PORTLAND PRESS

Review Article

Integrating –omics data into genome-scale metabolic network models: principles and challenges

Charlotte Ramon[1,2,*], Mattia G. Gollub[1,*] and Jörg Stelling[1]

[1] Department of Biosystems Science and Engineering and SIB Swiss Institute of Bioinformatics, ETH Zurich, Basel 4058, Switzerland; [2] PhD Program Systems Biology, Life Science Zurich Graduate School, Zurich, Switzerland

Correspondence: Jörg Stelling (joerg.stelling@bsse.ethz.ch)

At genome scale, it is not yet possible to devise detailed kinetic models for metabolism because data on the *in vivo* biochemistry are too sparse. Predictive large-scale models for metabolism most commonly use the constraint-based framework, in which network structures constrain possible metabolic phenotypes at steady state. However, these models commonly leave many possibilities open, making them less predictive than desired. With increasingly available –omics data, it is appealing to increase the predictive power of constraint-based models (CBMs) through data integration. Many corresponding methods have been developed, but data integration is still a challenge and existing methods perform less well than expected. Here, we review main approaches for the integration of different types of –omics data into CBMs focussing on the methods' assumptions and limitations. We argue that key assumptions – often derived from single-enzyme kinetics – do not generally apply in the context of networks, thereby explaining current limitations. Emerging methods bridging CBMs and biochemical kinetics may allow for –omics data integration in a common framework to provide more accurate predictions.

Introduction

Metabolism is a set of biochemical reactions that allow the conversion of nutrients into building blocks necessary for life. In human diseases such as cancer or diabetes, metabolism shifts in a specific way, and elucidating the underlying mechanisms is crucial to understand and treat these diseases. However, this is far from trivial given the complexity of metabolic networks. They include typically thousands of biochemical reactions catalyzed by different enzymes with various properties to convert hundreds of metabolites. This motivates the need for computational modeling methods to transform data into insight [1].

Biochemical processes in metabolic networks can be described by kinetic models, represented as a system of ordinary differential equations (ODEs) in the form

$$\frac{d\mathbf{c}(t)}{dt} = \mathbf{N} \cdot \mathbf{v}(\mathbf{c}(t), \boldsymbol{\theta}), \quad \mathbf{c}(0) = \mathbf{c}_0,$$

where $\mathbf{N}$ is the stoichiometric matrix of the system and $\mathbf{v}$ is the vector of reaction rates computed according to their rate laws from the species concentrations $\mathbf{c}$ and the rate law parameters $\boldsymbol{\theta}$. Integration of the ODE system returns unique trajectories of concentrations and reaction rates that depend on the parameters and on the initial concentrations, $\mathbf{c}_0$. Thereby, kinetic models provide a high level of detail. However, they depend on knowing enzyme-kinetic rate laws and parameters, which hinders their application to genome-scale metabolic networks, where most reactions are still not characterized *in vivo* [2].

Because of these difficulties, constraint-based models (CBMs) are a favorite approach for modeling metabolic networks. Instead of simulating a trajectory in the concentration space, CBMs describe the
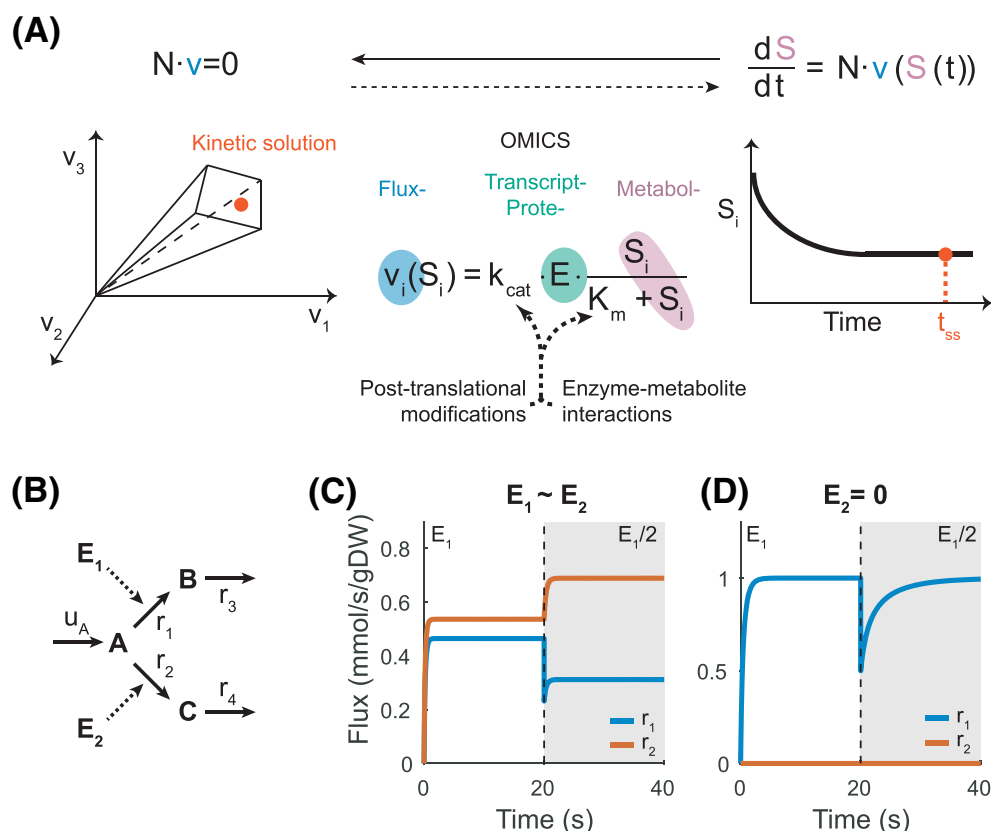
Figure 1. Reaction kinetics link CBMs and ODE models

(**A**) CBMs capture possible steady-state solutions in the flux space (left), corresponding to ODE solutions at steady state (right). Kinetics explain the relationship between fluxes (blue) and concentrations of enzymes (green), as measured by transcriptomics or proteomics, and of metabolites (purple) as measured by metabolomics. This relationship is affected by post-translational modification of the enzymes and by enzyme–metabolite interactions. (**B**) Example is biochemical network with constant input $u_A = 1$ and reactions r_1 and r_2 catalyzed by enzymes E_1 and E_2 according to Michaelis–Menten kinetics. We assume that enzymes are not at saturation. The concentration of E_1 is halved at $t = 20$ s. The resulting steady-state flux distribution (**C**) is generally non-trivial and depends on the kinetics of the whole network. In particular, the flux of r_2 changes even if the concentration of E_2 remains constant. In extreme cases such as a knockout of r_2 (**D**), the accumulation of metabolite A after the perturbation causes the flux of r_1 to stabilize at the same level as in the unperturbed system. Currently, CBMs cannot describe such effects.

entire space of feasible flux (rate) distributions **v** of the network given a set of flux constraints (Figure 1A). Commonly, CBMs assume the system to be at steady state, leading to the stoichiometric constraint $\mathbf{N} \cdot \mathbf{v} = \mathbf{0}$; reaction directionality and capacity constraints are incorporated as bounds on the reaction rates $\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}$. Although the constraint-based formulation does not model metabolite concentrations and time evolution, it empowers several methods that help answer biological questions that are more difficult to address with kinetic models [3,4]. A central method is flux balance analysis (FBA) [5], which identifies an optimal flux distribution given an assumed cellular objective such as maximal growth for microorganisms. However, for multicellular systems such as human cells or microbial communities, the cellular objective is unknown, preventing the use of optimization methods. The main challenge for CBMs then is a potentially under-constrained flux space because the tighter the constraints are, the closer the feasible flux space will be to an ODE solution obtained with perfect knowledge of the kinetic parameters, and therefore the more predictive the CBM will be [3].

When CBMs are insufficiently predictive, integrating –omics data is a possible avenue to reduce the flux space and to obtain more precise predictions [6]. While this concept is intuitive and many computational methods based on it have been developed, the heterogeneity of –omics data makes them difficult to integrate, sometimes even more difficult than generating the data itself [7]. Perhaps more importantly, the mapping of –omics data on to CBMs is non-trivial. As Figure 1A shows, reaction fluxes are determined by a combination of factors, including enzyme and

substrate concentrations, enzyme parameters, and regulatory interactions. Integrating concentration data ultimately reduces the question of what the influence of changes in enzyme or metabolite concentrations on the fluxes is. It is not easy to answer in a network context, despite the apparent simplicity of enzyme kinetic rate laws. Because reaction rates influence substrate concentrations, changes in the flux through one reaction also affect the fluxes of other reactions consuming the same metabolite. Consider the example in Figure 1B–D, a branching network with constant input where the two pathways are regulated by enzymes E_1 and E_2 with similar turnover numbers and substrate affinities. At steady state, the enzyme concentrations and parameters determine the distribution of fluxes between the two branches. When we perturb reaction r_1 by halving the concentration of E_1, its rate immediately halves accordingly, but subsequent substrate accumulation leads to the fluxes in both branches reaching a new steady state. If the same perturbation is applied to a linear pathway, the accumulation of substrate possibly leads to the restoration of the original steady state flux distribution.

Interdependencies and concentration effects, hence, have profound implications for data integration into the constraint-based framework, which does not natively capture the effect of concentration changes. Most state of the art methods for –omics data integration in CBMs must ultimately resort to heuristics to integrate experimental data – here, we focus on the methods' underlying principles, explicit or implicit assumptions, and corresponding limitations. Our central argument is that the absence of kinetic information in such methods is a main limitation, and that it will be necessary to systematically exploit (available) kinetic information in the data integration process.

Data integration methods

Kinetic rate laws can be decomposed into different components to emphasize the effect of different physical effects on the resulting flux. For instance, the irreversible Michaelis–Menten rate law can be decomposed into the product of three terms: the enzyme activity, the substrate saturation level, and the thermodynamic driving force [8]. In the following, we will review how one can integrate experimental data related to each of these terms of the rate law in order to obtain more predictive flux distributions, and we will specify what assumptions each approach entails.

Reaction rates

In vivo metabolic fluxes constitute the integrated response of the network to, for example external perturbations. However, these intracellular fluxes cannot be estimated directly from metabolite concentrations without knowing all relevant kinetic details such as rate laws, their parameters, and effective enzyme concentrations. ^{13}C-based flux analysis addresses this problem. Its principle is that ^{13}C-labeled substrates are fed to growing cells and the labeled substrates distribute throughout the network. The fluxes can then be estimated from the measured isotope distribution and a structural model of the metabolic network (see [9-11] for details on the method). However, the complexity of ^{13}C-based flux analysis limits its use to the estimation of 50–100 fluxes in central metabolism, preventing its use for the estimation of fluxes in a large-scale metabolic network [12]. Still, these fluxes can be easily integrated as constraints in a genome-scale model (GSM) by constraining the flux values to their experimentally measured values [13]. Alternatively, ^{13}C-based flux estimates are often used to validate *a posteriori* flux predictions generated using other methods [14].

Extracellular (net uptake or secretion) fluxes can be easily measured and integrated into the metabolic network. The measurement of metabolite uptake or secretion rates depends on the types of culture conditions used, which imply different assumptions. For chemostat cultures, both the growth rate and the extracellular fluxes are constant, such that one can estimate extracellular fluxes by comparing the metabolite concentrations in the culture with those of the fresh medium and by correcting for the cell dry weight. In batch culture conditions, both the growth rate and the extracellular fluxes vary over time [15]. Fluxes are usually estimated from dynamic extracellular metabolomics (exo-metabolomic) data, generated in a typical experiment that measures metabolite concentrations in the extracellular medium at different times (Figure 2A). Note, however, that the estimation usually assumes constant growth, uptake, and excretion rates over time, and that estimates could be incomparable between experiments with different dry weights per cell.

Exo-metabolomics data represent a metabolic footprint of the cell that can help to understand the phenotype [16]. For example, absolute metabolite concentrations measured in different cancer cell lines showed that glycine consumption correlates with the rate of cancer proliferation [17]. However, often ion intensities from MS-based metabolite analysis provide only relative data on metabolite abundances. For each metabolite, the ratio of the ion intensities in two conditions is then considered to be equal to the ratio of the respective metabolite concentrations. For example, this approach was used to study mycobacteria in normal and hypoxic conditions [18], where GSMs (specifically, their reaction graph representations) facilitated the interpretation of exo-metabolomics data by generating hypotheses
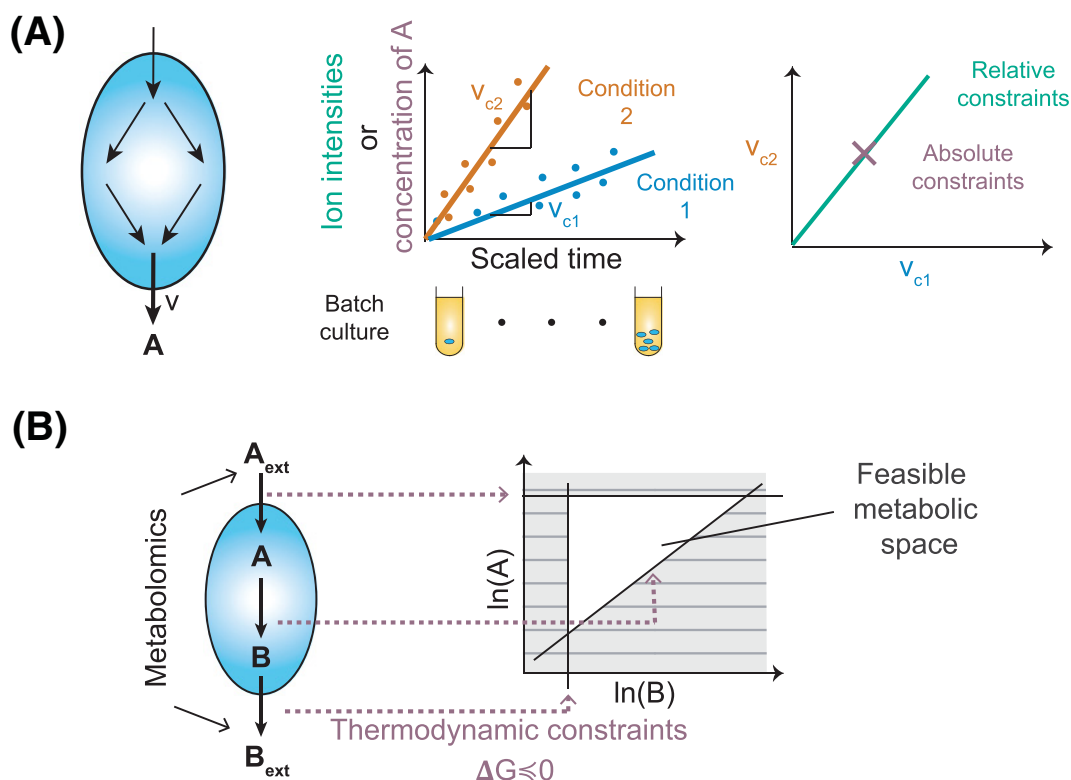
Figure 2. General strategies for integrating different types of –omics data into GSMs
(**A**) Integrating extracellular metabolomics data: in a batch culture, the supernatant is collected at different time points, and uptake or secretion rates are estimated based on ion intensities or absolute metabolite concentrations, yielding relative or absolute constraints on fluxes between experimental conditions, respectively. (**B**) Integrating metabolite concentration using thermodynamic FBA. Irreversible reactions impose specific constraints on metabolite concentrations by requiring a non-positive Gibbs free energy of reaction ΔG. Measuring the metabolite concentrations both intracellularly and extracellularly allows to map them to the metabolite concentration space and thereby to analyze the feasibility of network operation.

on which intracellular pathways are used in each condition. Measured extracellular fluxes can also be directly integrated as a set of quantitative or qualitative constraints in the GSM to constrain the flux space (Figure 2A). If absolute rates are available, the fluxes are constrained to the experimentally measured rates. With relative rates, one can apply overflow constraints, that is, constrain the ratios of exchange fluxes to the measured values. Flux spaces for different conditions can then be characterized by random sampling to compare metabolic network operation [19]. While these approaches constrain otherwise free exchange fluxes by measured rates, one can constrain a model further by only allowing measured exchange fluxes, and blocking all others. However, the set of measured external metabolites may be incomplete and the resulting model might not be viable. A recent method addresses this problem by finding a minimal number of additional exchange fluxes that need to be active for the model to sustain minimal cell growth. Exo-metabolomics data can then be used to create condition-specific models tailored to specific metabolomics profiles [20]. Overall, thus, it is relatively straightforward to integrate (dynamic) exo-metabolomics data into GSMs to increase their predictive power. However, the extracellular metabolic profile of cells is sometimes not available, for instance when analyzing human cells *in vivo*.

Metabolite concentrations

The accuracy and resolution of intracellular metabolomics measurements heavily depends on the experimental method used as well as the extraction method [21,22]. A more fundamental challenge to integrate this type of data, however, is that CBMs do not represent metabolite concentrations explicitly. As a consequence, intracellular metabolite concentrations cannot be integrated directly. The quasi steady-state assumption in the constraint-based framework states that metabolite concentrations are constant. However, this neglects that the cell can accumulate or deplete intracellular metabolites to adapt in dynamic situations such as during responses to changing nutrient conditions.

Recently, a novel method referred to as unsteady-state FBA (uFBA) has been developed to allow for changes in intracellular metabolite concentrations, by relaxing the assumption of quasi steady-state for some metabolites [23]. It explained the utilization of extracellular citrate by red blood cells stored in an anticoagulant solution by the large intracellular malate pool. Interestingly, to cope with non-linearity in time-course metabolomics data, the data were discretized into time-intervals where it is possible to assume linear profiles for piece-wise simulation.

A second aspect of the influence of intracellular metabolite concentrations on fluxes relates to the thermodynamic driving force: the product to substrate concentration ratio of a biochemical reaction can influence both the flux directionality and the rate of the reaction. In the thermodynamic framework, a chemical reaction is feasible in a certain direction if the Gibbs free energy change of the reaction, ΔG, is negative. In turn, ΔG depends on the free energy change of the reaction in standard conditions and the ratio of the concentrations of the reaction's products and educts [24]. As shown in Figure 2B, each irreversible biochemical reaction imposes a constraint on the product and substrate concentrations. The intersection of all these constraints constitutes the feasible metabolic (concentration) space for the network. Thermodynamic-based metabolic flux analysis (TMFA) exploits this principle to predict a flux distribution consistent with the thermodynamic constraints. Specifically, TMFA tries to find an optimal flux distribution such that the free energies of the irreversible reactions are negative and such that there are feasible metabolite concentrations that can satisfy feasibility of a network flux distribution [25]. Building on this method, Hoppe and co-workers [26] constrained the metabolic space to plausible metabolite concentrations based on literature to constrain flux directionalities. This approach provides a link between metabolite concentrations and flux directionality, without requiring any additional assumption. However, it is not yet applicable to large-scale metabolic models for two main reasons: the mixed integer linear programming approach (MILP) is computationally expensive, and one needs to know the standard Gibbs free energy change constants, which are difficult to measure in practice [27].

Enzyme concentrations

For the analysis of metabolic network operation in multicellular systems *in vivo*, often transcriptomics or proteomics data are more readily available than metabolomics data. The integration of transcriptomics or proteomics data therefore has been a major direction of methods development for GSMs. A variety of different goals was addressed, for example to compare cell states affected by genetic perturbations to understand key genes involved in cancer [28] or ageing [29], to create tissue-specific models by knowing which enzymes are expressed in the different tissues [30-33], and to understand the impact of environmental perturbations in microbiota on the human host's transcriptome [34]. Here, through some examples, we focus on the main concepts behind methods that integrate transcriptomic data into CBMs to increase the accuracy of flux predictions. Extensive summaries of the methods can be found in other reviews [14,35,36].

Fundamentally, transcriptomics and proteomics data both approximate enzyme concentrations in metabolic networks; we focus on transcriptomics in the following, and the same logic applies to proteomics data. Transcriptomics data can be integrated either by discretizing the gene expression data (the switch approach) or by deriving quantitative continuous constraints (the valve approach) [37]. For instance, iMAT uses the switch approach. Absolute gene expression data allow to classify reactions into those with highly or lowly expressed enzymes. iMAT favors (disfavors) flux through highly (lowly) expressed reactions, without requiring a cellular objective [38]. E-Flux is an example of methods based on the valve approach. Absolute gene expression data are normalized to a reference condition, and these data are used to directly constrain the fluxes, the rationale being that the higher the concentration of the enzyme, the higher is the flux capacity through this reaction [39]. However, the performance of such data integration methods was evaluated systematically only recently. In comparisons between predicted and experimentally measured flux distributions for different knockout mutants and overexpression strains of *Escherichia coli* and *Saccharomyces cerevisiae*, these data integration methods performed consistently less well than parsimonious enzyme usage FBA (pFBA) [14]. pFBA is an FBA variant that assumes minimal enzyme usage as a part of a cell's objective, and it does not integrate experimental data [40]. This raises the question of how data integration could make GSMs *less* predictive. Only linear-bound FBA (LBFBA) recently reported slight improvements over pFBA in certain conditions [41].

Consider first methods in the valve category, which set bounds on metabolic fluxes according to quantitative measures of enzyme concentrations. This assumes that mRNA abundances correlate with protein concentrations, which correlate with enzyme activities, which correlate with metabolic fluxes [42]. It is known that cellular mRNA and protein concentrations can correlate poorly [43], but this does not explain the methods' performance in the comparative study because integrating proteomics instead of transcriptomics data did not lead to significant improvements [14]. One possible explanation for proteomics data not improving predictions is that the methods compared in [14] are

tailored to transcriptomics data and, for example different discretizations may be required for the two data types. In our view, however, a more likely reason is that all current methods integrate transcriptomic or proteomic data without integrating the associated kinetic parameters. For instance, E-Flux limits the flux capacity of a reaction based on the amount of enzyme present. Enzyme kinetics justifies this principle in an isolated reaction, but not in a network. Consider the linear pathway shown in Figure 1B–D, where all fluxes are entirely determined by the input flux. There is no change in flux upon halving the enzyme concentration because the enzyme does not operate at saturation, and the substrate concentration adapts to the change in enzyme concentration. Inappropriate scaling of the maximal flux in this example would identify the reaction as the bottleneck and predict a flux change upon change in the enzyme concentration. Fundamentally, even with absolute proteomics data, the limitation is a potentially wrong scaling of constraints on the flux space when one does not consider the different turnover rates of metabolic enzymes. Unfortunately, these vary by several orders of magnitude between different enzymes [44]. Also, *in vivo* experimental evidence indicates that direct correlations between transcriptomics or proteomics data and metabolic fluxes are weak. Chubukov and co-workers [34] showed that changes in *E. coli's* central carbon metabolism fluxes upon changes in the carbon source available in the medium could not be explained alone by the transcriptome, and that changing substrate concentrations explained only a small fraction of flux changes.

There is, however, a special case in which kinetic parameters and metabolite concentrations do not affect the flux through enzyme-catalyzed reactions: without enzymes, there should be no flux through the catalyzed reactions. Hence, one could use transcriptomics data to derive which enzymes are absent, for example from specific tissues or cell lines. This is the rationale behind the switch approach: it sets a threshold to determine whether gene products are present or not, and it forces the flux through reactions that lack a catalyzing enzyme to zero [45]. Indeed, cell line-specific models developed with the switch approach showed improved gene essentiality predictions. However, we do not have principled ways of defining the threshold's value and this value substantially influences the model predictions. For example, cell line specification via a threshold sometimes removed important metabolic functionalities by falsely considering a gene product as absent [31]. An arbitrary threshold for microarray data can be avoided by a discretization workflow specific for this data type [46], used by recent methods integrating microarray data to construct tissue-specific models [47,48]. Overall, thus, while the integration of transcriptomics or proteomics data into GSMs appears intuitive and straightforward, this is not necessarily the case when either evaluating the performance of the corresponding methods, or when mapping the methods' assumptions to the underlying biochemical kinetics.

Hybrid frameworks
Integrating structure and kinetics
The constraint-based framework predicts metabolic fluxes with very limited information about the biological systems studied, but the omission of kinetics prevents a rigorous integration of metabolite and enzyme concentration data. While the limited availability of kinetic parameters motivates the wide adoption of heuristic-based methods for data integration, we argue that it is critical for data integration frameworks to re-consider reaction kinetics in the context of the reaction network. Because developing realistic genome-scale kinetic (ODE) models is still an ambitious goal even for the best-studied model organisms, hybrid approaches are promising. Hybrid models use kinetic information only to the extent that it is available or reliable, while the remaining network is modeled in the constraint-based framework. This allows to combine the descriptive power of an ODE model with the exploratory capabilities of constraint-based modeling.

In this framework, static approaches for –omics data integration constitute one main class. They use the ODE formulation and concentration measurements for a part of the network to generate condition-specific flux constraints (Figure 3A). The population FBA method [49] improves over E-Flux [39] by using published enzyme turnover rates as well as protein counts inferred from fluorescence measurements to compute flux bounds. Modeling individual cells with protein copy numbers sampled from the experimental distributions allowed to predict cell-to-cell variability and the Crabtree effect in a yeast population [50]. FBA with kinetic constraints (KFBA) [51] estimates kinetic parameters from a multi-omics dataset including flux, protein abundance, and metabolite abundance measurements [52]. For each test condition, enzyme and metabolite concentrations together with the estimated parameters and their confidence intervals are used to compute condition-specific flux constraints for a subsequent FBA. KFBA showed improved predictions over classical FBA, but the method was allowed to violate some constraints to find a feasible solution, highlighting the challenges in the construction of complete and consistent models. k-OptForce [53] uses metabolite concentrations and an ODE model to compute steady-state fluxes for those reactions with kinetic information; these fluxes act as constraints, for example in strain design [54]. Without explicitly computing flux constraints, integrative omics-metabolic analysis (IOMA) [55] intends to find the flux distribution that best agrees with proteomics and
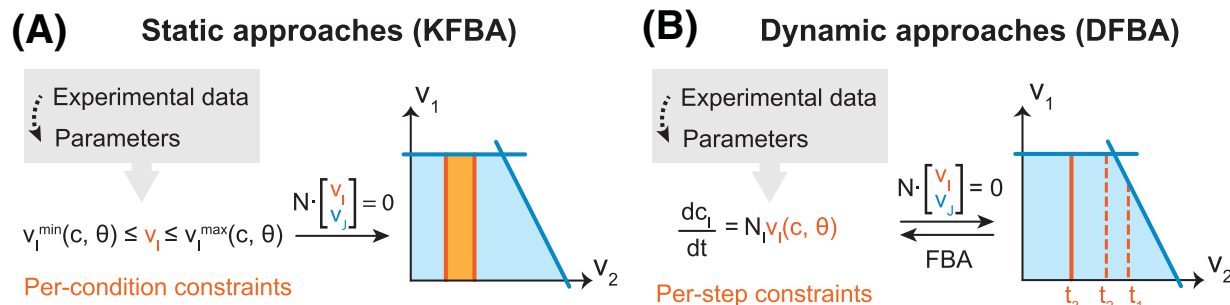
(A)   **Static approaches (KFBA)**   **(B)**   **Dynamic approaches (DFBA)**

Figure 3. Two strategies for the kinetics-based integration of -omics data in CBMs
In static approaches such as KFBA (**A**), kinetics and experimental data are used to determine flux constraints for each measured condition. In DFBA-like dynamic approaches (**B**), a part of the system is simulated with an ODE model while the remainder of the network is modeled with CBM methods. Iteratively, the fluxes computed at each time step of the integration are enforced in the CBM, whose FBA solution is then used as an input for the next step. Abbreviation: DFBA, dynamic FBA.

metabolomics data and with the kinetics of a set of well-characterized reactions. Minimizing the sum of per-reaction error variances over different conditions makes IOMA less sensitive to systematic bias due to regulation mechanisms not included in the model, but it also implicitly favors smaller fluxes, akin to pFBA [40].

The other main class of dynamic approaches is most relevant when experimental data are not available for the conditions of interest, or when the environmental conditions vary over time, for example due to external substrate consumption. Dynamic approaches rely on dynamic FBA (DFBA) [56] (Figure 3B; see also [57] for a recent review). DFBA models a subset of reactions whose dynamics are deemed relevant, or for which kinetics are available, as an ODE system. At each time step of the integration for solving the ODE system, simulated fluxes from the ODE model constrain an FBA problem over the remainder of the network, and the FBA results enter the ODE model for the next time step. Here, experimental data are used to compute uptake rates or kinetic parameters. In principle, the ODE model can have arbitrary coverage of the network, ranging from a single uptake reaction to more complex (possibly regulated) subnetworks as in integrated FBA (iFBA) [58]. However, DFBA simulations have to solve both the ODE system and the CBM in one pass at each time step, without multiple levels of flux dependencies. This is possible if the ODE system covers only the network boundaries such as biomass production and nutrient uptake reactions. Despite these constraints, DFBA can help analyze the properties of highly interdependent hybrid models. For example, an ODE model of central carbon metabolism integrated into a GSM revealed bistability in the metabolism of *E. coli* [59]. Overall, however, selecting the relevant kinetics and inferring the corresponding kinetic parameters from experimental data remain key challenges of hybrid frameworks.

Finally, in particular for hybrid models, metabolic regulation complicates the prediction of flux distributions. Transcriptomics and proteomics data may replace explicit models for the effects of transcriptional and post-translational enzyme regulation, but even together with metabolomics data for substrate concentrations of metabolic reactions, they are not sufficient to explain metabolic fluxes [34]. Hence, a significant portion of metabolic regulation must be attributed to other metabolite–enzyme interactions such as allosteric regulation [60]. Indeed, most metabolic enzymes in *E. coli* are regulated by metabolites, with more than a third of the inhibitors' concentrations being above the K_I value (inhibitor-enzyme dissociation constant) [61,62], and methods for the reconstruction of small molecule regulatory networks are becoming available [63,64]. For hybrid models, interdependencies via allosteric regulators, especially when currency metabolites such as ATP or NADH are involved, directly affect the selection of reactions to be modeled by ODE systems that, in contrast with CBMs, predict metabolite concentrations, given the concentrations in the intracellular or extracellular environment. In the constraint-based modeling framework, to our knowledge, only a single method accounts for regulatory enzyme–metabolite interactions. Allosteric regulation FBA (arFBA) [65] uses known regulatory interactions to minimize the discrepancy between the changes in the regulated flux and in the turnover rate of the effector with respect to a reference condition. It relies on the assumption that the change ratios are linearly correlated, which is not justifiable in general. Again, only the kinetics of a reaction can predict the effect of regulation at different levels of enzyme saturation or effector concentration.

Probabilistic approaches

Uncertain knowledge of reaction kinetics, their parameters, and their regulation is pervasive in systems biology, not restricted to CBMs and metabolic networks [66]. For smaller scale dynamic models of biochemical networks,
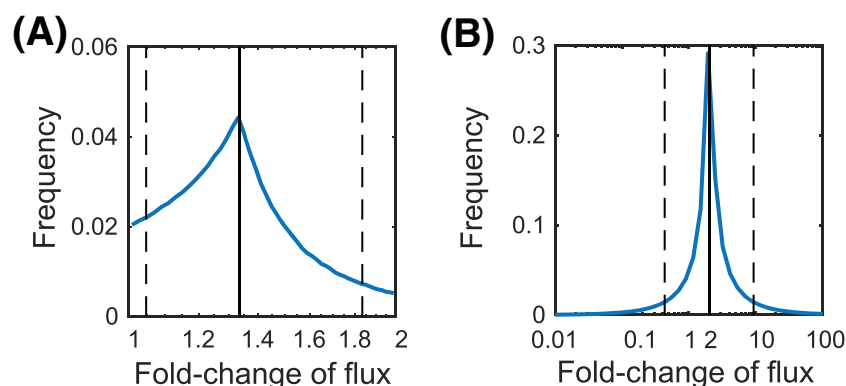
Figure 4. Bayesian analysis for CBMs

Example of flux sampling for a Michaelis–Menten kinetics with a substrate S, $v(S) = k_{cat} \cdot E \cdot \frac{S}{K_m + S}$. **(A)** Distribution of fold-changes of the flux $v(S)$ upon doubling the substrate concentration S. S as well as the parameters k_{cat} and K_M were independently and uniformly sampled in [0,1], while $E = 1$ was fixed. We computed the flux per sample i, with the sampled S_i and $2 \cdot S_i$. **(B)** Distribution of fold-changes of $v(S)$ upon doubling the enzyme concentration E when substrate concentrations can vary due to network effects. Using the samples for k_{cat} and K_M, we computed fluxes per sample with $E = 1$ and S_i, and with $E = 2$ and an independently sampled (uniform in [0,1]) substrate concentration S_i'. Vertical solid and dashed lines indicate the median and the 90% confidence interval, respectively.

probabilistic approaches that quantitate uncertainty, its propagation through networks, and it effects on phenotypes are increasingly employed [67]. In particular, Bayesian approaches gain importance because they can account for uncertainties in networks and experimental observations simultaneously [68,69]. For example, Bayesian inference was successful in real-life applications that simultaneously aimed to elucidate biological mechanisms and parameter values by integrating prior knowledge and experimental data for dynamic models of cell signaling [70,71]. We argue that developing equivalent computational methods for Bayesian data integration and metabolic flux analysis for the substantially larger CBMs is a challenge worthwhile to be addressed.

In a consistent Bayesian approach for data integration into CBMs, the network structure is the main prior information. It restricts the space of feasible fluxes, and instead of reasoning about (single) optimal fluxes or boundaries on fluxes, one can aim to derive probability distributions over fluxes given the network structure. Recent developments of sampling algorithms [72-74] as well as approximation algorithms [75] make the characterization of probability distributions over feasible fluxes at genome-scale a realistic endeavor. Then, the combination with –omics data (via their likelihood) can, in principle, generate probabilistic predictions of metabolic fluxes (the Bayesian posterior), without requiring assumptions on biological objectives. This may allow an unbiased investigation of metabolic flux distributions that are consistent with experimental observations, but substantial conceptual and technical challenges such as defining and efficiently sampling omics-based probability distributions on the flux space need to be solved.

To illustrate how Bayesian approaches could help integrating –omics data into CBMs probabilistically, consider the example of Michaelis–Menten kinetics in Figure 4. Simulations in Figure 4A suggest that relative metabolite concentrations can be informative for metabolic fluxes – the effect of a two-fold increase in substrate concentration is captured probabilistically, without knowing exact parameter values. In contrast, with an enzyme abundance increased by two-fold, the flux is expected to double, but with large variability due to potentially changing substrate concentrations when the reaction is embedded in a network (Figure 4B). Thus, assuming types of reaction kinetics, and using known distributions of enzyme parameters, Bayesian inference for –omics data integration appears to be a promising avenue for metabolic flux analysis, despite the large uncertainties in enzyme kinetics and regulation.

Conclusions and perspectives

At genome-scale, it is currently impossible to comprehensively measure (through ^{13}C experiments) or precisely predict (through kinetic models) metabolic flux distributions, which makes the constraint-based framework so attractive for the analysis of metabolic networks. In particular, CBMs are powerful at elucidating the entire feasible flux space given well-characterized constraints derived from reaction stoichiometries, mass balances, and thermodynamics. To constrain behaviors further – and thereby make models more predictive – the intuitively appealing idea of obtaining

constraints based on measured concentration data, however, has proven challenging. We argue that the main limitation of many current methods for data integration is to rely on assumptions derived from single-enzyme kinetics that do not necessarily hold in a network context. As simple examples show (see Figure 1), especially the links between enzyme concentrations and fluxes in a reaction network are not trivial in general, and further complicated by several layers of regulation *in vivo* [34].

As a consequence, we need to account for reaction kinetics in the integration of experimental data and in the description of many metabolic mechanisms such as substrate accumulation or allosteric regulation. We do not expect genome-scale kinetic models to replace the simple and conservative constraint-based approaches in the near future, but rather argue that it is worth exploiting the complementary strengths of the two frameworks: obtaining precise fluxes or constraints from well-characterized reactions, and exploring the resulting feasible flux space with constraint-based approaches. Methods such as KFBA and k-OptForce can set omics-based constraints on the individual fluxes using kinetics, but a key open question is which minimal network parts should be augmented with kinetics to obtain more predictive and still consistent models. Importantly, it is not necessary to constrain all fluxes by kinetics because the steady-state assumption propagates flux constraints to other reactions in the network. This applies to linear pathways, where all fluxes are equal at steady state, and to branched pathways as in Figure 1, where knowing two fluxes suffices to infer the remaining one. Minimal sets of relevant fluxes should ideally be identified from structural information only, for example using concepts from metabolic control analysis (MCA) [76]. The introduction of kinetics, however, must be approached with care: the example of KFBA consistently needing to violate constraints to obtain a feasible steady-state solution suggests that incomplete model structures or inaccurate parameters can easily lead to inconsistencies in the integration of experimental data.

More generally, we believe that probabilistic methods are under-explored in model-based methods for –omics data integration. In particular for applying Bayesian approaches, by now well-established in other domains of systems biology, CBMs are lagging behind. With recent technical developments in sampling in large networks, it is now possible to develop Bayesian approaches for flux predictions that consistently integrate multiple types of –omics data. Note that this approach differs from emerging machine learning methods that, for example were recently used to integrate a multi-omics compendium of 26 databases into one multiscale model for *E. coli* [36]. Unfortunately, machine learning methods usually have a weak connection to the underlying biochemistry, which hinders reasoning about the connection between biological processes and predicted phenotypes. Finally, because many metabolic reactions are regulated to maintain stable fluxes and to switch pathways depending on the environmental conditions, new methods that account for regulation will likely have a great impact on the quality of predictions.

Summary

- Data integration into CBMs of metabolism promises to increase the models' predictive powers.

- This integration is challenging, depending on the type of data.

- Many existing methods rely on assumptions derived from enzyme kinetics that do not necessarily hold in a network context.

- Emerging methods bridging CBMs and biochemical kinetics are expected to reduce these limitations.

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Author contribution

C.R., M.G.G., and J.S. wrote the manuscript.

Abbreviations

CBM, constraint-based model; DFBA, dynamic flux balance analysis; FBA, flux balance analysis; GSM, genome-scale model; IOMA, integrative omics-metabolic analysis; KFBA, FBA with kinetic constraint; ODE, ordinary differential equation; pFBA, parsimonious enzyme usage FBA; TMFA, thermodynamic-based metabolic flux analysis.

References

1 Nielsen, J. (2017) Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275, https://doi.org/10.1146/annurev-biochem-061516-044757

2 Vasilakou, E., Machado, D., Theorell, A., Rocha, I., Nöh, K., Oldiges, M. et al. (2016) Current state and challenges for dynamic metabolic modeling. *Curr. Opin. Microbiol.* **33**, 97–104, https://doi.org/10.1016/j.mib.2016.07.008

3 Terzer, M., Maynard, N.D., Covert, M.W. and Stelling, J. (2009) Genome-scale metabolic networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **1**, 285–297, https://doi.org/10.1002/wsbm.37

4 Lewis, N.E., Nagarajan, H. and Palsson, B.O. (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* **10**, 291–305, https://doi.org/10.1038/nrmicro2737

5 Orth, J.D., Thiele, I. and Palsson, B. (2010) What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248, https://doi.org/10.1038/nbt.1614

6 O'Brien, E.J., Monk, J.M. and Palsson, B.O. (2015) Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987, https://doi.org/10.1016/j.cell.2015.05.019

7 Palsson, B. and Zengler, K. (2010) The challenges of integrating multi-omic data sets. *Nat. Chem. Biol.* **6**, 787–789, https://doi.org/10.1038/nchembio.462

8 Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A. and Milo, R. (2013) A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett.* **587**, 2772–2777, https://doi.org/10.1016/j.febslet.2013.07.028

9 Cascante, M. and Marin, S. (2008) Metabolomics and fluxomics approaches. *Essays Biochem.* **45**, 67–81, https://doi.org/10.1042/bse0450067

10 Sauer, U. (2006) Metabolic networks in motion: 13C-based flux analysis. *Mol. Syst. Biol.* **2**, 62, https://doi.org/10.1038/msb4100109

11 Zamboni, N. and Sauer, U. (2009) Novel biological insights through metabolomics and 13C-flux analysis. *Curr. Opin. Microbiol.* **12**, 553–558, https://doi.org/10.1016/j.mib.2009.08.003

12 Niedenfuhr, S., Wiechert, W. and Noh, K. (2015) How to measure metabolic fluxes: a taxonomic guide for (13)C fluxomics. *Curr. Opin. Biotechnol.* **34**, 82–90, https://doi.org/10.1016/j.copbio.2014.12.003

13 Stewart, B.J., Navid, A., Turteltaub, K.W. and Bench, G. (2010) Yeast dynamic metabolic flux measurement in nutrient-rich media by HPLC and accelerator mass spectrometry. *Anal. Chem.* **82**, 9812–9817, https://doi.org/10.1021/ac102065f

14 Machado, D. and Herrgard, M. (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* **10**, e1003580, https://doi.org/10.1371/journal.pcbi.1003580

15 Griffiths, J.P. and Pirt, S.J. (1967) The uptake of amino acids by mouse cells (strain LS) during growth in batch culture and chemostat culture: the influence of cell growth rate. *Proc. R. Soc. Lond. B Biol. Sci.* **168**, 421–438, https://doi.org/10.1098/rspb.1967.0073

16 Pinu, F.R. and Villas-Boas, S.G. (2017) Extracellular microbial metabolomics: the state of the art. *Metabolites* **7**, https://doi.org/10.3390/metabo7030043

17 Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A.L. et al. (2012) Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* **336**, 1040–1044, https://doi.org/10.1126/science.1218595

18 Zimmermann, M., Kuehne, A., Boshoff, H.I., Barry 3rd, C.E., Zamboni, N. and Sauer, U. (2015) Dynamic exometabolome analysis reveals active metabolic pathways in non-replicating mycobacteria. *Environ. Microbiol.* **17**, 4802–4815, https://doi.org/10.1111/1462-2920.13056

19 Mo, M.L., Palsson, B.O. and Herrgard, M.J. (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **3**, 37, https://doi.org/10.1186/1752-0509-3-37

20 Aurich, M.K., Paglia, G., Rolfsson, Ó., Hrafnsdóttir, S., Magnúsdóttir, M., Stefaniak, M.M. et al. (2015) Prediction of intracellular metabolic states from extracellular metabolomic data. *Metabolomics* **11**, 603–619, https://doi.org/10.1007/s11306-014-0721-3

21 Canelas, A.B., ten Pierick, A., Ras, C., Seifar, R.M., van Dam, J.C., van Gulik, W.M. et al. (2009) Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. *Anal. Chem.* **81**, 7379–7389, https://doi.org/10.1021/ac900999t

22 Goodacre, R., Vaidyanathan, S., Dunn, W.B., Harrigan, G.G. and Kell, D.B. (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **22**, 245–252, https://doi.org/10.1016/j.tibtech.2004.03.007

23 Bordbar, A., Yurkovich, J.T., Paglia, G., Rolfsson, O., Sigurjonsson, O.E. and Palsson, B.O. (2017) Elucidating dynamic metabolic physiology through network integration of quantitative time-course metabolomics. *Sci. Rep.* **7**, 46249, https://doi.org/10.1038/srep46249

24 Alberty, R.A. (1994) Biochemical thermodynamics. *Biochim. Biophys. Acta* **1207**, 1–11, https://doi.org/10.1016/0167-4838(94)90045-0

25 Henry, C.S., Broadbelt, L.J. and Hatzimanikatis, V. (2007) Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805, https://doi.org/10.1529/biophysj.106.093138

26 Hoppe, A., Hoffmann, S. and Holzhutter, H.G. (2007) Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.* **1**, 23

27 Noor, E., Haraldsdottir, H.S., Milo, R. and Fleming, R.M. (2013) Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* **9**, e1003098, https://doi.org/10.1371/journal.pcbi.1003098

28 Auslander, N., Cunningham, C.E., Toosi, B.M., McEwen, E.J., Yizhak, K., Vizeacoumar, F.S. et al. (2017) An integrated computational and experimental study uncovers FUT9 as a metabolic driver of colorectal cancer. *Mol. Syst. Biol.* **13**, 956, https://doi.org/10.15252/msb.20177739

29 Yizhak, K., Gabay, O., Cohen, H. and Ruppin, E. (2013) Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* **4**, 2632, https://doi.org/10.1038/ncomms3632

30 Asgari, Y., Khosravi, P., Zabihinpour, Z. and Habibi, M. (2018) Exploring candidate biomarkers for lung and prostate cancers using gene expression and flux variability analysis. *Integr. Biol. (Camb.)* **10**, 113–120, https://doi.org/10.1039/C7IB00135E

31 Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D.C. and Lewis, N.E. (2017) A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Syst.* **4**, 318–329e6

32 Jerby, L., Shlomi, T. and Ruppin, E. (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* **6**, 401, https://doi.org/10.1038/msb.2010.56

33 Folger, O., Jerby, L., Frezza, C., Gottlieb, E., Ruppin, E. and Shlomi, T. (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501, https://doi.org/10.1038/msb.2011.35

34 Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R.J., Jules, M., Link, H. et al. (2013) Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis. Mol. Syst. Biol.* **9**, 709, https://doi.org/10.1038/msb.2013.66

35 Blazier, A.S. and Papin, J.A. (2012) Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* **3**, 299, https://doi.org/10.3389/fphys.2012.00299

36 Kim, M., Rai, N., Zorraquino, V. and Tagkopoulos, I. (2016) Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli. Nat. Commun.* **7**, 13090, https://doi.org/10.1038/ncomms13090

37 Hyduke, D.R., Lewis, N.E. and Palsson, B.O. (2013) Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* **9**, 167–174, https://doi.org/10.1039/C2MB25453K

38 Zur, H., Ruppin, E. and Shlomi, T. (2010) iMAT: an integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142, https://doi.org/10.1093/bioinformatics/btq602

39 Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R. et al. (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput. Biol.* **5**, e1000489, https://doi.org/10.1371/journal.pcbi.1000489

40 Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D. et al. (2010) Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390, https://doi.org/10.1038/msb.2010.47

41 Tian, M. and Reed, J.L. (2018) Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics*, https://doi.org/10.1093/bioinformatics/bty445

42 Hoppe, A. (2012) What mRNA abundances can tell us about metabolism. *Metabolites* **2**, 614–631, https://doi.org/10.3390/metabo2030614

43 Liu, Y., Beyer, A. and Aebersold, R. (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550, https://doi.org/10.1016/j.cell.2016.03.014

44 Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F. and Schomburg, D. (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem. Sci.* **27**, 54–56, https://doi.org/10.1016/S0968-0004(01)02027-8

45 Akesson, M., Forster, J. and Nielsen, J. (2004) Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* **6**, 285–293, https://doi.org/10.1016/j.ymben.2003.12.002

46 McCall, M.N., Uppal, K., Jaffee, H.A., Zilliox, M.J. and Irizarry, R.A. (2011) The Gene Expression Barcode: leveraging public data repositories to begin cataloging the human and murine transcriptomes. *Nucleic Acids Res.* **39**, D1011–D1015, https://doi.org/10.1093/nar/gkq1259

47 Pacheco, M.P., John, E., Kaoma, T., Heinaniemi, M., Nicot, N., Vallar, L. et al. (2015) Integrated metabolic modelling reveals cell-type specific epigenetic control points of the macrophage metabolic network. *BMC Genomics* **16**, 809, https://doi.org/10.1186/s12864-015-1984-4

48 Pacheco, M.P., Pfau, T. and Sauter, T. (2015) Benchmarking procedures for high-throughput context specific reconstruction algorithms. *Front. Physiol.* **6**, 410

49 Labhsetwar, P., Melo, M.C.R., Cole, J.A. and Luthey-Schulten, Z. (2017) Population FBA predicts metabolic phenotypes in yeast. *PLoS Comput. Biol.* **13**, e1005728, https://doi.org/10.1371/journal.pcbi.1005728

50 Crabtree, H.G (1928) The carbohydrate metabolism of certain pathological overgrowths. *Biochem. J.* **22**, 1289–1298, https://doi.org/10.1042/bj0221289

51 Cotten, C. and Reed, J.L. (2013) Mechanistic analysis of multi-omics datasets to generate kinetic parameters for constraint-based metabolic models. *BMC Bioinformatics* **14**, 32, https://doi.org/10.1186/1471-2105-14-32

52 Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A. et al. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* **316**, 593–597, https://doi.org/10.1126/science.1132067

53 Chowdhury, A., Zomorrodi, A.R. and Maranas, C.D. (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* **10**, e1003487, https://doi.org/10.1371/journal.pcbi.1003487

54 Ranganathan, S., Suthers, P.F. and Maranas, C.D. (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* **6**, e1000744, https://doi.org/10.1371/journal.pcbi.1000744

55 Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E. and Shlomi, T. (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **26**, i255–60, https://doi.org/10.1093/bioinformatics/btq183

56 Mahadevan, R., Edwards, J.S. and Doyle, F.J. (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli. Biophys. J.* **83**, 1331–1340, https://doi.org/10.1016/S0006-3495(02)73903-9

57 yås, O. and Stelling, J. (2018) Genome-scale metabolic networks in time and space. *Curr. Opin. Syst. Biol.* **8**, 51–58, https://doi.org/10.1016/j.coisb.2017.12.003

58 Covert, M.W., Xiao, N., Chen, T.J. and Karr, J.R. (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli. Bioinformatics* **24**, 2044–2050, https://doi.org/10.1093/bioinformatics/btn352

59 Mannan, A.A., Toya, Y., Shimizu, K., McFadden, J., Kierzek, A.M. and Rocco, A. (2015) Integrating kinetic model of E. coli with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism. *PLoS ONE* **10**, e0139507, https://doi.org/10.1371/journal.pone.0139507

60 Fenton, A.W. (2008) Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem. Sci.* **33**, 420–425, https://doi.org/10.1016/j.tibs.2008.05.009

61 Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J. and Rabinowitz, J.D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599, https://doi.org/10.1038/nchembio.186

62 Park, J.O., Rubin, S.A., Xu, Y.F., Amador-Noguez, D., Fan, J., Shlomi, T. et al. (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* **12**, 482–489, https://doi.org/10.1038/nchembio.2077

63 Link, H., Kochanowski, K. and Sauer, U. (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity *in vivo*. *Nat. Biotechnol.* **31**, 357–361, https://doi.org/10.1038/nbt.2489

64 Reznik, E., Christodoulou, D., Goldford, J.E., Briars, E., Sauer, U., Segrè, D. et al. (2017) Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity. *Cell Rep.* **20**, 2666–2677, https://doi.org/10.1016/j.celrep.2017.08.066

65 Machado, D., Herrgård, M.J. and Rocha, I. (2015) Modeling the contribution of allosteric regulation for flux control in the central carbon metabolism of *E. coli*. *Front. Bioeng. Biotechnol.* **3**, 154, https://doi.org/10.3389/fbioe.2015.00154

66 Geris, L. and Gomez-Cabrero, D. (2016) *Uncertainty in Biology*, Springer

67 Babtie, A.C. and Stumpf, M.P.H. (2017) How to deal with parameters for whole-cell modelling. *J. R. Soc. Interface* **14**, 20170237, https://doi.org/10.1098/rsif.2017.0237

68 Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202, https://doi.org/10.1098/rsif.2008.0172

69 Chen, W.W., Niepel, M. and Sorger, P.K. (2010) Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* **24**, 1861–1875, https://doi.org/10.1101/gad.1945410

70 Xu, T.-R., Vyshemirsky, V., Gormand, A., von Kriegsheim, A., Girolami, M., Baillie, G.S. et al. (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Signal.* **3**, ra20, https://doi.org/10.1126/scisignal.2000517

71 Sunnaker, M., Zamora-Sillero, E., Dechant, R., Ludwig, C., Busetto, A.G., Wagner, A. et al. (2013) Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key Msn2 control mechanism. *Sci. Signal.* **6**, ra41, https://doi.org/10.1126/scisignal.2003621

72 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, https://doi.org/10.1093/bioinformatics/btw132

73 Haraldsdottir, H.S., Cousins, B., Thiele, I., Fleming, R.M.T. and Vempala, S. (2017) CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics* **33**, 1741–1743, https://doi.org/10.1093/bioinformatics/btx052

74 De Martino, D., Mori, M. and Parisi, V. (2015) Uniform sampling of steady states in metabolic networks: heterogeneous scales and rounding. *PLoS ONE* **10**, e0122670, https://doi.org/10.1371/journal.pone.0122670

75 Braunstein, A., Muntoni, A.P. and Pagnani, A. (2017) An analytic approximation of the feasible space of metabolic networks. *Nat. Commun.* **8**, 14915, https://doi.org/10.1038/ncomms14915

76 Reder, C. (1988) Metabolic control theory: a structural approach. *J. Theor. Biol.* **135**, 175–201, https://doi.org/10.1016/S0022-5193(88)80073-0