OXFORD

Systems biology

Probabilistic thermodynamic analysis of metabolic networks

Mattia G. Gollub (iD) **, Hans-Michael Kaltenbach** (iD) **and Jörg Stelling** (iD) *

Department of Biosystems Science and Engineering and SIB Swiss Institute of Bioinformatics, ETH Zurich, Basel 4058, Switzerland

*To whom correspondence should be addressed.

Associate Editor: Lenore Cowen

Abstract

Motivation: Random sampling of metabolic fluxes can provide a comprehensive description of the capabilities of a metabolic network. However, current sampling approaches do not model thermodynamics explicitly, leading to inaccurate predictions of an organism's potential or actual metabolic operations.

Results: We present a probabilistic framework combining thermodynamic quantities with steady-state flux constraints to analyze the properties of a metabolic network. It includes methods for probabilistic metabolic optimization and for joint sampling of thermodynamic and flux spaces. Applied to a model of *Escherichia coli*, we use the methods to reveal known and novel mechanisms of substrate channeling, and to accurately predict reaction directions and metabolite concentrations. Interestingly, predicted flux distributions are multimodal, leading to discrete hypotheses on *E.coli*'s metabolic capabilities.

Availability and implementation: Python and MATLAB packages available at https://gitlab.com/csb.ethz/pta.

Contact: joerg.stelling@bsse.ethz.ch

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

1 Introduction

Constraint-based models (CBMs) aim to characterize the metabolic capabilities of an organism, potentially at genome-scale, by predicting metabolic fluxes based on constraints arising, for example, from the structure of the metabolic network and the steady-state assumption (Orth *et al.*, 2010). Because many degrees of freedom exist for the network's (potential, if not practical) operation, a central object of the analysis is the flux space. Formally, the flux space $\mathcal{F}$ of a metabolic network with n reactions and m metabolites is the set of flux distributions $\mathbf{v} \in \mathbb{R}^n$ that satisfy steady state and additional flux constraints:

$$\mathbf{S} \cdot \mathbf{v} = 0, \quad \mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}, \tag{1}$$

where $\mathbf{S} \in \mathbb{R}^{m \times n}$ is the stoichiometric matrix (encoding the network structure) and $\mathbf{lb} \in \mathbb{R}^n$ and $\mathbf{ub} \in \mathbb{R}^n$ are measured or assumed lower and upper bounds on fluxes (encoding, e.g. reaction reversibilities).

The most common approaches to analyzing $\mathcal{F}$ predict particular flux solutions, for example, by assuming cellular objectives in flux balance analysis (FBA) (Orth *et al.*, 2010). To explore and characterize all possible solutions that lead to observed phenotypes or desired goals, uniform sampling (US) (Haraldsdóttir *et al.*, 2017; Price *et al.*, 2004) and its extensions in the form of loopless (Desouki *et al.*, 2015; Saa and Nielsen, 2016) and non-uniform sampling (Binns *et al.*, 2015; Heinonen *et al.*, 2019; Keaty and Jensen, 2020) provide a comprehensive description of the flux space.

Compared to characterizations by metabolic pathway analysis (Terzer and Stelling, 2008), sampling predictions directly relate to experimental flux distributions.

However, available constraints often restrict the flux space insufficiently. Different -omics data from high-throughput measurement technologies can in principle be used to derive additional constraints, but data integration is challenging and the resulting constraints are condition-specific (Ramon *et al.*, 2018). In contrast, constraints based on thermodynamics derive from first principles; they are absolute and independent of specific kinetic properties or regulatory mechanisms. Therefore, modeling thermodynamic laws has the potential to reveal the options available to the cell.

In a metabolic model, thermodynamics quantifies the favorability of a set Γ of γ reactions (excluding unbalanced pseudo-reactions such as biomass and exchange reactions) with Gibbs reaction energies $\Delta_{\mathbf{r}}\mathbf{G}' \in \mathbb{R}^{\gamma}$:

$$\Delta_{\mathbf{r}}\mathbf{G}' = \Delta_{\mathbf{r}}\mathbf{G}'^{\circ} + RT \cdot \mathbf{S}_{\Gamma}^{\top} \cdot \ln \mathbf{c}, \tag{2}$$

where $\mathbf{c} \in \mathbb{R}^m$ are the compartment-specific metabolite concentrations, $\Delta_{\mathbf{r}}\mathbf{G}'^{\circ} \in \mathbb{R}^{\gamma}$ the standard Gibbs reaction energies [corrected for pH, ionic strength and transport between compartments (Haraldsdóttir *et al.*, 2012)], R the gas constant, T the temperature and $\mathbf{S}_{\Gamma}^{\top}$ is the transpose of the stoichiometric sub-matrix corresponding to Γ. The second term of (2) (here called the *activity term*) maps the contribution of metabolite concentrations to the reactions they

participate in. Reaction energies constrain fluxes via the second law of thermodynamics, which implies that positive (negative) net flux can only occur if the free energy of a reaction is negative (positive). This ensures that biochemical processes result in net energy dissipation.

How well thermodynamics constrain a model depends on the precision of the estimates of the variables in (2). Standard reaction energies can be estimated from measured equilibrium constants (Goldberg *et al.*, 2004) by the group contribution method and its extensions (Du *et al.*, 2018a; Jankowski *et al.*, 2008; Noor *et al.*, 2013). Because multiple reactions may affect the same chemical groups, their estimates are often linearly dependent and thus lie in a lower-dimensional subspace (Fig. 1A). Metabolomics data can provide precise estimates or approximate distributions (Park *et al.*, 2016) for metabolite concentrations (Fig. 1B).

Thermodynamics-based Metabolic Flux Analysis (TMFA) (Henry *et al.*, 2007) integrates thermodynamics into CBMs with (i) a set of linear constraints describing (2) and bounds on its quantities and (ii) integer constraints enforcing the second law of thermodynamics. TMFA has been used to predict metabolite concentrations (Henry *et al.*, 2007; Park *et al.*, 2016) and pathway favorability (Chiappino-Pepe *et al.*, 2017; Du *et al.*, 2018b). However, it describes the uncertainty by independent error bounds for each variable. This facilitates the construction of optimization problems, but also over-approximates the uncertainty. Figure 1C illustrates a space of thermodynamic variables in TMFA: independent constraints on $\ln \mathbf{c}$ and $\Delta_\mathbf{r}\mathbf{G}'^\circ$ constrain $\Delta_\mathbf{r}\mathbf{G}'$ as well. In addition, the steady-state condition (1) excludes orthants that imply directions without a feasible flux distribution in $\mathcal{F}$. This approach discards any correlation between the $\Delta_r G_i'^\circ$ of different reactions and the probability of a particular solution.

The common construction of CBMs for pure flux-based analysis poses a further challenge. Model curators or automatic pipelines regularly need to prevent unrealistic internal or ATP-generating cycles (Schilling *et al.*, 2000). Because the flux balance framework does not resolve this problem consistently, assumptions or condition-specific biological knowledge are employed to assign reaction reversibilities. Importantly, missing knowledge or inaccurate irreversibilities can impose false thermodynamic constraints that do not affect flux balance and thus remain undetected.
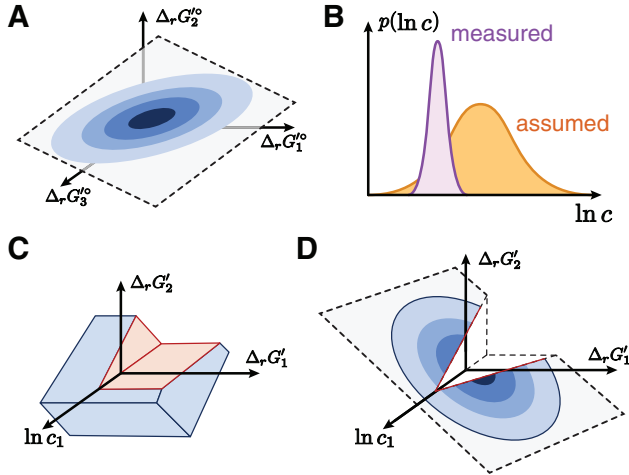
To address these challenges, here we propose *Probabilistic Thermodynamic Analysis* (PTA), a framework in which the uncertainty of free energies and concentrations is modeled with a joint probability distribution (Fig. 1D). We develop optimization and sampling methods based on PTA and show their application to predict substrate channeling, reaction directions and metabolite concentrations, as well as to explore the capabilities of the metabolic network.

2 Materials and methods

2.1 The steady-state thermodynamic space

To account for correlations between the uncertainties of reaction energies of different reactions, we assume that metabolite concentrations follow a log-normal distribution based on measurements or an assumed physiological distribution (Park *et al.*, 2016)

$$\ln \mathbf{c} \sim \mathcal{N}(\mu_\mathbf{c}, \boldsymbol{\Sigma}_\mathbf{c}). \tag{3}$$

From the estimates of a group contribution method (Noor *et al.*, 2013), we obtain a normal distribution for the standard reaction energies

$$\Delta_\mathbf{r}\mathbf{G}'^\circ \sim \mathcal{N}(\mu_\circ, \boldsymbol{\Sigma}_\circ). \tag{4}$$

From (2), it follows that the reaction energy estimates are also normally distributed as

$$\Delta_\mathbf{r}\mathbf{G}' \sim \mathcal{N}\left(\mu_\mathbf{r} = \mathbf{R} \cdot \begin{bmatrix} \mu_\mathbf{c} \\ \mu_\circ \end{bmatrix}, \ \boldsymbol{\Sigma}_\mathbf{r} = \mathbf{R} \cdot \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{c} & 0 \\ 0 & \boldsymbol{\Sigma}_\circ \end{bmatrix} \cdot \mathbf{R}^\top\right), \tag{5}$$

where $\mathbf{R} = [RT \cdot \mathbf{S}_\Gamma^\top, \mathbf{I}]$. For computational convenience, we define the vector $\mathbf{t} = [\ln \mathbf{c}, \Delta_\mathbf{r}\mathbf{G}'^\circ, \Delta_\mathbf{r}\mathbf{G}']^\top$. It is again normally distributed with mean $\mu_\mathbf{t} = [\mu_\mathbf{c}, \mu_\circ, \mu_\mathbf{r}]^\top$ and covariance matrix

$$\boldsymbol{\Sigma}_\mathbf{t} = \begin{bmatrix} \boldsymbol{\Sigma}_\mathbf{c} & 0 & RT \cdot \boldsymbol{\Sigma}_\mathbf{c} \cdot \mathbf{S}_\Gamma \\ 0 & \boldsymbol{\Sigma}_\circ & \boldsymbol{\Sigma}_\circ \\ RT \cdot \mathbf{S}_\Gamma^\top \cdot \boldsymbol{\Sigma}_\mathbf{c} & \boldsymbol{\Sigma}_\circ & \boldsymbol{\Sigma}_\mathbf{r} \end{bmatrix}. \tag{6}$$

We use this distribution to constrain $\mathbf{t}$ to values that are compatible with the estimated concentrations and standard free energies up to a confidence level α, such that

$$(\mathbf{t} - \mu_\mathbf{t})^\top \cdot \boldsymbol{\Sigma}_\mathbf{t}^+ \cdot (\mathbf{t} - \mu_\mathbf{t}) \leq \chi^2_{q;\alpha}, \tag{7}$$

where $\boldsymbol{\Sigma}_\mathbf{t}^+$ is a pseudo-inverse of $\boldsymbol{\Sigma}_\mathbf{t}$, of rank $q = \text{rank}(\boldsymbol{\Sigma}_\mathbf{t}^+) = \text{rank}(\boldsymbol{\Sigma}_\mathbf{t})$ and $\chi^2_{q;\alpha}$ is the α-quantile of the χ^2-distribution with q degrees of freedom. We exploit that q is much smaller than the dimension of $\mathbf{t}$ by using a q-dimensional standard normal distribution to express $\mathbf{t}$ as

$$\mathbf{t} = \mu_\mathbf{t} + \mathbf{Q} \cdot \mathbf{m}, \tag{8}$$

where $\mathbf{Q} = \boldsymbol{\Sigma}_\mathbf{t}^{1/2}$ is a matrix square root such that $\boldsymbol{\Sigma}_\mathbf{t} = \mathbf{Q}\mathbf{Q}^\top$.

Lastly, we enforce the second law of thermodynamics and require that all reactions in Γ have a well-defined direction, that is, non-zero flux. For each i, $1 \leq i \leq \gamma$:

$$v_{\Gamma(i)} \cdot \Delta_r G_i' < 0. \tag{9}$$

Note that this requires reactions that are structurally blocked (e.g. by involving dead-end metabolites) to be removed in advance; they would conflict with (9).

We define the *steady-state thermodynamic space* (or short: *thermodynamic space*) $\mathcal{T}$ as the set of $\mathbf{t}$ that fulfill (7) and satisfy the steady-state constraints (1), (9). In other words, $\mathcal{T}$ defines the possible reaction energies and metabolite concentrations at steady-state.

2.2 Probabilistic metabolic optimization (PMO)

The definitions of $\mathcal{F}$ and $\mathcal{T}$ allow us to extend the FBA framework to include thermodynamic quantities similar to TMFA. We translate (1) and (7) into linear and quadratic constraints and rewrite (9) with

Fig. 1. Integration of thermodynamics in CBMs. (**A**) Group contribution methods provide estimates of standard reaction energies. The uncertainties of multiple estimates may be correlated and lie in a lower dimensional subspace. (**B**) Probability distributions of metabolite concentrations can be constructed from measured or assumed data. (**C**) The space of reaction energies and metabolite concentrations defined by TMFA, with regions that are not feasible at steady-state removed (red). The space is defined by independent bounds on each variable of $\ln \mathbf{c}$ and $\Delta_\mathbf{r}\mathbf{G}'^\circ$, leading to an overapproximation of the uncertainty. (**D**) PTA models the complete uncertainty information probabilistically, leading to a joint probability distribution over the reaction energies of the entire network. Steady-state flux constraints restrict it to feasible orthants

integer constraints in the *big-M* formulation (Williams, 2013). The resulting Mixed-Integer Quadratically Constrained Program (MIQCP) can be used to perform FBA-like analyses under thermodynamic constraints.

Since we know the probability distribution of $\mathbf{t}$, we can also determine the most probable reaction energies and metabolite concentrations by setting a quadratic objective that maximizes the probability of $\mathbf{t}$ under steady-state constraints:

$$\mathbf{m}^* = \operatorname*{argmin}_{\mathbf{m},\mathbf{v}} \quad \|\mathbf{m}\|_2^2 \tag{10}$$
$$\text{s.t. Eq. (1), Eq. (7), Eq. (9).}$$

We call $\mathbf{t}^* = \mu_{\mathbf{t}} + \mathbf{Q} \cdot \mathbf{m}^*$ the *mode* of the distribution of $\mathbf{t}$.

2.3 Thermodynamic assessment of metabolic models

Structural assessment aims to detect *thermodynamic inconsistencies* in a model, that is, groups of reactions that, given their irreversibility annotation, cannot satisfy thermodynamic constraints with any assignment of concentrations and standard free energies. For example, two stoichiometrically equivalent paths can be irreversible in opposite directions; then, all non-zero flux solutions contain a thermodynamically infeasible internal cycle.

To resolve such cases, we first block all reactions related to exchanges, biomass and ATP maintenance and enumerate all internal Elementary Flux Modes (EFMs) (Terzer and Stelling, 2008). Next, we select all the *forced internal cycles*, that is, the internal EFMs that must be active in all non-zero flux distributions. We find these cycles using linear programming to verify that changing any of the signs of an EFM in the original model always results in an unfeasible problem. We then inspect the reactions in the forced internal cycles to find and resolve inconsistent irreversibilities.

Quantitative assessment of a model aims to find features in PMO predictions indicative of knowledge gaps or model inaccuracies. First, we construct a condition-specific model with available flux and concentration constraints from a base model and use PMO to estimate $\mathbf{t}^*$. Next, we compute the *z-score* z_i for each predicted value t_i^*, which measures the deviation between t_i^* and μ_i in units of standard deviation (see Supplementary Information). We focus on metabolite concentrations and classify predictions with $|z_i|$ greater than a threshold θ (or $c_i \geq 10$mM for non-intracellular metabolites) as *anomalies* and investigate whether those indicate missing or unknown mechanisms or model inaccuracies. Finally, we curate the base model with the results of the search. The values in $\mathbf{t}^*$ represent the unlikely case in which predicted standard free energies are as close as possible to the mean estimate and reactions can operate close to the thermodynamic equilibrium. Because the predicted deviations likely underestimate the true deviations, we use a conservative $\theta = 1$. We repeat this process for a selected set of representative conditions. Note that curation actions are applied to the base model—further condition-specific models can be constructed without needing to curate each condition manually.

2.4 Thermodynamics and flux sampling (TFS)

We use sampling to characterize $\mathcal{T}$ in terms of probability distributions of metabolite concentrations as well as probabilities for the different flux modes (see algorithm 1). Geometrically, $\mathcal{T}$ is the intersection between an ellipsoid and the orthants with at least one steady-state flux distribution satisfying flux and thermodynamic constraints. This space cannot be defined explicitly, and is usually neither convex nor connected. In fact, an n-dimensional space has 2^n orthants, each corresponding to a possible combination of reaction directions in a metabolic network. Clearly, the enumeration—let alone verification of the steady-state condition—of all orthants of $\mathcal{T}$ is infeasible with many reversible reactions. We therefore developed a Markov Chain Monte Carlo (MCMC) method based on the Hit-and-Run sampler for General Target Distributions (Bélisle *et al.*, 1993) to sample $\mathcal{T}$ (Fig. 2).

Choice of initial points. For MCMC simulations it is generally advised to simulate multiple chains, choosing the starting points

---

Algorithm 1: Sampling the steady state thermodynamic space $\mathcal{T}$

Input : $\mathbf{S}$, $\mathbf{lb}$, $\mathbf{ub}$, $\mathbf{Q}$, $\mu_{\mathbf{t}}$, ellipsoid $\mathcal{E}$, linear inequalities $\mathcal{L}$
Output: Set of $\Delta_{\mathbf{r}}\mathbf{G}'$ samples Δ, orthant counts W

$\Delta \leftarrow W \leftarrow \emptyset$;
$\mathbf{s} \leftarrow$ find point $\mathbf{m}$ satisfying Eq. (1), Eq. (7), Eq. (8) and Eq. (9) ;
repeat
   $\mathbf{d} \leftarrow$ random direction;
   $\mathbf{e} \leftarrow$ componentwise division of $\mathbf{Qs} + \mu_{\mathbf{t}}$ by $\mathbf{Qd}$;
   $\mathbf{e} \leftarrow [-\infty,\ \mathrm{sort}(\mathbf{e}),\ \infty]$;
   $I \leftarrow I_w \leftarrow \emptyset$;
   for $i \leftarrow 0$ **to** $|\mathbf{e}| - 2$ **do**
      $(u, v) \leftarrow$ intersect (e_i, e_{i+1}) with $\mathcal{E}$ and $\mathcal{L}$;
      if (u, v) *exists* **then**
         $\Delta_{\mathbf{r}}\mathbf{G}'_{uv} \leftarrow \mathbf{Q}\left(\mathbf{s} + \frac{u+v}{2} \cdot \mathbf{d}\right) + \mu_{\mathbf{t}}$;
         if $\Delta_{\mathbf{r}}\mathbf{G}'_{uv}$ *allows solution in* $(\mathbf{S}, \mathbf{lb}, \mathbf{ub})$ **then**
            $I \leftarrow I \cup \{(u, v)\}$;
            $I_w \leftarrow I_w \cup \{\ \text{normal CDF in } (u, v) \text{ on ray } (\mathbf{s}, \mathbf{d})\ \}$;

   $(u, v) \leftarrow$ sample from I with weights I_w;
   $\mathbf{s} \leftarrow$ sample from truncated normal in (u, v) on ray $(\mathbf{s}, \mathbf{d})$;
   $\Delta \leftarrow \Delta \cup \{\mathbf{Qs} + \mu_{\mathbf{t}}\}$;
   $\mathbf{o} \leftarrow \mathrm{sign}(\mathbf{Qs} + \mu_{\mathbf{t}})$;
   $W[\mathbf{o}] \leftarrow W[\mathbf{o}] + 1$;
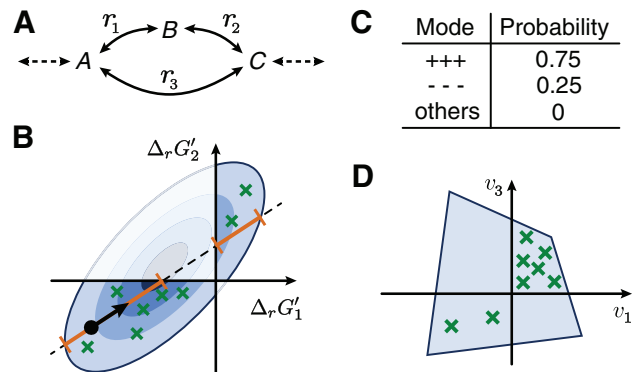**until** *stopping condition reached* ;

Fig. 2. Overview of TFS. (**A**) Example network, where left-to-right fluxes are positive. At steady state $v_2 = v_1$ and $\Delta_{\mathbf{r}}\mathbf{G}'_3 = \Delta_{\mathbf{r}}\mathbf{G}'_1 + \Delta_{\mathbf{r}}\mathbf{G}'_2$. (**B**) We sample the thermodynamic space using a modified Hit-and-Run algorithm. At each step, we limit sampling to portions of the space that allow steady-state flux distributions (orange). (**C**) As a result, we estimate the probability of each mode. (**D**) Finally, we sample the flux space drawing from each mode a number of samples proportional to its probability

---

from an over-dispersed distribution, to detect if the chains explore the entire space. For each reversible reaction, we use PMO to find two sets of reaction directions over the network: one that maximizes the flux through the reaction, and one that minimizes it. Then, again using PMO, for each set of directions we search for a point $\mathbf{s_i}$ in $\mathcal{T}$ that satisfies the directionality constraints. We require each $\mathbf{s_i}$ to have maximal distance from the boundary of the respective space to reduce the risk of numerical errors. Finally, we start simulating one chain from each $\mathbf{s_i}$.

Finding the intersection I. At each step of Hit-and-Run we first intersect the ray with the ellipsoid defined by (7) and a set of linear inequalities constraining the signs of free energies of irreversible reactions. Without a close-form description of the boundary of the space, finding the orthants that allow steady-state flux solutions is challenging. Despite the exponential number of orthants, a ray in an n-dimensional space only crosses at most $n + 1$ of them. Thus, at each step, we identify all orthants intersected by the ray and discard

those that do not allow a steady-state flux distribution. We verify the steady-state condition with a linear program that checks the existence of a flux distribution satisfying (1) and the directionality constraints of the orthant. For each feasible orthant, we add its intersection with the ray to I.

Sampling from I. We sample the next point along a ray proportionally to the target probability distribution, which is a Multivariate Normal (MVN). First, we restrict the MVN to a normal distribution along the 1-dimensional ray. We then compute the Cumulative Distribution Functions (CDFs) for each valid segment of I to determine its weight, sample a valid segment according to these weights and finally sample a point in the segments according to the associated truncated normal distribution on the segment. We optionally store the sample and the samples count for the selected orthant to estimate the orthant's probability.

Reducing the dimensionality. Our method for sampling $\mathcal{T}$ so far neglects that only reaction free energies determine flux directions, and that free energies are determined only by concentration ratios, not absolute concentrations. Thus, we can reduce the dimensionality of the problem by not representing metabolite concentrations explicitly and solving (4) using μ_r and Σ_r instead of μ_t and Σ_t. This allows faster convergence of the chains. Afterwards, for each sample of $\Delta_r G_i'$, we sample the corresponding $\ln c$ and $\Delta_r G'^\circ$ from the distribution of t conditioned on $\Delta_r G_i'$. This distribution can be constructed from the *Schur complement* of Σ_r in Σ_t and has the form of an MVN, which can be sampled efficiently:

$$\begin{bmatrix} \ln c \\ \Delta_r G'^\circ \end{bmatrix} \bigg| \Delta_r G_i' \sim \mathcal{N}\left(\begin{bmatrix} \mu_c \\ \mu_\circ \end{bmatrix} + B\Sigma_r^{-1}(\Delta_r G_i' - \mu_r), \begin{bmatrix} \Sigma_c & 0 \\ 0 & \Sigma_\circ \end{bmatrix} - B\Sigma_r^{-1}B^\top \right),$$
(11)

where $B = [RT\Sigma_c S_\Gamma \; \Sigma_\circ]^\top$ is a submatrix of Σ_t.

Convergence. Since each sampled orthant is full-dimensional and all orthants are reachable, asymptotic convergence to arbitrary target distributions is guaranteed (Bélisle *et al.*, 1993). However, there is no theoretical bound on the convergence rate. We therefore asses the quality of the sampled chains using recommended criteria on the chains' Potential Scale Reduction Factor (PSRF) and Effective Sample Size (ESS) (Gelman *et al.*, 2013). Several hit-and-run samplers reformulate the problem in isotropic parametrization to improve convergence (Haraldsdóttir *et al.*, 2017). Since the distribution of m is a standard MVN, the parametrization already has a potentially high isotropy but this may not be optimal because flux constraints exclude parts of the solution space. To improve convergence, we initially run shorter chains and compute the PSRF of each reaction energy. We then adjust the parametrization such that the space is better explored along dimensions showing high PSRFs.

Flux sampling with thermodynamic prior. Finally, we sample the flux space using the estimated distributions of reaction directions as prior. We assume that the probability of an orthant is determined by its probability in the thermodynamic space, for which we already obtained estimates. For each direction sample $d_i \in \{-1, 1\}^\gamma$, one can easily construct and sample the flux space defined by (1) and constrained by d_i using Coordinate Hit-and-Run with Rounding (CHRR) (Haraldsdóttir *et al.*, 2017). However, this is computationally prohibitive for models with millions of thermodynamically realistic orthants. We therefore approximate the distribution in flux space by selecting n_o random orthants according to their relative probability. For each orthant, we then use CHRR to draw a number of flux samples proportional to its probability.

2.5 Data, models and implementation

We validated our methods on data for *Escherichia coli* growing on different carbon sources (Gerosa *et al.*, 2015) with models based on iML1515 (Monk *et al.*, 2017). The dataset contains measured growth rates, uptake and secretion rates for 10 metabolites, absolute concentrations for 42 metabolites and 13C estimates for 26 fluxes in core metabolism. Measurements were performed in minimal media with eight different carbon sources. iML1515 did not admit solutions satisfying the measured rates on galactose and gluconate,

potentially due to its fixed biomass composition and ATP maintenance. Thus, we only performed simulations on acetate, fructose, glucose, glycerol, pyruvate and succinate.

We generated models in three steps: (i) We used *NetworkReducer*'s lossy reduction (Erdrich *et al.*, 2015) to simplify the subsystems related to lipids and cofactors metabolism. (ii) We applied the results of the thermodynamic curation and integrated condition-specific constraints such as media composition and measured growth, uptake and secretion rates. The distribution of each intracellular metabolite was set to the log-normal distribution of all metabolites in the acetate and glucose conditions. We used the same mean, but a 5-fold larger standard deviation, for periplasmic and extracellular metabolites (except the ones provided in the media, for which we used the actual concentrations). Measured intracellular concentrations were added only for thermodynamic assessment and when specified in Results. (iii) We applied *NetworkReducer*'s lossless compression and manual simplifications to obtain a reduced model. We focused on keeping carbon, amino acid and nucleotide metabolism untouched; hence we will refer to the generated models as iML1515-CAN. The models used for TMFA and US simulations were generated with the same approach, but without integrating the results of thermodynamic curation. For PTA and TMFA simulations, the set of thermodynamically constrained reactions Γ consisted of all reactions except for biomass, exchange, water transport and export of all metabolites other than CO2. The dimensions of the models were $881 \leq n \leq 883$, $500 \leq m \leq 503$ and $715 \leq \gamma \leq 721$.

We implemented the TFS and CHRR (used also for comparisons against US) samplers in C++ and validated them against established approaches (see Supplementary Information). We provide Python and MATLAB (Mathworks, Natick, MA) interfaces for all methods, using COBRApy (Ebrahim *et al.*, 2013) and the COBRA Toolbox (Heirendt *et al.*, 2019) for model manipulation and the Gurobi Optimizer (Gurobi Optimization, Beaverton, OR) to solve PMO problems.

3 Results

3.1 Thermodynamic assessment of iML1515-CAN

First, we searched for inconsistent irreversibilities in our reduced version of the *E.coli* model iML1515 (Monk *et al.*, 2017; see Section 2). Seven irreversible reactions had to be set to reversible to make the model thermodynamically feasible. Manual inspection of the standard reaction energies confirmed that all these reactions could be reversed at physiological conditions. For example, we identified a forced internal cycle in propionate metabolism (Fig. 3A), which we addressed by making acetate-CoA ligase (ACCOAL) and propanoyl-CoA: succinate CoA-transferase (PPCSCT) reversible and by allowing secretion of propionate, as it must be themodynamically favorable.

After resolving the remaining reversibility conflicts, we searched for quantitative anomalies. We selected one glycolytic (glucose) and one gluconeogenic (actetate) growth condition and applied PMO to the corresponding models constrained by the measured concentrations.

We identified 36 unique anomalies in the two models, summarized in Table 1 and detailed in Supplementary Information. The most common reason for anomalies (16 metabolites) is the possible occurrence of *substrate channeling*. Thermodynamics-based modeling assumes a well-mixed system. If two or more enzymes catalyzing consecutive steps of a pathway assemble in a complex, reaction products may find themselves close to the binding pockets of the subsequent enzymes, which increases the reaction probability. In such cases, the intermediate is said to be *channeled* between the two enzymes, breaking the assumption of homogeneity. As a result, reactions that would require unusually high or low concentrations of a reactant can become feasible even at unfavorable intracellular concentrations. An example highlighted by PMO are the first two steps of proline synthesis, *ProB* and *ProA* (Fig. 3D). Constrained by the measured concentrations of glutamate and energy cofactors, PMO
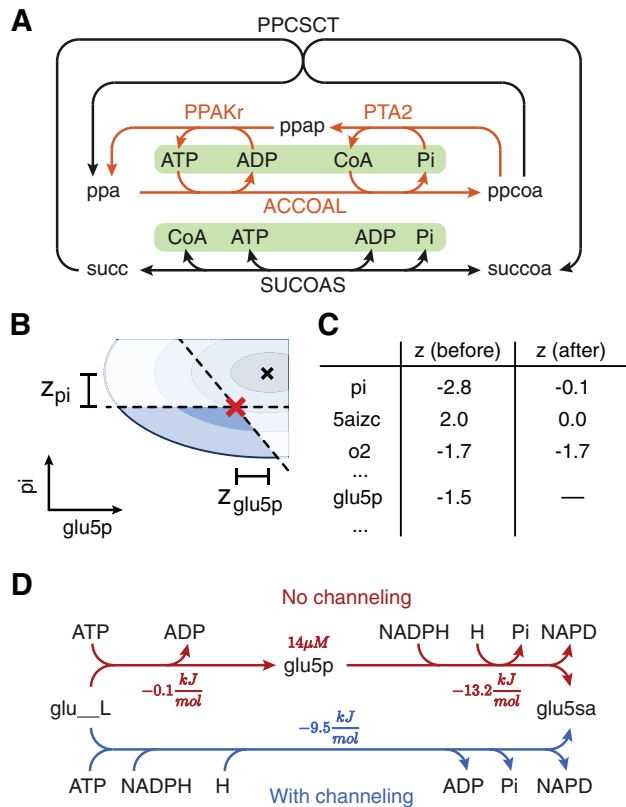
Fig. 3. Thermodynamic assessment of iML1515-CAN. (A) Forced internal cycle in propionate metabolism (orange). The irreversibility of ACCOAL conflicts with the direction of phosphate acetyltransferase (PTA2) and propionate kinase (PPAKr). Moreover, due to the direction of PPCSCT and a shared pool of cofactors (green), it also incorrectly implies that conversion of succinyl-CoA to succinate in the TCA by succinyl-CoA synthetase (SUCOAS) is unfavorable. (B) After using PMO to find t^* (red cross), we compute the z-scores of each metabolite and metabolites, i.e. the normalized distance from the mean (black cross). (C) Metabolites with high z-score ('before') are selected. After literature search, 29 out of the 36 flagged by the analysis led to improvements in the model, as shown by smaller z-scores in the curated model ('after'). (D) A symptomatic case suggesting substrate channeling. An unfavorable reaction produces an intermediate (glutamate-5-phosphate) that must have low concentration to maintain thermodynamic feasibility and is followed by a favorable reaction. The total reaction energies with and without channeling are not identical because the estimated concentrations differ

Table 1. Summary of the anomalies in the concentrations predicted by PMO and the resulting modifications to the model in the curation (third column: number of reactions added, removed or converted to lumped reactions)

Reasons for anomalies	Anomalies	Reactions
Substrate channeling with literature support	9	16
Substrate channeling in gluconeogenesis	3	10
Substrate channeling without literature support	4	0
Incomplete stoichiometry	2	2
Inaccurate irreversibility annotations	6	9
Inaccurate group contribution estimates	4	3
Artifacts of steady-state constraints	5	5
False positives	3	0
Total	36(7.2%)	45(6.2%)

Note: Percentage values are relative to m and γ, respectively.

predicts a best-case maximum concentration of $14\mu M$ for the intermediate glutamate-5-phosphate, 1.4 standard deviations below the mean estimate. Thus, we hypothesized that glutamate-5-phosphate is channeled between *ProB* and *ProA*, relieving the thermodynamic

constraint on its intracellular concentration, which is consistent with literature (Marco-Marín *et al.*, 2007). Consequently, we updated the model by replacing the two steps with a single net reaction.

We predicted and found literature support for substrate channeling in several other pathways (see Supplementary Information for details). Most interestingly, with available metabolomics data, four steps of gluconeogenesis (phosphoglycerate mutase, phosphoglycerate kinase, triose-phosphate isomerase and fructose-biphosphate aldolase) appear unfavorable, suggesting enzyme complexes with more favorable steps (phosphoenolpyruvate carboxykinase and fructose-bisphosphatase) that promote substrate channeling. This is consistent with 13C and kinetic studies showing channeling of glycolytic compounds (Abernathy *et al.*, 2019; Rakus *et al.*, 2004).

In addition, two anomalies revealed that the first step of serine synthesis, phosphoglycerate dehydrogenase, is unfavorable with NAD as cofactor. It is known that a more complex mechanism couples the oxidation of 3-phosphoglycerate to the reduction of ubiquinone, which is more favorable (Zhang *et al.*, 2017). We attributed 15 anomalies to inaccurate $\Delta_r G°$ estimates, inaccurate irreversibility annotations and cases in which dilution effects likely dominate the steady-state constraint. Only three anomalies were false positives in which concentrations far from the average are expected, as for intracellular oxygen. Overall, 91.7% of the metabolites flagged after PMO reflected model inaccuracies or knowledge gaps, leading to the curation of 45 reactions, 6.2% of the reactions with thermodynamic constraints.

3.2 Sampling modes and directions in iML1515-CAN

We used TFS to sample the thermodynamic space of iML1515-CAN in different growth conditions, both with ($M+$, all conditions) and without ($M-$, all conditions except glucose and acetate, where metabolomics data have been used to curate the model) integrating metabolomics data. Specifically, we collected 10^8 samples of orthants and 10^5 samples of $\Delta_r G'$. In each condition, PMO successfully found initial points and the sampler reached convergence to the target distribution according to common metrics (see Section 2). Interestingly, TFS only needed 401 variables to represent $\Delta_r G'$ of the 717 reactions in Γ. In models constrained only with measured growth and exchange rates, we discovered 16.5×10^6 orthants, out of which 11.5×10^6 suffice to cover 95% of the thermodynamic space. These numbers decreased to 11.3×10^6 and 7.2×10^6 once we integrated measured metabolite concentrations. Overall, this represents a substantial reduction over the average theoretical maximum number of reversible reactions (computed from the number of reversible reactions) in the TMFA-constrained models ($M-$: 5.8×10^{22}, $M+$: 3.1×10^{20}). Hence, correlations in thermodynamic space may impose important constraints on possible combinations of flux directions.

We compared the predicted reaction directions of TFS, TMFA and US on TMFA-constrained models and validated them against 13C estimates. For TFS and US, a prediction was considered irreversible when at least 95% of the samples represented the same direction. On average, TMFA could only constrain a small fraction of the models' 71–79 reversible reactions, in particular without metabolomics data (Fig. 4A). US yielded the most constrained predictions, while TFS predicted irreversibilities for about half of the 102–105 reversible reactions in the PMO-curated models. Despite having the highest precision, US also showed the lowest accuracy: it incorrectly predicted the direction of part of the 11–15 reversible fluxes with available 13C estimates. Incorrect predictions focus on two main pathways in gluconeogenic conditions: upper glycolysis, where US systematically prefers fructose 6-phosphate aldolase over fructose-bisphosphate aldolase, and the non-oxidative pentose phosphate pathway. In one condition, these choices result from TMFA incorrectly predicting fructose-bisphosphate aldolase to be irreversible. In all conditions, the irreversibilities predicted by TFS were consistent with the 13C estimates.

Beyond the validated fluxes, we found cases ($M-$: 11–18, $M+$: 9–13) where both US and TFS predicted a reaction to be irreversible, but in different directions, mostly for reactions that can participate
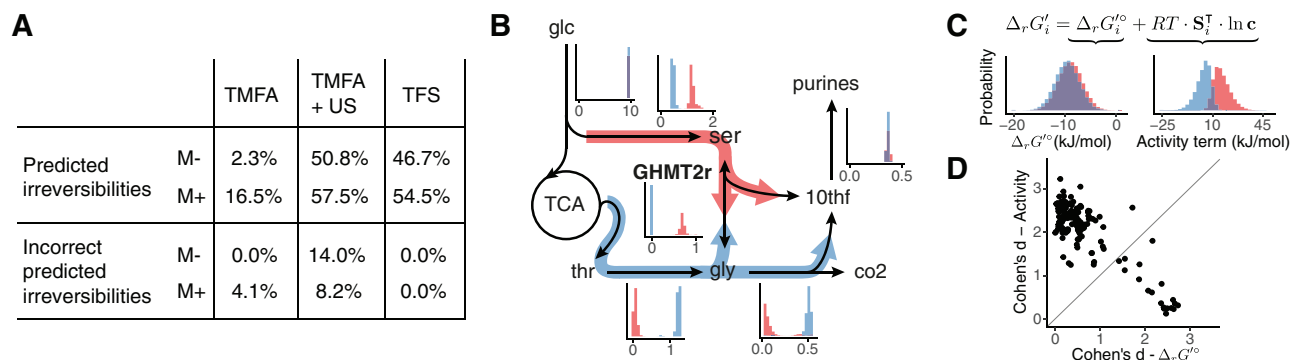
A

		TMFA	TMFA + US	TFS
Predicted irreversibilities	M−	2.3%	50.8%	46.7%
	M+	16.5%	57.5%	54.5%
Incorrect predicted irreversibilities	M−	0.0%	14.0%	0.0%
	M+	4.1%	8.2%	0.0%

B

C

$$\Delta_r G_i' = \underline{\Delta_r G_i'^\circ} + \underline{RT \cdot \mathbf{S}_i^\intercal \cdot \ln \mathbf{c}}$$

D

Fig. 4. Reaction directions and fluxes. (**A**) Precision and accuracy of predicting directionalities with different methods. Predicted irreversibilities are given as percentage of the average number of reversible reactions in the different conditions. The percentage of incorrectly predicted irreversibilities is relative to the average number of fluxes that are reversible in the models and have 13C estimates. (**B**) Example of multimodal predictions of fluxes with TFS for the synthesis of purines and glycine. Orange: serine is converted to glycine, donating one carbon to form 10-formyltetrahydrofolate (10thf), a cofactor for purine synthesis. Blue: glycine is synthesized from threonine and is converted to CO2 and 10thf. Histograms show predicted flux distributions, clustered by the direction of GHMT2r. (**C**) Distribution of $\Delta_r G'^\circ$ (left) and the activity term (right) of GHMT2r, clustered by the direction of the reaction. $\mathbf{S}_i^\intercal$ is the i-th row of $\mathbf{S}_\Gamma^\intercal$. (**D**) Cohen's d for the distributions of the standard reaction energies and of the activity terms in the two clusters for all reactions in Γ

in internal cycles and when alternative paths exist for the synthesis of the same compound. For example, 5-phospho-alpha-D-ribose 1-diphosphate is generally synthesized directly from ribose 5-phosphate (as predicted by TFS), but US predicts a conversion in the opposite direction and 5-phospho-alpha-D-ribose 1-diphosphate synthesis through a longer and stoichiometrically equivalent path via ribose 1-phosphate and ribose 1,5-diphosphate. While literature suggests that the directions predicted by TFS are correct (see Supplementary Information), the available data does not allow for a systematic validation. These differences highlight the effect of using different spaces to determine reaction directions. In US, the probability of a direction is determined by the fraction of flux space that allows it; this favors directions that impose minimal constraints on the rest of the network. In contrast, TFS predicts directions based on the probability of having a favorable reaction energy.

3.3 Exploration of metabolic modes
Using the samples of orthants and their weights we sampled the flux space of 10^4 orthants for each condition. TFS predicted millions of modes in iML1515-CAN, but focusing on the predicted flux distribution of individual reactions one observes fewer than 100 *marginal modes*, which have simpler interpretation. An example are the two modes defined by the direction of glycine hydroxymethyltransferase (GHMT2r) (Fig. 4B). Physiologically, serine is used to synthesize glycine and feed 1C-metabolism, required for purine synthesis. US only predicts this path. TFS predicts an additional mode where glycine and 1C are synthesized from threonine. This *threonine bypass* is not commonly used by *E.coli*, likely because of higher resource requirements. However, it has been successfully used to enhance PHB production (Lin *et al.*, 2015). Interestingly, most reactions in glycolysis and the TCA display several marginal modes, depending on the direction of reactions in the rest of the network. While the effect of reversing a reaction on the fluxes can be analyzed with classical FBA or US, TFS provides additional information: the probability (and feasibility) of each direction and the correlation between the direction and the thermodynamic quantities. In the case of GHMT2r, the two modes can be better distinguished by the activity term—more precise estimates of $\Delta_r \mathbf{G}'^\circ$ would contribute little information about the reaction direction (Fig. 4C). Conversely, knowing the direction of the reaction would constrain the ratio of serine and glycine, and thus their concentrations. We computed Cohen's d for the distribution of $\Delta_r \mathbf{G}'^\circ$ and of the activity term in the two modes for each reversible reaction (Fig. 4D). In most cases, the modes are better separated by the activity term, suggesting that uncertainties in the standard reaction energy estimates have limited impact. Additional measurements of absolute metabolite concentrations will likely constrain more reaction directions.

3.4 Prediction of metabolite concentrations
Finally, we compared the metabolite concentrations predicted by TFS and TMFA to the experimentally measured values in the four $M-$ conditions. TFS predictions generally agree with the measured values and highlight strong thermodynamic constraints in the concentrations of some metabolites, such as the NAD/NADH ratio and organic phosphate (Fig. 5A). The mean concentrations predicted by TFS correlate with the experimental measurements ($r = 0.32$, Fig. 5B). For several metabolites, the predicted distribution was very close to the prior distribution [Kullback-Leibler divergence (KL) smaller than 0.2], suggesting that they participate in reactions that are always favorable. Reducing the coverage of the predictions to metabolites with KL ≥ 0.2 led to a higher correlation ($r = 0.63$), showing that their concentration is thermodynamically constrained. When we focused on the 95% Confidence Interval (CI) of the samples and compared them to the ranges predicted by TMFA (with 95% CI on the distributions of $\Delta_r \mathbf{G}'^\circ$ and ln $\mathbf{c}$), TFS always predicted narrower ranges (Fig. 5C), with an average 4.5-fold reduction (0.65 on the $\log_{10}$-scale). We quantified the overlap of the predicted distributions (uniform within the predicted interval for TMFA) with the distribution of the measurements using the Hellinger distance (Fig. 5D). In 66% of the cases, TFS agreed better with the measurements than TMFA, further supporting its overall higher predictive power for metabolite concentrations.

4 Discussion
Realistic models must obey the laws of thermodynamics: quantitative observations inform on the model structure, and mechanistic knowledge constrains parameters. Here, we show how this connection can be exploited in both directions to analyze metabolism. Our PTA framework for thermodynamic-based modeling includes optimization and sampling methods for the evaluation and exploration of metabolic networks. As demonstrated for *E.coli*, it allows to combine metabolic models with experimental data to discover quantitative and structural model inaccuracies and generate hypotheses on mechanisms such as substrate channeling. We expect this capability to be particularly important for less well-characterized organisms. Conversely, we showed that the stoichiometry of a network, coupled with thermodynamic constraints and few physiological observations can significantly reduce the space of feasible metabolic modes and metabolite concentrations.

In contrast to other methods, PTA interprets uncertainties in the thermodynamic parameters probabilistically to account for correlations between uncertainties of multiple estimates. Moreover, we require all reactions to have a well-defined direction. The model cannot silently resolve thermodynamic inconsistencies by setting a flux to zero as in TMFA—problems are visible to the modeler,
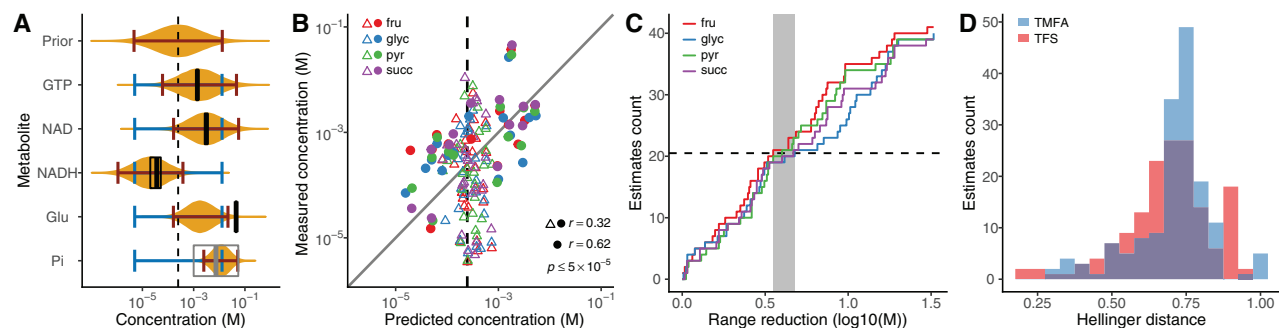
Fig. 5. Metabolite concentrations. (**A**) Examples of distributions predicted by TFS (orange), the respective 95% confidence intervals (red), ranges predicted by TMFA (blue) and mean and 95% confidence intervals of the metabolomics measurements (black, range from literature in gray). (**B**) Mean concentrations predicted by TFS agree with measured values (Pearson's $r = 0.32$, Root Mean Square Error RMSE = 0.87 in $\log_{10}$ scale). Reducing the coverage to metabolites with a KL between prior and posterior distribution above 0.2 (full circles) increases the agreement to $r = 0.63$, RMSE = 0.72. (**C**) Cumulative distribution of the reduction in the width of the predicted interval of TFS over TMFA for the validated metabolites. (**D**) Hellinger distances between the distribution of the predicted and measured metabolite concentrations for TMFA and TFS

guiding model curation and the formulation of novel hypotheses. Using the thermodynamic space as prior for sampling the flux space changed the exploration principle from the one used by US. Instead of using the volume of the flux space only, TFS determines the probability of a mode based on the probability of having a set of reaction energies allowing it. This excludes thermodynamically infeasible cycles from the solution. It also prevents distributed thermodynamic bottlenecks (Mavrovouniotis, 1993) and highlights feasible modes that would be missed by US, such as the threonine bypass.

However, PTA comes at the cost of higher computational complexity. Enforcing non-zero flux for each reaction makes optimization problems more difficult to solve, and sampling the non-convex thermodynamic space requires long simulation times; the asymptotic complexity is potentially exponential. Here, we successfully analyzed models with up to 880 reactions. For each condition, enumeration of internal EFMs and solution of PMO required less than five minutes, while sampling took approximately one day. To make larger models tractable, PTA can trade speed for coverage by limiting thermodynamic constraints to compartments or subsystems of interest.

To discover new biology, PMO proved effective at identifying occurrences of substrate channeling. While TMFA suggested such occurrences (Chiappino-Pepe *et al.*, 2017), predictions were limited to cases where predicted concentrations in an essential pathway exceed physiological ranges. Most of our predictions were supported by at least another source in *E.coli* or other organisms, and we also obtained novel hypotheses. That gluconeogenesis appears unfavorable is puzzling: several observations suggest channeling of glycolytic compounds. Yet, this observation could be attributed to inaccurate estimates of standard reaction energies, for example, because magnesium ions, which affect the thermodynamics of glycolysis (Vojinović and von Stockar, 2009), are insufficiently accounted for. Additionally, the analysis showed that the curation of reaction directions for FBA applications leads to thermodynamic inaccuracies. As coverage and precision of standard reaction energies and metabolite concentrations increases, we therefore encourage to model thermodynamics explicitly.

Finally, while TFS predicts significantly fewer modes than contained in the flux space, these modes cover a wide range of behaviors. Thermodynamics alone does not determine what a cell does, but it restricts what options are available to achieve a certain phenotype. The final choice probably follows from other constraints or principles such as enzyme cost and minimal resource usage, whose integration could further increase the predictive power of thermodynamically constrained models.

Acknowledgements

Funding

References

Abernathy,M.H. *et al.* (2019) Comparative studies of glycolytic pathways and channeling under in vitro and in vivo modes. *AIChE J.*, **65**, 483–490.

Bélisle,C.J. *et al.* (1993) Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.*, **18**, 255–266.

Binns,M. *et al.* (2015) Sampling with poling-based flux balance analysis: optimal versus sub-optimal flux space analysis of actinobacillus succinogenes. *BMC Bioinformatics*, **16**, 49.

Chiappino-Pepe,A. *et al.* (2017) Bioenergetics-based modeling of plasmodium falciparum metabolism reveals its essential genes, nutritional requirements, and thermodynamic bottlenecks. *PLoS Comput. Biol.*, **13**, e1005397.

Desouki,A.A. *et al.* (2015) Cyclefreeflux: efficient removal of thermodynamically infeasible loops from flux distributions. *Bioinformatics*, **31**, 2159–2165.

Du,B. *et al.* (2018a) Temperature-dependent estimation of Gibbs energies using an updated group-contribution method. *Biophys. J.*, **114**, 2691–2702.

Du,B. *et al.* (2018b) Thermodynamic favorability and pathway yield as evolutionary tradeoffs in biosynthetic pathway choice. *Proc. Natl. Acad. Sci. USA*, **115**, 11339–11344.

Ebrahim,A. *et al.* (2013) Cobrapy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.*, **7**, 74–76.

Erdrich,P. *et al.* (2015) An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Syst. Biol.*, **9**, 48.

Gelman,A. *et al.* (2013) *Bayesian Data Analysis*. CRC Press, Boca Raton, FL.

Gerosa,L. *et al.* (2015) Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Syst.*, **1**, 270–282.

Goldberg,R.N. *et al.* (2004) Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics*, **20**, 2874–2877.

Haraldsdóttir,H. *et al.* (2012) Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophys. J.*, **102**, 1703–1711.

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

Heinonen,M. *et al.* (2019) Bayesian metabolic flux analysis reveals intracellular flux couplings. *Bioinformatics*, **35**, i548–i557.

Heirendt,L. *et al.* (2019) Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. *Nat. Protoc.*, **14**, 639–702.

Henry,C.S. *et al.* (2007) Thermodynamics-based metabolic flux analysis. *Biophys. J.*, **92**, 1792–1805.

Jankowski,M.D. *et al.* (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.*, **95**, 1487–1499.

Keaty,T.C. and Jensen,P.A. (2020) Gapsplit: efficient random sampling for non-convex constraint-based models. *Bioinformatics*, **36**, 2623–2625.

Lin,Z. *et al.* (2015) Metabolic engineering of *Escherichia coli* for poly (3-hydroxybutyrate) production via threonine bypass. *Microb. Cell Factories*, **14**, 185.

Marco-Marín,C. *et al.* (2007) A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of *Escherichia coli* glutamate 5-kinase. *J. Mol. Biol.*, **367**, 1431–1446.

Mavrovouniotis,M.L. (1993) Identification of localized and distributed bottle-necks in metabolic pathways. *ISMB*, **93**, 273–283.

Monk,J.M. *et al.* (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.*, **35**, 904–908.

Noor,E. *et al.* (2013) Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.*, **9**, e1003098.

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

Park,J.O. *et al.* (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.*, **12**, 482–489.

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.

Rakus,D. *et al.* (2004) Interaction between muscle aldolase and muscle fructose 1, 6-bisphosphatase results in the substrate channeling. *Biochemistry*, **43**, 14948–14957.

Ramon,C. *et al.* (2018) Integrating –omics data into genome-scale metabolic network models: principles and challenges. *Essays Biochem.*, **62**, 563–574.

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

Schilling,C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, **203**, 229–248.

Terzer,M. and Stelling,J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–2235.

Vojinović,V. and von Stockar,U. (2009) Influence of uncertainties in pH, pMg, activity coefficients, metabolite concentrations, and other factors on the analysis of the thermodynamic feasibility of metabolic pathways. *Biotechnol. Bioeng.*, **103**, 780–795.

Williams,H.P. (2013) *Model Building in Mathematical Programming*. John Wiley & Sons, Hoboken, NJ.

Zhang,W. *et al.* (2017) Coupling between d-3-phosphoglycerate dehydrogenase and d-2-hydroxyglutarate dehydrogenase drives bacterial l-serine synthesis. *Proc. Natl. Acad. Sci. USA*, **114**, E7574–E7582.