correlation. The flux PGI: D-Glucose 6-phosphate ⟺ D-Fructose 6-phosphate and carbon dioxide exchange have no correlation, but the maximal growth still pinpoints to a carbon dioxide exchange value of approximately −21. The dependency of glyoxylate cycle related MALS and citric acid cycle (TCA) cycle related SUCOAS: ATP + CoA + Succinate ⟺ HPO_4^{-2} + ADP + Succinyl-CoA on maximal growth requirement can be observed in Figure 3. Under maximal growth a negative correlation between the two fluxes emerges. The conventional FBA solution pinpoints optimal values as zero MALS with SUCOAS around 5, while the Bayesian model reveals that MALS can still have a flux value of around 4 as long as SUCOAS tends towards 1.5 simultaneously.

The patterns of G6PDH2r and FBA and FUM indicate a linear inequality for these fluxes. Especially with FUM this is natural since the pentose-phosphate pathway flux G6PDH2r limits the TCA cycle flux FUM. The same effect is also seen with G6PDH2r and ICDHyr: Isocitrate + NADP ⟺ 2-Oxoglutarate + CO_2 + NADPH, another TCA cycle flux.

To get more insight into the biology behind the flux couplings, the flux pair patterns can also be illustrated in the metabolic network (Fig. 4). Figure 4 shows the samples of the flux distributions for several example pairs of reactions. These scatter plots indicate the dependency of the flux configurations between two reactions

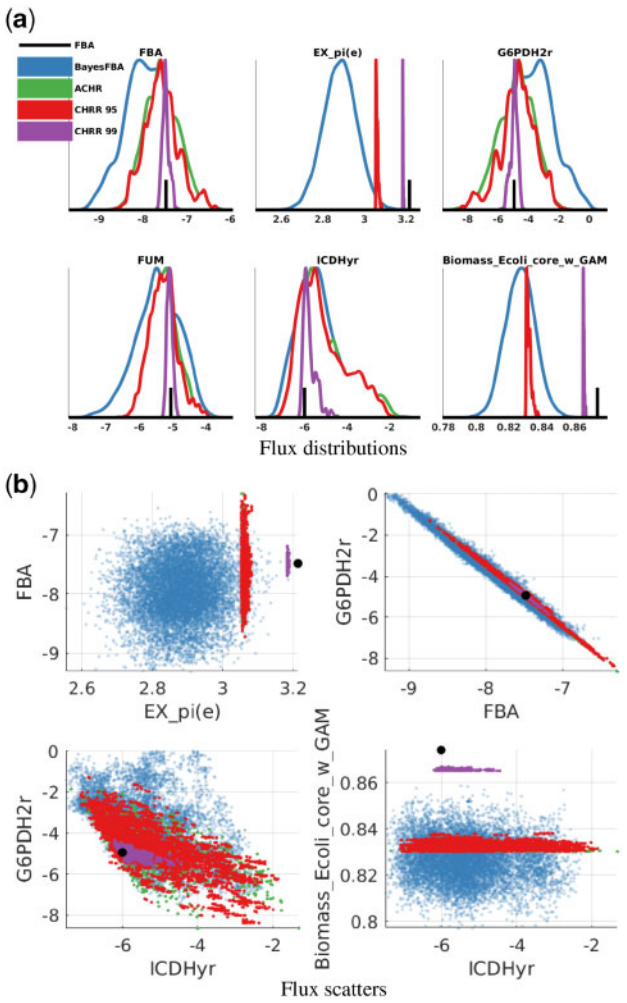

Fig. 5. Sampler comparison on the *E.coli* core model around the FBA solution (black). The ACHR (green), CCHR 95% (red), CCHR 99% (purple) and BayesFBA (blue) distributions are shown for six example flux distributions (a), and for four example pairwise flux distributions (b). The green and red overlap heavily

across the reaction. There is natural correlation between adjacent or subsequent fluxes but also correlation between fluxes in different pathways, such as glycolysis and TCA cycle (See the SUCOAS and TKT1 pair).

3.4 Comparison to hit-and-run samplers

We compare our Bayesian model and the Gibbs sampler to the competing flux sampling methods of ACHR and CCHR. The CCHR sampler has the ability to sample uniformly from the optimal flux solution space, while both samplers have the ability to relax the optimality. The CCHR draws hard FVA bounds around all fluxes, while ACHR can deviate softly from the optimal flux solution. We run ACHR and CCHR methods on the *E.coli* core model around the conventional FBA solution. For Bayesian FBA, the biomass flux observation distribution has its mean in the center point of FVA solution that gives at least 95% of the optimal growth. The distribution has mean 0.852 and variance 0.01. For ACHR experiments, the lower and upper bounds for biomass flux were set as the 95% FVA bounds. In CHRR experiments, random draws were obtained from FVA solution that gives at least 95 or 99% of the optimal growth. ACHR and CHRR were run with their default parameters, 10 independent chains were drawn for all samplers, thinning was 10 and the number of samples per chain was 1000.

Figure 5 compares the flux distributions of BayesFBA, ACHR and CCHR with 95% FVA bounds, and CHRR with 99% FVA bounds. The samples obtained by ACHR and CHRR for the Biomass flux concentrate near the lower bound of the 95 or 99% FVA solution, whereas BayesFBA gives wide Normal-like distribution with mean 0.83. Similar effect is also seen on the scatter plots with both ACHR and CCHR inducing hard bounds on the phosphate exchange EX_pi(e). BayesFBA is not constrained on FVA bounds of relaxed growth, thus we avoid specifying the growth relaxation. Instead, BayesFBA uses the uncertainty in the measured biomass production. The variances of distributions obtained by BayesFBA are larger than for the samples obtained by ACHR and CHRR; ACHR and CHRR likely underestimate the variance. We do not find significant differences in the samples obtained by ACHR and CHRR. Moreover, the dependencies between reactions obtained by different sampling methods are similar, for example, as seen in Figure 5b for reactions ICDHyr and G6PDH2r.

3.5 Intracellular flux elucidation of *C.acetobutylicum*

We consider the results obtained from ^{13C}MFA of *C.acetobutylicum* grown in chemostat, i.e. in continuous cultivation maintaining steady-state, with reference condition, glucose limited condition and butanol stimulus with the goal of inferring the internal fluxes. We effectively repeat the study of Wallenius *et al.* (2016), where FBA and FVA were performed and constrained on 12 intracellular fluxes determined by ^{13C}MFA and 7 exchange fluxes. The model for *C.acetobutylicum* consists of 451 metabolites and 604 reactions, and is given as an.xml file in the Supplementary Material of Wallenius *et al.* (2016).

The data of measured fluxes in glucose limited condition are shown in Table 2. For the reference and the butanol stimulated conditions, see Supplementary Table S2. The internal fluxes are obtained from ^{13C}MFA analysis, whereas the exchange fluxes were measured by chromatographical methods or transferred from the ^{13C}MFA results. Flux values were normalized to the specific growth rate which was given the value of 1, except for the reference condition, the measured growth is 0.95. Exchange fluxes measured from the cultivations were given to Bayesian MFA as mean $\mathbf{v}_o$ and standard deviations $\Omega_0 = 0.05 \cdot \mathbf{v}_o$ and fluxes obtained from ^{13C}MFA were given as ranges. In all Bayesian MFA experiments, the steady-state relaxation was $\sigma_{\dot{x}} = 0.01$. Finally, 500 samples were drawn from the posterior with thinning 1000.

To study the FBA's FVA's and BMFA's performance in predicting the measured ranges for 12 internal fluxes obtained by ^{13C}MFA, three sets of models were generated: (A) a model with reaction directions for the 12 ^{13C}MFA determined internal fluxes set according to the data (bounds in Table 2), (B) a set of 12 models, each model defined as unconstraining 1 of 12 ^{13C}MFA determined internal flux at a time, (the reaction direction is still constrained) while the rest 11 fluxes are constrained according to the measurements, this is the leave-one-out (LOO) setting. Finally, we define model (C) with all 12 ^{13C}MFA determined internal fluxes constrained to their measured values. Model (A) is the most relaxed with only reaction directions specified, and for model set (B) we tested how well we can estimate the true flux value for 1 internal reaction while the rest 11 reactions specified (LOO setting). In all three cases, we constrain the model with measured values for the six exchange reactions and the measured growth. For each set of models, the standard MFA with Taxicab penalty, FVA, and Bayesian MFA were performed. The

Table 2. The 6 exchange flux, the growth and exopolysaccharide (EPS) production and the 12 internal flux measurements

Reaction	KEGG ID	Bounds	Glucose limited	F1 scores % ex only FVA	ex only BMFA	LOO FVA	LOO BMFA	ex + 13C FVA	ex + 13C BMFA
Exchange fluxes									
Glucose exchange[a]	C00031	≤ 0	-73.3 ± 3.7						
Acetate exchange[b]	C00033	$0 \leq$	[12.96 .. 13.016]						
Acetone exchange[a]	C00207	$0 \leq$	12.5 ± 0.06						
Butanol exchange[b]	C06142	$0 \leq$	[29.62 .. 30.67]						
Butyrate exchange[b]	C00246	$0 \leq$	[0 .. 3.23]						
Ethanol exchange[a]	C00469	$0 \leq$	6.13 ± 0.31						
EPS production[b]		$0 \leq$	[10.01 .. 10.26]						
Growth[a]		$0 \leq$	1.00 ± 0.05						
Malate DHase	R00342		[−112 .. −5.94]	19	7	65	19	68	23
3P-glycerate DHase	R01513	$0 \leq$	[1.53..4.85]	1	53	12	59	100	100
Acetaldehyde DHase	R00228		[−27.4 .. 50.3]	8	56	12	73	95	92
Triosephosphate DHase	R01061	$0 \leq$	[77.2 .. 132]	10	95	5	75	63	85
Acetolactate synthase	R04672		[−95.2 .. 99.4]	17	67	17	68	68	68
Aspartate transaminase	R00355	≤ 0	[−8.95 .. −0.80]	1	32	14	29	69	69
5, 10-CH=THF hydrolyase	R01655		[−2.24 .. 0.05]	0	3	0	3	100	100
Malate hydrolyase	R01082	≤ 0	[−10.2 .. −0.78]	2	83	59	54	62	67
Ribulose-5P epimerase	R01529		[−4.39 .. −1.27]	1	0	1	0	100	100
Pyruvate carboxylase	R00344	$0 \leq$	[13.6 .. 119]	19	36	65	33	66	26
Carbonate hydrolyase	R10092		[26.2 .. 75.7]	10	57	97	31	100	53
C-acetyl transferase	R00212	$0 \leq$	[66.3 .. 154]	16	2	84	13	100	81
Average				9	41	36	38	83	72

Note: Measurements from the cultivations include SDs, while fluxes determined by ^{13C}MFA are given as a range. The unit for flux ranges, flux means and flux SD is: g^{-1}(CDW).

[a]Measured by chromatographic methods.

[b]Obtained from ^{13C}MFA.

EPS, exopolysaccharide.

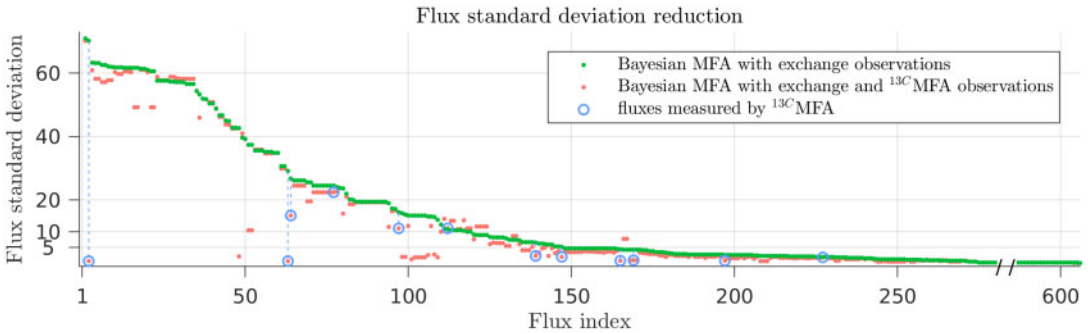

Fig. 6. Global flux standard deviation reduction due to addition of twelve ^{13C}MFA internal flux measurements in glucose limited condition. The green points indicate the standard deviation of fluxes given only exchange measurements. The red points indicate the corresponding standard deviations after inclusion of twelve ^{13C}MFA intracellular flux measurements. The blue circles highlight the ^{13C}MFA measured fluxes

MFA and FVA were performed by The Cobra Toolbox's `optimizeCbModel` and `fluxVariability` functions by maximizing growth with the growth lower and upper bounds set to the measured value 1.

We study how the flux variances from Bayesian MFA results decrease when adding ^{13C}MFA constraints. The Figure 6 shows the reduction of standard deviations of flux distributions of all fluxes of *C.acetobutylicum* in glucose limited condition when all 12 ^{13C}MFA constraints are added to the model (model set C). When adding the ^{13C}MFA constraints, the variance of most unmeasured internal fluxes decreases, demonstrating how Bayesian MFA propagates the information about 12 measured fluxes to several tens of other internal fluxes. The reduction of standard deviations of flux distributions for the reference and butanol stimulated conditions are shown in Supplementary Figures S2 and S3.

To quantify the performance of FVA and Bayesian MFA to predict the ^{13C}MFA measured range of flux values, precision, recall and the F1 score were computed for each reaction and model set (see Supplementary Methods). The F1 score values are shown in Table 2 as percentage and the precision and recall values are shown in the Supplementary Table S3. From Table 2 and Supplementary Table S3, it can be concluded that in the glucose limited condition, the Bayesian MFA outperforms FVA, as for model (A) 9 of 12 reactions has higher precision and F1 score, and for both models (A) and

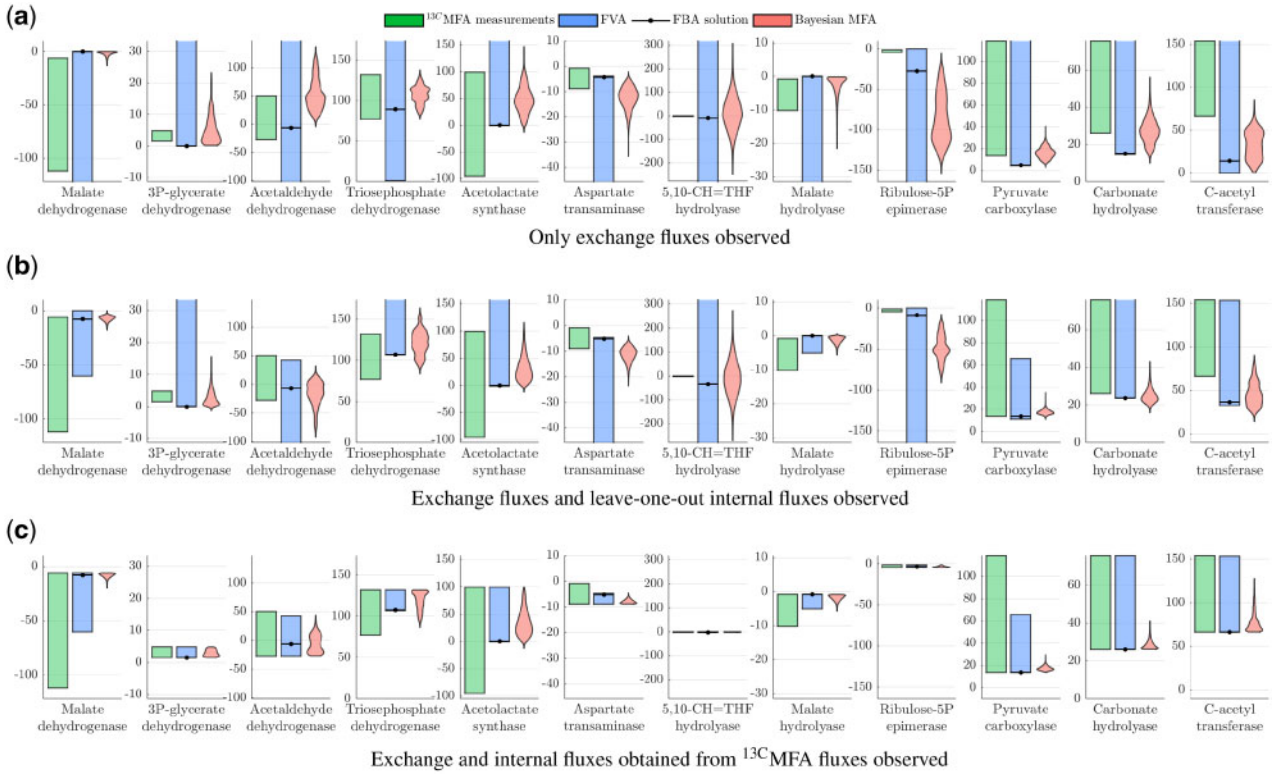

Fig. 7. For the glucose limited condition, flux distributions of the 12 internal fluxes predicted solely from exchange fluxes (**a**), and distributions after seeing ^{13C}MFA data in LOO experiment (**b**) and after seeing all 12 internal reactions (**c**)

model set (B), the average precision and F1 score is higher for BMFA than for FVA. Similar results are obtained for butanol stimulated (Supplementary Table S4) and reference condition (Supplementary Table S5). In butanol stimulated condition, BMFA has for 8 out of 12 reactions higher precision, and for 7 out of 12 reactions higher F1 score for model set (A), for 10 out of 12 reactions higher precision and for 7 out of 12 reactions higher F1 score for model set (B). For reference condition, BMFA has for 8 out of 12 reactions higher precision and F1 score for model set (A), and for half of the reactions higher precision for model set B). The average F1 score and precision over all 12 reactions are always higher for BMFA than FVA.

The ^{13C}MFA measured ranges, together with FVA, FBA and BMFA predictions for model sets (A), (B) and (C) are shown for the 12 ^{13C}MFA determined reactions in Figure 7. In Figure 7, for the model sets (A) and (B), the FVA gives a wide range of solutions as it doesn't take into account the flux couplings, whereas the distribution from BMFA is narrower, and closer to the range of true fluxes; this is also seen in the higher precision and F1 scores for BMFA compared with FVA (see Table 2). For example, Figure 7b shows the results for the model set (B): for reactions 3P-glycerate dehydrogenase, Acetaldehyde dehydrogenase, Triosephosphate dehydrogenase and Acetolactate synthase the posterior distribution obtained by BMFA resembles more the range obtained by ^{13C}MFA, whereas the FVA gives wider ranges. In Figure 7c, the resulting flux ranges for FVA and BMFA distributions are always within the true measured range, but the Bayesian MFA captures the probability in the flux values. Similar results are obtained for the butanol stimulated and reference conditions (see Supplementary Tables S4 and S5 and Supplementary Figs S4 and S5).

4 Discussion

The conventional FBA formalism is a powerful framework for flux analysis that however assumes several unrealistic simplifying model approximations. Several approaches from robust flux analysis and sampling to flux variability analyses indicate the need to alleviate the approximations towards a more principled model.

We proposed the Bayesian flux analysis formalism that considers fluxes as distributions instead of point estimates. The model learns a posterior distribution of fluxes given prior information, flux measurements, upper and lower bounds and steady-state assumptions into account. The degree of belief in the measurements and steady-state can be adjusted via measurement noise variances and biological knowledge as encoded in (subjective) priors. The model characterizes the complete space of possible flux configurations by modeling the uncertainties of fluxes and flux combinations. The Bayesian formalism can be seen as a drop-in replacement for deterministic flux analysis tools—such as FBA and FVA—at the cost of added running time necessary to properly characterize the flux distributions. The runtime can be effectively alleviated by only considering the core parts of the metabolic model or by running multiple MCMC chains in parallel.

Our results show that the conventional FBA and FVA tools provides an overly simplistic view of the flux capabilities of the cellular system under study, while the Bayesian model expresses the full variance in the flux configurations. The Bayesian model of metabolism opens doors for building flux analysis models in a Bayesian way. We will leave experimental design, knock-outs and strain design using the Bayesian modeling basis for future work. In future we expect the Bayesian formalism to provide an alternative statistical approach for majority of current FBA- and MFA-based

tools with the benefit of rigorous uncertainty modeling and improved interpretation.

Funding

This work has been supported by the Academy of Finland Center of Excellence in Systems Immunology and Physiology, the Academy of Finland [grant numbers 299915 and 313271], the Finnish Funding Agency for Innovation Tekes [grant number 40128/14, Living Factories] and the Finnish Cultural Foundation.

Conflict of Interest: none declared.

References

Almaas,E. *et al.* (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, **427**, 839–843.

Altmann,Y. *et al.* (2014) Sampling from a multivariate gaussian distribution truncated on a simplex: a review. In: *Statistical Signal Processing*.

Becker,S.A. *et al.* (2007) Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox. *Nat. Protoc.*, **2**, 727.

Bhadra,S. *et al.* (2018) Principal metabolic flux mode analysis. *Bioinformatics*, **34**, 2409–2417.

Bordbar,A. *et al.* (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.*, **15**, 107–120.

Bordel,S. *et al.* (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.*, **6**, e1000859. doi: 10.1371/journal.pcbi.1000859.

Botev,Z. (2016) The normal law under linear restrictions: Simulation and estimation via minimax tilting. *J. R. Stat. Soc. B.* **79**, 125–148.

Carreira,R. *et al.* (2014) CBFA: phenotype prediction integrating metabolic models with constraints derived from experimental data. *BMC Syst. Biol.*, **8**, 123.

Chopin,N. (2011) Fast simulation of truncated Gaussian distributions. *Stat. Comput.*, **21**, 275–288.

Dubois,D. *et al.* (1996) Possibility theory in constraint satisfaction problems: handling priority, preference and uncertainty. *Appl. Intell.*, **6**, 287–309.

Emery,X. *et al.* (2014) Simulating large Gaussian random vectors subject to inequality constraints by Gibbs sampling. *Math. Geosci.*, **46**, 265–283.

Feist,A. and Palsson,B. (2010) The biomass objective function. *Curr. Opin. Microbiol.*, **13**, 344–349.

Gelman,A. and Rubin,D. (1992) Inference from iterative simulation using multiple sequences. *Stat. Sci.*, **7**, 457–511.

Gelman,A. *et al.* (2013) *Bayesian Data Analysis, Chapter 11.4–11.5*, 3rd edn. CRC Press, Boca Raton, Florida.

Geweke,J. (1991) Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In: *Computing Science and Statistics*, pp. 571–578.

Gudmundsson,S. and Thiele,I. (2010) Computationally efficient flux variability analysis. *BMC Bioinformatics*, **11**, 489.

Haraldsdóttir,H.S. *et al.* (2017) CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, **33**, 1741–1743.

Heino,J. *et al.* (2007) Bayesian flux balance analysis applied to a skeletal muscle metabolic model. *J. Theor. Biol.*, **248**, 91–110.

Heino,J. *et al.* (2010) Metabolica: a statistical research tool for analyzing metabolic networks. *Comput. Methods Programs Biomed.*, **97**, 151–167.

Horrace,W. (2005) Some results on the multivariate truncated Normal distribution. *J. Multivariate Anal.*, **94**, 209–221.

Kadirkamanathan,V. *et al.* (2006) Markov chain Monte Carlo algorithm based metabolic flux distribution analysis on *Corynebacterium glutamicum*. *Bioinformatics*, **22**, 2681–2687.

Kaufman,D. and Smith,R. (1998) Direction choice for accelerated convergence in hit-and-run sampling. *Oper. Res.*, **46**, 84–95.

Kim,H. *et al.* (2008) Metabolic flux analysis and metabolic engineering of microorganisms. *Mol. Biosyst.*, **4**, 113–120.

Kotecha,J. and Djuric,P. (1999) Gibbs sampling approach for generation of truncated multivariate gaussian random variables. In: *Acoustics, Speech, and Signal Processing*.

Li,Y. and Ghosh,S. (2015) Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints. *J. Stat. Theory Pract.*, **9**, 712–732.

Llaneras,F. *et al.* (2009) A possibilistic framework for constraint-based metabolic flux analysis. *BMC Syst. Biol.*, **3**, 79.

MacGillivray,M. *et al.* (2017) Robust analysis of fluxes in genome-scale metabolic pathways. *Sci. Rep.*, **7**, 268. doi: 10.1038/s41598-017-00170-3.

Mahadevan,R. and Schilling,C. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, **5**, 264–276.

Megchelenbrink,W. *et al.* (2014) optGPSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS One*, **9**, 1–8.

Mo,M. *et al.* (2010) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.*, **3**, 37. doi: 10.1186/1752-0509-3-37.

Murray,I. *et al.* (2010) Elliptical slice sampling. *J. Mach. Learn. Res.*, **9**, 541–548.

Orth,J. *et al.* (2010) What is flux balance analysis. *Nat. Biotechnol.*, **28**, 245–248.

Pakman,A. and Paninski,L. (2014) Exact hamiltonian Monte Carlo for truncated multivariate Gaussians. *J. Comput. Graph. Stat.*, **23**, 518–542.

Pakula,T.M. *et al.* (2016) Genome wide analysis of protein production load in *Trichoderma reesei*. *Biotechnol. Biofuels*, **9**, 132.

Palsson,B. (2015) *Systems Biology: Constraint-Based Reconstruction and Analysis*. Cambridge University Press, Cambridge, UK.

Saa,P. and Nielsen,L. (2016a) ll-ACHRB: a scalable algorithm for sampling the feasible solution space of metabolic networks. *Bioinformatics*, **32**, 2330–2337.

Saa,P.A. and Nielsen,L.K. (2016b) Construction of feasible and accurate kinetic models of metabolism: a Bayesian approach. *Sci. Rep.*, **6**, 29635.

Schellenberger,J. *et al.* (2010) BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**,

Schellenberger,J. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2. 0. *Nat. Protoc.*, **6**, 1290.

Smith,R. (1984) Efficient Monte-Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.*, **32**, 1296–1308.

Theorell,A. *et al.* (2017) To be certain about the uncertainty: Bayesian statistics for ^{13}C metabolic flux analysis. *Biotechnol. Bioeng.*, **114**, 2668–2684.

Wallenius,J. *et al.* (2016) Carbon 13-metabolic flux analysis derived constraint-based metabolic modelling of *Clostridium acetobutylicum* in stressed chemostat conditions. *Bioresource Technol.*, **219**, 378–386.

Zavlanos,M. and Julius,A. (2011) Robust flux balance analysis of metabolic networks. In: *American Control Conference (ACC)*, 2011. IEEE, pp. 2915–2920.