Bioinformatics, 32(15), 2016, 2330–2337
doi: 10.1093/bioinformatics/btw132
Advance Access Publication Date: 11 March 2016
Original Paper

Systems biology

ll-ACHRB: a scalable algorithm for sampling the feasible solution space of metabolic networks

Pedro A. Saa and Lars K. Nielsen*

Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, St Lucia, QLD, Australia

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on October 30, 2015; revised on February 7, 2016; accepted on March 3, 2016

Abstract

Motivation: Random sampling of the solution space has emerged as a popular tool to explore and infer properties of large metabolic networks. However, conventional sampling approaches commonly used do not eliminate thermodynamically unfeasible loops.

Results: In order to overcome this limitation, we developed an efficient sampling algorithm called loopless Artificially Centered Hit-and-Run on a Box (ll-ACHRB). This algorithm is inspired by the Hit-and-Run on a Box algorithm for uniform sampling from general regions, but employs the directions of choice approach of Artificially Centered Hit-and-Run. A novel strategy for generating feasible warmup points improved both sampling efficiency and mixing. ll-ACHRB shows overall better performance than current strategies to generate feasible flux samples across several models. Furthermore, we demonstrate that a failure to eliminate unfeasible loops greatly affects sample statistics, in particular the correlation structure. Finally, we discuss recommendations for the interpretation of sampling results and possible algorithmic improvements.

Availability and implementation: Source code for MATLAB and OCTAVE including examples are freely available for download at http://www.aibn.uq.edu.au/cssb-resources under Software. Optimization runs can use Gurobi Optimizer (by default if available) or GLPK (included with the algorithm).

Contact: lars.nielsen@uq.edu.au

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

1 Introduction

A diversity of computational methods referred to as constrained-based methods has been developed to study the characteristics and capabilities of Genome-scale Models (GEMs). Uniform random sampling of the flux solution space is a popular choice for interrogation of GEMs without introducing optimality criteria. From evaluation of the impact of physiological constraints on the sampled flux distributions (Bordbar *et al.*, 2010; Lewis *et al.*, 2010; Mo *et al.*, 2009; Price *et al.*, 2004; Shoaie *et al.*, 2013; Thomas *et al.*, 2014; Wiback *et al.*, 2004) to identification of transcriptional regulation (Bordbar *et al.*, 2014; Bordel *et al.*, 2010), definition of a flux backbone (Almaas *et al.*, 2004) and correlated reaction sets (Dal'Molin *et al.*, 2015; Price *et al.*, 2004, 2006; Reed and Palsson, 2004), uniform sampling has been applied to explore achievable states and emergent properties of networks. More importantly, uniform sampling has emerged as an unbiased tool to assess the capabilities of metabolic reconstructions (Lewis *et al.*, 2012).

For this type of analysis to be valid, a key—and often overlooked—assumption is that the flux samples are thermodynamically feasible. GEMs are mathematically described by the stoichiometry matrix $\mathbf{S}_{m\times n}$, which encodes the mass balances for m metabolites reacting in n biochemical reactions. The capacities of these reactions are phenomenologically limited by thermodynamic and kinetic constraints in the form of lower **lb** and upper **ub** bounds. Assuming steady-state for metabolites accumulation, the vector of reaction

fluxes $\mathbf{v}$ consistent with mass conservation is defined by the following set of constraints (Eqs. 1 and 2),

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \tag{1}$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub} \tag{2}$$

The resulting flux space $\boldsymbol{\Omega}$ defines a convex polytope describing the achievable states of the network. Sampling from this multi-dimensional body can be efficiently achieved using Markov Chain Monte Carlo (MCMC) methods such as the 'Hit-and-Run' (HR) (Smith, 1984) and the 'Artificially Centered Hit-and-Run' (ACHR) (Kaufman and Smith, 1998) algorithms. The latter method has been particularly preferred for sampling the irregularly shaped solution space of metabolic networks, as it explores elongated directions of $\boldsymbol{\Omega}$ enabling in principle faster mixing. There are, however, additional constraints on $\mathbf{v}$ that ensure its thermodynamic feasibility. Let us denote $\mathbf{N}_{\text{int}}$ the matrix storing the null space vectors of the internal reactions described by $\mathbf{S}_{\text{int}}$, and $\Delta\boldsymbol{\mu}_{\text{int}}$ the vector of chemical potentials for these reactions. According to the first and second law of thermodynamics, $\mathbf{v}$ is thermodynamically feasible if for $\mathbf{v}_{\text{int}} \subseteq \mathbf{v}$ (Beard *et al.*, 2002),

$$\mathbf{N}_{\text{int}}^{\text{T}} \cdot \Delta\boldsymbol{\mu}_{\text{int}} = \mathbf{0} \tag{3}$$

$$\text{diag}(\Delta\boldsymbol{\mu}_{\text{int}}) \cdot \mathbf{v}_{\text{int}} \leq \mathbf{0} \tag{4}$$

Equations (3) and (4) guarantee respectively that the global potential energy of the network is balanced (first law), and that reactions proceed in the opposite direction of chemical potential change (second law). These two equations can be satisfied simultaneously only if the net flux around all closed loops is equal to zero, and hence it is commonly referred as the 'loopless condition'. Imposition of this condition generally causes the solution space $\boldsymbol{\Omega}_{\text{loopless}}$ to be non-convex, which is much harder to sample and so far no algorithm has been developed to do so. Instead, *ad-hoc* methods using conventional samplers have been used to sample from $\boldsymbol{\Omega}$ and then either remove samples with 'loopy' patterns (Price *et al.*, 2006), or find the nearest feasible flux distribution (Schellenberger *et al.*, 2011). Both approaches have disadvantages. The first approach may yield few feasible samples making it inefficient in high-dimensional models, whereas the second approach is impractical in large models as it requires the solution of a non-polynomial time MILP problem for each flux distribution.

Here we present a scalable, direct sampling algorithm called loopless Artificially Centered Hit-and-Run on a Box (ll-ACHRB). This algorithm is motivated by a recent variation of the Hit-and-Run algorithm for uniform sampling from general regions (convex and non-convex) termed 'Hit-and-Run on a Box' (HRB) (Kiatsupaibul *et al.*, 2011), and employs directions of choice as in ACHR. The algorithm shows overall better performance than current strategies to generate thermodynamically feasible flux samples in several metabolic models ranging from small-size models to GEMs. Finally, we report significant discrepancies in feasible and unfeasible sample statistics, highlighting the consequences of employing thermodynamically unfeasible samples for statistical inference.

2 Algorithm

Our aim is to sample from the thermodynamically feasible solution space defined by Eqs. (1–4). To this end, we have developed a general Hit-and-Run algorithm capable of sampling from this commonly non-convex space by enclosing $\boldsymbol{\Omega}_{\text{loopless}}$ in a hyper-rectangle from which step sizes can be readily chosen following HRB. Execution of subsequent moves is achieved by randomly choosing directions using warmup points as in ACHR. As opposed to conventional warmup methods, our strategy pulls warmup points from the interior of $\boldsymbol{\Omega}_{\text{loopless}}$, improving both mixing and acceptance rates (see Results). Importantly, the proposed algorithm enables sampling of exactly N feasible flux samples as opposed to the strategies reported to date.

Application of the three-step sampling workflow is illustrated with a toy model initially comprised of 6 internal reactions, three exchange reactions and five metabolites (Fig. 1). Initially, the pre-processing step takes place using loopless-Flux Variability Analysis (ll-FVA) (Schellenberger *et al.*, 2011) to remove blocked reactions from the model (v_6), as well as fix reaction directionalities imposed by the loopless condition (v_1, v_3 and v_4) (Fig. 1A). The resulting model has $r = 8$ reactions, whereof two are reversible (v_2 and v_5). Execution of ll-FVA also defines the minimum enclosing box on $\boldsymbol{\Omega}_{\text{loopless}}$ required for step size selection (see below). After pre-processing, a warmup phase is used to generate feasible points spanning $\boldsymbol{\Omega}_{\text{loopless}}$ (Fig. 1B). As noted in Megchelenbrink *et al.* (2014), it is important that these points lie in the interior of the target space, otherwise the allowed step sizes can become very small hereby hampering convergence. In order to generate these points, we cap the lower and upper bounds of each reaction one-at-a-time to a fraction $\alpha < 1$ (e.g. $\alpha = 0.95$) of the maximum flux range, and next the flux through the reaction with modified bounds is minimized and maximized subject to thermodynamic feasibility. In this way, this step yields $2r$ feasible interior solutions spanning the desired space.

Once warmup points have been generated, the sampling phase is initiated following the HRB sampling scheme (Fig. 1C). Let us denote $\mathbf{v}_k$ an arbitrary flux vector in $\boldsymbol{\Omega}_{\text{loopless}} \in R^n$, $\mathbf{v}_{\text{center}}$ a flux vector describing the 'centroid' of $\boldsymbol{\Omega}_{\text{loopless}}$, $\mathbf{u}_k$ a direction on the surface of the n-dimensional unit hypersphere $\partial D = \{\mathbf{u} \in R^n | \|\mathbf{u}\| = 1\}$ and $\boldsymbol{\Theta}(\mathbf{v}_{\text{center}}, \mathbf{v}_{\text{warmup}})$ the direction probability distribution on ∂D as defined in ACHR for accelerated convergence (Kaufman and Smith, 1998). Without loss of generality, let us also assume that $\boldsymbol{\Omega}_{\text{loopless}}$ is contained in a known n-dimensional box B and that a set of feasible points $\mathbf{v}_{\text{warmup}}$ are known. The ll-ACHRB proceeds as follows. Starting from $\mathbf{v}_k$, a random direction $\mathbf{u}_k$ is drawn from $\boldsymbol{\Theta}$ and the line $\ell_k = \{\lambda \in R | \mathbf{v}_k + \lambda \mathbf{u}_k \in B\}$ is constructed. Next, the maximum steps sizes in the negative $\left(\lambda_i^{\max-}\right)$ and positive $\left(\lambda_i^{\max+}\right)$ directions are found by intersecting ℓ_k with $\boldsymbol{\Omega}_{\text{loopless}}$ (i.e. longest cord in B), and a random step size λ_k is generated by uniformly sampling from that interval. The new sample candidate is defined as $\mathbf{v}_{k+1} = \mathbf{v}_k + \lambda_k \mathbf{u}_k$. If $\mathbf{v}_{k+1} \notin \boldsymbol{\Omega}_{\text{loopless}}$, the enclosing box B is 'shrinked' by redefining $\left(\lambda_i^{\max-}, \lambda_i^{\max+}\right)$ according to the sign of λ_k (Neal, 2003), and another λ_k is uniformly sampled until $\mathbf{v}_{k+1} \in \boldsymbol{\Omega}_{\text{loopless}}$. Once a feasible sample is found, the centroid is updated according to the ACHR update scheme and the entire process is reiterated until N uniform samples have been collected. The details of the algorithm are summarized in Algorithm 1.

The HRB algorithm repeatedly samples step sizes from the intersection of ℓ_k and $\boldsymbol{\Omega}_{\text{loopless}}$ until a feasible flux distribution is found. This rejection step is the main modification to the traditional HR scheme, where a uniform feasible sample on $\boldsymbol{\Omega}$ can be readily obtained (Smith, 1984). Importantly, as shown by Kiatsupaibul *et al.* (2011) the resulting chain $\{v_k | k \geq 0\}$ from HRB is also uniform on the target space.

The computational efficiency of the algorithm depends on the recurring two main queries: (i) find the length of the longest cord in B in the direction of ℓ_k (line 14) and (ii) check the feasibility of

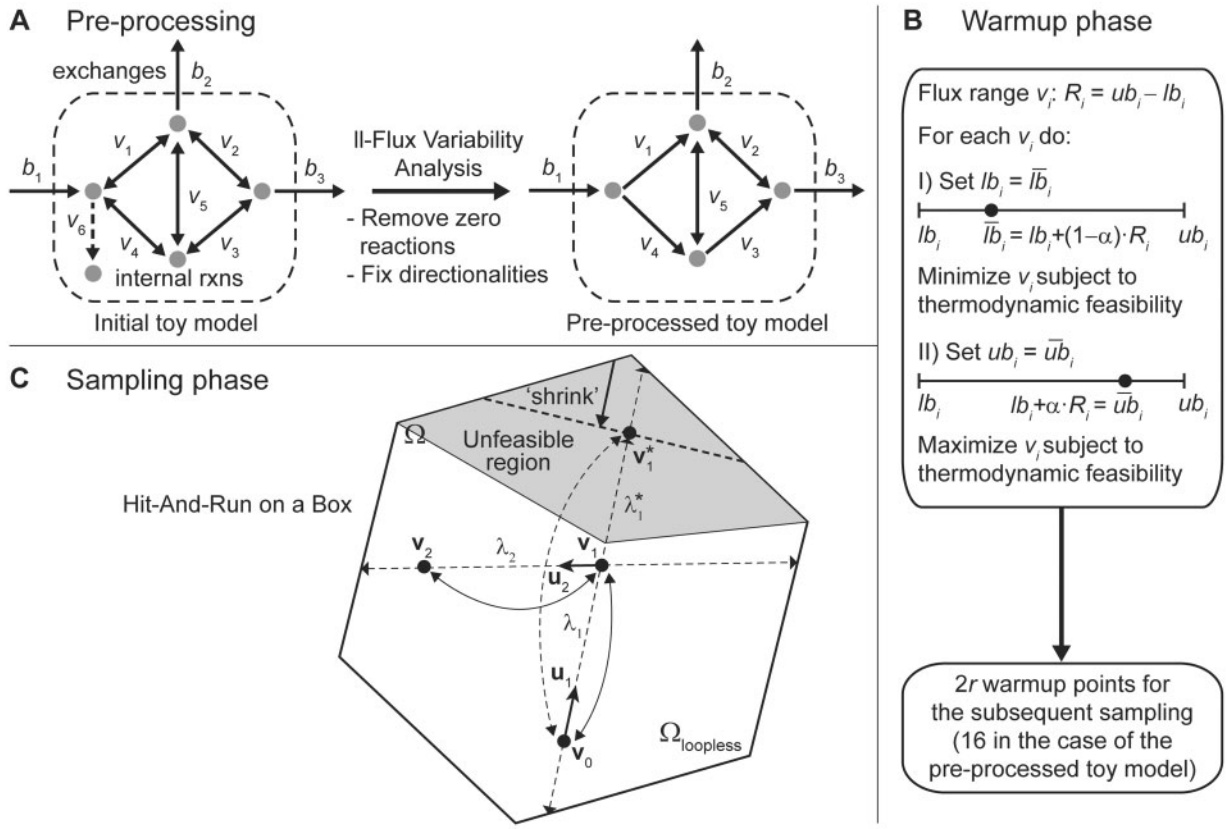

Fig. 1. Application of the sampling workflow in a toy model. (**A**) Pre-processing phase. The toy model consist of 6 internal reactions (v_i, $i = 1,\ldots,6$), 3 exchange reactions (b_j, $j = 1,2,3$) (9 reactions in total) and 5 metabolites. Using ll-FVA for pre-processing, reaction v_6 (dashed arrow) is found to be blocked and is removed from the network. Additionally, 3 of the 5 internal reactions are found to be irreversible and thus their directions are fixed. The pre-processed model has $r = 8$ reactions, whereof two are reversible (v_2 and v_5). (**B**) Warmup phase. $2r$ points in the interior of the feasible solution space are generated by minimizing and maximizing each reaction subject to thermodynamic feasibility and modified bounds. (**C**) Sampling phase. The Hit-and-Run on a Box (HRB) sampling scheme is used to uniformly sample from the feasible nonconvex space ($\boldsymbol{\Omega}_{\text{loopless}}$) by 'shrinking' the sampling space until a feasible flux distribution is accepted

candidate sample $\mathbf{v}_k$ (line 16). The former task is readily performed using Eq. (5), taking into account the convexity of $\boldsymbol{\Omega}$ and that $\boldsymbol{\Omega}_{\text{loopless}} \subseteq \boldsymbol{\Omega}$. Even if initially the longest cord in $\boldsymbol{\Omega}$ is not fully supported in $\boldsymbol{\Omega}_{\text{loopless}}$, the shrinking step (lines 19–26) decreases the size of the initial box eventually matching the length of B on $\boldsymbol{\Omega}_{\text{loopless}}$. The second query, i.e. evaluation of the thermodynamic feasibility of the candidate sample, is more challenging and for this we developed a faster method (see next section).

$$\lambda_{k,i}^{-} = \begin{cases} \dfrac{lb_i - v_{k,i}}{u_{k,i}}, & u_{k,i} > 0 \\ \dfrac{ub_i - v_{k,i}}{u_{k,i}}, & u_{k,i} < 0 \end{cases}, \quad \forall i \in \text{reactions}$$

$$\lambda_{k,i}^{+} = \begin{cases} \dfrac{ub_i - v_{k,i}}{u_{k,i}}, & u_{k,i} > 0 \\ \dfrac{lb_i - v_{k,i}}{u_{k,i}}, & u_{k,i} < 0 \end{cases} \tag{5}$$

$$\lambda_k^{\max-} = \max\ \boldsymbol{\lambda}_k^{-}$$
$$\lambda_k^{\max+} = \min\ \boldsymbol{\lambda}_k^{+}$$

3 Methods

3.1 Verification of thermodynamic feasibility

Thermodynamic feasibility is ensured by Eqs. (3 and 4). For a given flux distribution $\mathbf{v}$ with $\mathbf{v}_{\text{int}}$ denoting its internal reactions, resolution of the following LP problem with decision variables $\Delta\boldsymbol{\mu}_{\text{int}}$ (chemical potential variables) and blank objective function provides a test for thermodynamic feasibility (Schellenberger *et al.*, 2011),

$$\begin{aligned} &\min\ \mathbf{0} \\ &\text{s.t.} \\ &\mathbf{N}_{\text{int}}^{\text{T}} \cdot \Delta\boldsymbol{\mu}_{\text{int}} = \mathbf{0} \\ &-K \cdot (\mathbf{1} + \text{sgn}(\mathbf{v}_{\text{int}})) - \text{sgn}(\mathbf{v}_{\text{int}}) \le \Delta\boldsymbol{\mu}_{\text{int}} \le K \cdot (\mathbf{1} - \text{sgn}(\mathbf{v}_{\text{int}})) - \text{sgn}(\mathbf{v}_{\text{int}}) \end{aligned} \tag{6}$$

where K denotes a large constant (e.g. 10^3). If the above problem has a solution for $\Delta\boldsymbol{\mu}_{\text{int}}$, then $\mathbf{v}$ is thermodynamically feasible. However, this approach is very time-consuming for large models. To overcome this limitation, we have developed a more efficient method to detect unfeasible loops making use of the loop law topology encoded in $\mathbf{N}_{\text{int}}^{\text{T}}$ and the flux signs of the candidate sample. $\mathbf{N}_{\text{int}}$ stores a null space basis of internal reaction vectors describing loop laws where energy is conserved. For $\mathbf{v}$ to be unfeasible, at least one of the loop laws encoded in the rows of $\mathbf{N}_{\text{int}}^{\text{T}}$ $\left(\mathbf{n}_{\text{int},i}^{\text{T}}\right)$ has to be unfeasible, i.e. $\mathbf{n}_{\text{int},i}^{\text{T}} \cdot \Delta\boldsymbol{\mu}_{\text{int}} > 0 \ \vee\ \mathbf{n}_{\text{int},i}^{\text{T}} \cdot \Delta\boldsymbol{\mu}_{\text{int}} < 0$. Notably, if a reaction is zero then the associated chemical potential variable is free

Algorithm 1: loopless Artificially-Centered Hit-and-Run on a Box.

Inputs: stoichiometric matrix **S**, lower and upper bounds (**lb**,**ub**), number of samples N
Outputs: N feasible flux samples $\mathbf{v}_k$

(**lb**$_{feasible}$,**ub**$_{feasible}$) ← PreProcessModel(**S**,**lb**,**ub**) //Pre-processing
lb ← **lb**$_{feasible}$, **ub** ← **ub**$_{feasible}$
(**S**,**lb**,**ub**) ← RemoveZeroReactions(**S**,**lb**,**ub**)
r ← GetReactions(**S**) //Warmup phase
for i ← 1 to $2r$ do
  $\mathbf{v}_{warmup,i}$ ← GenerateRandomInteriorPoint(**S**,**lb**,**ub**)
end for
$\mathbf{v}_{center}$ ← FindFeasibleCentroid(**S**,**lb**,**ub**,$\mathbf{v}_{warmup}$)
$\mathbf{v}_0$ ← $\mathbf{v}_{center}$ //Sampling phase
for k ← 0 to N do
  $\mathbf{u}_k$ ← SampleDirectionFromΘ($\mathbf{v}_{center}$,$\mathbf{v}_{warmup}$)
  i ←1, *condition* ← true
  while *condition*
    $(\lambda_i^{max-}, \lambda_i^{max+})$ ← FindLengthLongestCord($\mathbf{v}_k$,$\mathbf{u}_k$,**lb**,**ub**)
    $\mathbf{v}_{k,i}$ ← $\mathbf{v}_k$+$\lambda_{k,i}\mathbf{u}_k$
    if IsFeasible($\mathbf{v}_{k,i}$) then
      $\mathbf{v}_{k+1}$ ← $\mathbf{v}_{k,i}$, k ← k+1, *condition* ← false
      $\mathbf{v}_{center}$ ← UpdateCenter($\mathbf{v}_{center}$,k)
    else //Shrinking step
      i ← i+1
      if $\lambda_{k,i} > 0$ then
        λ_i^{max+} ← $\lambda_{k,i}$
      Else
        λ_i^{max-} ← $\lambda_{k,i}$
      end if
    end if
  end while
end for

and the loop laws containing this reaction always hold, i.e. zero reactions are loop breakers. Hence, the problem of determining the feasibility of a flux distribution can be formulated as a search for any laws in which the terms of $\mathbf{n}_{int,i}^{T} \cdot \Delta\boldsymbol{\mu}_{int}$ are all positive or negative (Nigam and Liang, 2007). This search only requires the signs, not the actual values, of $\Delta\boldsymbol{\mu}_{int}$ and the former are readily derived from $\mathbf{v}_{int}$ using Eq. (4). Importantly, given the directionalities of the network many loop laws always hold, i.e. $\mathbf{n}_{int,i}^{T} \cdot \Delta\boldsymbol{\mu}_{int}$ can be zero, and the search can be narrowed to only those potentially unfeasible. Compared to solving Eq. (6), this strategy yielded substantial speed increases up to two orders of magnitude (Supplementary Fig. S1).

3.2 Initial centroid estimation

Estimation of the centroid $\mathbf{v}_{center}$ in ACHR is typically achieved by taking the mean of $\mathbf{v}_{warmup}$. However, given the non-convexity of $\boldsymbol{\Omega}_{loopless}$, $\mathbf{v}_{center}$ might not be contained in the feasible space. In order to ensure the feasibility of the initial $\mathbf{v}_{center}$, we compute the nearest feasible flux centroid $\mathbf{v}_{center}^{*}$ using the following MIQP formulation,

$$\begin{aligned}
&\min_{\mathbf{v}_{center}^{*}} \ \|\boldsymbol{\delta}\|^2 \\
&\text{s.t.} \\
&\mathbf{S} \cdot \mathbf{v}_{center}^{*} = \mathbf{0} \\
&\mathbf{N}_{int}^{T} \cdot \Delta\boldsymbol{\mu}_{int}^{*} = \mathbf{0} \\
&\mathbf{v}_{center}^{*} - \mathbf{v}_{center} = \boldsymbol{\delta} \\
&\mathbf{1} - \boldsymbol{\varepsilon}_{int}^{*} \cdot (K+1) \ \le \ \Delta\boldsymbol{\mu}_{int}^{*} \ \le \ \mathbf{K} - \boldsymbol{\varepsilon}_{int}^{*} \cdot (\mathrm{K}+1) \\
&-K \cdot \left(\mathbf{1} - \boldsymbol{\varepsilon}_{int}^{*}\right) \ \le \ \mathbf{v}_{int}^{*} \ \le \ K\boldsymbol{\varepsilon}_{int}^{*} \\
&\mathbf{lb} \ \le \ \mathbf{v}_{center}^{*} \ \le \ \mathbf{ub} \\
&\boldsymbol{\varepsilon}_{int}^{*} \in \{0,1\}, \ \boldsymbol{\delta} \in \mathbb{R}^{n}, \ \mathrm{K \ large \ (e.g., \ } 10^3)
\end{aligned} \tag{7}$$

where $\boldsymbol{\varepsilon}_{int}^{*}$ represents a binary vector associated with the internal reactions of $\mathbf{v}_{center}^{*}$ and $\boldsymbol{\delta}$ denotes the distance vector from $\mathbf{v}_{center}$ to $\mathbf{v}_{center}^{*}$. The closest feasible flux distribution to $\mathbf{v}_{center}^{*}$ is found by minimizing the norm of $\boldsymbol{\delta}$.

3.3 Effective sample size and mixing

Given a stationary discrete Markov Chain $\{v_k | \ k \geq 0\}$, we would like to estimate the time after which the points can be considered uncorrelated. The autocorrelation time describes the mixing of the chain and can be employed for this purpose (Givens and Hoeting, 2012),

$$\tau = 1 + 2\sum_{k=1}^{\infty} \widehat{\rho}_k(\theta) \tag{8}$$

where $\widehat{\rho}_k(\theta)$ denotes the autocorrelation of lag k for parameter θ and τ is the (integrated) autocorrelation time. A closely related measure of mixing is the Effective Sample Size (ESS) (Kass *et al.*, 1998).

$$\mathrm{ESS} = \frac{N}{\tau} \tag{9}$$

where N is the total sample size. Low effective sample size, i.e. high autocorrelation time, is indicative of poor mixing of the chain. A common procedure to compute the sum in Eq. (8) and determine ESS is to find a cut-off point beyond which the autocorrelation is very close to zero, and then sum $\widehat{\rho}_k$ up to that point. Common choices for the cut-off point k are such that $\widehat{\rho}_k < 0.01$ or $\widehat{\rho}_k < 2s_k$, where s_k is the estimated standard deviation of τ up to k (Givens and Hoeting, 2012). We have used the former criterion to estimate $\widehat{\rho}_k$.

3.4 Computational implementation

The algorithm was implemented in MATLAB 2013a (The MathWorks, Natick, MA) using the Gurobi Optimizer 5.6 (Gurobi Optimization, Inc., Houston, Texas). Simulations were run on a 16-CPU 64-GB ram Virtual Machine running CentOS Linux 6.5. The VM is hosted on the QRIScloud Polaris cell. Accelerated sampling convergence was achieved by exploiting the parallelizable structure of the algorithm and executing it in multiple cores employing the Parallel Computing Toolbox of MATLAB.

4 Results

4.1 ll-ACHRB performs and scales better than alternative approaches

The computational performance of the proposed algorithm was evaluated in different models covering a wide spectrum of metabolic capabilities. The considered models include the toy model shown in Figure 1A, an *E.coli* core model (Orth *et al.*, 2011), the Red Blood Cell GEM iRBC283 (Bordbar *et al.*, 2011), the *H.pylori* GEM iIT341 (Thiele *et al.*, 2005), the *T.maritima* GEM iLJ478 (Zhang *et al.*, 2009), the *G.metallireducens* GEM iAF987 (Feist *et al.*, 2014), and the latest *E.coli* GEM iJO1366 (Orth *et al.*, 2011). The features of each model after pre-processing are shown in Table 1. We have assumed growth in minimal media and no particular biological objective for the models. In the case of iIT341, iLJ478, iAF987 and iJO1366, we stripped the biomass equation of trace elements and in the case of iJO1366, we further compacted the model by lumping linear pathways into single reactions (see Supplementary Tables S1–S4 for omitted elements). The resulting models preserved

the same fundamental metabolic capabilities of the original with the difference that they are now numerically tractable for MILP optimization.

In order to assess the behaviour of ll-ACHRB, its sampling performance was compared against ACHR—the most used sampler in the literature. The testing procedure was as follows. For each model and algorithm, three independent Hit-and-Run dynamics were run until 10^5 samples were collected. Only one point every 100 steps (thinning) was collected to reduce serial correlation and the first 10^4 samples were discarded to reduce bias (burn-in). Also, in order to ensure a fair comparison, each algorithm was started from an initial feasible point and centroid.

Table 2 shows the performance of each sampler in each case. Inspection of the effective sampling time shows that ll-ACHRB is both more effective and efficient than ACHR, displaying a robust and efficient performance across all models. In contrast, the efficiency of ACHR for generating feasible samples depends strongly on the model characteristics. In fact, the results suggest that the overall sampling performance depends less on the size of the network than on its topological complexity, i.e. number of potentially unfeasible loop laws (Table 1). The smallest model had a single potentially unfeasible loop and this reduced the feasible sample fraction using ACHR to around 0.5. The RBC GEM was much larger than the toy model but had no potential unfeasible loops, and hence generated only feasible samples with both samplers. In contrast, no feasible samples were found employing ACHR in the *E.coli* core, iIT341, iLJ478, iAF987 and JO1366 models. Unlike ACHR, ll-ACHRB was capable of generating feasible samples for all models in reasonable time (see effective sampling time, Table 2). These results highlight that application of *ex post* screening of feasible samples may not only be inefficient, but could be impossible for more complex models, i.e. fail to generate any feasible samples.

Table 1. Characteristics of the benchmark models used in this study

Model	Mets	Rxns	Reversible reactions	Internal reversible reactions	Potentially unfeasible loop laws[a]
toy model	4	8	2	2	1(2)
E.coli core	68	87	15	13	1(13)
iRBC283	207	283	49	29	0(69)
iIT341	321	370	24	22	2(33)
iLJ478	318	370	37	35	6(48)
iAF987	547	656	43	43	4(125)
iJO1366[b]	497	1007	78	73	7(497)

[a]Number in parenthesis denotes the total number of loop laws present in the model.

[b]This model was further compacted as indicated in the text.

We also evaluated the performance of our warmup strategy. To this task, the ll-ACHRB sampling scheme was employed but now using warm-up points generated using the conventional approach of minimizing and maximizing each reaction flux (Schellenberger *et al.*, 2011). Although the conventional approach was much faster in generating the required warmup points—particularly in the larger models—the sampling algorithm was incapable of effectively generating feasible samples from iAF987 and iJO1366 using these points (rejection rates ~1.0, Supplementary Table S5). Furthermore, the ll-ACHRB algorithm using the proposed warmup strategy displayed lower rejection rates and typically reduced effective sampling times compared to the case when the conventional warmup strategy was employed. The above results support the suitability of ll-ACHRB for the generation of feasible flux samples. Next, we considered the quality of the samples, i.e. the mixing behaviour of ll-ACHRB.

4.2 ll-ACHRB displays good mixing behaviour

Efficient mixing is particularly difficult to achieve in large metabolic models encompassing highly heterogeneous scales (De Martino *et al.*, 2015). In order to assess the mixing properties of ll-ACHRB, we compared the ESS distribution of each sampler in every model (Fig. 2). ESS represents the effective number of independent samples generated from the Markov Chain, and thus is of great importance for statistical inference. The mixing behaviour was identical for the *E.coli* core and iRBC283 models. ll-ACHRB mixed substantially slower for iLJ478 and iJO1366, displaying an ESS median reduction of 8.5-fold and 2-fold respectively compared to ACHR ($P = 10^{-49}$ and $P = 10^{-22}$ respectively, Wilcoxon rank sum test). In contrast, ll-ACHRB showed slightly better mixing performance in the toy model (0.2% ESS median increase, $P = 2{\cdot}10^{-3}$ Wilcoxon rank sum test), and important mixing improvements in both iIT341 and iAF987 (14.6% and 48.8% ESS median increase respectively, $P = 5{\cdot}10^{-3}$ and $P = 6{\cdot}10^{-5}$ Wilcoxon rank sum test). Altogether, ll-ACHRB displays similar mixing properties as ACHR, but with the guarantee of thermodynamic feasibility of the samples. Ultimately, the features of each model will determine the ll-ACHRB mixing quality.

Particular subspaces may be difficult to reach in a complicated space. For example, a reaction with a non-zero flux range may never appear with non-zero flux in the sample, and thus we may lose essential qualitative information (Reimers, 2015). We evaluated the

Table 2. Performance indicators of tested samplers in different models[a]

Model	Warmup time (min)		Mean rejection rate		Fraction of feasible samples		Effective sampling time[b] (min)	
	ACHR	ll-ACHRB	ACHR	ll-ACHRB	ACHR	ll-ACHRB	ACHR	ll-ACHRB
toy model	3.47×10^{-4}	3.47×10^{-4}	0.0	0.160	0.499	1.0	4.43	2.68
E.coli core	4.86×10^{-3}	7.45×10^{-3}	0.0	0.486	0	1.0	n.d.	4.27
iRBC283	2.03×10^{-2}	7.74×10^{-2}	0.0	0.0	1.0	1.0	2.71	2.75
iIT341	3.51×10^{-2}	2.45×10^{-1}	0.0	0.149	0	1.0	n.d.	34.03
iLJ478	3.41×10^{-2}	1.22	0.0	0.361	0	1.0	n.d.	54.89
iAF987	9.96×10^{-2}	8.81	0.0	0.106	0	1.0	n.d.	112.20
iJO1366	4.70×10^{-1}	2.61×10^{2}	0.0	0.143	0	1.0	n.d.	151.38

[a]Quantities in this table are the average of three independent Hit-And-Run dynamics run in parallel in 15 cores.

[b]The effective sampling time was calculated as total sampling time/feasible sample fraction.

individual sample sets from ll-ACHRB for the presence of such reactions. All reactions with a non-zero flux range were sampled with non-zero flux in the toy example, iRBC283, iLJ478 and iAF987 models. The *E. coli* core model and iJO1366 each had one blocked reaction, whereas iIT341 had two. These reactions and the implications for these models will be discussed later.

4.3 Imposition of loopless condition significantly affects sample statistics

Conventional samplers do not eliminate samples containing thermodynamically unfeasible loops. Inclusion of unfeasible samples is expected to affect sample statistics. This is dramatically illustrated for the toy example in Figure 3. Inclusion of unfeasible flux samples (generated with ACHR) not only causes a change in marginal distributions, but also major differences in flux patterns, e.g. v_2 and v_5 phase plane. This suggests that greater discrepancies may be expected in statistics more suggestive of the underlying joint probability distribution (e.g. correlation coefficients), rather than in marginal statistics (e.g. mean). In the following we explore the extent of these differences.

Figure 4 compares the estimated sample statistics for the iRBC283 and iJO1366 models ignoring (ACHR) and enforcing (ll-ACHRB) the loopless condition. As expected given the absence of potentially active unfeasible loops (Table 2), there is no significant differences between the normalized means and correlation coefficients in iRBC283 using either method ($P = 0.30$ and $P = 0.42$ respectively, paired *t*-test). In contrast, there are significant differences in the sample statistics from the iJO1366 model (Fig. 3B). Both normalized means and correlation coefficients show very poor consensus between the two algorithms ($P = 3 \times 10^{-2}$ and $P = 8 \times 10^{-99}$, paired *t*-test). These results highlight the effect of the loopless condition on the sample statistics. Imposition of this condition limits the metabolic network to operate only in particular directional modes free of closed loops. These modes exert non-trivial constraints on

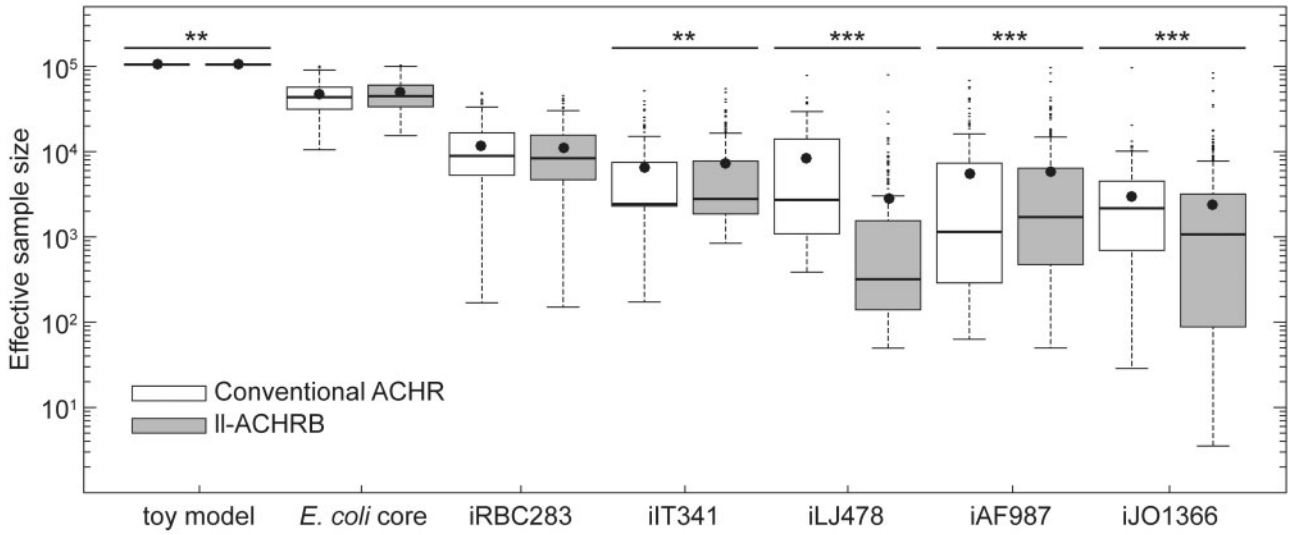

Fig. 2. Mixing properties of the tested samplers in different stoichiometric models. Box plots represent the Effective Sample Size (ESS) distribution estimated from 3×10^5 samples derived from each sampler for different metabolic models. Significant differences between the ESS medians for each case were determined using the nonparametric Wilcoxon rank sum test and are indicated by ** (*P*-value $< 10^{-3}$) and *** (*P*-value $< 10^{-4}$)

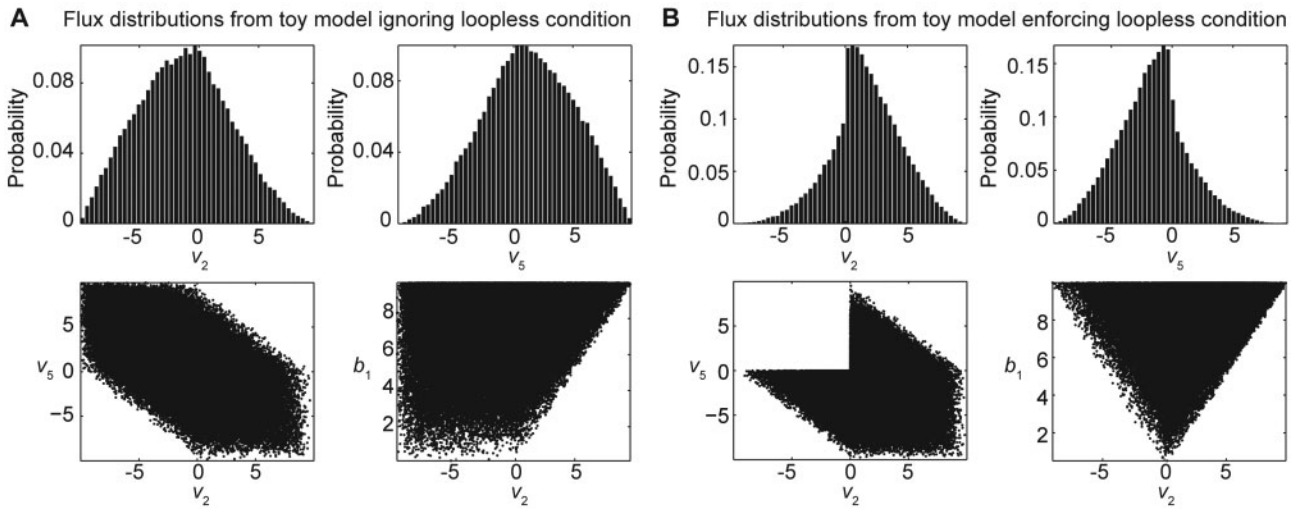

Fig. 3. Comparison of 2×10^4 samples from the toy model with $ub_1 = 10$ ignoring (ACHR) and enforcing (ll-ACHRB) the loopless condition. Inclusion of unfeasible flux samples by ACHR not only causes a reshape of marginal distributions, but also major differences in the observed flux patterns, e.g. v_2 and v_5 phase plane

each reaction, conditionally narrowing the directional space to fewer feasible combinations (hence the MILP formulation for optimizing) and yielding a non-convex feasible space. Consequently, greater differences in the estimated correlation structure are observed in the more complex iJO1366 model.

Finally, we assessed the impact of the loopless condition on co-sets of iJO1366. Co-sets are defined as groups of enzymatic reactions that are perfectly correlated ($|\rho| = 1$) and represent functional modules of the network (Xi *et al.*, 2011). We computed 10 and 21 co-sets for iJO1366 from 5×10^5 samples ignoring and enforcing the loopless condition, respectively (Supplementary Table S6). All co-sets (10) found while enforcing the loopless condition are also found when ignoring it. Interestingly, of the additional 11 unique co-sets found when ignoring the loopless condition, only one (1) can be directly linked to an unfeasible loop law. The remaining unique co-sets (10) are not only unrelated, but they involve very different pathways. This simple example illustrates some of the non-obvious and unexpected effects of the loopless condition on sample statistics.

5 Discussion

Exact calculation of the volume of a large polytope is an #P-hard problem (Dyer and Frieze, 1988), which leads to polynomial-time approximate algorithms for convex bodies (Simonovits, 2003). These algorithms typically rely on Monte Carlo sampling methods, and have proven to be particularly useful in the analysis of metabolic networks (Schellenberger and Palsson, 2009). However, rigorous enforcement of thermodynamic feasibility of flux samples has so far been overlooked, as it greatly complicates sampling due to the imposition of non-convex constraints (NP-hard problem). The significant differences in sample statistics when excluding unfeasible samples (Figs 3 and 4) highlights that thermodynamic feasibility cannot be ignored without introducing significant errors.

Here, we developed a scalable algorithm inspired by mature Monte Carlo samplers to incorporate this condition and generate only feasible samples. ll-ACHRB outperforms the conventional ACHR where unfeasible loop laws are present, and even for models without potential unfeasible loop laws, ll-ACHRB was as fast as ACHR (Table 2).

Mixing is a concern when sampling in complicated spaces. Reimers (2015) reported a high frequency of reactions with only zero-flux samples despite a non-zero flux range using ACHR to sample iJO1366. Using ll-ACHRB, we only found one blocked reaction in the iJO1366 model involved in an unfeasible loop. Apart from collecting a bigger sample, the main difference in our approach is the pre-processing of iJO1366, which first (a) compacted linear pathways and (b) removed trace elements from the biomass equation. The former reduces the dimension of the problem, while the latter should make the problem better conditioned.

The inaccessible reaction in the core *E.coli* model (fumarate reductase) highlights that caution still is required when analysing sampling results. The loopless condition dictates that fumarate reductase can only be active under anaerobic conditions and this is indeed consistent with experimental data (Unden and Bongaerts, 1997). In this instance, fumarate reductase activity under anaerobic conditions can only be accurately captured, if the model is deliberately constrained for anaerobic growth, i.e. zero oxygen uptake (Supplementary Fig. S2 Panel A). Another illustrative example of the caution needed when analysing sampling results from ll-ACHRB are

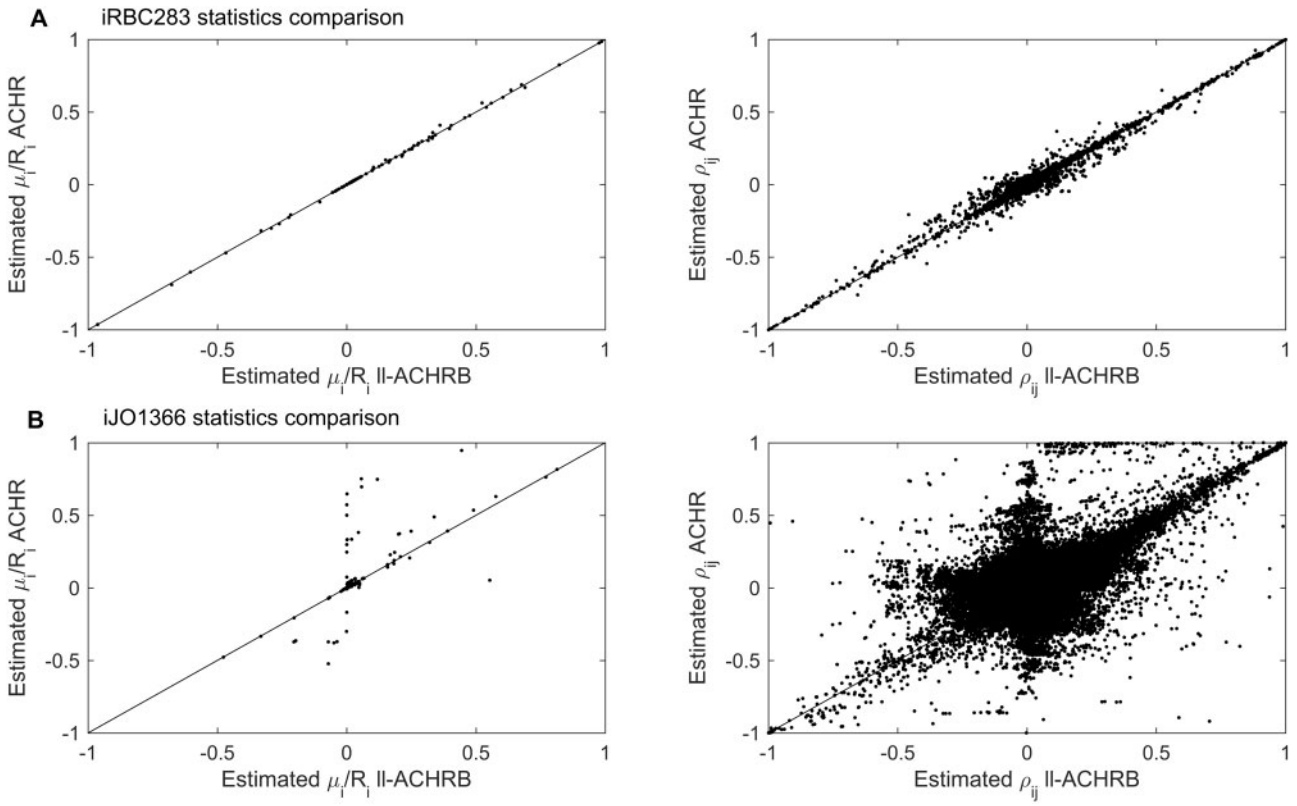

Fig. 4. Comparison of sample statistics for the iRBC283 (**A**) and iJO1366r (**B**) models estimated respectively from 5×10^5 flux samples generated ignoring (ACHR) and enforcing (ll-ACHRB) the loopless condition. Left-hand side panels compare the normalized flux means for reaction *i*, i.e. mean divided by flux range range (μ_i/R_i), whereas the right-hand side panels compare non-redundant pairwise correlation coefficients between reactions *i* and *j* (ρ_{ij})

the two blocked reactions from iIT341 (reversible O-succinylhomoserine lyase—SHSL4r—and L-threonine deaminase—THRD_L) (Supplementary Fig. S2 Panel B). Samples from ll-ACHRB show these two perfectly coupled reactions to be blocked. By constraining these two reactions to zero, the unfeasible loop is avoided. Notably, Price *et al.* (2006) have also studied this loop and have proposed to manually fix the directionalities of another two perfectly coupled reactions involved in it, namely, phosphotransacetylase (PTAr) and acetate kinese (ACKr). This measure unblocks THRD_L and SHSL4r and allows all reactions to carry flux avoiding the loop, however it massively constrains the flux range of PTAr and ACKr from 65.926 to 0.0169 mmol/gdcw/h. In this particular case, more experimental evidence is needed to decide between the alternatives. More importantly, these cases underscore the need for careful analysis of sampling results from ll-ACHRB.

We finish by mentioning possible improvements to the algorithm. De Martino *et al.* (2015) have recently reported efficient procedures for improving the performance of HR by finding an ellipsoid that closely matches the sampling space. By extracting directions from the surface of the matching ellipsoid, the performance of the traditional HR can be greatly improved. This strategy is convenient as HR guarantees convergence to the uniform distribution on the target space (Lovasz, 1999) as opposed to the non-Markovian ACHR directions choice. A combination of this strategy with the present algorithm would confer this desired guarantee; however, it is unclear how well it would scale in the case of a non-convex space. This will be subject for future research.

Acknowledgements

This research was undertaken with the assistance of resources from the Queensland Cyber Infrastructure Foundation (http://www.qcif.edu.au).

Funding

P.S. was supported by Becas-Chile, UQ Centennial and IPRS Scholarship Programs.

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.

Beard,D. *et al.* (2002) Energy balance for analysis of complex metabolic networks. *Biophys. J.*, **83**, 79–86.

Bordbar,A. *et al.* (2010) Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Molecular systems biology*, **6**, 422.

Bordbar,A. *et al.* (2011) iAB-RBC-283: a proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and patho-physiological states. *BMC Syst. Biol.*, **5**, 110.

Bordbar,A. *et al.* (2014) Minimal metabolic pathway structure is consistent with associated biomolecular interactions. *Mol. Syst. Biol.*, **10**, 737.

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

Dal'Molin,C.G.D. *et al.* (2015) A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. *Front. Plant Sci.*, **6**, 4.

De Martino,D. *et al.* (2015) Uniform sampling of steady states in metabolic networks: heterogeneous scales and rounding. *Plos One*, **10**, e0122670.

Dyer,M. and Frieze,A. (1988) On the complexity of computing the volume of a polyhedron. *SIAM J. Comput.*, **17**, 967–974.

Feist,A.M. *et al.* (2014) Constraint-based modeling of carbon fixation and the energetics of electron transfer in geobacter metallireducens. *PLOS Comput. Biol.*, **10**, e1003575.

Givens,G. and Hoeting,J. (2012) *Computational Statistics*. 2nd edition. John Wiley & Sons Inc, Hoboken, NJ.

Kass,R. *et al.* (1998) Markov Chain Monte Carlo in Practice: a roundtable discussion. *Am. Stat.*, **52**, 93–100.

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

Kiatsupaibul,S. *et al.* (2011) An analysis of a variation of hit-and-run for uniform sampling from general regions. *ACM Trans. Model. Comput. Simul.*, **21**, 3.

Lewis,N.E. *et al.* (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.*, **10**, 291–305.

Lewis,N.E. *et al.* (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.*, **28**, 1279. U1291.

Lovasz,L. (1999) Hit-and-run mixes fast. *Math. Prog.*, **86**, 443–461.

Megchelenbrink,W. *et al.* (2014) optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *Plos One*, **9**, e86587.

Mo,M.L. *et al.* (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.*, **3**, 37.

Neal,R.M. (2003) Slice sampling. *Ann. Stat.*, **31**, 705–741.

Nigam,R. and Liang,S. (2007) Algorithm for perturbing thermodynamically infeasible metabolic networks. *Comput. Biol. Med.*, **37**, 126–133.

Orth,J.D. *et al.* (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Mol. Syst. Biol.*, **7**, 535.

Price,N.D. *et al.* (2004) Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys. J.*, **87**, 2172–2186.

Price,N.D. *et al.* (2006) Candidate states of Helicobacter pylori's genome-scale metabolic network upon application of "loop law" thermodynamic constraints. *Biophys. J.*, **90**, 3919–3928.

Reed,J.L. and Palsson,B.O. (2004) Genome-scale in silico models of *E.coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res.*, **14**, 1797–1805.

Reimers,A. (2015) Obstructions to sampling qualitative properties. *Plos One*, **10**, e0135636.

Schellenberger,J. *et al.* (2011) Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.*, **100**, 544–553.

Schellenberger,J. and Palsson,B.O. (2009) Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.*, **284**, 5457–5461.

Schellenberger,J. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.*, **6**, 1290–1307.

Shoaie,S. *et al.* (2013) Understanding the interactions between bacteria in the human gut through metabolic modeling. *Sci. Rep. UK*, **3**, 2532.

Simonovits,M. (2003) How to compute the volume in high dimension? *Math. Prog.*, **97**, 337–374.

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

Thiele,I. *et al.* (2005) Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.*, **187**, 5818–5830.

Thomas,A. *et al.* (2014) Network reconstruction of platelet metabolism identifies metabolic signature for aspirin resistance. *Sci. Rep.*, **4**, 3925.

Unden,G. and Bongaerts,J. (1997) Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim. Biophys. Acta Bioenerg.*, **1320**, 217–234.

Wiback,S.J. *et al.* (2004) Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.*, **228**, 437–447.

Xi,Y.P. *et al.* (2011) Comparative study of computational methods to detect the correlated reaction sets in biochemical networks. *Brief. Bioinf.*, **12**, 132–150.

Zhang,Y. *et al.* (2009) Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science*, **325**, 1544–1549.