

Published in final edited form as:

J Theor Biol. 2007 September 7; 248(1): 91–110.

Bayesian flux balance analysis applied to a skeletal muscle metabolic model

Jenni Heino^{a,*}, Knarik Tunyan^a, Daniela Calvetti^b, and Erkki Somersalo^a

^a*Institute of Mathematics, Helsinki University of Technology, P.O. Box 1100, FI-02015, Finland* ^b*Department of Mathematics and Center for Modeling Integrated Metabolic Systems (MIMS), Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA*

Abstract

In this article, the steady state condition for the multi-compartment models for cellular metabolism is considered. The problem is to estimate the reaction and transport fluxes, as well as the concentrations in venous blood when the stoichiometry and bound constraints for the fluxes and the concentrations are given. The problem has been addressed previously by a number of authors, and optimization based approaches as well as extreme pathway analysis have been proposed. These approaches are briefly discussed here. The main emphasis of this work is a Bayesian statistical approach to the flux balance analysis (FBA). We show how the bound constraints and optimality conditions such as maximizing the oxidative phosphorylation flux can be incorporated into the model in the Bayesian framework by proper construction of the prior densities. We propose an effective Markov Chain Monte Carlo (MCMC) scheme to explore the posterior densities, and compare the results with those obtained via the previously studied Linear Programming (LP) approach. The proposed methodology, which is applied here to a two-compartment model for skeletal muscle metabolism, can be extended to more complex models.

Keywords

Flux balance analysis; steady state; skeletal muscle metabolism; linear programming; Bayesian statistics; Markov Chain Monte Carlo; Gibbs sampler

1 Introduction

Computational models for cellular metabolism play an important role in the quest for understanding the complex biochemical interactions between different organs, different cell types within an organ and the biochemistry within the cells. As our understanding of the functioning of the cells increase, the models become more complex and the number of parameters needed to identify them increase. The determination of a unique set of model parameters based on scarce and uncertain data is usually not possible, and the estimation of the parameters has to rely on additional information concerning the system. The model identification can thus be seen as a statistical inference problem and, in this context, a Bayesian

* Corresponding author. Address: Institute of Mathematics, Helsinki University of Technology, P.O. Box 1100, FI-02015, Finland. Email: jenni.heino@tkk.fi. Telephone: +358 9 451 5722. Fax: +358 9 451 3070.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

statistical framework has turned out to be useful as it allows the incorporation of various levels of uncertain information into the model (Calvetti & Somersalo, 2006; Calvetti et al., 2006b).

In the study of the dynamic response of a metabolic system to varying conditions such as hypoxia or ischemia, it is common to assume that initially the system is at rest, corresponding to a biochemical steady state. The flux balance analysis (FBA) is used to determine the initial fluxes and consequently, a set of parameter values is often manually adjusted to satisfy the steady state. Due to the non-uniqueness of the parameter values identifying a steady state, the analysis of dynamic response based on the chosen values may depend strongly on how the values are chosen. Therefore, it is useful to understand to what extent the flux balance equations are able to identify the steady state and to what extent the steady state is stable with respect to perturbations in quantities which are assumed known. The present paper addresses this issues. The analysis is based on the Markov Chain Monte Carlo (MCMC) techniques, commonly used in statistical inference problems (Gilks et al., 1996; Kaipio & Somersalo, 2004; Liu, 2003).

In the literature it has been proposed to identify a feasible steady state by minimizing or maximizing a suitable objective function (Bonarius et al., 1997; Ramakrishna et al., 2001; Varma & Palsson, 1996; Ramakrishna et al., 2001). While such approach can usually single out a reasonable, feasible steady state, it is important to also investigate the stability of such solutions with respect to perturbations in the input parameters and to assess how representative statistically such solutions are.

The paper is organized as follows. In Section 2, we discuss the two-compartment metabolic model for skeletal muscle and the constraints that are appropriate for the fluxes. In Section 3, we review the steady state flux balance analysis from the classical optimization point of view. More specifically, we briefly describe the Linear Programming (LP) solution to FBA proposed in the literature, that will serve as a benchmark for the subsequent statistical analysis, and provides a feasible initial value for statistical simulations. In addition, we discuss the stability of the LP estimators estimators, i.e., the dependency of the result on the bound constraints and on uncertainties in the model input values. In Section 4, the statistical framework is presented, and the problem is recast in the form of Bayesian inference problem. Conclusions and comment on future work are found in Section 5. The details concerning the model are presented in Appendix A.

2 Multi-compartment models

The simplest example of the general multi-compartment models for cell metabolism is a two-compartment model consisting only of blood and lumped tissue compartments, see, e.g., Calvetti & Somersalo (2006); Calvetti et al. (2006b); Salem et al. (2004); Zhou et al. (2005). More complex models take into account the partitioning of the cells into cytosol and organelles such as mitochondria, and may differentiate between different cell types such as neurons and astrocytes in the brain metabolism models. Although in this paper we apply our Bayesian approach only to the case of a two-compartment model, the analysis can be extended to more complex models with the same structure.

2.1 Dynamic model and steady state

The model of metabolic system that we consider here consists of two domains, blood and a single cell type, where the cytosol and mitochondria are lumped together (see Appendix A, Figure 8). The state of the system at time t is described by vectors $\mathbf{C}_c(t)$ and $\mathbf{C}_b(t)$ that contain the concentrations of the biochemical species in the cell domain and the venous blood, respectively. As some of the species are interchanged between the compartments, their respective concentrations change. In the blood domain the concentrations change also due to convection, while in the cell domain, biochemical reactions consume or produce the

metabolites and the intermediates. Thus, the dynamics of the concentrations are described by a system of differential equations of the form

$$V_c \frac{dC_c}{dt} = \underbrace{S\Phi(t)}_{\text{reaction}} + \underbrace{M(\mathbf{J}_{b \rightarrow c}(t) - \mathbf{J}_{c \rightarrow b}(t))}_{\text{transport}}, \quad (1)$$

$$V_b \frac{dC_b}{dt} = \underbrace{\frac{Q(t)}{F}(C_{a,\text{tot}}(t) - C_{b,\text{tot}}(t))}_{\text{convection}} + \underbrace{\mathbf{J}_{c \rightarrow b}(t) - \mathbf{J}_{b \rightarrow c}(t)}_{\text{transport}}. \quad (2)$$

The convection term depends on the difference of total arterial $C_{a,\text{tot}}$ and venous $C_{b,\text{tot}}$ concentrations of the species, where “total” means that in the oxygen and carbon dioxide concentrations, the oxy-hemoglobin, carbamino-hemoglobin and bicarbonate concentrations are taken into account in addition to the free dissolved concentrations (Dash & Bassingthwaite, 2006; Lai et al., 2006). The factor $Q(t)$ represents the blood flow, and F is the mixing ratio. The transport flux vectors $\mathbf{J}_{c \rightarrow b}$ and $\mathbf{J}_{b \rightarrow c}$ contain the non-negative transport fluxes of the species from cell to blood and blood to cell, respectively, and the matrix M describes which compounds participating in the metabolic processes in the cell domain are exchanged with the blood domain. Hence, if the j th compound in C_c is not transported to the blood domain, the corresponding row in the matrix M vanishes, otherwise the row contains a one in an appropriate place to pick the flux of the j th species. The reaction term consists of the product of a vector Φ of reaction fluxes, with each component accounting for one biochemical reaction, multiplied by the stoichiometric matrix S , whose nonzero components s_{ij} indicate how many units of compound i is created ($s_{ij} > 0$) or consumed ($s_{ij} < 0$) in reaction j . The coefficients V_c and V_b are the virtual volumes of the compartments, and since we represent them as diagonal matrices, we may assign different volumes to different species, e.g., to correct uneven distribution of the compounds or to take into account the hemoglobin/myoglobin concentrations.

In this study, we are interested in flux balance analysis at steady state condition, i.e., $Q(t) = \text{constant}$, $C_{a,\text{tot}} = \text{constant}$ and $dC_c/dt = 0$, $dC_b/dt = 0$.

We write the steady state condition as the matrix equation

$$\underbrace{\begin{bmatrix} S & M & -M & 0 \\ 0 & I & -I & (Q/F)I \end{bmatrix}}_{=A} \underbrace{\begin{bmatrix} \Phi \\ \mathbf{J}_{b \rightarrow c} \\ \mathbf{J}_{c \rightarrow b} \\ C_b \end{bmatrix}}_{=u} = \underbrace{\begin{bmatrix} 0 \\ (Q/F)C_a \end{bmatrix}}_{=r},$$

or briefly,

$$Au=r, \quad A \in \mathbb{R}^{m \times n}, \quad u \in \mathbb{R}^n, \quad r \in \mathbb{R}^m. \quad (3)$$

Note that we have dropped the subindex “total” to simplify the notation. For later reference, we partition the matrix A as

$$A = \begin{bmatrix} S & M & -M & 0 \\ 0 & I & -I & (Q/F)I \end{bmatrix} = \begin{bmatrix} A_c \\ A_b \end{bmatrix} \in \begin{matrix} \mathbb{R}^{k \times n} \\ \mathbb{R}^{(m-k) \times n} \end{matrix}. \quad (4)$$

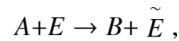
The steady state flux balance analysis is concerned with estimating the vector u when the vector r is known either exactly or approximately. Although the problem can be stated formally as a

linear algebraic equation, the matrix does not have full rank and the entries of the solution need to obey a set of constraints which will be discussed later.

2.2 Parametric model: Michaelis–Menten formulas

The net transport fluxes between the blood and tissue domains are related to the substrate uptake that can be estimated from blood concentrations, for which some estimates have been presented in the literature, see, e.g., Calvetti et al. (2006a) and references therein. Reaction fluxes, on the other hand, are harder to estimate and information concerning them is often based on *in silico* experiments.

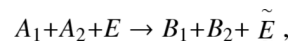
The reaction fluxes in our model are expressed in Michaelis–Menten form. If Φ represents the reaction flux of a single substrate facilitated reaction,



where A, B are metabolites and E, \tilde{E} are the facilitators and assuming that the reaction coefficients are unity for simplicity, we express the flux in the form

$$\Phi = V_{\max} \frac{P}{\mu + P} \frac{C_A}{K + C_A}, \quad P = \frac{C_E}{C_{\tilde{E}}}, \quad (5)$$

where V_{\max} is the maximum reaction velocity, the C 's are the corresponding concentrations and μ and K are reaction specific affinity coefficients. Similarly, for a facilitated bi-substrate reaction



we use the modified Michaelis-Menten form

$$\Phi = V_{\max} \frac{P}{\mu + P} \frac{C_{A_1} C_{A_2}}{K + C_{A_1} C_{A_2}}, \quad P = \frac{C_E}{C_{\tilde{E}}}.$$

If the reaction is not facilitated, the factor $P/(\mu + P)$ is set to unity. The transport fluxes can be modeled similarly. The rate of a carrier facilitated transport of substrate A from compartment x to compartment y is expressed on the form

$$J_{A,x \rightarrow y} = T \frac{C_{A,x}}{M + C_{A,x}}, \quad (6)$$

and for passive, diffusive transport, the rate is expressed as

$$J_{A,x \rightarrow y} = \gamma C_{A,x}.$$

These approximations were used earlier in the dynamic context, see Calvetti & Somersalo (2006); Calvetti et al. (2006b); Salem et al. (2004); Zhou et al. (2005). For a justification of such approximations, we refer to the textbooks Keener & Sneyd (1998); Marangoni (2003). If we stack all the model parameters V_{\max} , K , μ , T , M and γ in a long vector that is denoted by θ , we can express the mapping \mathbf{F} from these parameters to the flux vectors as

$$\mathbf{u} = \mathbf{F}(\theta, \mathbf{C}), \quad (7)$$

where \mathbf{C} is a vector of the cell and blood concentrations of all the species.

2.3 Constraints

The components of the vector \mathbf{u} need to satisfy a set of constraints. The most natural constraint coming from the assumptions in our model formulation, that the fluxes and concentrations are non-negative, $\mathbf{u} \geq 0$, already complicates the flux balance analysis, since it may exclude the pseudo-inverse solution of (3). Furthermore, since nonnegativity alone is not enough to specify a meaningful solution as it does not exclude the trivial solution where $\mathbf{C}_b = \mathbf{C}_a$ and all fluxes vanish, we need to impose strictly positive lower bounds for some vital fluxes. Also, the reaction fluxes have natural upper bounds, which can be seen in the expressions of Michaelis–Menten forms: equations (5) and (6) imply that $\Phi \leq V_{\max}$ and $J_{A,x \rightarrow y} \leq T$, respectively. We can express the bound constraints in vectorial form as $\mathbf{u}_{\min} \leq \mathbf{u} \leq \mathbf{u}_{\max}$. We remark that finding physiologically meaningful bounds is not a trivial issue, and that too strict and biasing bounds may result in misleading results.

In addition to the bound constraints described above, in general additional constraints on the fluxes arise from the requirement that the metabolic system respects the Second Law of Thermodynamics (Beard et al (2002); Nelson & Cox (2005); Siesjö (1978)). Checking the thermodynamic constraints requires that estimates of the concentrations of the substrates participating in reversible reactions are available. Since these concentrations are not estimated in the present work, we will not consider the thermodynamic constraints.

2.4 Optimality

The steady state condition (3) together with bound constraints is, in general, not enough to uniquely identify a steady state, and additional conditions are needed to single out the physiologically feasible steady state. In the literature, different criteria for optimality have been proposed to identify particular steady states (see, e.g., Bonarius et al. (1997); Ramakrishna et al. (2001); Varma & Palsson (1996)). In this article, following the optimality condition proposed by Ramakrishna et al. (2001), we assume that the metabolic chain is driven by the energetic efficiency principle: the system prefers a steady state where the mitochondrial ATP production via oxidative phosphorylation is at a maximum. Thus, we may pose the steady state problem as follows: *find the solution \mathbf{u} that satisfies the steady state condition (3), obeys a priori constraints and maximizes the reaction flux of the oxidative phosphorylation reaction.*

3 Constrained Optimization Approach

In this section, we study the flux balance analysis from the classical optimization point of view, considering the optimality condition introduced in Section 2.4. If u_j is the reaction flux of oxidative phosphorylation Φ_{21} , (see Appendix A, Table 2), the objective function h to be maximized is

$$h(\mathbf{u}) = \mathbf{e}_j^T \mathbf{u}, \quad \mathbf{e}_j = [0 \dots 0 \underbrace{1}_j 0 \dots 0]^T, \quad j=21, \quad (8)$$

and we seek to maximize the objective function $h(\mathbf{u})$ among those vectors \mathbf{u} satisfying

$$\mathbf{A}\mathbf{u} = \mathbf{r}, \quad \mathbf{u}_{\min} \leq \mathbf{u} \leq \mathbf{u}_{\max}. \quad (9)$$

The solution of this linear constrained optimization problem can be computed by *Linear Programming* (LP), a methodology that is briefly reviewed in the following section.

3.1 Linear Programming Solution

The Linear Programming problem can be formulated in its standard form as follows. Given $B \in \mathbb{R}^{m \times n}$, $z \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, find the vector $x \in \mathbb{R}^n$, which maximizes $c^T x$ subject to constraints

$$Bx \leq z, \quad (10)$$

$$x \geq 0,$$

see, e.g., Vanderbei (1996). The steady state flux estimation problem with the objective function (8) and the constraints (9) can be expressed as an LP problem of this form by using standard techniques. The details of the transformation to standard form can be found in the literature (see, e.g., Vanderbei (1996)) and therefore are omitted here. Numerous algorithms for the solution of the LP problems have been proposed in the literature, including interior point methods and the classical simplex algorithm. The flux balance problem has been solved by linear programming, see, e.g., Kauffman et al. (2003). In the present work, the LP solution will be used as a benchmark to which we compare to the results of the Bayesian FBA proposed in this article. In general, the data defining the bounds in LP problems are assumed to be exact. This is not necessarily the case in the FBA application that we consider here, in view of the natural fluctuations of the measurements over a population and the limitations of the collection and measurements procedure. Therefore, we will start with analyzing the stability of the LP solution with respect to uncertainties in the data, in our application consisting of the arterial concentration values. This sensitivity analysis will be a precursor to more extensive FBA that follows.

The sensitivity analysis of the LP solution contains two aspects: the sensitivity of the optimal solution to perturbations in the right-hand side of (10), and the possibility of having several optimal solutions without changing the value of the objective function. In this work, we are mostly interested in the former question.

Before computed examples, we conclude this section by mentioning that the LP approach is closely related to the *extreme pathway analysis*. In the past decade, a significant amount of research has been done to understand and describe a metabolic system using extreme pathway analysis, see Papin et al. (2002b, 2003, 2004); Schilling et al. (2000); Schilling & Palsson (1998). Similar to the flux balance analysis, the extreme pathways analysis imposes a steady state condition, but instead of a single optimal solution, it produces a set of feasible solutions in finite dimensional space. This is done by finding a system of basis pathways, called *extreme pathways*, that is unique for the metabolic network under consideration and which fully describes its properties. All possible steady-state flux solutions can then be represented as nonnegative linear combinations of these extreme pathways. Further analysis using a number of tools, such as SVD and α -spectrum analyses, is performed, see Price et al. (2003a); Wiback et al. (2003) for details.

A number of metabolic models, e.g., *Haemophilus influenzae* (Schilling & Palsson, 2000), the human red blood cell (Wiback & Palsson, 2002), *Escherichia coli* (Wiback et al., 2004), have been investigated with extreme pathway analysis. Since the number of extreme pathways can be very large for large-scale metabolic network systems (Papin et al., 2002a, 2003), a set of improved tools was developed to solve these problems (Barrett et al., 2006; Price et al., 2003b; Wiback et al., 2003).

3.2 Computed examples

In this section, we solve the steady state flux estimation problem for the skeletal muscle model (1)-(2) using the LP approach with the objective function (8) under the constraints (9). Note

that, in general, several simultaneous objective functions can be considered, see, e.g., Vo et al. (2004).

To demonstrate the sensitivity of the LP solution to the upper and lower bounds, we calculate the solution with two different sets of lower bounds \mathbf{u}_{\min} and upper bounds \mathbf{u}_{\max} . The values of bound vectors are listed in Appendix B, Table 4. Figure 1 displays these two sets of bounds and the corresponding LP solutions. A more detailed discussion of the bounds, and in particular, their relation to bounds concerning the concentrations of the metabolites as well as model parameters will be discussed later in this work. At this stage, the bounds are applied without questioning how they were obtained. Figure 1 clearly shows that the LP solution depends on the boundary constraints. We remark that the two solutions obtained by using the two different sets of bounds yield different values for the objective function; more specifically, $h(\mathbf{u}) = 3.9377$ for the wider bounds and $h(\mathbf{u}) = 1.2903$ for the tighter bounds.

In addition to the bounds for the entries of the solution vector, the LP solution depends on the input values, i.e., on the concentrations of the biochemical compounds in arterial blood, whose values, in turn, may be contaminated by measurement noise and fluctuate over a population. We model this uncertainties in the input by replacing the equation (3) by $\mathbf{r} = \mathbf{A}\mathbf{u} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^m$ is a noise vector. In order for (3) to hold in the mean value sense, we may assume that \mathbf{e} is a zero mean random vector. In our numerical experiments, we shall assume that \mathbf{e} is normally distributed with mutually independent components, $\mathbf{e} \sim N(0, \Gamma)$, with covariance matrix $\Gamma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$, where the diagonal element σ_j^2 is the variance of the j th component. Observe that the k first equations are related to the steady state condition in the cell domain. If we assume that the only uncertainties are in the input arterial concentrations, we must choose $\sigma_1^2 = \dots = \sigma_k^2 = 0$, hence the steady state condition $A_c \mathbf{u} = 0$ in the cell domain is strictly enforced.

To numerically investigate the stability of the LP solution, we generated a sample of 1000 normally distributed random realizations of the arterial values around the known mean value $\mathbf{r}_{b,\text{mean}} = (Q/F)\mathbf{C}_a$ with a given variance, and calculate the relative discrepancy of the corresponding LP solutions,

$$d_i = \frac{\|\mathbf{u}_i - \mathbf{u}_0\|}{\|\mathbf{u}_0\|}, \quad 1 \leq i \leq 1000,$$

where \mathbf{u}_0 is the LP solution corresponding to the mean value $\mathbf{r}_{b,\text{mean}}$, and \mathbf{u}_i is the LP solution corresponding to the i th realization of the noise vector. The left panel of Figure 2 shows the histogram of the relative discrepancies d_i when the standard deviation σ_j of the noise is 5% of the corresponding component of the noiseless vector $\mathbf{r}_{b,\text{mean}}$. In the calculation of the LP solutions, which was done by using the built-in Matlab function `linprog`, we used the wider bound constraints. For each noise level ranging from 0.5% to 20% with the step 0.5%, we computed the mean of the discrepancies, as displayed in the right panel of Figure 2. Clearly the discrepancy of the LP solutions increases in a linearlike fashion as a function of the noise level.

In some cases it was not possible for the LP solver to find a feasible solution within the given bounds, and the frequency of non-feasible problems increased with the noise level. The tighter the bound constraints are, the more easily the problem fails to have a solution. For example, when using the tighter bounds of Table 4, with 5% noise level in the right-hand side the LP solver could not find a feasible solution in 83% of the cases.

4 Bayesian approach

In this section, we propose a Bayesian alternative for the constrained optimization based FBA. The advantage of a statistical framework is that it is naturally suited for modelling the uncertainties and variabilities inherent to a metabolic steady-state.

We begin by giving a brief overview of the basic principles of Bayesian parameter estimation.

4.1 General framework

In the Bayesian framework, we model uncertainties by random variables and encode the available information in terms of probability distributions. If $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ denote two multivariate random variables whose joint probability density is denoted by $\pi(\mathbf{x}, \mathbf{y})$, their marginal probability densities are related to the joint density by

$$\pi(\mathbf{x}) = \int \pi(\mathbf{x}, \mathbf{y}) d\mathbf{y}, \quad \pi(\mathbf{y}) = \int \pi(\mathbf{x}, \mathbf{y}) d\mathbf{x}.$$

The marginal densities express the probability distributions of the values of \mathbf{x} and \mathbf{y} , respectively, provided that no information about the other random variable is available. If \mathbf{y} is a directly observable variable and \mathbf{x} is the variable of primary interest, we call the marginal density of \mathbf{x} the *prior density* and denote it by $\pi(\mathbf{x}) = \pi_{\text{prior}}(\mathbf{x})$. The prior density expresses the belief of the value of \mathbf{x} prior to any observations of \mathbf{y} .

The *conditional probability densities* $\pi(\mathbf{x} | \mathbf{y})$ and $\pi(\mathbf{y} | \mathbf{x})$ are defined via the formula

$$\pi(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y} | \mathbf{x}) \pi(\mathbf{x}) = \pi(\mathbf{x} | \mathbf{y}) \pi(\mathbf{y}).$$

The conditional density expresses the distribution of one variable provided that the other takes on a given value. The above formula yields immediately the Bayes formula,

$$\pi(\mathbf{x} | \mathbf{y}) = \frac{\pi(\mathbf{x}) \pi(\mathbf{y} | \mathbf{x})}{\pi(\mathbf{y})}. \quad (11)$$

In statistical parameter estimation, this formula is fundamental, because it expresses the probability density of the variable \mathbf{x} of primary interest given the observation of \mathbf{y} and all possible prior information concerning \mathbf{x} . The conditional density (11) is called the *posterior density*.

Since the posterior density is a probability density over a space whose dimension equals the number of components in the unknown \mathbf{x} it cannot be immediately visualized. Instead, it is used for computing statistical estimates. One of the most commonly used estimates is the *posterior mean*,

$$\mathbf{x}_{\text{pm}} = \int \mathbf{x} \pi(\mathbf{x} | \mathbf{y}) d\mathbf{x},$$

which is the expectation of \mathbf{x} with respect to the posterior probability density. The posterior mean is optimal in the sense that it is the estimator with the minimum estimation error variance.

The computation of the posterior mean from the posterior density requires an integration which, in high dimensional space, cannot be done by numerical quadratures. The most popular method for computing this integral and more generally, to explore the posterior probability density, is to use Markov Chain Monte Carlo techniques. The basic idea is to generate a Markov chain whose realizations are asymptotically distributed according to the posterior probability density. This is achieved by defining a rule q of random draw: given the current point \mathbf{x}_j , draw \mathbf{x}_{j+1}

from the distribution $t \mapsto q(t|x_j)$. Repeating the random draw with some initial point \mathbf{x}_0 , one generates a large sample, $S = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N\}$. If the drawing rule q is judiciously constructed, this ensemble is distributed asymptotically, as $N \rightarrow \infty$, according to $\pi(\mathbf{x} | \mathbf{y})$. The most commonly used MCMC strategies are the variants of the Metropolis-Hastings scheme and Gibbs sampler. In this work we use the latter one. To describe the drawing strategy, we need only to specify the move $\mathbf{x}_j \rightarrow \mathbf{x}_{j+1}$. Denoting the posterior density by $\pi_{\text{post}}(\mathbf{x}) = \pi(\mathbf{x} | \mathbf{y})$, the Gibbs sampler is described as follows. Given $\mathbf{x}_j = (x_{j,1}, x_{j,2}, \dots, x_{j,n-1}, x_{j,n})$,

draw $x_{j+1,1}$ from $t \mapsto \pi_{\text{post}}(t, x_{j,2}, \dots, x_{j,n-1}, x_{j,n})$,
 draw $x_{j+1,2}$ from $t \mapsto \pi_{\text{post}}(x_{j+1,1}, t, \dots, x_{j,n-1}, x_{j,n})$,
 \vdots
 draw $x_{j+1,n-1}$ from $t \mapsto \pi_{\text{post}}(x_{j+1,1}, x_{j+1,2}, \dots, t, x_{j,n})$,
 draw $x_{j+1,n}$ from $t \mapsto \pi_{\text{post}}(x_{j+1,1}, x_{j+1,2}, \dots, x_{j+1,n-1}, t)$.

In other words, we update one component at a time, conditioning on the old, or the currently updated values of the remaining components. Details of the updating scheme are discussed later.

When basing estimates on an MCMC chain, it is important to assess whether the chain is long enough to guarantee convergence of computed estimates.

Sample based estimates of the type

$$E\{g(\mathbf{x})\} = \int g(\mathbf{x})\pi_{\text{post}}(\mathbf{x})d\mathbf{x} \approx \frac{1}{N} \sum_{j=1}^N g(\mathbf{x}_j),$$

typically rely on convergence results derived from the Law of Large Numbers or from the Central Limit Theorem. The latter one asserts that if the samples \mathbf{x}_j are independent and identically distributed, the convergence rate is $\sim 1/\sqrt{N}$. MCMC methods, however, do not necessarily yield independent realizations. Therefore, a complete analysis of the chain requires that we also estimate the correlation length. If the correlation length of the sampled values $g(\mathbf{x}_j)$ is k , every k th sample is approximately independent and we may assume that the rate of convergence is $\sim 1/\sqrt{N/k}$. For large k , the rate may be painfully slow; a slow convergence is seen also by looking at the sample history: rather than resembling a realization of white noise, the sample value keeps drifting.

To obtain an estimate for the correlation length, denote by $g_c(\mathbf{x}_j)$ the centered sample point,

$$g_c(\mathbf{x}_j) = g(\mathbf{x}_j) - \frac{1}{N} \sum_{i=1}^N g(\mathbf{x}_i),$$

and define the normalized autocovariance of the sample by the formula

$$\gamma_\ell^2 = \frac{1}{\|g_c\|^2(N-\ell)} \sum_{j=1}^{N-\ell} g_c(\mathbf{x}_j)g_c(\mathbf{x}_{j+\ell}), \quad \ell \geq 0,$$

where $\|g_c\|^2 = \frac{1}{N} \sum_{j=1}^N g_c(\mathbf{x}_j)^2$.

The correlation length of the sample $\{g(\mathbf{x}_j)\}$ is then defined as the smallest integer for which γ_ℓ^2 drops below a given threshold.

4.2 Setting up the statistical model

The Bayesian analysis consists of three key steps: (1) setting up the prior density that reflects our belief of the unknowns of primary interest, (2) definition of the likelihood function that links the unknown to any possible observed quantity, and, (3) exploration of the posterior density, including the computation of statistical estimates. In this section, we discuss the first two steps.

4.2.1 Likelihood—We start the discussion by defining the likelihood function. In the present context, the vector \mathbf{u} represents the unknown of primary interest, and it is modeled as a random variable to reflect our lack of information of its value. The equation (3) is the starting point for observation equation, i.e., the vector \mathbf{r} is interpreted as observation. When discussing the LP solutions, we already indicated that the vector \mathbf{r} may be poorly known. We may not be sure about how strictly steady state condition holds, nor may we assume that the arterial concentration values are known exactly. In fact, it is quite reasonable to assume that concentrations of biochemical compounds fluctuate over a population.

Therefore, we write a stochastic extension of the equation (3), namely

$$\mathbf{r} = \mathbf{A}\mathbf{u} + \mathbf{e}, \quad \mathbf{e} \sim \pi_{\text{noise}}(\mathbf{e}), \quad (12)$$

where the probability density π_{noise} reflects our degree of uncertainty about the data and strictness of steady state. Here, we model the uncertainty \mathbf{e} as noise. From now on, we shall assume for simplicity that the vectors \mathbf{e} and \mathbf{u} are mutually independent, thus leading to a likelihood model

$$\pi(\mathbf{r}|\mathbf{u}) = \pi_{\text{noise}}(\mathbf{r} - \mathbf{A}\mathbf{u}).$$

To further simplify the discussion, we shall assume that \mathbf{e} is zero mean Gaussian with covariance matrix Γ , yielding a likelihood function of the form

$$\pi(\mathbf{r}|\mathbf{u}) \propto \exp\left(-\frac{1}{2}(\mathbf{r} - \mathbf{A}\mathbf{u})^T \Gamma^{-1}(\mathbf{r} - \mathbf{A}\mathbf{u})\right). \quad (13)$$

Observe that here we are implicitly assuming that Γ is symmetric and positive definite. However, if we are confident that the steady state condition holds strictly and that the only uncertainty is in the arterial concentration values, the matrix Γ becomes singular. To cope with this situation, the model needs to be reduced so as to ensure that the steady state is strictly enforced. We shall discuss the details of such model reduction later in this section.

4.2.2 Prior—The prior probability density should contain all available information that we may have of the primary unknown before observations. In the application at hand, the prior contains our belief concerning the bounds and the belief that the cell metabolism is driven by the principle of maximizing ATP production.

The bound constraints $\mathbf{u}_{\min} \leq \mathbf{u} \leq \mathbf{u}_{\max}$ can be encoded into the prior by requiring that

$$\pi_{\text{prior}} \propto \prod_{j=1}^n \chi_{[u_{\min,j}, u_{\max,j}]}(u_j), \quad (14)$$

where $\chi_{[u_{\min,j}, u_{\max,j}]}$ denotes the characteristic function of the admissible interval $[u_{\min,j}, u_{\max,j}]$ of the j th component. Observe that the form of the prior indicates that we are not expecting *a priori* any interdependence between the components of \mathbf{u} .

Since in the Bayesian framework we are not interested in finding a single estimator, the principle of maximizing the ATP production cannot be enforced as an optimization objective. Rather, according to the general Bayesian philosophy, we take the energy principle as a prior belief and seek to construct a probability density that favors solutions with high oxidative phosphorylation flux, i.e., we let $\mathbf{u} \sim \pi_{\text{phos}}(\mathbf{u})$, where the density π_{phos} is concentrated on energetically correct solutions. In particular, we choose the density to be

$$\pi_{\text{phos}}(\mathbf{u}) \propto \left(\frac{u_j - u_{\min,j}}{u_{\max,j} - u_{\min,j}} \right)^2, \quad (15)$$

where u_j is the reaction flux corresponding to oxidative phosphorylation. This density attains its maximum at $u_j = u_{\max,j}$ and minimum at $u_j = u_{\min,j}$, so it favors high ATP production.

4.3 Exploring the posterior

In this section, we explore the posterior distribution of the fluxes by generating representative samples via Gibbs samplers. Once the samples are available, we analyze the histograms, compute the posterior mean and compare it with the corresponding LP solution.

We first consider the case of a uniform prior (14), without including the part that favors high aerobic ATP production.

Traditionally, in FBA the steady state is strictly enforced, i.e., the concentrations of the biochemical species are assumed constant. This means that in the likelihood model above the matrix Γ is singular and therefore the model must be reduced. More precisely, consider the matrix A in the partitioned form (4). Strictly enforcing steady state implies, in particular, that

$$A_c \mathbf{u} = 0, \quad (16)$$

i.e., \mathbf{u} must be in the null space of A_c .

To find a basis for the null space of A_c , we introduce its singular value decomposition $A_c = U_c D_c V_c^T \in \mathbb{R}^{k \times n}$, where $U_c \in \mathbb{R}^{k \times k}$ and $V_c \in \mathbb{R}^{n \times n}$ are orthonormal matrices and $D_c \in \mathbb{R}^{k \times n}$ is a diagonal matrix whose nonnegative diagonal entries, $d_{c,1} \geq d_{c,2} \geq \dots \geq d_{c,k} \geq 0$ are called singular values (Golub & Van Loan, 1989). Assume that $d_{c,r}$, $r \leq k$ is the last numerically nonzero singular value of A_c . Partitioning the matrix V_c accordingly,

$$V_c = \begin{bmatrix} v_{c,1} & \dots & v_{c,r} & \vdots & v_{c,r+1} & \dots & v_{c,n} \end{bmatrix} = \begin{bmatrix} V_{c,1} & V_{c,2} \end{bmatrix},$$

and observing that the columns of $V_{c,2} \in \mathbb{R}^{n \times (n-r)}$ form an orthonormal basis for the null space of A_c , in order for (16) to hold, \mathbf{u} must be of the form

$$\mathbf{u} = V_{c,2} \mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^\ell, \quad \ell = n - r.$$

This representation reduces the dimension of the unknown vector, since we may now write the likelihood model for the ℓ -dimensional vector \mathbf{z} .

Consider now the observation model where we assume that the arterial concentrations are corrupted by noise,

$$\mathbf{r}_b = A_b \mathbf{u} + \mathbf{e}_b = A_b V_{c,2} \mathbf{z} + \mathbf{e}_b.$$

Here $\mathbf{r}_b = (Q/F)C_a \in \mathbb{R}^{m-k}$ is the observed vector, and $\mathbf{e}_b \in \mathbb{R}^{m-k}$ is the noise vector which we assume normally distributed, with zero mean and covariance matrix $\Gamma_b \in \mathbb{R}^{(m-k) \times (m-k)}$. The reduced likelihood model for the new variable \mathbf{z} is then

$$\pi(\mathbf{r}_b|\mathbf{z}) \propto \exp\left(-\frac{1}{2}(\mathbf{A}_b V_{c,2} \mathbf{z} - \mathbf{r}_b)^T \Gamma_b^{-1} (\mathbf{A}_b V_{c,2} \mathbf{z} - \mathbf{r}_b)\right).$$

Since Γ_b^{-1} admits a symmetric factorization of the form $\Gamma_b^{-1} = R_b^T R_b$, we can write the likelihood function as

$$\pi(\mathbf{r}_b|\mathbf{z}) \propto \exp\left(-\frac{1}{2}\|R_b \mathbf{A}_b V_{c,2} \mathbf{z} - R_b \mathbf{r}_b\|^2\right).$$

If $m - k \leq \ell$, as is the case in the application at hand, even after the reduction the problem remains underdetermined. Replacing $R_b \mathbf{A}_b V_{c,2} \in \mathbb{R}^{(m-k) \times \ell}$ by its singular value decomposition

$$R_b \mathbf{A}_b V_{c,2} = U_b D_b V_b^T, \quad D_b = \text{diag}(d_{b,1}, d_{b,2}, \dots, d_{b,m-k}) \in \mathbb{R}^{(m-k) \times \ell},$$

and using the orthogonality of $U_b \in \mathbb{R}^{(m-k) \times (m-k)}$, it follows that

$$\begin{aligned} \pi(\mathbf{r}_b|\mathbf{z}) &\propto \exp\left(-\frac{1}{2}\|D_b V_b^T \mathbf{z} - U_b^T R_b \mathbf{r}_b\|^2\right) \\ &= \exp\left(-\frac{1}{2}\sum_{j=1}^{m-k} (d_{b,j} y_j - b_j)^2\right), \end{aligned} \quad (17)$$

where $\mathbf{b} = U_b^T R_b \mathbf{r}_b \in \mathbb{R}^{m-k}$ and $\mathbf{y} = V_b^T \mathbf{z} \in \mathbb{R}^\ell$. The original variable \mathbf{u} can be expressed in terms of \mathbf{y} as

$$\begin{aligned} \mathbf{u} &= \underbrace{V_{c,2} V_b}_{=V \in \mathbb{R}^{n \times \ell}} \mathbf{y} = V \mathbf{y}. \end{aligned}$$

In the remainder of this paper, we refer to this as the *minimal likelihood model*. Without any bound constraints on the components of \mathbf{u} , the y_j are independent Gaussian random variables, and the components y_j , with $j > m - k$, have infinite variance since the likelihood (17) does not restrict them in any way. It is with the introduction of the prior that we restrict their variance.

We now outline a componentwise Gibbs sampling algorithm for this model. Starting from an initial vector \mathbf{y}_0 that satisfies the componentwise bounds

$$\mathbf{u}_{\min} \leq V \mathbf{y}_0 \leq \mathbf{u}_{\max},$$

we generate a sample $S = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_N\}$ through the following componentwise updating scheme. The initial value \mathbf{y}_0 can be found, e.g., by the LP solver, assuming that the system is consistent and a solution exists. Indeed, if \mathbf{u}_0 is the LP solution, then $\mathbf{A}_c \mathbf{u}_0 = 0$ and we may set $\mathbf{y}_0 = V^T \mathbf{u}_0$. Given the current vector \mathbf{y}_k in the chain, generate \mathbf{y}_{k+1} componentwise by drawing $y_{k+1,j}$, for $1 \leq j \leq \ell$, from the distribution

$$t \mapsto \pi_{\text{post}}(y_{k+1,1}, \dots, y_{k+1,j-1}, t, y_{k,j+1}, \dots, y_{k,\ell}).$$

Denote by $\mathbf{y}' \in \mathbb{R}^{\ell-1}$ and $V' \in \mathbb{R}^{\ell \times (\ell-1)}$ the arrays

$$\mathbf{y}' = \begin{bmatrix} y_{k+1,1} \\ \vdots \\ y_{k+1,j-1} \\ y_{k,j+1} \\ \vdots \\ y_{k,\ell} \end{bmatrix}, \quad \mathbf{V}' = \begin{bmatrix} v_1 & \dots & v_{j-1} & v_{j+1} & \dots & v_\ell \end{bmatrix}.$$

and find the maximal interval $[t_{\min}, t_{\max}]$ for which the system of inequalities

$$\mathbf{u}_{\min} - \mathbf{V}'\mathbf{y}'_k \leq t\mathbf{v}_j \leq \mathbf{u}_{\max} - \mathbf{V}'\mathbf{y}'_k$$

holds. For $t \in [t_{\min}, t_{\max}]$, the vector $\mathbf{v}(t) = t\mathbf{v}_j + \mathbf{V}'\mathbf{y}'$ has positive prior probability. Then, if $j \leq m - k$ and $d_{b,j} > 0$, draw $y_{k+1,j}$ from the truncated Gaussian distribution

$$t \mapsto \chi_{[t_{\min}, t_{\max}]} \exp\left(-\frac{1}{2}(d_{b,j}t - b_j)^2\right),$$

while if $d_{b,j} = 0$ or $j > m - k$, draw it from the uniform distribution over the interval $[t_{\min}, t_{\max}]$.

4.3.1 Relaxing the steady state condition—The strict enforcement of the steady state, although convenient from the computational complexity perspective, might not always be desirable or fully justified. In the present application, for example, it may not be known if the muscle is under constant energy demand during the experiment. Further, the initial value \mathbf{y}_0 for the reduced model may not be available, as we do not know if the strict steady state is consistent with the imposed bound constraints.

By not requiring that the steady state condition holds strictly, we no longer restrict the vector \mathbf{u} to the null space of A_c . As a result, in the additive noise model (12), the error covariance matrix Γ in the likelihood model is non-singular. Note that under these assumptions, in our application the components of the noise vector corresponding to the matrix A_b contain the uncertainty of the steady state plus possible measurement errors of the arterial concentrations. The Gibbs sampling can be applied to the whole model by relaxing the condition that the solution is in the null space of A_c . Introducing a symmetric decomposition for the inverse of the error covariance $\Gamma^{-1} = \mathbf{R}^T\mathbf{R}$, we can express the likelihood function (13) in the form

$$\pi(\mathbf{r}|\mathbf{u}) \propto \exp\left(-\frac{1}{2}\|\mathbf{R}\mathbf{r} - \mathbf{R}\mathbf{A}\mathbf{u}\|^2\right).$$

From the singular value decomposition $\mathbf{R}\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_m)$, by using the orthogonality of \mathbf{U} it follows that the likelihood function can be written in the form

$$\pi(\mathbf{r}|\mathbf{u}) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^m (b_j - d_j y_j)^2\right),$$

where $\mathbf{b} = \mathbf{U}^T\mathbf{R}\mathbf{r}$ and $\mathbf{y} = \mathbf{V}^T\mathbf{u}$. From here on, the componentwise Gibbs sampling with the bound constraints proceeds as before. In this case the dimensions of the problem are significantly larger than for the minimal model, thus requiring a greater computational load for the exploration of the posterior probability density via MCMC analysis.

Finally, we note that the Gibbs sampling of the full model can be done directly using the original coordinates \mathbf{u} without passing to the new coordinates using the SVD.

4.3.2 Implementation of energy principle: rejection sampling—So far, our prior has only contained the upper and lower bound information on the unknowns, ignoring the portion that favors high oxidative phosphorylation flux values. The inclusion of (15) into the prior density is straightforward, but the Gibbs sampler algorithm requires modifications, in particular because the sampling is done in the \mathbf{y} -space. One possibility is to implement the prior information sequentially: having computed the sample $\{\mathbf{u}_1, \dots, \mathbf{u}_N\}$ using the bound constraint prior, we perform a resampling: for each vector \mathbf{u}_j , we calculate the acceptance ratio, $\alpha = \pi_{\text{phos}}(\mathbf{u}_j)$, $0 \leq \alpha \leq 1$, and accept \mathbf{u}_j , with probability α , into the final sample. The resulting sample will be, in general, smaller and richer in vectors for which $\pi_{\text{phos}}(\mathbf{u})$ is higher, i.e., that comply better with the energy principle. The above strategy is equivalent to what is known in the literature as *rejection sampling*, see, e.g., Liu (2003).

4.3.3 Parametric prior distributions—One of the motivations for adopting a Bayesian perspective for the flux estimation problem is that it allows the inclusion of a priori beliefs about the distribution of the vector \mathbf{u} into the estimation. This prior belief may be based on preliminary studies, additional data sets or assumptions about the mathematical model.

For example, in Calvetti & Somersalo (2006) the probability density of the Michaelis-Menten parameters, collected in the model vector θ , was estimated by Bayesian sampling techniques using ischemic biopsy data from the literature (Katz (1988)). Since this information does not concern directly the fluxes at steady state, the question becomes then how to enter this available information about the Michaelis-Menten parameters in the form of a prior for the fluxes, which are the variables of primary interest in the present investigation. Observe that although in the cited article the steady state condition was included in the model before the onset of the ischemia, the flux balance was not enforced, and the steady state was assumed to be a resting steady state. Therefore, the FBA can provide complementary information of the system.

We propose the following straightforward procedure based on sampling. Consider the Michaelis-Menten model (7), and assume that we have a probability density $\tilde{\pi}(\theta, \mathbf{C})$ available for the model parameters and the steady state concentrations. We may draw a representative sample $\{(\theta_1, \mathbf{C}_1), \dots, (\theta_M, \mathbf{C}_M)\}$ from this distribution, and then calculate the corresponding flux sample, $\{\mathbf{u}_1, \dots, \mathbf{u}_M\}$, where $\mathbf{u}_i = \mathbf{F}(\theta_i, \mathbf{C}_i)$. The flux sample can then be used, e.g., to estimate upper and lower bounds for the fluxes or, more generally, to fit a given parametric model for the random variable \mathbf{u} . Since in the Gibbs sampling methods presented above the sampling is done for the variable \mathbf{y} , we can calculate the sample $\{\mathbf{y}_1, \dots, \mathbf{y}_M\}$ using the formula $\mathbf{y} = \mathbf{V}^T \mathbf{u}$, and use it to fit a parametric model for the random variable \mathbf{y} .

4.4 Results

This section presents the results of computed examples illustrating the Bayesian flux estimation method and compares them to the optimization based results discussed earlier. All computations were performed using Matlab.

4.4.1 Prior distributions and bounds—In order to set suitable bounds or parametric prior distributions for the components of the vector \mathbf{u} , we proceed as described in Section 4.3.3. In the article Calvetti & Somersalo (2006), a Markov Chain Monte Carlo sample of the vector $\mathbf{x} = (\mathbf{C}_0, \theta)$ was calculated based on biopsy data of Katz (1988). Although we could use directly this sample to calculate a sample of vector \mathbf{u} via the model (7), to avoid using a too committal prior, we follow a different approach. First, we assume that the components of the vector \mathbf{x} are mutually independent. Then, using the sample generated in Calvetti & Somersalo (2006), we

estimate a parametric distribution for each component separately, using a log-normal distribution for the Michaelis-Menten parameters, collected into the vector θ , and a uniform distribution for the initial concentrations over the interval $[(1/4)\bar{C}_{0,j}, 4\bar{C}_{0,j}]$, where $\bar{C}_{0,j}$ is the componentwise mean. Similarly, for the log-normal distribution, we first compute the mean $\bar{\theta}_j$ and then select a lognormal density which preserves $\bar{\theta}_j$ and is such that 97.5% of the sample is below $4\bar{\theta}_j$. Having the parametric distributions for the components of \mathbf{x} , we then generate a sample of 10 000 independently drawn vectors, calculate the corresponding flux vectors, and set the bounds for the fluxes so that 95% of the sample vectors are within the confidence interval.

We remark that fitting a log-normal prior to the sample in the maximum likelihood sense would lead to considerably tighter prior for the Michaelis-Menten parameters and for the flux vector \mathbf{u} . The narrower bounds used in the LP computations are a result of the latter approach.

Before discussing the results of the MCMC sampling, a few comments concerning the priors for different parameters are in order. The first one concerns committal and non-committal priors. For simplicity, consider a Michaelis-Menten flux model without facilitators,

$$\Phi = V_{\max} \frac{C}{K+C},$$

assuming that V_{\max} and K are fixed, and denote $\Phi = \Phi(C)$. In general, it is commonly believed that the wider the prior of the variable C is, the more uninformative and less committal it is. This is no longer obvious if the variable of primary interest is Φ instead of C . To understand why, denote by $\pi_C(C)$ the probability density of C . After a change of variables, the probability density π_Φ of Φ is

$$\pi_\Phi(\Phi) = \frac{1}{\Phi'(C)} \pi_C(C), \quad C = C(\Phi) = K \frac{\Phi}{V_{\max} - \Phi}.$$

Since the mapping $C \mapsto \Phi(C)$ saturates for large values of C , the derivative tends to zero and hence, the probability density of Φ is large for large values of C . Therefore, the wider (noninformative) the probability density of C is, the more confident we are that Φ is close to its saturation value as $C \rightarrow \infty$. This phenomenon is illustrated in Figure 3, where we have probability densities of Φ corresponding to uniform distributions of C over two different intervals.

Our second point is that the uniform priors for the components of \mathbf{u} could be replaced with other parametric priors. We found that, by using the Weibull distributions (see, e.g., Miller & Freund (1977)), we could get a relatively good fit to the histograms of the components of \mathbf{u} , and further, for histograms of the minimal model vector \mathbf{y} . However, since the use of different parametric priors had little effect on the final results, we will not discuss this issue any further.

4.4.2 Gibbs sampling computations—In our computed examples we applied Gibbs sampling MCMC for the analysis for the steady state to both the whole and the minimal model. For each model, we computed a sample of 200 000 elements.

The values for the arterial concentrations C_a were taken from Dash & Bassingthwaighe (2006) (see also Appendix A, Table 3). To set up the likelihood model, we assumed a standard deviation of 5% for the arterial concentration values. For the whole model, where we also needed to set uncertainties for the steady state condition, the standard deviation for the steady state conditions was chosen to be 1% of the value of the right-hand side of $A_b \mathbf{u}$ and $A_c \mathbf{u}$, respectively, calculated at the mean of the prior density. The initial point for the sampling was

calculated from the mean of the sample obtained in Calvetti & Somersalo (2006) (this sample was used also for the priors) using the Michaelis-Menten model (7).

We use the correlation length of the component sample as a measure of convergence. Thus, the correlation lengths are calculated using for g a function that picks a given component p from the vector \mathbf{x} , i.e., $g(\mathbf{x}) = x_p$ (see Section 4.1). In our calculations, we define the correlation length using a threshold value of 0.25, i.e., $\gamma_\ell < 0.5$. The calculation of the autocovariance is repeated up to the value $\ell=1000$.

We recall that for both the minimal and the whole model, the sampling is done for \mathbf{y} , while the variable of primary interest is \mathbf{u} . From our numerical simulations, we saw that while some of the components of \mathbf{y} converge fast, for others the convergence can be painfully slow. We also noticed that the correlation length of a \mathbf{y} component clearly depends on the value of the corresponding singular value of the system matrix, with large singular values displaying low correlation length, thus good convergence rate, of the coefficient of the corresponding singular vector. This phenomenon, which we have observed for both the minimal and the whole model, is illustrated in Figure 4.

The singular values for the whole model that are clearly distinct from zero are very markedly separated from the near vanishing ones, and it is in correspondence of the gap that the correlation length increases dramatically. This phenomenon is independent of the Gibbs sampling strategy in the sense that it is present whether we perform the updating starting from the component corresponding to highest or lowest singular value.

In order to get a measure of the reliability of the estimates of \mathbf{u} , we computed the correlation lengths for the components of \mathbf{u} . Table 5 in Appendix B lists the correlation lengths of the components of the Gibbs samples for the whole and minimal model, respectively.

In addition to the posterior mean estimates, the histograms of the components of \mathbf{u} from the samples $\{\mathbf{u}_i\}$, $i = 1, \dots, N$, contain valuable information on how well the respective quantities are defined by the system. Figure 5 shows the histograms of some of the components of the solution \mathbf{u} for the whole and minimal model. The posterior mean estimates and the corresponding LP solution are also shown. It should also be pointed out that different sampling strategies may lead into slightly different results.

The initial point of the sampling may also affect the results in particular for the components that have not yet converged. To investigate the effect of the initial point, we calculated the Gibbs samples using two other initial points. For both the whole model and minimal model the LP solution was used. For the whole model, also the midpoint between the bounds was used, for the minimal model this natural choice for a starting point would not work as the initial point has to lie in the null space of A_c . The results show that for the components for which the correlation length is small, two different initial points produce in practice the same histograms, whereas for poorly converged components (high correlation length) the effect of the initial point is clearly visible. Figure 6 demonstrates the effect of the initial point.

In our computed simulations we implemented the prior belief that the system favors steady states with a high oxidative phosphorylation reaction flux by rejection sampling, i.e., we resampled the existing sample $\{\mathbf{u}_1, \dots, \mathbf{u}_N\}$, corresponding to the uniform prior, favoring those \mathbf{u} for which $\pi_{\text{prior}}(\mathbf{u})$ is large. Figure 7 shows the histograms from a Gibbs sample generated using the minimal model sampler before and after rejection sampling. We observe that the optimality criterion, as implemented here, has a visible but not dramatic effect on the posterior distribution, shifting the histograms slightly towards the energetically more efficient (higher aerobic ATP production) direction.

5 Discussion and further studies

By looking at the histograms, it is clear that some of them have identical shapes. For instance, the histograms of the glycolytic reactions are all identical up to scaling, since at steady state the fluxes should indeed coincide. The same is true for the TCA cycle reaction fluxes. Mathematically, these fluxes depend linearly on the same underlying independent components, and it is possible to calculate correlation coefficients to find out the correlated fluxes. Note, that although the independent components are not necessarily equal to the singular vectors of the system matrix, since the coefficients of the singular vectors get correlated via the bound constraints in a non-trivial way, the singular vectors provide a good first approximation to the independent components.

The histograms of the reversible reactions, glycogen synthesis and utilization and phosphocreatine synthesis and breakdown, are identical and spread over the feasible region, indicating that the net flux vanishes while the one-way fluxes are not determined.

It is interesting to observe that the prior favoring high oxidative phosphorylation flux has a marginal effect on the means of the distributions. The LP solution, however, may differ dramatically from the mean, and it is often at a tail of the distributions. While it is not surprising that the optimization approach leads to a solution of high TCA fluxes, it is maybe less obvious that the glycolytic fluxes tend to the minimum in our computed example. This phenomenon may be due to the instability of the LP solution: from Figure 1, we observe that with the narrower bound constraints, the glycolytic fluxes ($j = 1, 4, 5, 6$) are closer to their maximum. This observation suggests that a distribution based estimation strategy, being less prone to instabilities, may be preferable to computing single estimates. Notice also that the LP solution requires that the steady state condition and the bound constraints are consistent, which may not be easy to check, in particular when bound constraints from different sources are used. The Bayesian approach is more flexible in the sense that since the steady state is not enforced strictly, inconsistency with the prior bounds leads only to low likelihoods.

The trade-off in using sampling methods is that sampling strategies are generally slower than methods that produce single point estimates such as LP solution. However, the obtained solution is also much more informative; the estimate of the probability density of the solution, rather than a single point estimate that may be rather unstable. To give more practical feeling on computational effort, the computing time of a sample of 200 000 elements was approximately 160 minutes for the whole model, which amounts to roughly 0.05 seconds per sample point, and 72 minutes for the minimal model, which amounts to roughly 0.02 seconds per sample point, on a PC with 2 GHz processor and 2 Gb of RAM. The computing time of a single LP solution at the same setup was of the order of a fraction of a second.

There are several directions where this analysis can be extended. The most straightforward one is the study of more complex models, including sub-compartments of different cell types and sub-domains within them. Another natural extension is the analysis of independent components, since they define subsystems in the metabolic networks that can be estimated separately. The study of conditional distributions is also of interest if we have partial information about the system available. Furthermore, a Bayesian FBA which take into account concentration data (Katz, 1988; Korth et al., 2000) as well as thermodynamic constraints, will be the topic of future work.

Acknowledgements

The research of Jenni Heino and Knarik Tunyan was supported by the Finnish Funding Agency for Technology and Innovation (Tekes), Modeling and Simulation Technology Program (MASI). The work of Daniela Calvetti was supported in part by NIH grant GM-66309 to establish the Center for Modeling Integrated Metabolic Systems at Case Western Reserve University.

Appendix A. Skeletal muscle model

This appendix gives the details of the two-compartment skeletal muscle model considered in this article. Table 1 lists the biochemical species included in the model with their abbreviations. All 30 species are present in the tissue domain, and the first 9 also in the blood domain. The 26 biochemical reactions implemented in the model are listed in Table 2. Table 3 specifies the type of transport flux (facilitated or passive) between blood and tissue for each species in the blood domain, and gives the numerical values (from Dash & Bassingthwaight (2006)) for the arterial blood concentrations. In our calculations, the units for the concentrations are mmol/l, for the reaction and transport fluxes mmol/min and for the blood flow l/min. The numerical values used for the blood flow $Q(t)$ and the mixing ratio F are 0.9 l/min and 2/3, respectively. Finally, Figure 8 presents a schematic diagram of the two-compartment skeletal muscle model.

Appendix B. Bounds and correlation lengths

Table 4 presents the numerical values for the bounds for the solution vector \mathbf{u} . The LP solution was calculated with both the set of tight bounds and the set of wide bounds. In the Bayesian statistical computations, the set of wide bounds was used. Table 5 presents the numerical values of the correlation lengths for the components of \mathbf{u} from the Gibbs sampling.

References

- Barrett CL, Price ND, Palsson BO. Network-level analysis of metabolic regulation in the human red blood cell using random sampling and singular value decomposition. *BMC Bioinformatics* 2006;7(132):10.1186/1471-2105-7-132
- Beard DA, Liang S, Quian H. Energy balance for analyzing complex metabolic networks. *Biophys J* 2002;83(1):79–86. [PubMed: 12080101]
- Bonarius HPJ, Schmid G, Tramper J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *TRENDS Biotechnol* 1997;15:308–314.
- Calvetti D, Dash RK, Somersalo E, Cabrera ME. Local regularization method applied to estimating oxygen consumption during muscle activities. *Inverse Problems* 2006a;22:229–243.10.1088/0266-5611/22/1/013
- Calvetti D, Hageman R, Somersalo E. Large-scale Bayesian parameter estimation for a three-compartment cardiac metabolism model during ischemia. *Inverse Problems* 2006b;22:1797–1816.10.1088/0266-5611/22/5/016
- Calvetti D, Somersalo E. Large-scale statistical parameter estimation in complex systems with an application to metabolic models. *Multiscale Modelling and Simulation* 2006;5(4):1333–1366.
- Dash RK, Bassingthwaight JB. Simultaneous blood-tissue exchange of oxygen, carbon dioxide, bicarbonate and hydrogen ion. *Ann Biomed Eng* 2006;34(7):1129–1148.10.1007/s10439-005-9066-4 [PubMed: 16775761]
- Gilks, WR.; Richardson, S.; Spiegelhalter, DJ. *Markov Chain Monte Carlo in Practice*. Chapman & Hall; 1996.
- Golub, GH.; Van Loan, CF. *Matrix Computations*. The Johns Hopkins University Press; 1989.
- Kaipio, J.; Somersalo, E. *Statistical and Computational Inverse Problems*. Springer-Verlag; 2004.
- Katz A. G-1,6-P₂, glycolysis, and energy metabolism during circulatory occlusion in human skeletal muscle. *Am J Physiol Cell Physiol* 1988;255(24):C140–C144.
- Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Current Opinion in Biotechnology* 2003;14:491–496.10.1016/j.copbio.2003.08.001 [PubMed: 14580578]
- Keener, J.; Sneyd, J. *Mathematical Physiology*. Springer-Verlag; 1998.
- Korth U, Merkel G, Fernandez FF, Jandewerth O, Dogan G, Koch T, van Ackern K, Weichel O, Klein J. Tourniquet-induced changes of energy metabolism in human skeletal muscle monitored by microdialysis. *Anesthesiology* 2000;93(6):1407–1412. [PubMed: 11149434]

- Lai N, Dash RK, Nasca MM, Saidel GM, Cabrera ME. Relating pulmonary oxygen uptake to muscle oxygen consumption at exercise onset: in vivo and in silico studies. *Eur J Appl Physiol* 2006;97:380–394.10.1007/s00421-006-0176-y [PubMed: 16636861]
- Liu, JS. Monte Carlo Strategies in Scientific Computing. Springer-Verlag; 2003.
- Marangoni, AG. Enzyme Kinetics. A Modern Approach. Wiley Interscience; Hoboken: 2003.
- Miller, I.; Freund, JE. Probability and Statistics for Engineers. 2nd. Prentice-Hall, Inc.; Englewood Cliffs, NJ: 1977.
- Nelson, DL.; Cox, MM. Lehninger Principles of Biochemistry. 4th. W.H. Freeman & Company; New York: 2005.
- Papin JA, Price ND, Edwards JS, Palsson BO. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J Theor Biol* 2002a;215(1):67–82. [PubMed: 12051985]
- Papin JA, Price ND, Palsson BO. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Research* 2002b;12:1889–1900.10.1101/gr.327702 [PubMed: 12466293]
- Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO. Metabolic pathways in the post-genome era. *TRENDS in Biochemical Sciences* 2003;28(5):250–258.10.1016/S0968-0004(03)00064-1 [PubMed: 12765837]
- Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO. Comparison of network-based pathway analysis methods. *TRENDS in Biotechnology* 2004;22(8):400–405. [PubMed: 15283984]
- Price ND, Reed JL, Papin JA, Famili I, Palsson BO. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophysical Journal* 2003a;84:794–804. [PubMed: 12547764]
- Price ND, Reed JL, Papin JA, Wiback SJ, Palsson BO. Network-based analysis of metabolic regulation in the human red blood cell. *J Theor Biol* 2003b;225:185–194.10.1016/S0022-5193(03)00237-6 [PubMed: 14575652]
- Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J of Physiol Regulatory, Integrative and Comp Physiol* 2001;280:R695–R704.
- Salem JE, Cabrera ME, Chandler MP, McElfresh TA, Huang H, Sterk JP, Stanley WC. Step and ramp induction of myocardial ischemia: comparison of *in vivo* and *in silico* results. *J Physiol Pharmacol* 2004;55(3):519–536. [PubMed: 15381824]
- Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 2000;203:229–248.10.1006/jtbi.2000.1073 [PubMed: 10716907]
- Schilling CH, Palsson BO. The underlying pathway structure of biochemical reaction networks. *Proc Natl Acad Sci USA* 1998;95:4193–4198. [PubMed: 9539712]
- Schilling CH, Palsson BO. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J Theor Biol* 2000;203:249–283.10.1006/jtbi.2000.1088 [PubMed: 10716908]
- Siesjö, BK. Brain Energy Metabolism. Wiley & Sons; 1978.
- Vanderbei, RJ. Linear Programming: Foundations and Extensions. Kluwer Academic Publishers; Boston: 1996.
- Varma A, Palsson BO. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology* 1994;12:994–998.10.1038/nbt1094-994
- Vo TD, Greenberg HJ, Palsson BO. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem* 2004;279(38):39532–39540.
- Wiback SJ, Mahadevan R, Palsson BO. Reconstructing metabolic flux vectors from extreme pathways: defining the α -spectrum. *J Theor Biol* 2003;224:313–324.10.1016/S0022-5193(03)00168-1 [PubMed: 12941590]
- Wiback SJ, Mahadevan R, Palsson BO. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The *Escherichia coli* spectrum. *Biotechnology and Bioengineering* 2004;86(3):317–331.10.1002/bit.20011 [PubMed: 15083512]

Wiback SJ, Palsson BO. Extreme pathway analysis of human red blood cell metabolism. *Biophysical Journal* 2002;83:808–818. [PubMed: 12124266]

Zhou L, Salem JE, Saidel GM, Stanley WC, Cabrera ME. Mechanistic model of cardiac energy metabolism predicts localization of glycolysis to cytosolic subdomain during ischemia. *Am J Physiol Heart Circ Physiol* 2005;288:2400–2411.10.1152/ajpheart.01030.2004

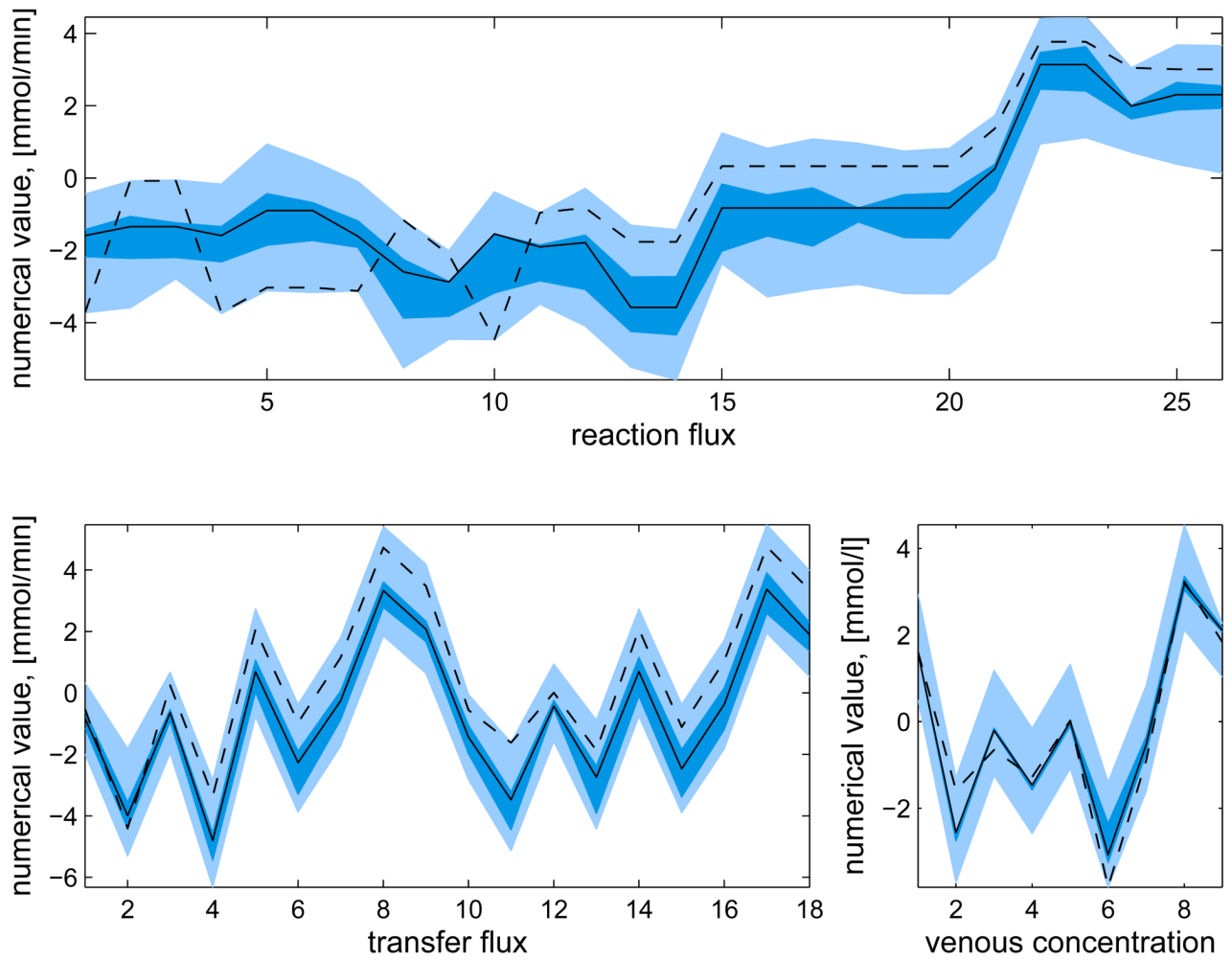
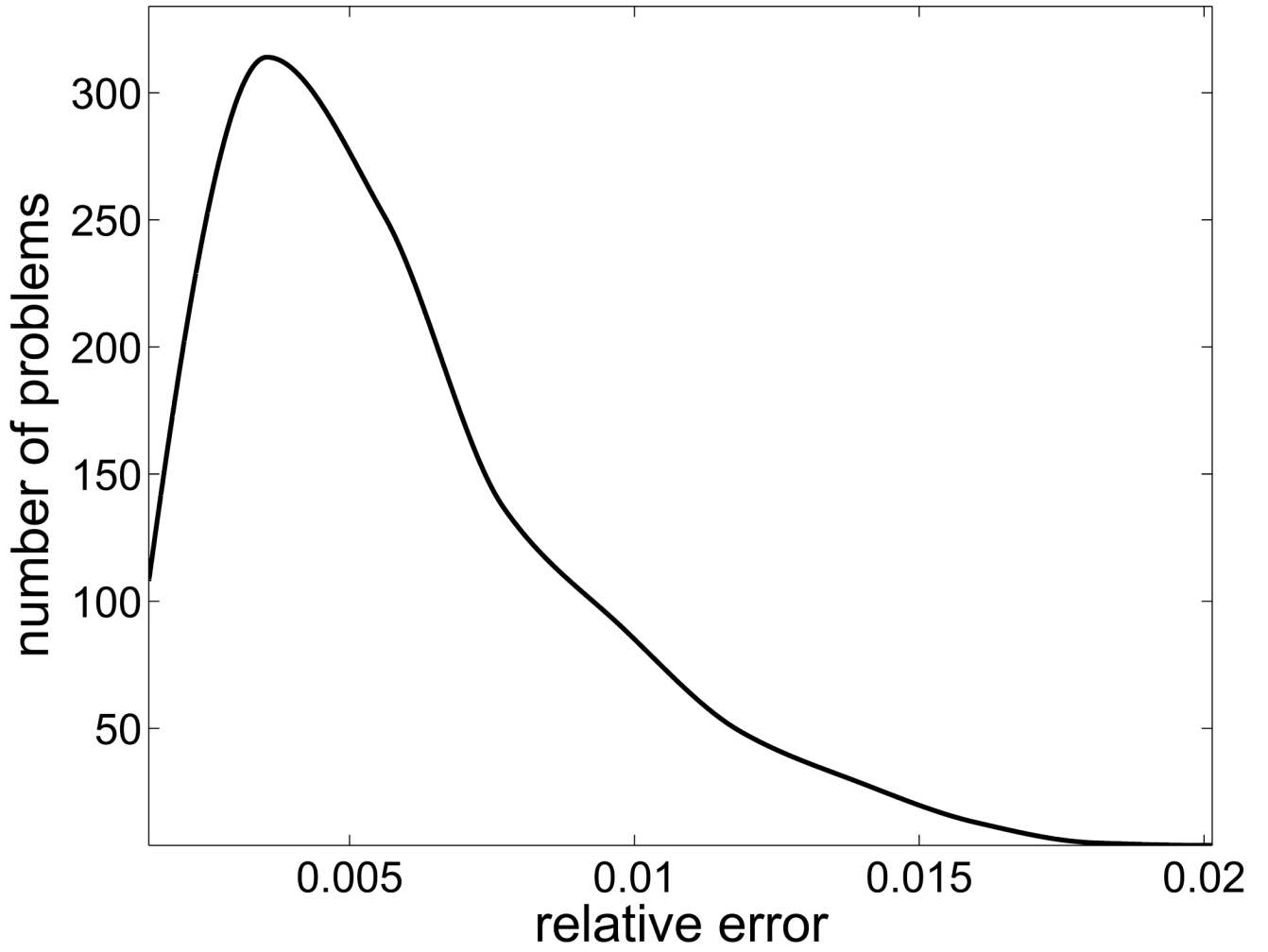


Fig. 1. Two sets of the bounds and the corresponding solutions. The dotted curve is the LP solution for the wider bounds, the solid line is the LP solution for the tighter bounds. The plots are in logarithmic scale.



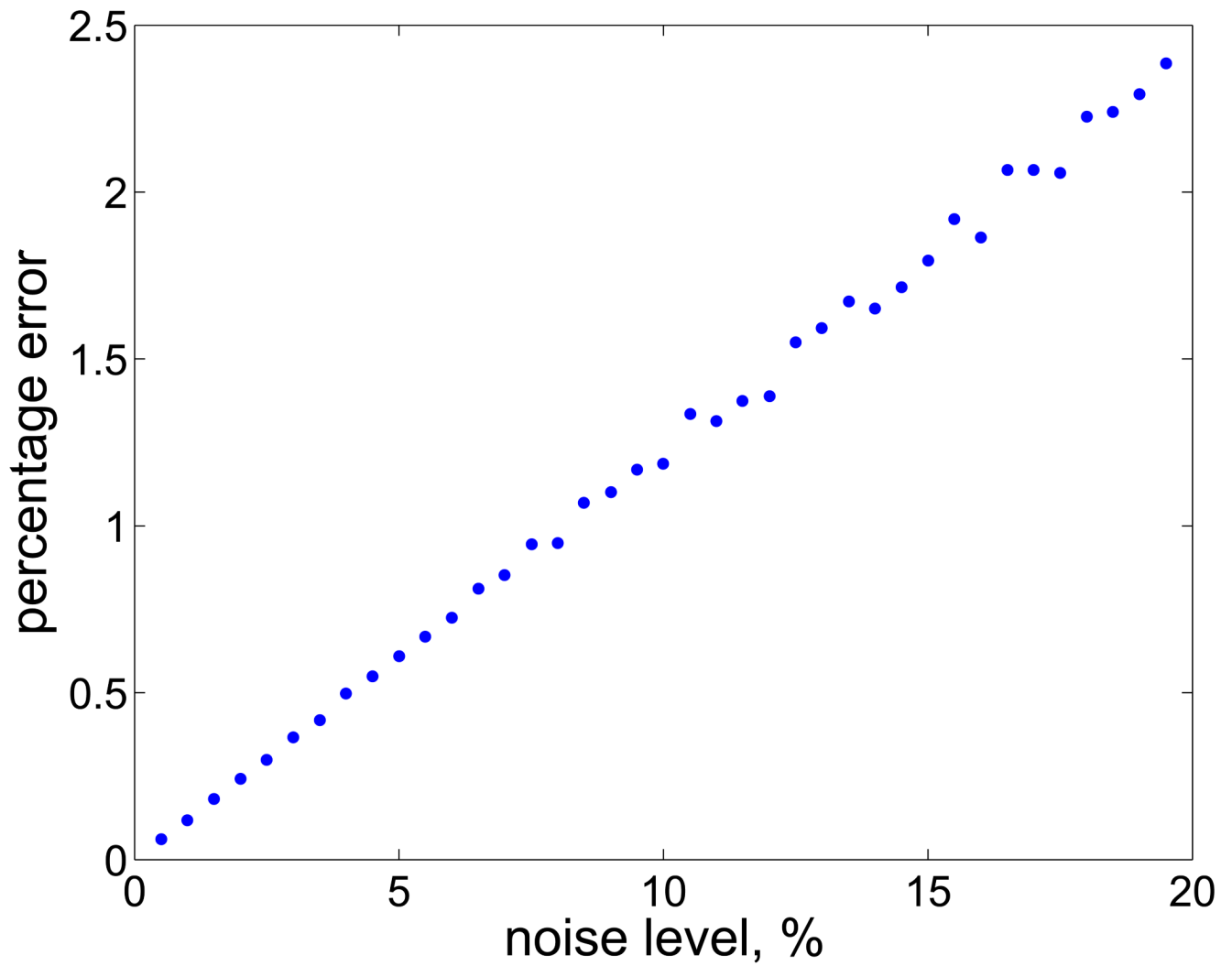


Fig. 2. Histogram of the discrepancies d_i with 5% noise level (left), and the dependency of the mean discrepancy on the noise level of the arterial concentration values (right).

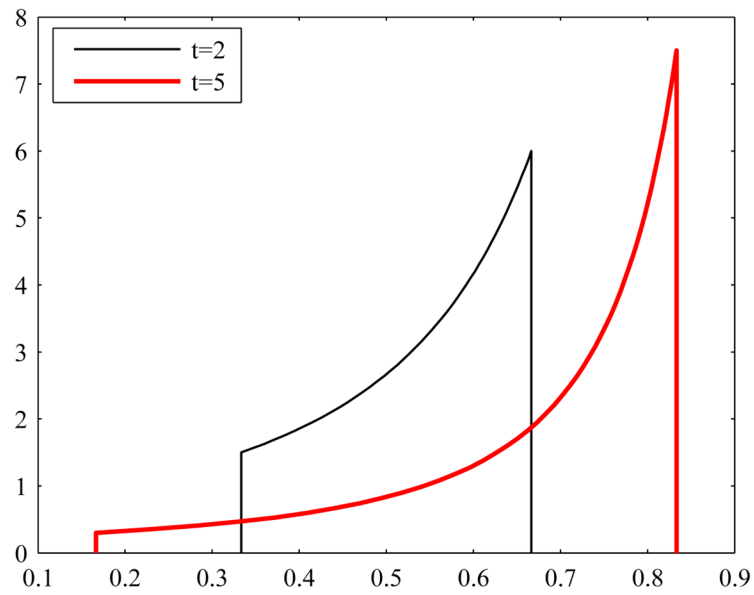


Fig. 3. Probability densities of Φ corresponding to uniform distributions of C over an interval $[1/t, t]$, with two different values of t . Here, $V_{\max} = 1$ and $K = 1$.

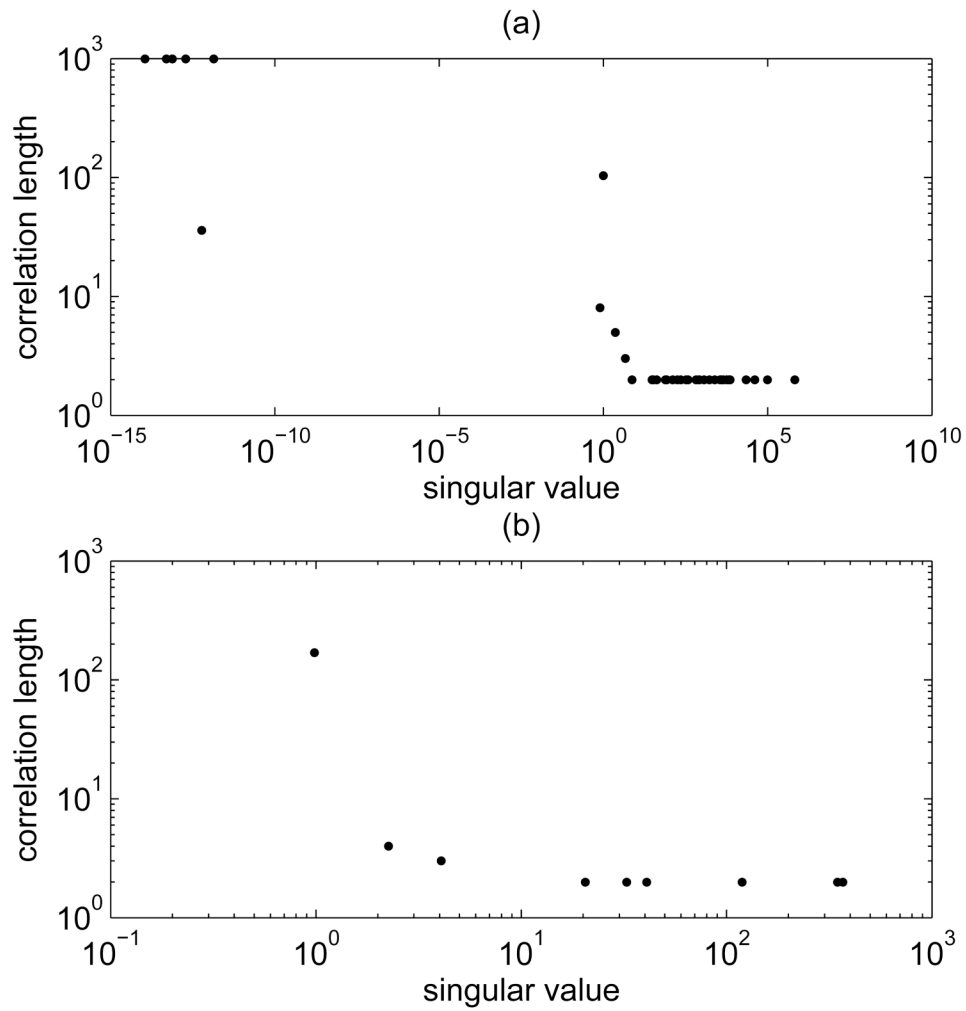


Fig. 4. Singular values and correlation lengths for (a) the whole model and (b) the minimal model sampler. The values are based on samples with 200 000 elements. The correlation length was calculated as explained in Section 4.1, using a value $\ell = 1000$ for the maximum value of the correlation length.

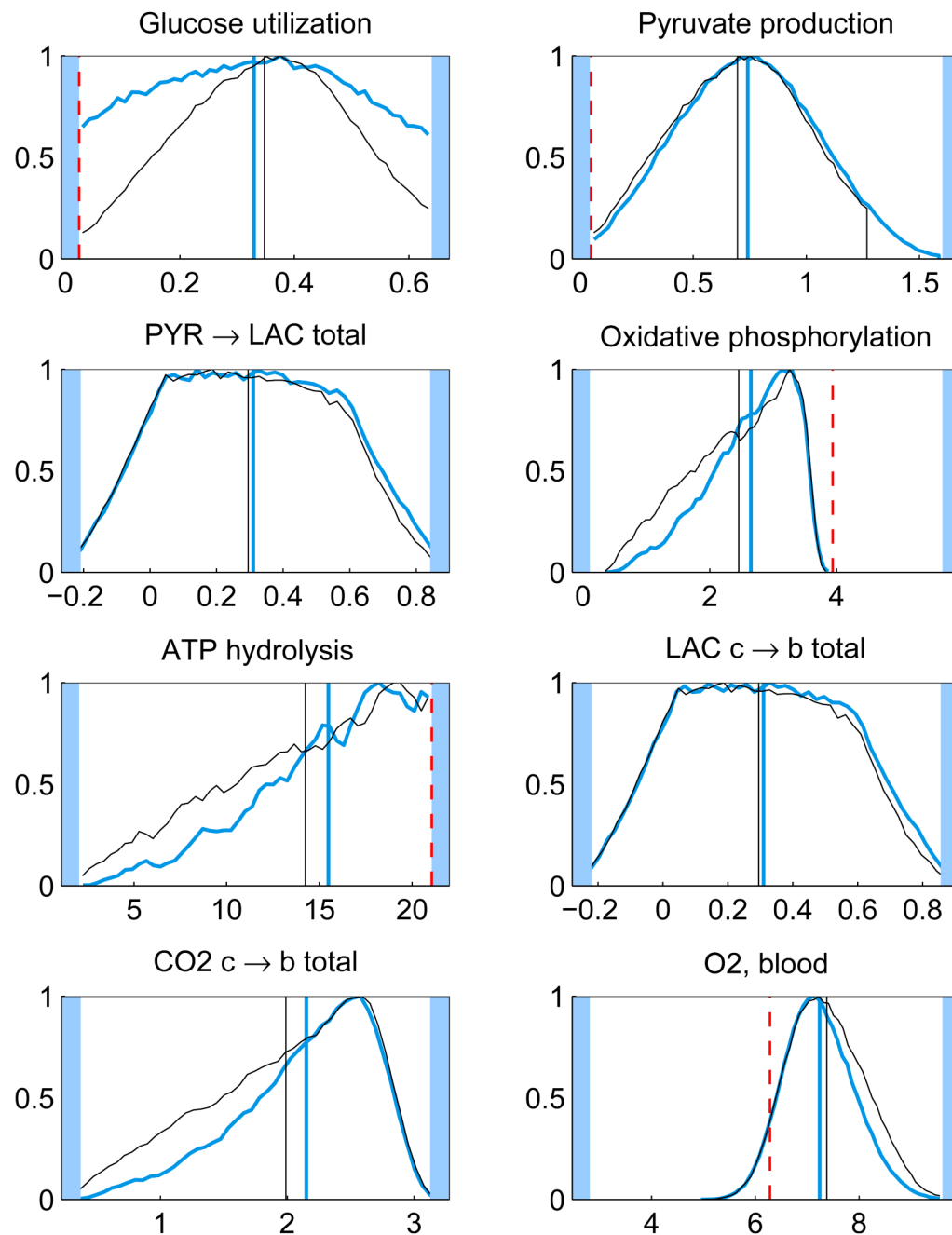


Fig. 5. The histograms and corresponding posterior mean estimates of some of the reaction and transport fluxes and the concentration of oxygen with Gibbs sampling using the whole model (thicker blue) and minimal model (thinner black). For reversible reactions and transport fluxes, the net flux is plotted. The maximum value of each histogram is normalized to unity. The units for the reaction and transport fluxes are mmol/min and for the concentrations mmol/l. The LP solution (dashed red) is also depicted (except for net fluxes). The highlighted area defines the bounds, or in case of net fluxes, the span of the histogram.

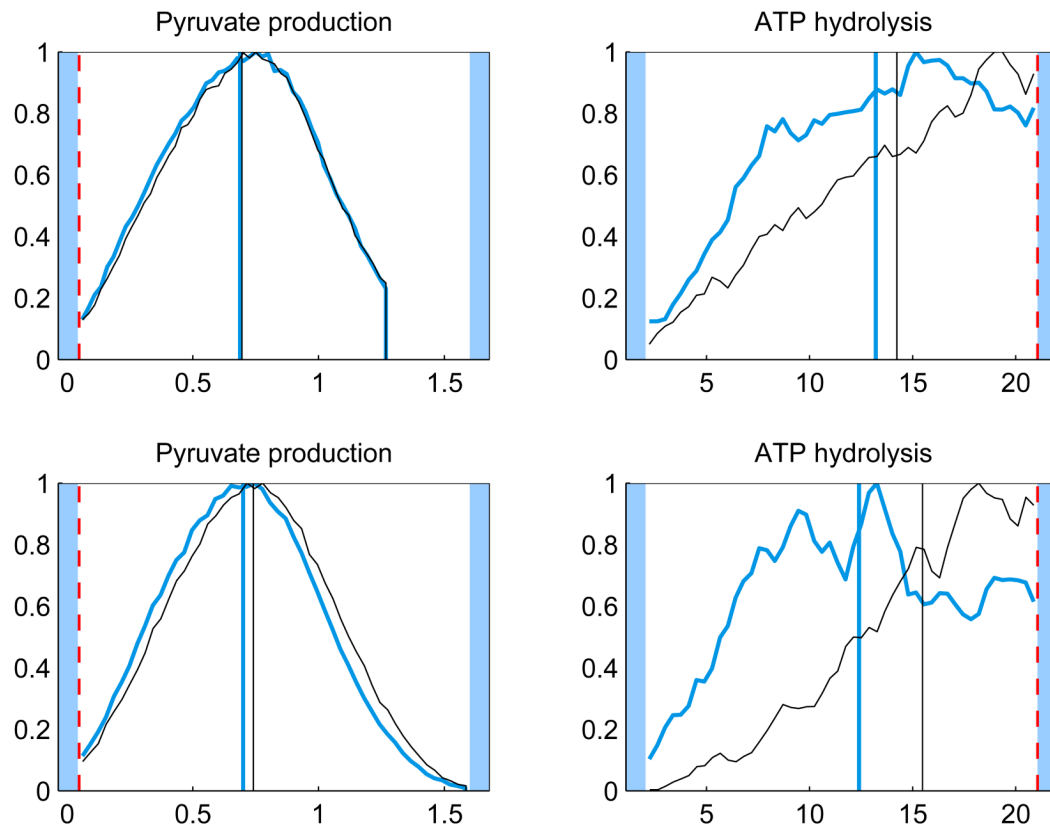
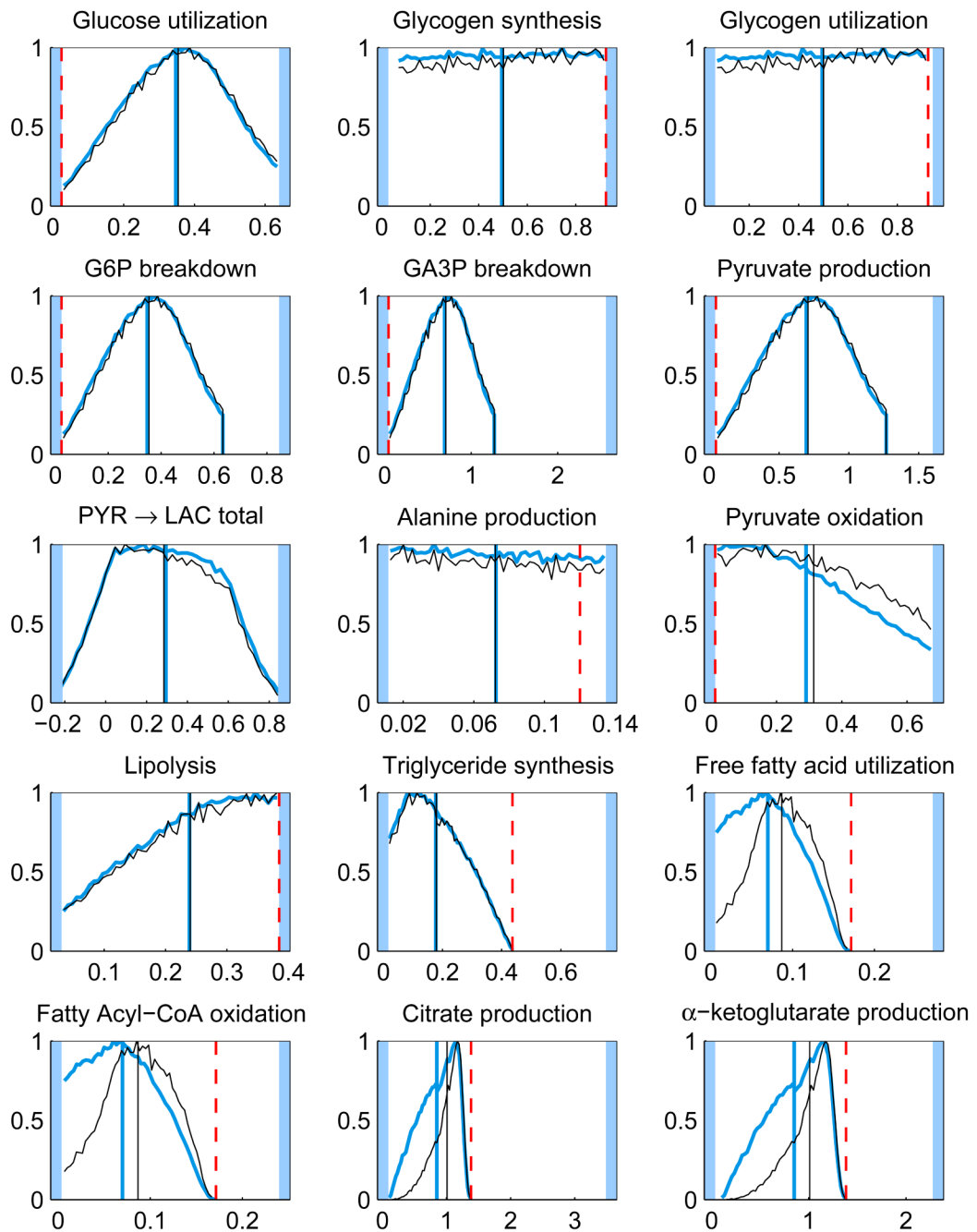
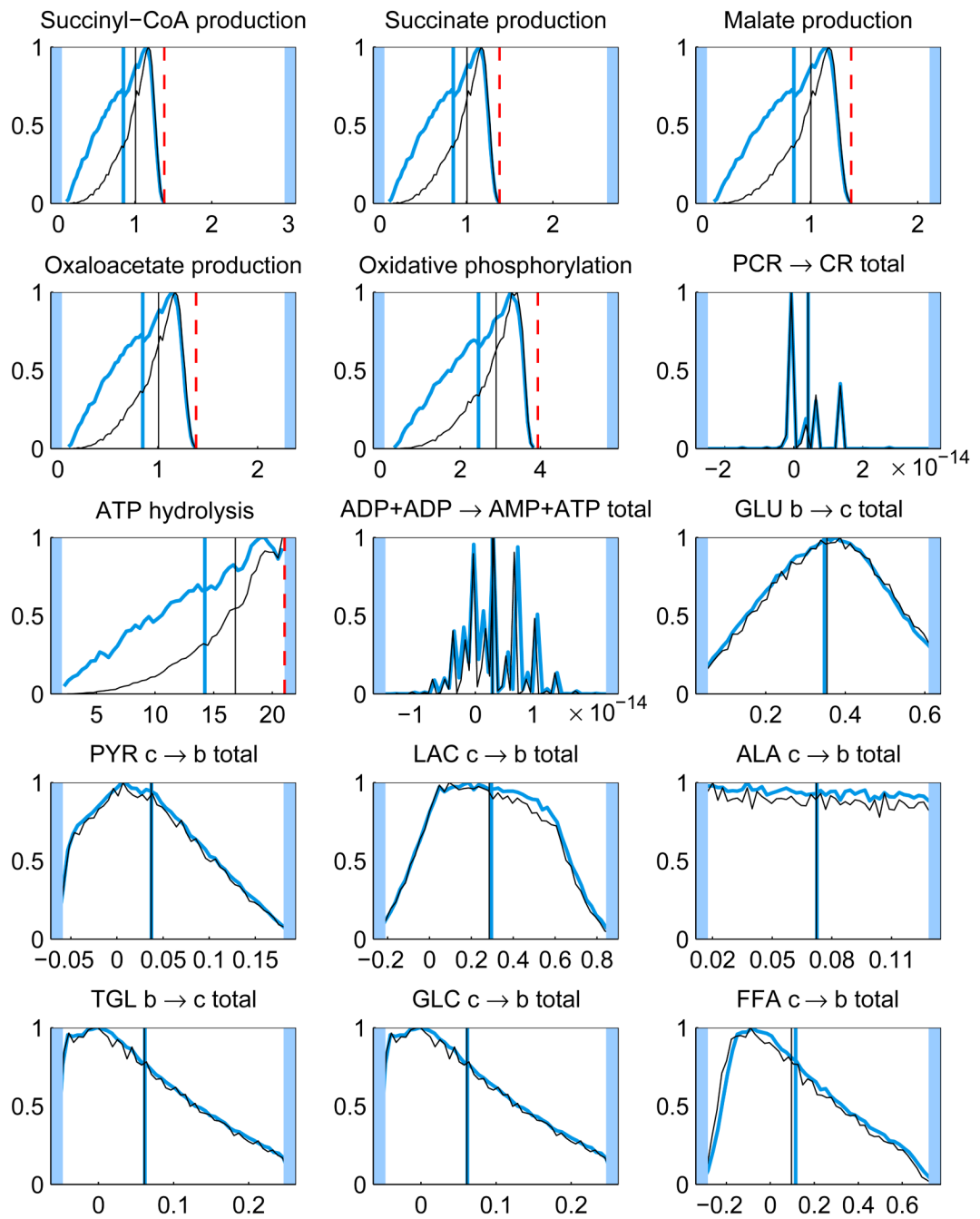


Fig. 6. The histograms of a well converged (left panels) and a poorly converged (right panels) component of \mathbf{u} using the minimal model (top) and whole model (bottom) for different starting points of the sampling. For the minimal model the starting points were the LP solution (thicker blue) and the mean of the sample used for priors (thinner black). For the whole model the starting points were the midpoint between bounds (thicker blue) and the mean of the sample used for priors (thinner black). The LP solution (dashed red) is also shown. The units for the reaction fluxes are mmol/min and the maximum value of each histogram is normalized to unity.





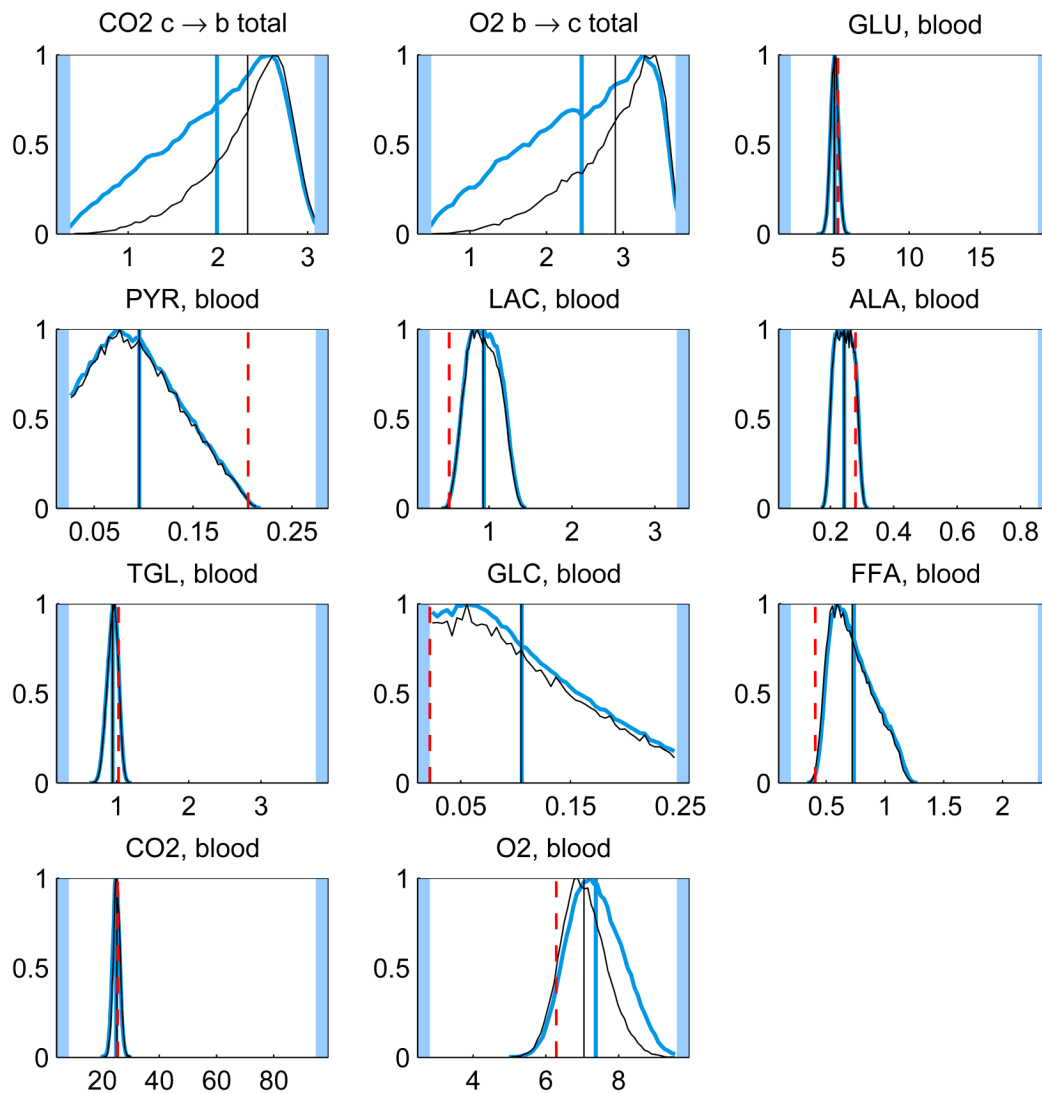


Fig. 7. The histograms and the corresponding CM estimates of all the components of the solution \mathbf{u} from Gibbs sampling using the minimal model before (thicker blue) and after (thinner black) applying rejection sampling. For reversible reactions and transport fluxes, the net flux is plotted. The maximum of each histogram is normalized to unity. The units for the reaction and transport fluxes are mmol/min and for the concentrations mmol/l. The LP solution (dashed red) is also depicted (except for net fluxes). The highlighted area in the plots defines the bounds, or in case of net fluxes, the span of the histogram.

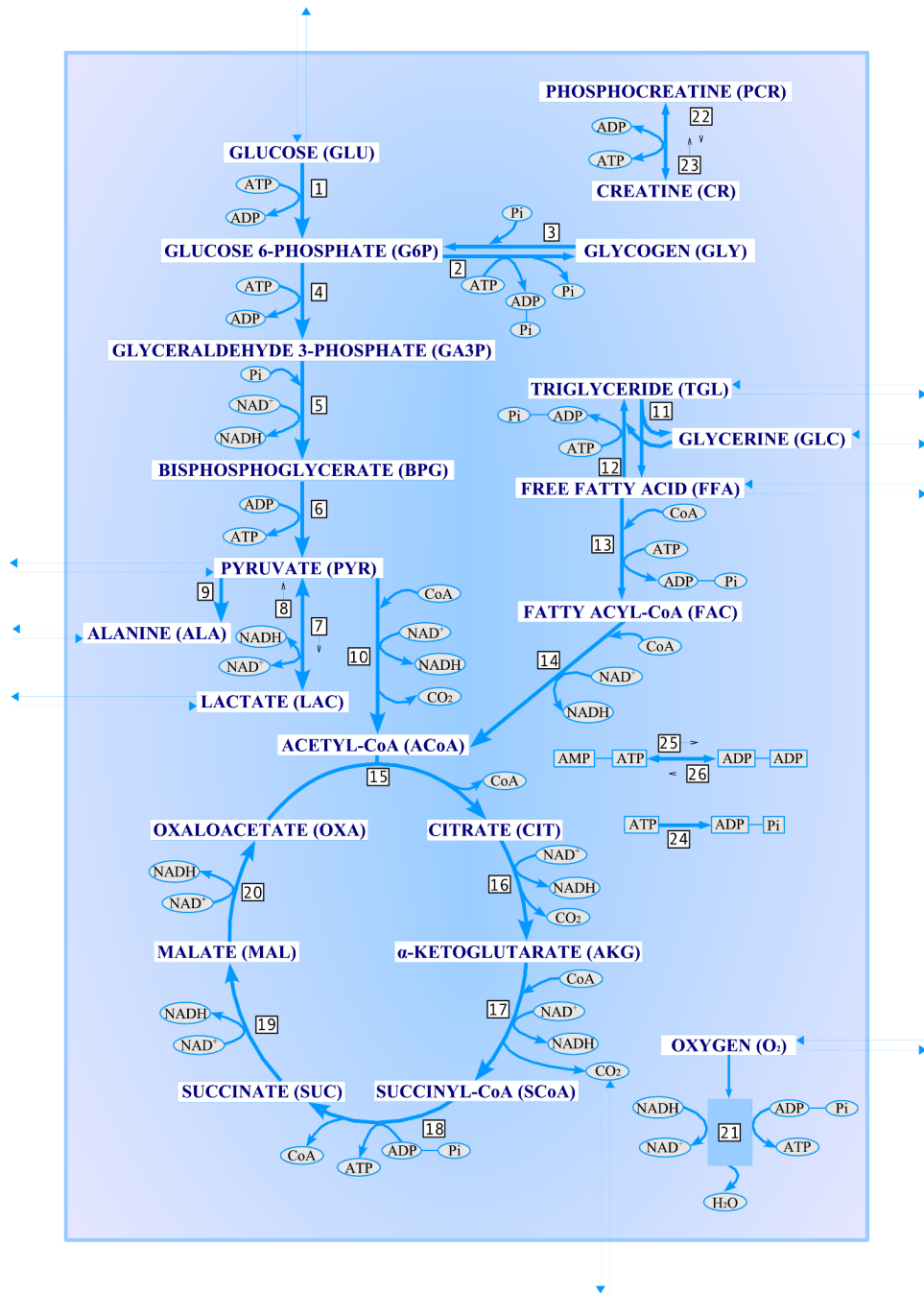


Fig. 8. A schematic diagram of the biochemical pathways in the two-compartment skeletal muscle model.

Table 1

Biochemical compounds.

1. GLU (glucose)	16. CIT (citrate)
2. PYR (pyruvate)	17. AKG α -ketoglutarate
3. LAC (lactate)	18. SCoA (succinyl-CoA)
4. ALA (alanine)	19. SUC (succinate)
5. TGL (triglyceride)	20. MAL (malate)
6. GLC (glycerine)	21. OXA (oxaloacetate)
7. FFA (free fatty acid)	22. PCR (phosphocreatine)
8. CO ₂ (carbon dioxide)	23. CR (creatine)
9. O ₂ (oxygen)	24. P _i (inorganic phosphate)
10. G6P (glucose 6-phosphate)	25. CoA (coenzyme A)
11. GLY (glycogen)	26. NADH
12. GA3P (glyceraldehyde 3-phosphate)	27. NAD ⁺
13. BPG (bisphosphoglycerate)	28. ATP
14. FAC (fatty acyl-CoA)	29. ADP
15. ACoA (acetyl-CoA)	30. AMP

Table 2

Biochemical reactions. The non-integer stoichiometry of oxidative phosphorylation corrects the effect of lumping together the concentrations in cytosol and mitochondria.

Reaction	Stoichiometry
1. Glucose Utilization	$GLU+ATP \rightarrow G6P+ADP$
2. Glycogen synthesis	$G6P+ATP \rightarrow GLY + ADP + 2 P_i$
3. Glycogen utilization	$GLY + P_i \rightarrow G6P$
4. Glucose 6-phosphate breakdown	$G6P+ATP \rightarrow 2 GA3P+ADP$
5. Glyceraldehyde 3-phosphate breakdown	$GA3P+P_i+NAD^+ \rightarrow BPG + NADH$
6. Pyruvate production	$BPG + 2 ADP \rightarrow PYR + 2 ATP$
7. Pyruvate reduction	$PYR + NADH \rightarrow LAC + NAD^+$
8. Lactate oxidation	$LAC + NAD^+ \rightarrow PYR + NADH$
9. Alanine production	$PYR \rightarrow ALA$
10. Pyruvate oxidation	$PYR + CoA + NAD^+ \rightarrow ACoA + NADH + CO_2$
11. Lipolysis	$TGL \rightarrow GLC + 3 FFA$
12. Triglyceride synthesis	$GLC+3 FFA + 3 ATP \rightarrow TGL+3 ADP + 3 P_i$
13. Free fatty acid utilization	$FFA + CoA + 2 ATP \rightarrow FAC + 2 ADP + 2 P_i$
14. Fatty Acyl-CoA Oxidation	$FAC + 7 CoA + (35/3) NAD^+ \rightarrow 8 ACoA + (35/3) NADH$
15. Citrate production	$ACoA + OXA \rightarrow CIT + CoA$
16. α -ketoglutarate production	$CIT + NAD^+ \rightarrow AKG + NADH + CO_2$
17. Succinyl-CoA production	$AKG + CoA + NAD^+ \rightarrow SCoA + NADH + CO_2$
18. Succinate production	$SCoA + ADP + P_i \rightarrow SUC + CoA + ATP$
19. Malate production	$SUC + (2/3) NAD^+ \rightarrow MAL + (2/3) NADH$
20. Oxaloacetate production	$MAL + NAD^+ \rightarrow OXA + NADH$
21. Oxidative phosphorylation	$O_2 + 5.64 ADP + 5.64 P_i + 1.88 NADH \rightarrow 2 H_2O + 5.64 ATP + 1.88 NAD^+$
22. Phosphocreatine breakdown	$PCR + ADP \rightarrow CR + ATP$
23. Phosphocreatine synthesis	$CR + ATP \rightarrow PCR + ADP$
24. ATP hydrolysis	$ATP \rightarrow ADP + P_i$
25. AMP utilization	$AMP + ATP \rightarrow ADP + ADP$
26. AMP production	$ADP + ADP \rightarrow AMP + ATP$

Transport fluxes of species in the blood and arterial concentration values. For the transport fluxes, (f) refers to facilitated and (p) for passive transport.

Table 3

Transport fluxes		Species in blood	C_{arterial} value [mmol/l]
blood→cell	cell→blood		
27. $J_{\text{GLU},b \rightarrow c}$	(f)	45. GLU	5.000
28. $J_{\text{PYR},b \rightarrow c}$	(p)	46. PYR	0.0680
29. $J_{\text{LAC},b \rightarrow c}$	(f)	47. LAC	0.7200
30. $J_{\text{ALA},b \rightarrow c}$	(p)	48. ALA	0.1900
31. $J_{\text{TGL},b \rightarrow c}$	(p)	49. TGL	0.9900
32. $J_{\text{GLC},b \rightarrow c}$	(p)	50. GLC	0.0600
33. $J_{\text{FFA},b \rightarrow c}$	(p)	51. FFA	0.6500
34. $J_{\text{CO}_2,b \rightarrow c}$	(p)	52. CO ₂	23,4066
35. $J_{\text{O}_2,b \rightarrow c}$	(p)	53. O ₂	9,2000
	36. $J_{\text{GLU},c \rightarrow b}$		
	37. $J_{\text{PYR},c \rightarrow b}$	(f)	
	38. $J_{\text{LAC},c \rightarrow b}$	(f)	
	39. $J_{\text{ALA},c \rightarrow b}$	(p)	
	40. $J_{\text{TGL},c \rightarrow b}$	(p)	
	41. $J_{\text{GLC},c \rightarrow b}$	(p)	
	42. $J_{\text{FFA},c \rightarrow b}$	(p)	
	43. $J_{\text{CO}_2,c \rightarrow b}$	(p)	
	44. $J_{\text{O}_2,c \rightarrow b}$	(p)	

Table 4

Bounds for the components u_j of the solution vector \mathbf{u} .

j	Tight bounds		Wide bounds		j	Tight bounds		Wide bounds	
	lower values	upper values	lower values	upper values		lower values	upper values	lower values	upper values
1	0.114	0.239	0.024	0.640	28	0.012	0.028	0.005	0.163
2	0.108	0.345	0.028	0.924	29	0.402	0.584	0.142	1.961
3	0.111	0.290	0.062	0.944	30	0.004	0.011	0.002	0.059
4	0.098	0.261	0.024	0.847	31	1.037	2.890	0.463	15.440
5	0.156	0.650	0.044	2.557	32	0.037	0.150	0.021	0.683
6	0.177	0.505	0.042	1.601	33	0.417	1.057	0.177	6.041
7	0.147	0.305	0.044	0.908	34	16.182	36.760	6.391	223.867
8	0.021	0.105	0.005	0.313	35	5.339	10.267	1.898	65.720
9	0.022	0.057	0.012	0.135	36	0.139	0.346	0.059	0.950
10	0.042	0.214	0.012	0.679	37	0.012	0.040	0.006	0.200
11	0.058	0.157	0.031	0.385	38	0.591	0.784	0.208	2.563
12	0.046	0.206	0.017	0.753	39	0.020	0.093	0.012	0.414
13	0.014	0.065	0.005	0.271	40	0.913	3.168	0.468	15.336
14	0.013	0.065	0.004	0.240	41	0.034	0.157	0.020	0.691
15	0.133	0.846	0.093	3.488	42	0.305	1.173	0.162	5.556
16	0.201	0.627	0.037	2.275	43	13.293	49.600	7.016	238.170
17	0.153	0.759	0.046	2.952	44	3.922	9.779	1.695	53.412
18	0.296	0.437	0.053	2.627	45	4.844	4.930	1.664	19.069
19	0.192	0.630	0.041	2.112	46	0.064	0.077	0.024	0.274
20	0.189	0.660	0.040	2.271	47	0.805	0.864	0.289	3.264
21	0.707	1.464	0.108	5.667	48	0.207	0.232	0.075	0.855
22	11.674	32.281	2.558	82.562	49	0.904	1.018	0.339	3.759
23	11.068	37.744	3.058	85.473	50	0.039	0.095	0.022	0.250
24	5.131	7.481	2.040	21.063	51	0.484	0.705	0.200	2.300
25	6.554	14.031	1.460	39.760	52	20.990	28.660	8.363	94.716
26	6.870	12.806	1.148	38.816	53	7.921	8.712	2.810	9.603
27	0.319	0.549	0.139	1.387					

Table 5 Correlation lengths of the components u_j of \mathbf{u} from samples of 200000 elements calculated using the minimal model and the whole model with Gibbs sampling.

j	Minimal model	Whole model	j	Minimal model	Whole model
1	27	4	28	2	2
2	12	5	29	54	29
3	12	5	30	2	2
4	27	27	31	>1000	>1000
5	27	27	32	38	64
6	27	27	33	765	257
7	21	23	34	>1000	>1000
8	4	2	35	>1000	>1000
9	3	2	36	19	12
10	17	12	37	3	2
11	4	5	38	47	34
12	5	7	39	4	2
13	>1000	>1000	40	>1000	>1000
14	>1000	>1000	41	39	4
15	>1000	>1000	42	754	252
16	>1000	>1000	43	>1000	>1000
17	>1000	>1000	44	>1000	>1000
18	>1000	>1000	45	4	4
19	>1000	>1000	46	3	3
20	>1000	>1000	47	22	24
21	>1000	>1000	48	3	2
22	>1000	>1000	49	4	3
23	>1000	>1000	50	6	7
24	>1000	>1000	51	6	7
25	>1000	>1000	52	234	133
26	>1000	>1000	53	>1000	>1000
27	24	13			