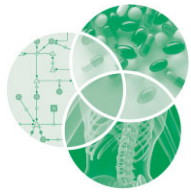
Focus Article



The application of flux balance analysis in systems biology

Erwin P. Gianchandani, Arvind K. Chavali and Jason A. Papin*

An increasing number of genome-scale reconstructions of intracellular biochemical networks are being generated. Coupled with these stoichiometric models, several systems-based approaches for probing these reconstructions *in silico* have been developed. One such approach, called flux balance analysis (FBA), has been effective at predicting systemic phenotypes in the form of fluxes through a reaction network. FBA employs a linear programming (LP) strategy to generate a flux distribution that is optimized toward a particular 'objective,' subject to a set of underlying physicochemical and thermodynamic constraints. Although classical FBA assumes steady-state conditions, several extensions have been proposed in recent years to constrain the allowable flux distributions and enable characterization of dynamic profiles even with minimal kinetic information. Furthermore, FBA coupled with techniques for measuring fluxes *in vivo* has facilitated integration of computational and experimental approaches, and is allowing pursuit of rational hypothesis-driven research. Ultimately, as we will describe in this review, studying intracellular reaction fluxes allows us to understand network structure and function and has broad applications ranging from metabolic engineering to drug discovery. © 2009 John Wiley & Sons, Inc. *WIREs Syst Biol Med* 2010 2 372–382

Systems biology is pioneering the study of intracellular biochemical networks. Whereas network components (e.g., metabolites, enzymes, etc.) and interactions (e.g., phosphorylation, inhibition, etc.) until recently have been studied individually, systems-based approaches are aiming to integrate an emerging wealth of experimental data and interrogate genome-scale biology holistically. Ultimately, it is the collection of components and interactions underlying a network that gives rise to the phenotypic behavior that we observe.

An increasing number of genome-scale reconstructions of metabolic systems have been published in recent years (see Refs 1–7 for examples). A key goal of these reconstructions is to catalog gene-protein-reaction (GPR) relationships underlying the networks in a quantitative, structured, and chemically consistent manner.[8] Methods for subsequently interrogating these reconstructions have also been advanced in tandem.

One such approach for predicting phenotypes from these network reconstructions is called flux balance analysis (FBA). FBA has been effective at calculating flux distributions through biochemical networks, and has enabled *in silico* characterization of growth yields, environmental conditions, and robustness under gene knockouts or knockdowns.[9,10] Briefly, FBA employs a linear programming (LP) strategy to generate a flux distribution that is optimized toward a particular 'objective,' subject to a set of governing constraints.[9,10] These constraints enforce physicochemical and thermodynamic principles.[9,10] Importantly, classical FBA assumes steady-state conditions and is often confounded by a large feasible solution space.[9,10]

In this review, we will first describe in detail the formulation of a classical FBA problem. Subsequently, we will summarize several recent extensions to FBA that have allowed the approach to be more broadly applicable to problems in systems biology. For example, some of these advances have led to the prediction of dynamic profiles of biological systems with minimal kinetic detail. Finally, we will report how FBA, coupled with experimental measurements, is advancing biological knowledge and

*Correspondence to: papin@virginia.edu

Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA

DOI: 10.1002/wsbm.60

driving discovery in drug development and metabolic engineering.

FORMULATION OF FLUX BALANCE ANALYSIS

FBA is a constraint-based approach that facilitates *in silico* prediction of systemic phenotypes in the form of reaction fluxes through an intracellular biochemical network.[9,10] FBA takes the form of an LP problem, in which an objective function is optimized subject to a set of underlying constraints (see Figure 1).

Defining the optimization framework

FBA requires a stoichiometric network reconstruction, often represented in the form of a stoichiometric matrix, **S**.[8,11] For a given biochemical network, the set of components and interactions underlying that network is assimilated from genome annotation, curated databases, and published literature sources and represented in **S**[(8,11)] (see Figure 1(a)). The rows of **S** correspond to various network components, while the columns of **S** delineate the reactions, or how these components interact with one another. At the intersection of every row and column within **S** lies a stoichiometric coefficient that quantitatively captures the precise interaction. Inputs to a given reaction are designated with negative values, while outputs from a reaction have positive values. For example, in Figure 1(b), the conversion of 1 mol of fructose-6-phosphate (F6P) to 1 mol of fructose-1,6-bisphosphate (FDP) by the enzyme phosphofructokinase 1 (PFK) is represented in the first column of **S**, with '–1' in the row corresponding to F6P and '+1' in the row corresponding to FDP. Cofactors are also included within **S**. In this manner, the stoichiometric network reconstruction enables explicit, quantitative, and chemically consistent accounting of a biochemical network. In recent years, genome-scale reconstructions of metabolism have been published for many organisms, including the prototypic prokaryote *Escherichia coli*[1,2] and eukaryote *Saccharomyces cerevisiae*.[3–5] In addition, recently, the human metabolic network was reported.[6,7]

Once a biochemical network is reconstructed, the rows of **S** can be multiplied by a column vector **v** containing the fluxes through the reactions, forming a system of linear equations.[9,10] Under steady-state conditions, the product of this multiplication must equal zero in order to ensure that no mass is consumed or produced within the system. A key challenge in predicting flux distributions is that most biological systems are underdetermined, i.e., there exist fewer components (and consequently fewer equations) than reactions whose fluxes are to be predicted. FBA overcomes this issue by using LP to optimize for a particular flux, all the while ensuring that the resultant flux distribution adheres to the constraint that mass must be balanced within the system under steady-state conditions (i.e., $\mathbf{S} \bullet \mathbf{v} = 0$)[9,10] (see Figure 1(c)). Note that this constraint thus serves to narrow the solution space of feasible flux distributions. The bounded solution space can be further narrowed by specifying additional constraints, such as reaction capacities that define minimum and maximum allowable flux values through individual reactions as well as other physicochemical properties[9] (see Figure 1(c)).

Examples of objective functions used in the literature include maximizing growth rate or the rate of synthesis of biomass,[2–4,12–20] maximizing (or minimizing) ATP production,[21–24] maximizing (or minimizing) the rate of production of a particular metabolic product,[25–30] and maximizing (or minimizing) the rate of nutrient uptake.[29,31,32] For example, as we will describe below, maximization of by-product synthesis is a common objective in metabolic engineering, as *in silico* stoichiometric models are used to predict how best to engineer an organism to maximize its yield of a desired output. Objective functions take on the form:

$$Z = \mathbf{c} \bullet \mathbf{v} \tag{1}$$

where **c** denotes a row vector of coefficients that multiply into the column vector **v**[9,10] (of fluxes) (see Figure 1(c)). Although generally only one reaction is optimized with FBA, multiple reactions may be simultaneously optimized, each with its own weighting, as specified in **c**.[9,10] As FBA constitutes a LP problem, linear objective functions have been proposed.

EXTENSIONS TO FLUX BALANCE ANALYSIS

FBA has demonstrated reasonable agreement with experimental data (e.g., 86% consistency in the prediction of gene essentiality for *E. coli* genome-scale metabolism,[33] and more recently, 85% and approximately 70% consistency for the prediction of gene essentiality in the metabolic networks of the pathogens *Pseudomonas aeruginosa*[17] and *Leishmania major*,[18] respectively). To improve the predictive ability of FBA, there have been attempts in recent years to further define the physicochemical constraints that govern the approach. In addition, several different modifications to FBA have been proposed to facilitate its use in predicting dynamics of

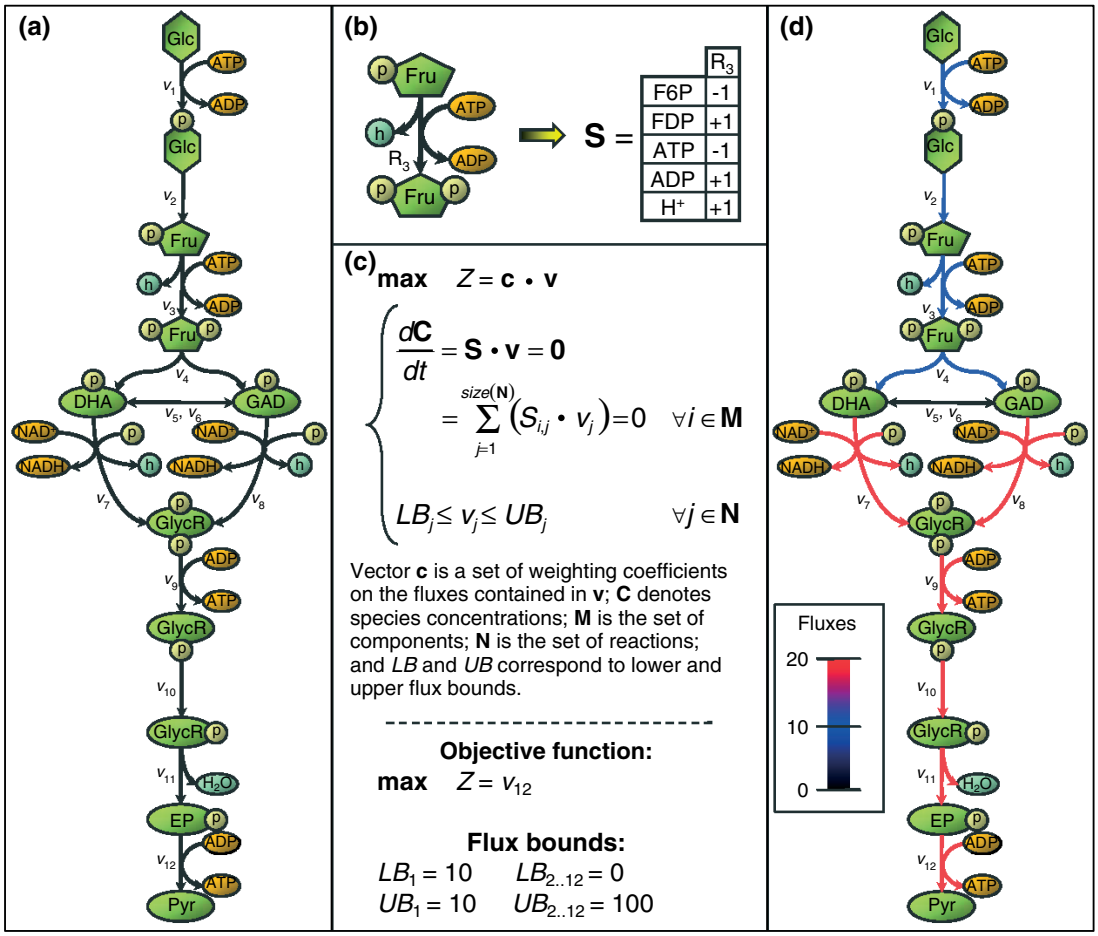


FIGURE 1 | The classical flux balance analysis (FBA) approach. Panel (a) illustrates the glycolytic pathway as a representative biochemical network. The reactions that convert glucose to pyruvate are labeled v_1 through v_{12}. In panel (b), the process of generating a stoichiometric matrix, S, for a given reaction is demonstrated. Specifically, the conversion of 1 mole of fructose-6-phosphate (F6P) to 1 mole of fructose-1,6-bisphosphate (FDP) by the enzyme phosphofructokinase 1 (PFK) (reaction v_3 in panel (a)) is represented as a column within S. The rows of S describe the components of the network, while the columns represent the underlying interactions. The stoichiometric coefficient at the intersection of a row and column quantitatively captures the precise interaction, with inputs delineated with negative coefficients (e.g., '–1′ F6P) and outputs delineated with positive coefficients (e.g., '+1′ FDP). Panel (c) depicts the classical FBA optimization framework, including the objective function and flux bounds that may be applied to the glycolytic network. Note that exchange reactions are excluded from this rendering for simplicity. Finally, the FBA-derived flux distribution through the network is shown in panel (d).

large-scale biochemical networks. In the following section we describe some of these strategies in detail.

Incorporating additional constraints

While classical FBA enforces the law of conservation of mass upon a system, an extension of FBA called energy balance analysis (EBA) extends the concept to incorporate general principles of thermodynamics.[34] EBA eliminates thermodynamically infeasible results associated with FBA by incorporating energy-balance loop equations that are analogous to Kirchhoff's voltage laws for electrical networks.[34,35] Equations balancing the global potential energy of a network are overlaid on the underlying network stoichiometry, and inequality constraints for each flux are also added to ensure that entropy production is positive for every reaction.[34] When applied to *E. coli* central metabolism, EBA yielded the same optimal growth rate as FBA, suggesting that EBA indeed appropriately constrains the feasible solution space.[34] In addition, EBA also provided explanations for why certain genes that were deemed inessential through FBA were in fact essential as had been experimentally characterized.[34]

Separately, gene regulatory constraints were incorporated into metabolic models, leading to a modification of FBA called regulatory flux balance analysis (rFBA).[36–38] Effectively, rFBA requires that

Boolean rules governing gene transcription are assimilated from literature and overlaid on an existing stoichiometric model of metabolism.[38] Subsequently, the Boolean rules define which gene products (or proteins) are turned 'ON' and which ones remain 'OFF' in response to the availability of environmental cues.[38] The resultant gene transcription states serve to constrain bounds on the corresponding fluxes within metabolism. For example, if the transcription of a particular gene is turned 'OFF' in response to a specified environment, the flux for the reaction catalyzed by the gene product is constrained to zero. The Boolean formalism can be implemented using commercially available software packages, such as Microsoft Excel (Microsoft Corporation, Redmond, WA). rFBA has demonstrated improved consistency with experimental measurements in stoichiometric models of *E. coli* and *S. cerevisiae* metabolisms.[38,39] For example, rFBA yielded 91% agreement for *E. coli* genome-scale metabolism, up from 86% for FBA without the gene regulatory constraints.[38]

Recently, a matrix formalism for representing transcriptional regulatory networks (TRNs) was developed.[40] The regulatory network matrix, called **R**, is similar to **S** in that its rows and columns correspond to network components and interactions, respectively.[40] Rows of **R** delineate the presence and absence of extracellular metabolites, genes, and gene products (proteins), and columns describe the Boolean regulatory relationships in a pseudo-stoichiometric form.[40] The **R** matrix is advantageous to Boolean formalisms because it is easily integratable with existing stoichiometric representations of metabolism and signaling.[40] An **R** matrix-based representation of the *E. coli lac* operon and a separate prototypic TRN quantified functional (expression) states in response to various environmental conditions.[40] In addition, an **R** matrix of the genome-scale *E. coli* TRN was constructed.[41] Future implementations of rFBA may utilize the **R** matrix to couple transcriptional regulation with metabolism and yield improved metabolic flux predictions.

Enhancing FBA to predict system dynamics

A key assumption underlying FBA is that the biological system being studied is operating under steady-state conditions.[9,10] Consequently, classical FBA does not take into account time and species concentrations. This assumption enables reasonable approximations of flux distributions for metabolic systems, as the time scale upon which structural changes of the metabolic network occur (such as the expression of a gene coding for an enzyme that was previously absent) is much slower than the time scale upon which metabolic system dynamics take place. However, longer-term dynamics of metabolic systems, particularly when coupled with gene regulation or cell signaling, necessitate the ability to handle 'stiff' systems that include both 'fast' and 'slow' interactions.[42,43] Several approaches to this end have been developed.

Under abnormal conditions such as single gene deletions or reaction knockouts, a biochemical network may undergo a period of adjustment during which time it does not optimize for a predetermined objective function.[44] *Minimization of metabolic adjustment (MOMA)* was developed to model such a modified metabolic network.[44] MOMA employs a quadratic programming (QP) framework to effectively project the FBA-based optimum flux distribution for the wild-type system onto the feasible flux space of the mutant system.[44] In this way, MOMA successfully predicted flux distributions during the transition period between wild-type and mutant steady-states.[9,45]

A variant of FBA called *dynamic flux balance analysis (DFBA)* provides a framework for assessing the transience of metabolism due to metabolic reprogramming.[46] *DFBA* was implemented in two distinct ways, namely, as a dynamic optimization approach that required solving a nonlinear programming (NLP) problem, as well as a static optimization approach that required solving a LP problem.[46] The dynamic approach entailed performing a single optimization spanning the time period of interest and yielded temporal profiles of fluxes as well as, correspondingly, metabolite concentrations.[46] The static approach, meanwhile, discretized the time window into uniform time intervals, and an instantaneous optimization problem was solved at the beginning of each interval to predict the fluxes at that time point, followed by integration over the interval to compute species concentrations over time.[46] *DFBA* simulated the batch growth of *E. coli* on glucose, and the predictions were found to be qualitatively consistent with experiments.[46] In addition, it was shown that the static approach resulted in more accurate predictions than the dynamic one.[46]

A more recent strategy coupling MOMA with *DFBA*, called *M-DFBA*, was extended to the energetic-metabolic network in mammalian myocardial cells.[21] By employing the *MOMA* hypothesis, the objective function of the myocardial metabolic network was altered from maximization of ATP production to minimization of fluctuation of metabolite concentrations between normal and ischemic conditions.[21] The results of *M-DFBA* demonstrated greater consistency with experimental data than those

of *DFBA* when the network was tested under ischemic conditions.[21] Furthermore, the results supported the hypothesis that metabolic systems revert to suboptimal states during a transient perturbation.[21]

Finally, a recent extension of *DFBA* called *integrated dynamic flux balance analysis (idFBA)* lays the foundation for the dynamic analysis of integrated biochemical networks spanning signaling, metabolic, and transcriptional regulatory processes.[42] As with *DFBA, idFBA* discretizes the time window into uniform time intervals and solves an instantaneous optimization problem at the beginning of each interval to predict the fluxes at that time point and in turn integrate over time to compute species concentrations.[42] However, *idFBA* uniquely assumes quasi-steady-state conditions for 'fast' reactions and incorporates 'slow' reactions, such as transcriptional and translational events, into the stoichiometric formalism in a time-delayed manner.[42] For example, whereas signaling events are typically very fast occurring on the order of seconds, delays in transcriptional and translational processes can be on the order of up to several hours.[43] *idFBA* generated comparable time-course predictions when contrasted with an equivalent kinetic (ordinary differential equation) model of the *S. cerevisiae* high-osmolarity glycerol (HOG) pathway in response to osmotic stress.[42]

Identifying objective reactions for FBA

A central remaining challenge in FBA is to define for a given biological system an objective function (with biological meaning) for which that system optimizes.[47] Several strategies have been proposed that attempt to infer an objective function for a given system from the underlying network stoichiometry and some experimentally measured internal state variables.[47] The experimental measurements are often in the form of isotopomer tracings (see Ref. 48 for an example of such flux measurements). For example, *ObjFind* is a bi-level optimization framework that assigns weightings to the reactions within a network, optimizing for the resultant objective function.[49] The weightings that result in an *in silico* flux distribution that is most consistent with the observed fluxes are assumed to describe the contribution of the corresponding reactions to the network objective: the higher the weighting, termed the coefficient of importance (CoI), the more important that reaction is in the network objective, and vice versa.[49] *ObjFind* successfully predicted that maximizing growth rate was the most likely objective for *E. coli* central metabolism under anaerobic and aerobic minimal media conditions.[49] A subsequent approach added Gibbs free energy terms to the set of mass-balance constraints, and using a similar inverse bi-level optimization framework that coupled FBA and EBA, predicted the objective function in the hypermetabolic human liver given time-series flux data.[50] Two other efforts have attempted to pick the reaction for which a system optimizes, using Bayesian-based probabilistic ranking [51] and Euclidean metric [52] schemes.

A limitation of *ObjFind* and its counterparts is that it is unable to correctly identify the network objective if it has not been experimentally characterized *a priori*.[47] Recently, a bi-level optimization framework called *biological objective solution search (BOSS)* attempted to overcome this limitation by adding a *de novo* reaction to the **S** matrix representing the stoichiometry of an underlying system and optimizing the flux through this reaction.[47] As opposed to assigning weightings to reactions, BOSS attempted to predict the stoichiometric coefficients of a *de novo* reaction for which a given system optimizes. Inputs to BOSS include the underlying stoichiometry of a system and experimental flux measurements under the condition for which the objective is to be determined. BOSS randomly samples the possible objective reaction space, all the while minimizing the difference between the resultant FBA-derived flux distribution for a given random objective and experimental flux measurements.[47] In this way, BOSS predicted that the objective of *S. cerevisiae* central metabolism is maximizing growth rate without any *a priori* knowledge of the biomass reaction or assumption of the network objective. Consequently, BOSS facilitates FBA-based phenotypic profiling of additional biological systems whose objective functions have not yet been experimentally characterized.

EXPERIMENTAL FLUX MEASUREMENTS

Computational predictions of flux distributions for a given biochemical network may be validated with experimentally measured flux data. Furthermore, coupling *in silico* predictions with *in vivo* measurements facilitates methodological advances and biological discoveries, as we describe below.

A common strategy for quantifying intracellular fluxes experimentally, called metabolic flux analysis (MFA), involves using ^{13}C isotopomers combined with separation techniques such as gas chromatography (GC), liquid chromatography (LC), or capillary electrophoresis (CE), as well as instruments for detecting elemental composition such as nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS).[53] Briefly, a metabolite with *n* carbon atoms can be specifically labeled (or unlabeled) at each carbon

atom position, meaning that there can be 2^n different labeling states or 2^n different 'isotopomers'.[53] The isotopomer distribution of a metabolite with n carbon atoms is characterized by the percentage of each isotopomer within the metabolite pool.[53] Each labeling state is separated by GC, LC, or CE, and subsequently detected by NMR or MS using statistical matching to reference samples.[53–57] For example, a ^{13}C-labeled substrate (e.g., [1−^{13}C]glucose) may be fed into a system of interest. These labeled carbon atoms are allowed to distribute throughout the underlying biochemical network until isotopic enrichment within individual intracellular metabolite pools can be detected.[53–57] The resulting data, plus extracellular flux measurements, enable quantitation of the intracellular fluxes for the system.[53]

The intracellular fluxes of many large-scale microbial,[58–60] yeast,[61–63] plant,[64,65] and mammalian [66] systems have been measured. Furthermore, in recent years, additional isotopic tracers have been developed [67,68] and combinations of isotopic tracers have been employed simultaneously to more precisely characterize intracellular metabolite pools.[68] For example, as *E. coli* feeds off of glucose, the [1−^{13}C]glucose tracer has been used to capture its metabolic dynamics. By contrast, a [U-^{13}C,$^{2}H_8$]glycerol tracer was combined with deuterated water $^{2}H_2O$ to estimate fluxes of the gluconeogenesis pathway in mice.[69,70]

Given the sheer number of isotopomers for a given metabolite with n carbon atoms, a key challenge associated with MFA is the large number of isotopomer equations that need to be solved in order to derive a flux distribution.[71] The problem becomes all the more complex when multiple types of isotopic tracers are fed into a system.[71] The elementary metabolite unit (EMU) framework is a novel approach that attempts to overcome the MFA limitation by identifying the minimum amount of information needed to simulate isotopic labeling within a reaction network.[71] Effectively, EMU takes as its input the set of atomic transitions that occur within a biochemical reaction network, applies a highly efficient decomposition algorithm, and identifies functional units (or EMUs) that form a new basis for generating system equations describing relationships between fluxes and stable isotope measurements.[71] The EMU framework demonstrated an order-of-magnitude reduction in the number of equations required.[71]

Coupling stoichiometric network reconstructions and FBA-based flux predictions with the types of experimental flux measurements described has led to novel system characterizations. For example, FBA of genome-scale *S. cerevisiae* metabolism identified reactions for which alternative pathways exist, i.e., key targets for MFA.[62] Subsequently, ^{13}C flux data for 30 of these FBA-identified reactions were used to further constrain FBA and improve the consistency between *in silico* predictions and observed *in vivo* behavior.[62] By iteratively integrating computational analyses with experiments, a more accurate flux distribution was obtained for genome-scale *S. cerevisiae* metabolism. Moreover, this approach quantified within *S. cerevisiae* the relative importance of 'genetic buffering' through alternative pathways as well as network redundancy through duplicate genes; these measures are indicators of the overall robustness of the *S. cerevisiae* metabolic network.[62]

APPLICATIONS OF FLUX METHODS

In recent years, FBA has demonstrated its value in furthering our biological knowledge. In particular, *in silico* flux predictions have been integrated with experimental methods to formulate novel hypotheses and yield valuable insights. For instance, as described below, FBA-based approaches have been applied to several challenging problems, including the generation of tissue-specific human metabolic reconstructions based on gene expression datasets, the identification of novel therapeutic targets against harmful and infectious agents, and the development of genetically engineered organisms for the production of desirable substrates.

Tissue-specific human metabolism

The human body consists of trillions of cells spanning a diverse array of cell types (e.g., macrophages, erythrocytes, and neurons) involved in wide-ranging functionality (e.g., innate immunity, oxygen transport, and action potential transduction). The metabolic requirements of a given cell are unique to its particular cell type and, ultimately, to the function of the tissue environment in which that cell type is commonly found. Developing 'tissue-specific' metabolic networks is vital in the study of many human diseases. For example, when modeling host-pathogen interactions in the study of infectious disease, it is critically important to evaluate pathogen intracellular networks in the context of host cell networks and how a host's metabolism contributes to the survival of a pathogen species. The trypanosomatid parasite *L. major*, a causative agent of cutaneous leishmaniasis, lacks the ability to synthesize key amino acids (e.g., arginine, lysine, valine, and others) and instead needs to scavenge them from the phagolysosome compartment of the host macrophage.[72] Accounting for the

environment that the pathogen faces by integrating a macrophage metabolic network with that of *L. major* may lead to novel insights in drug target discovery. Although existing *Homo sapiens* metabolic network reconstructions (see Refs 6, 7) capture the underlying stoichiometry for a generic human cell, there exists a need to develop methods for constraining a generalized metabolic network to study tissue-specific properties of a wide variety of human cells.[73,74]

Recently, computational methods utilizing gene expression data to constrain a generic human metabolic network have been developed.[73,74] Here, we provide details on two such methods. One approach, called gene inactivity moderated by metabolism and expression (GIMME), involves scoring each reaction based on gene expression data, and constraining the network by removing reactions that are below a certain scoring threshold.[73] A context-specific metabolic network is achieved if the reduced model is functional for a given metabolic objective (such as biomass synthesis). However, if the reduced model is not functional, then reactions that are necessary for achieving the metabolic objective are reintroduced. These reinserted reactions have expression scores that are below the specified scoring threshold, which results in an inconsistency with the expression data. An LP strategy is formulated to minimize the sum of reaction fluxes multiplied by deviations in expression scores from the specified scoring threshold. In other words, the total inconsistency score for the network is minimized. The GIMME algorithm was used to constrain the human metabolic network [6] given three gene expression datasets for skeletal muscle.[73]

A second approach involves separating the human metabolic network into highly- and lowly-expressed reaction sets based on available gene expression data.[74] A mixed-integer linear programming (MILP) algorithm with FBA constraints is implemented. The resulting steady-state flux distribution is then used to predict whether the genes in the expression set are up- or down-regulated. In other words, if a reaction is associated with a nonzero flux and corresponds to a lowly expressed gene in the expression set, then the gene is predicted to be post-transcriptionally up-regulated. Conversely, if a reaction is characterized by a zero flux but is associated with a highly expressed gene in the expression set, then the gene is predicted to be post-transcriptionally down-regulated.[74] Tissue-specific behavior was characterized using the human metabolic network[6] and expression data from 10 different tissues.[74] Notably, this method does not require *a priori* knowledge of tissue-specific metabolic objectives. Such an approach is advantageous because the objective functions of multi-cellular systems (e.g., *H. sapiens*) are not yet well characterized.[74]

Analysis of network perturbations

FBA of genome-scale metabolic networks also provides an ideal platform for large-scale evaluation of essential, growth-reducing, or non-essential genes, thereby allowing potential drug targets to be identified. For example, FBA was applied to identify antitubercular drug targets (against the pathogen *Mycobacterium tuberculosis*).[75] A highly interconnected network of the mycolic acid pathway in *M. tuberculosis* involving 197 metabolites, 219 reactions, and 28 proteins was reconstructed, and FBA was implemented to determine essential and nonessential genes.[75] From the group of essential genes, a prioritized list of genes without any known human orthologs (as determined by sequence similarity analysis) was generated.[75] The genes contained in this list were hypothesized to be viable drug targets against *M. tuberculosis*.

In addition to single gene knockouts, nontrivial lethal double- or triple-gene knockouts enable characterization of sets of genes that are lethal for the organism when deleted in combination but not lethal when knocked out individually. Moreover, *in silico* flux analysis also allows for constraining the flux of a reaction within a metabolic network to a fraction of its wildtype value in order to simulate the effects of partial inhibition. Performing FBA on the *L. major* metabolic network led to the identification of 69 single lethal gene deletions and 56 nontrivial lethal double-gene deletions.[18] In addition, the flux of mitochondrial F_0F_1-ATP synthase was systematically constrained and the resultant effects on growth rate were measured to gain insight into the robustness of the *L. major* metabolic network when simulated in the presence of an F_0F_1-ATP synthase inhibitor.[18]

Another FBA application area involves metabolic engineering of unicellular organisms for the synthesis of desirable by-products. By predicting flux distributions for various gene (and reaction) knockout or knockdown conditions, specific changes that may facilitate optimal by-product yields can be identified. The results can then be used to engineer strains (e.g., through up- or down-regulation of target genes) that have desired phenotypes. For instance, FBA of a central metabolic network reconstruction of the photosynthetic algae *Chlamydomonas reinhardtii* coupled changes in hydrogen production rates with reaction knockouts.[76] Flux methods have also been used to engineer *E. coli* strains to overproduce in high yields the amino acids threonine[28] and valine,[77] lactic acid,[78] and succinic acid.[79]

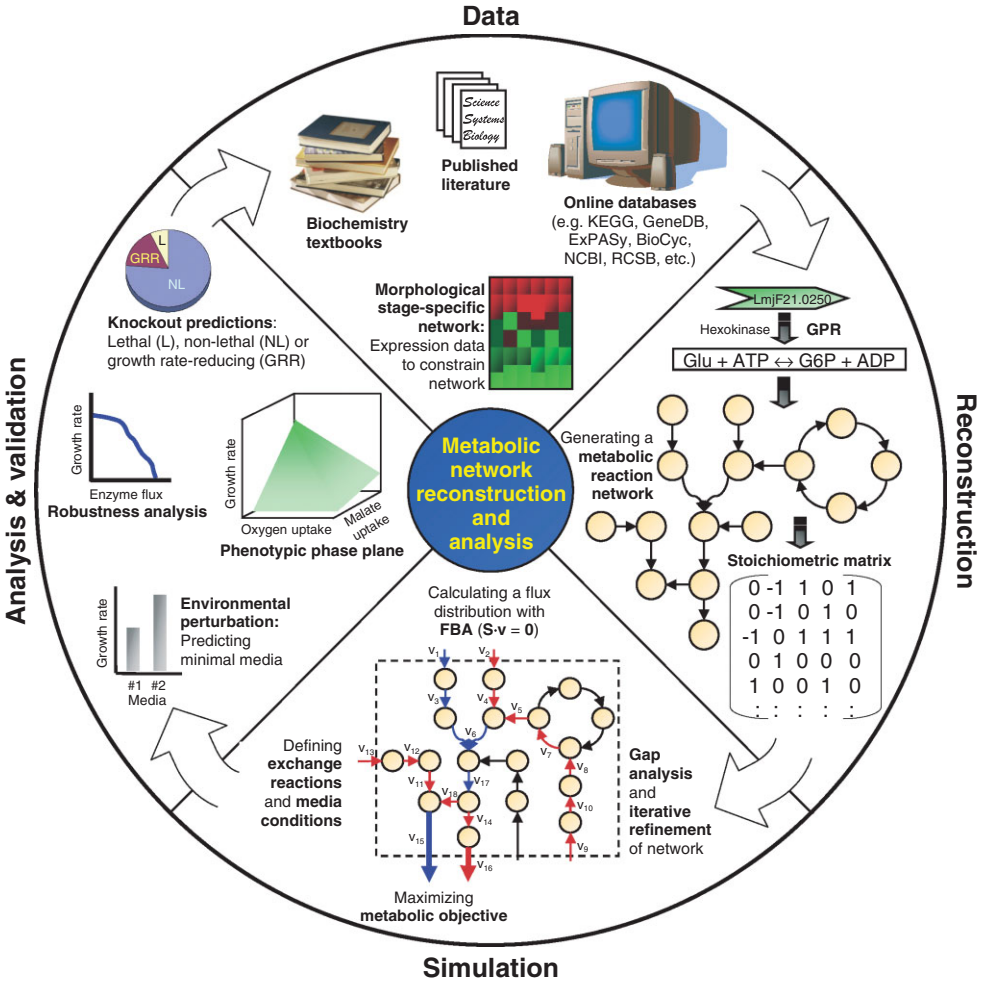


FIGURE 2 | Stoichiometric network reconstruction and analysis with FBA. A stoichiometric network reconstruction is assembled piece-by-piece, with gene-protein-reaction (GPR) relationships assimilated from experimental data available in published literature and online databases. FBA is then used to predict a steady-state flux distribution through the network, given certain underlying physicochemical and thermodynamic constraints. This prediction can be validated with independent experiments, and inconsistencies can lead to model refinement. In addition, with the stoichiometric network reconstruction and FBA, the network can be perturbed in different ways, ranging from differences in environmental stimuli to mutations in network structure (e.g., single- or double-gene knockouts) to assess properties of the system such as robustness.

CONCLUSION

These examples are merely a fraction of the success stories arising out of FBA of large-scale biological systems. Ultimately, a key advantage of FBA over alternate approaches such as more classical kinetic modeling is that it enables prediction of network phenotypes given far less knowledge of kinetic parameters. Consequently, FBA can generate hypotheses about global network structure as well as individual protein function without detailed kinetic experiments that are often tedious, expensive, and time-consuming.

As described throughout this review, recent extensions to the classical FBA formulation have enabled computation of dynamic profiles even when the steady-state assumption breaks down. For example, idFBA enables prediction of fluxes through a biochemical network spanning 'fast' and 'slow' events (and for which the steady-state assumption does not apply). Future extensions of idFBA to integrated stoichiometric network reconstructions spanning signaling, metabolic, and transcriptional regulatory activities will facilitate the characterization of whole-cell dynamics, leading to a fundamental understanding of how various diseases arise.

Furthermore, FBA constitutes an essential part of the iterative, systems-based experimental-computational paradigm (see Figure 2). Briefly, a stoichiometric network reconstruction assembled piece-by-piece from published literature may be interrogated with FBA, and the resulting *in silico* flux distribution may be validated with independent experiments (e.g., MFA). Discrepancies lead to model refinement. Once a model is validated, FBA may be used to predict flux distributions under various experimental conditions (such as the presence or absence of different carbon sources) and systemic perturbations (such as gene or reaction knockouts or knockdowns). Importantly, such *in silico* hypotheses require further experimental verification as part of an iterative experimental-computational approach that will pave the way for drug target identification and metabolic engineering strategies.

REFERENCES

1. Feist AM, Palsson BO. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat Biotechnol* 2008, 26(6):659–667.
2. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007, 3:121.
3. Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 2004, 14(7):1298–1309.
4. Forster J, Famili I, Fu P, Palsson BO, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 2003, 13(2):244–253.
5. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 2008, 26(10):1155–1160.
6. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 2007, 104(6):1777–1782.
7. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, et al. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* 2007, 3:135.
8. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 2009, 7(2):129–143.
9. Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of metabolomics. *Brief Bioinform* 2006, 7(2):140–150.
10. Price ND, Papin JA, Schilling CH, Palsson BO. Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* 2003, 21(4):162–169.
11. Gianchandani EP, Brautigan DL, Papin JA. Systems analyses characterize integrated functions of biochemical networks. *Trends Biochem Sci* 2006, 31(5):284–291.
12. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J Biol Chem* 2007, 282(39):28791–28799.
13. Senger RS, Papoutsakis ET. Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis. *Biotechnol Bioeng* 2008, 101(5):1036–1052.
14. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, et al. Genome-scale metabolic model of Helicobacter pylori 26695. *J Bacteriol* 2002, 184(16):4582–4593.
15. Oliveira AP, Nielsen J, Forster J. Modeling Lactococcus lactis using a genome-scale flux model. *BMC Microbiol* 2005, 5:39.
16. Beste DJ, Hooper T, Stewart G, Bonde B, Avignone-Rossa C, et al. GSMN-TB: a web-based genome-scale network model of Mycobacterium tuberculosis metabolism. *Genome Biol* 2007, 8(5):R89.
17. Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA. Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1. *J Bacteriol* 2008, 190(8):2790–2803.
18. Chavali AK, Whittemore JD, Eddy JA, Williams KT, Papin JA. Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Mol Syst Biol* 2008, 4:177.
19. Boyle NR, Morgan JA. Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. *BMC Syst Biol* 2009, 3:4.
20. Shastri AA, Morgan JA. Flux balance analysis of photoautotrophic metabolism. *Biotechnol Prog* 2005, 21(6):1617–1626.
21. Luo RY, Liao S, Tao GY, Li YY, Zeng S, et al. Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. *Mol Syst Biol* 2006, 2:2006.0031.
22. Teusink B, Wiersma A, Molenaar D, Francke C, de Vos WM, et al. Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model. *J Biol Chem* 2006, 281(52):40041–40048.
23. Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J Physiol Regul Integr Comp Physiol* 2001, 280(3):R695–R704.
24. Vo TD, Greenberg HJ, Palsson BO. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem* 2004, 279(38):39532–39540.
25. Varma A, Boesch BW, Palsson BO. Biochemical production capabilities of *Escherichia coli*. *Biotechnol Bioeng* 1993, 42(1):59–73.

26. Blank LM, Ebert BE, Buhler B, Schmid A. Metabolic capacity estimation of Escherichia coli as a platform for redox biocatalysis: constraint-based modeling and experimental verification. *Biotechnol Bioeng* 2008, 100(6):1050–1065.
27. Portnoy VA, Herrgard MJ, Palsson BO. Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain. *Appl Environ Microbiol* 2008, 74(24):7561–7569.
28. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol* 2007, 3:149.
29. Schwender J. Metabolic flux analysis as a tool in metabolic engineering of plants. *Curr Opin Biotechnol* 2008, 19(2):131–137.
30. Izallalen M, Mahadevan R, Burgard A, Postier B, Didonato R Jr, et al. Geobacter sulfurreducens strain engineered for increased rates of respiration. *Metab Eng* 2008, 10(5):267–275.
31. Cakir T, Efe C, Dikicioglu D, Hortacsu A, Kirdar B, et al. Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. *Biotechnol Prog* 2007, 23(2):320–326.
32. Kim TY, Kim HU, Park JM, Song H, Kim JS, et al. Genome-scale analysis of Mannheimia succiniciproducens metabolism. *Biotechnol Bioeng* 2007, 97(4):657–671.
33. Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* 2000, 97(10):5528–5533.
34. Beard DA, Liang SD, Qian H. Energy balance for analysis of complex metabolic networks. *Biophys J* 2002, 83(1):79–86.
35. Balabanian N, Bickart TA. *Electrical Network Theory*. Malabar, FL: R.E. Krieger Pub. Co.; 1983.
36. Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 2001, 213(1):73–88.
37. Covert MW, Palsson BO. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 2002, 277(31):28058–28064.
38. Covert MW, Palsson BO. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* 2003, 221(3):309–325.
39. Herrgard MJ, Lee BS, Portnoy V, Palsson BO. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* 2006, 16(5):627–635.
40. Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol* 2006, 2(8):e101.
41. Gianchandani EP, Joyce AR, Palsson BO, Papin JA. Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system. *PLoS Comput Biol* 2009, 5(6):e1000403.
42. Min Lee J, Gianchandani EP, Eddy JA, Papin JA. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol* 2008, 4(5):e1000086.
43. Gerdtzen ZP, Daoutidis P, Hu WS. Non-linear reduction for kinetic models of metabolic reaction networks. *Metab Eng* 2004, 6(2):140–154.
44. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 2002, 99(23):15112–15117.
45. Stephanopoulos G, Alper H, Moxley J. Exploiting biological complexity for strain improvement through systems biology. *Nat Biotechnol* 2004, 22(10):1261–1267.
46. Mahadevan R, Edwards JS, Doyle FJ III. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 2002, 83(3):1331–1340.
47. Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA. Predicting biological system objectives de novo from internal state measurements. *BMC Bioinform* 2008, 9:43.
48. Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J Bacteriol* 2001, 183(4):1441–1451.
49. Burgard AP, Maranas CD. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* 2003, 82(6):670–677.
50. Nolan RP, Fenley AP, Lee K. Identification of distributed metabolic objectives in the hypermetabolic liver by flux and energy balance analysis. *Metab Eng* 2006, 8(1):30–45.
51. Knorr AL, Jain R, Srivastava R. Bayesian-based selection of metabolic objective functions. *Bioinformatics* 2007, 23(3):351–357.
52. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* 2007, 3:119.
53. Wiechert W. An introduction to 13C metabolic flux analysis. *Genet Eng* 2002, 24:215–238.
54. Weckwerth W. Metabolomics in systems biology. *Annu Rev Plant Biol* 2003, 54:669–689.
55. Fiehn O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 2002, 48(1–2):155–171.
56. Sauer U. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol* 2004, 15(1):58–63.

57. Hellerstein MK. In vivo measurement of fluxes through metabolic pathways: the missing link in functional genomics and pharmaceutical research. *Annu Rev Nutr* 2003, 23:379–402.
58. Smedsgaard J, Nielsen J. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J Exp Bot* 2005, 56(410):273–286.
59. Fischer E, Zamboni N, Sauer U. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. *Anal Biochem* 2004, 325(2):308–316.
60. Fischer E, Sauer U. Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS. *Eur J Biochem.* 2003, 270(5):880–891.
61. Blank LM, Sauer U. TCA cycle activity in Saccharomyces cerevisiae is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology* 2004, 150(Pt 4):1085–1093.
62. Blank LM, Kuepfer L, Sauer U. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* 2005, 6(6):R49.
63. Blank LM, Lehmbeck F, Sauer U. Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res* 2005, 5(6–7):545–558.
64. Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, et al. Metabolite profiling for plant functional genomics. *Nat Biotechnol* 2000, 18(11):1157–1161.
65. Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L. Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J* 2000, 23(1):131–142.
66. Jansen JJ, Hoefsloot HC, Boelens HF, van der Greef J, Smilde AK. Analysis of longitudinal metabolomics data. *Bioinformatics* 2004, 20(15):2438–2446.
67. Landau BR, Wahren J, Chandramouli V, Schumann WC, Ekberg K, et al. Use of 2H2O for estimating rates of gluconeogenesis. Application to the fasted state. *J Clin Invest* 1995, 95(1):172–178.
68. Zamboni N, Sauer U. Model-independent fluxome profiling from 2H and 13C experiments for metabolic variant discrimination. *Genome Biol* 2004, 5(12): R99.
69. Antoniewicz MR, Aleman J, Kelleher JK, Stephanopoulos GN, eds. *Comprehensive Analysis of the Gluconeogenesis Pathways through the Use of Multiple Isotopic Tracers*. The 2006 Annual Meeting of the American Institute for Chemical Engineers (AIChE). San Francisco, CA: American Institute for Chemical Engineers; 2006.
70. Antoniewicz MR, Aleman J, Kelleher JK, Stephanopoulos GN, eds. *High Resolution Flux Estimation of Gluconeogenesis Using A Novel [U-13C,2H8]Glycerol Tracer in Mice*. The 2007 Annual Meeting of the American Institute for Chemical Engineers (AIChE). Salt Lake City, UT: American Institute for Chemical Engineers; 2007.
71. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab Eng* 2007, 9(1):68–86.
72. McConville MJ, de Souza D, Saunders E, Likic VA, Naderer T. Living in a phagolysosome; metabolism of Leishmania amastigotes. *Trends Parasitol* 2007, 23(8):368–375.
73. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* 2008, 4(5):e1000082.
74. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 2008, 26(9):1003–1010.
75. Raman K, Rajagopalan P, Chandra N. Flux balance analysis of mycolic acid pathway: targets for antitubercular drugs. *PLoS Comput Biol* 2005, 1(5):e46.
76. Manichaikul A, Ghamsari L, Hom EFY, Lin C, Murray RR, et al. Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nat Methods* 2009, 6(8):589–592.
77. Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci U S A* 2007, 104(19):7797–7802.
78. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, et al. In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 2005, 91(5):643–648.
79. Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, et al. Metabolic engineering of Escherichia coli for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. *Appl Environ Microbiol* 2005, 71(12):7880–7887.

