



Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation

Tunahan Çakır^{1*} and Mohammad Jafar Khatibipour^{1,2}

¹ Computational Systems Biology Group, Department of Bioengineering, Gebze Technical University (formerly known as Gebze Institute of Technology), Gebze, Turkey

² Department of Chemical Engineering, Gebze Technical University (formerly known as Gebze Institute of Technology), Gebze, Turkey

Edited by:

Daniel Machado, University of Minho, Portugal

Reviewed by:

Pierre Millard, University of Manchester, UK
Corrado Priami, University of Trento and COSBI, Italy
Huub C. J. Hoefsloot, University of Amsterdam, Netherlands

*Correspondence:

Tunahan Çakır, Department of Bioengineering, Gebze Technical University, Gebze 41400, Kocaeli, Turkey
e-mail: tcakir@gyte.edu.tr

The primary focus in the network-centric analysis of cellular metabolism by systems biology approaches is to identify the active metabolic network for the condition of interest. Two major approaches are available for the discovery of the condition-specific metabolic networks. One approach starts from genome-scale metabolic networks, which cover all possible reactions known to occur in the related organism in a condition-independent manner, and applies methods such as the optimization-based Flux-Balance Analysis to elucidate the active network. The other approach starts from the condition-specific metabolome data, and processes the data with statistical or optimization-based methods to extract information content of the data such that the active network is inferred. These approaches, termed bottom-up and top-down, respectively, are currently employed independently. However, considering that both approaches have the same goal, they can both benefit from each other paving the way for the novel integrative analysis methods of metabolome data and flux-analysis approaches in the post-genomic era. This study reviews the strengths of constraint-based analysis and network inference methods reported in the metabolic systems biology field; then elaborates on the potential paths to reconcile the two approaches to shed better light on how the metabolism functions.

Keywords: constraint-based models, metabolic network inference, active metabolic state, metabolome, network biology, reverse engineering, flux-balance analysis

INTRODUCTION

Metabolic network is the outmost layer of cellular activity from the genome. The genome of a cell is a comprehensive and condensed information base, defining a boundary for the biochemical capacity of the cell. The processing of genetic information passes through several layers of fabrication and regulation before reaching their end products. This is from information to the function, from genotype to phenotype. Metabolic enzymes count for a significant percentage of the end products of genes, and their activity sets the physiology of the cell. Since metabolic network activity is the major representative of cell functionality, it is of great importance to gain as much knowledge as possible on the active metabolic network at a specific cellular state.

Systems-based approach to molecular biology has contributed to an increased knowledge of metabolic pathways for an increasing number of organisms, and led to almost complete metabolic networks for a number of major organisms, from yeast to human. Such static networks are available in a condition-independent manner through web-based databases such as KEGG or MetaCyc (Altman et al., 2013), or reconstructed in a format suitable for simulation by several researchers at genome scale (Oberhardt et al., 2009; Kim et al., 2012). There are several mathematical approaches to process such networks to come up with condition-specific networks, the most common one being the Flux-Balance Analysis (FBA) framework (Orth et al., 2010). This is a bottom-up direction toward the active network since already-known “parts,”

interactions, are used as inputs (Bruggeman and Westerhoff, 2007; Petranovic and Nielsen, 2008).

In parallel to the developments on the knowledge of metabolic networks, techniques to measure metabolite levels at high throughput, termed metabolomics, have arisen (Kell, 2004; Dunn et al., 2005). Quantitative or semi-quantitative metabolome data, although one of the most challenging compared to other omic sciences, have come a long way in a decade, from the detection and quantification of about 50 metabolites (Devantier et al., 2005) to more than 1000 metabolites (Psychogios et al., 2011). Metabolome data are a snapshot of the condition-specific status of the investigated organisms. Reverse-engineering metabolome data to discover the underlying network structure is the goal behind metabolic network inference approaches (Srividhya et al., 2007; Çakır et al., 2009). The information content of metabolome data is revealed by processing it with correlation or optimization-based methods (Weckwerth et al., 2004; Hendrickx et al., 2011; Öksüz et al., 2013). Such an approach to discover metabolic network structure is termed top-down approach since the parts, interactions, are not known *a priori*, and predicted from the whole set of available biomolecules (Bruggeman and Westerhoff, 2007; Petranovic and Nielsen, 2008).

In this review, we will cover the basic developments in bottom-up and top-down approaches to discover active metabolic network, and then ponder over the possible ways of reconciling these two approaches for a better prediction of active

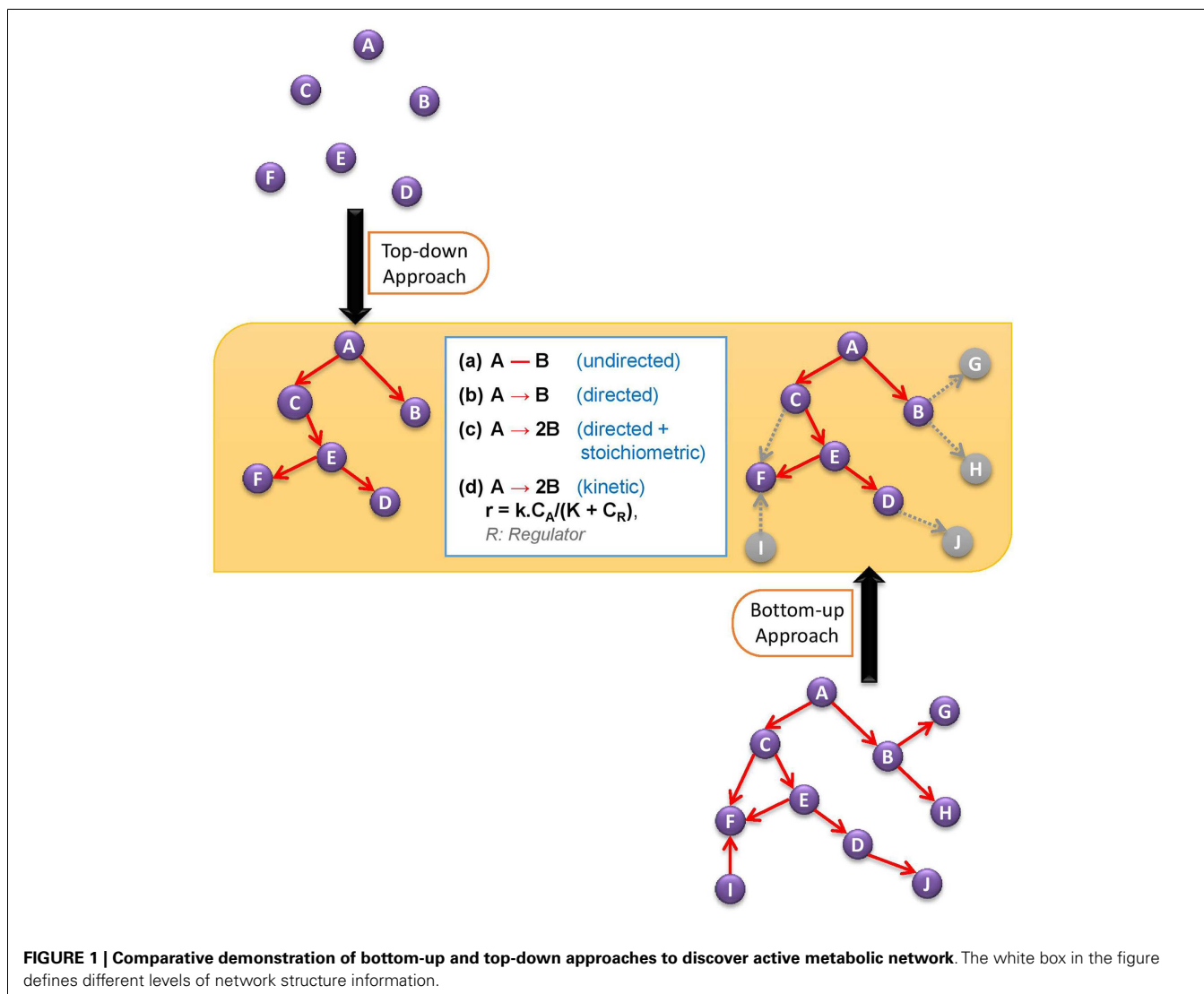


FIGURE 1 | Comparative demonstration of bottom-up and top-down approaches to discover active metabolic network. The white box in the figure defines different levels of network structure information.

network structure. **Figure 1** illustrates the two alternative network discovery approaches.

BOTTOM-UP APPROACHES TO DISCOVER CONDITION-SPECIFIC METABOLIC NETWORKS

Different methods and algorithms have been used for the discovery and characterization of active metabolic networks at different states of cells and culture environments. In the bottom-up approach, everything starts from an already available network of biochemical transformations that cover all possible scenarios in the distribution of metabolic fluxes, and sets an upper bound for the existence of reactions in the active metabolic network. Such a network is termed a static metabolic network. A static metabolic network can be provided either by a previously reconstructed genome-scale stoichiometric model or by a collection of all reactions whose existence in the organism of interest has been certified in literature and databases. Most popular among such databases are KEGG (Kanehisa et al., 2014), MetaCyc (Caspi et al., 2014), and Reactome (Croft et al., 2014). Other efforts with more curated

databases such as Rhea (Alcántara et al., 2012) and MetRxn (Kumar et al., 2012) are also available. A genome-scale stoichiometric model is reconstructed based on the annotation of all genes in the genome of one organism to their end products and then to the corresponding reactions, leading to a list of gene-protein-reaction rules (Thiele and Palsson, 2010). In this way, the minimum information content of a genome-scale model is (i) a list of reactions, and (ii) a list of gene-protein-reaction rules. The presence of gene-protein-reaction rules in stoichiometric models has enabled the opportunity for transcriptome and proteome data to be incorporated into the discovery methods of active metabolic networks (Blazier and Papin, 2012).

Given a genome-scale reaction network, the aim is to find the active reaction network at a specific condition or for a specific cell type in a multicellular organism (**Box 1**). The core of all such discovery approaches is a stoichiometric matrix. Each row of the stoichiometric matrix represents a metabolite and each column stands for a reaction, the corresponding element being the stoichiometric coefficient of that metabolite in that reaction. The relationship

Box 1 | Different levels of Metabolic Network Structure Information.

Our understanding of an active metabolic network can be sorted into several stages of information.

- (i) At the lowest level of information, we want to know what the structure of the network is, representing it with an undirected (or directed, if the reversibility information is available) graph in which each node stands for a metabolite and each edge stands for a biochemical transformation. Alternative to the retrieval from the metabolic reaction databases, the structure of the network – both directed and undirected – can also be estimated to some extent by analyzing and reverse engineering the metabolome data without the use of *a priori* database information on the reactions.
- (ii) At a higher level, the information on the stoichiometry of reactions can be incorporated, leading to a directed stoichiometric biochemical network.
- (iii) Having the stoichiometric structure of the network, we can characterize the metabolic state in more detail by quantifying the metabolic fluxes. In most cases, rather than a unique flux distribution, constraints are set on flux values to shrink the solution space. Such modeling approaches are known as Constraint-Based Modeling. This level of understanding the active metabolic network (structure + flux distribution) has been the area of focus in the research community for more than a decade. In most cases, the information provided at this level has been satisfactory for engineering research to design more efficient cell factories, and also, recently, for medical research to distinguish significant differences between healthy and disease states.
- (iv) There are, however, certain limitations at the above level although it provides a network activity structure weighted with fluxes. The dynamic behavior of the system cannot be captured, and the predictability power of such models is hampered mainly because they are not considering the role of regulatory mechanisms in controlling the rate of biochemical reactions. In some cases, the regulation of reaction rates plays such a dominant role that it would be hard to make any prediction by just considering the flux-based network activity structure. Here come the kinetic models into the picture, which take enzymatic regulations and metabolite concentrations into account for a dynamic and better prediction of network structure.

between the reaction rates in the network and the dynamic change in the concentration of metabolites is represented as given below:

$$\frac{d\mathbf{C}}{dt} = \mathbf{S} \times \mathbf{v} \quad (1)$$

where \mathbf{S} is the stoichiometric matrix, \mathbf{C} is the vector of intracellular metabolite concentrations, and \mathbf{v} is a column vector of metabolic reaction rates (fluxes) to be determined. Under the assumption of steady state, the concentration of each intracellular metabolite is not going to change with time, meaning the sum of rate of reactions producing that metabolite is equivalent to the sum of rate of reactions consuming that metabolite (metabolic fluxes around each metabolite are balanced). This is represented mathematically

as follows:

$$\mathbf{S} \times \mathbf{v} = \mathbf{0} \quad (2)$$

This is an algebraic system of linear equations with all fluxes being zero as a trivial solution. In order to escape from the trivial solution, the value of at least one of the fluxes must be set to a non-zero value, that flux usually being an exchange flux between the intracellular and extracellular environment since the experimental measurement of exchange fluxes is relatively easier. The system is almost always underdetermined with a large solution space, mainly because of the existence of branch points in the metabolic network. There are both experimental and computational approaches to estimate a condition-specific network for such a system.

The experimental approach is based on stable-isotope (mostly ^{13}C carbon) labeling of the major carbon source, and then tracing the propagation of the labeled carbon atoms down to protein-bound amino acids at isotopic steady state by using mass spectrometry or NMR spectroscopy (Wiechert et al., 2001; Sauer, 2006; Mueller and Heinzle, 2013). The qualitative isotopic labeling information is then used as an input to two alternative methods. In one method, termed isotopomer modeling, a total flux distribution is estimated based on the experimental labeling results through a computationally demanding non-linear optimization formulation, which employs global iterative fitting and statistical analysis (Wiechert et al., 2001; Antoniewicz et al., 2007). The other ^{13}C -labeling assisted method is based on the estimation of the local ratios of fluxes emerging from a branch point (Sauer, 2006; Zamboni et al., 2009) rather than the absolute quantification of all fluxes. These experimental flux split ratios can be used to shrink the solution space of Eq. 2 in a complementary flux calculation, leading to the discovery of a condition-specific network (Schuetz et al., 2007; Tarlak et al., 2014). Softwares are available for the rather sophisticated calculation of experimental fluxes (or flux ratios) from carbon labeling data for both methods (Zamboni et al., 2005; Quek et al., 2009; Weitzel et al., 2013). A new trend in this area is to collect data at the non-stationary phase of isotopic labeling rather than at the isotopic steady state, which was shown to be more informative in terms of predicting the flux-weighted active metabolic network structure (Schaub et al., 2008; Young et al., 2008; Wiechert and Nöh, 2013). Works on the tracing of intracellular metabolites rather than only 10–15 protein-bound amino acids have also appeared due to the higher coverage of metabolic pathways despite the inherent experimental difficulties in terms of higher turnover rates as well as stability issues (Van Winden et al., 2005; Toya et al., 2007; Millard et al., 2014).

The computational approach for the discovery of condition-specific metabolic network based on Eq. 2 is known as constraint-based modeling. Constraint-based modeling methods aim to shrink the solution space of the equation as much as possible by putting relevant constraints on the system. The most common method, FBA, treats the problem in Eq. 2 as an optimization problem and linear programming is applied to solve it. The stoichiometry of metabolic reactions (stoichiometric matrix), reaction directionality information, a physiologically relevant objective function, and the value of at least one of the exchange fluxes are all that

are required for FBA to return a condition-specific flux distribution. The flux distribution returned by FBA is not necessarily unique, and there may be a variety of flux distributions all leading to the same optimum value of the objective function. Therefore, Flux Variability Analysis (FVA) must be used together with FBA, to determine the variability, if any, on each metabolic flux in regard to the condition of interest (Mahadevan and Schilling, 2003; Müller and Bockmayr, 2013). The maximization of biomass production has been successfully applied as a reliable objective function for FBA to predict flux distributions in a variety of microorganisms (Varma and Palsson, 1994; Feist and Palsson, 2010). In some studies, it has been hypothesized that one objective function alone may not capture the metabolic behavior of the cell comprehensively. Therefore, multi-objective optimization platforms have been designed and utilized to come up with more specific flux distributions. Several modified versions of FBA including parsimonious FBA, pFBA (Lewis et al., 2010), and flexible-optimality FBA, flexoFBA (Tarlak et al., 2014), have been developed in this manner. On the other hand, some research groups have developed methods based on the availability of additional omics data, which are discussed below. For a thorough review of a number of FBA-derived flux calculation methods, the readers are referred to Lewis et al. (2012).

CONSTRAINTS BASED ON TRANSCRIPTOME OR PROTEOME DATA

The rate of an enzymatic reaction inside the cell is a function of several different factors, such as the concentration of substrates, products, and regulators of the enzyme and also the amount of available active enzyme for that reaction. Among these factors, the concentration of active enzymes can be related to the activity of genes through layers of transcription, translation, and post-translational modifications. Transcriptome data are much more accessible and comprehensive compared to the other omics data. Several different research groups have developed different strategies to incorporate transcriptome data into constraint-based models. The idea behind this is that the amount of mRNAs (gene activities) may be correlated with the concentration of active enzymes, and hence this can be utilized to provide additional constraints on metabolic fluxes. At the bottom line, if an enzyme coding gene is not transcribed at steady state, the corresponding reaction should be inactive at that steady state, if there is no other enzyme catalyzing that reaction. This idea was first used by Åkesson et al. to set the flux values to zero for those reactions whose corresponding genes were expressed at low levels (Åkesson et al., 2004). More sophisticated and structured versions of this approach appeared later, under the names of GIMME (Becker and Palsson, 2008) and iMAT (Shlomi et al., 2008). These approaches classify some reactions as inactive reactions based on the low expression levels of their associated genes. Then, they employ a computational framework which minimizes the contradiction between the classification and an active physiological flux distribution since some of these classifications may render the flux state unrealistic (such as zero growth rate). Several other alternative methods appeared recently to incorporate transcriptome data into the prediction of active metabolic network and flux distribution. In an interesting study, for example, mRNA levels from transcriptome data were used as weights for the corresponding reactions to predict a flux

distribution without using a conventional objective function such as the maximization of biomass growth (Lee et al., 2012). A recent study (Machado and Herrgård, 2014) evaluated these methods systematically for the prediction of flux distributions, and the results were compared to that of parsimonious FBA as a reference method that does not consider the transcriptome data. In general, none of the methods could significantly improve the results of pFBA and none of them outperformed the others for the tested cases (*S. cerevisiae* and *E. coli*). Instead of the prediction of flux distributions, these methods, however, may significantly help in the discovery of active metabolic networks in context/tissue-specific cells and in the conditions where a relevant objective function cannot be hypothesized.

Transcriptome data are not necessarily correlated with the rate of corresponding reactions. Inconsistency between mRNA levels and reaction rates is a result of influence of several other factors in the regulation of enzymatic reactions. Therefore, if proteome data are available, it can be used instead of transcriptome data as a better representative for the concentration of active enzymes since proteome is hierarchically closer to the enzyme states than transcriptome data. The methods that are developed to integrate transcriptome data with the FBA method can all be used for the purpose of integrating proteome data. For example, GIMMEp (Bordbar et al., 2012) is the proteome equivalent version of GIMME. Some of such integrative methods were primarily tested with proteome data. INIT (Agren et al., 2012), for example, was developed by using proteome abundance data from Human Protein Atlas database. However, it was shown that utilizing proteome data instead of transcriptome data could not improve the prediction of flux distributions for the tested cases (*S. cerevisiae* and *E. coli*) (Machado and Herrgård, 2014). In a study which used metabolome and proteome data in the flux calculation method, on the other hand, even the use of only proteome data were shown to improve the results compared to the traditional FBA (see the next section for more details) (Yizhak et al., 2010).

Substrate concentrations, the concentration of enzyme regulators, the turn over number of the catalyzing enzyme, and the concentration of the active enzyme are all playing significant roles in the determination of reaction rates, and among them only the concentration of the active enzyme may be represented by the corresponding protein or mRNA concentration. Translated proteins are not necessarily active enzymes, and they may need to undergo post-translational modifications (e.g., phosphorylation/acetylation) to become capable of catalyzing the reactions. This is one of the main reasons behind inconsistency between protein levels and reaction rates. On the other hand, the turn over number (catalytic power) of one enzyme may differ by several orders of magnitude from the turn over number of another enzyme (Hoppe, 2012). It means that although the concentration of one enzyme may be much less than the others in the network, the reaction catalyzed by that enzyme can proceed much faster than others. According to this fact, the use of the absolute concentrations of proteins or mRNAs to constrain reaction rates does not seem promising. However, the turn over number of one enzyme in an individual is an intrinsic parameter of the enzyme that does not change from one condition to another except by effective mutations that rarely occur. Because of this, the relative levels of proteins

or mRNAs can be utilized to overcome the problem of big differences in turn over numbers. One steady state with available data on flux values and protein/mRNA levels can be taken as the reference state, and then the relative/differential levels of proteins/mRNAs to the reference state can be used to predict the flux distributions at the new conditions. Based on this approach, algorithms have been developed to incorporate relative/differential transcriptome data into metabolic-flux analysis, among which are MADE (Jensen and Papin, 2011) and GX-FBA (Navid and Almaas, 2012). One other main reason for the inconsistency between protein levels and reaction rates is the distribution of flux control among different layers from genotype to phenotype. Metabolic fluxes can be regulated hierarchically (through gene expression levels) or metabolically (through metabolic interactions) (Daran-Lapujade et al., 2007; Postmus et al., 2008; Nikerel et al., 2012; Chubukov et al., 2013). Use of transcriptome or proteome data will not be helpful if the metabolic fluxes are controlled at the metabolic level.

CONSTRAINTS BASED ON METABOLOME DATA

One approach to find more specific and physiologically relevant flux distributions is to provide additional constraints by specifying the directionality of reversible reactions. This can be done by taking Gibbs free energies of metabolites into consideration. The Gibbs free energy change of a reversible biochemical transformation (one reaction or a series of reactions) determines the direction of that transformation and its departure from reversibility. The earlier studies assumed standard conditions (all metabolite concentrations were assumed to be 1 M), and did not explicitly consider metabolite concentrations in the calculation of Gibbs energy changes of reactions due to the scarcity of metabolome data (Henry et al., 2006). Recent studies, however, take the concentration of metabolites into account, when available, to perform thermodynamic-based metabolic-flux analysis, leading to more reliable predictions (Hoppe et al., 2007; Bennett et al., 2009; Soh and Hatzimanikatis, 2010; Hamilton et al., 2013).

Extracellular metabolome data can be used to constrain genome-scale metabolic models for the calculation of intracellular flux distributions by simply constraining the secretion and uptake rates of extracellular metabolites based on such data (Çakır et al., 2007; Mo et al., 2009). In a different approach, Michaelis–Menten-based kinetics was used for the estimation of reaction rates for the reactions for which appropriate intracellular metabolome (and proteome) data are available (Yizhak et al., 2010). The FBA framework was designed in such a way that the calculated fluxes are as consistent as possible with the kinetically derived reaction rates, if available. The simultaneous use of metabolome and proteome data for this purpose significantly improved the results. The use of metabolome data alone also resulted in better predictions than the traditional FBA. In a recent study, a kinetic platform was established based on Michaelis–Menten equation to bridge gene expression levels, metabolite concentrations and metabolic fluxes without requiring the knowledge of kinetic parameters (Zelezniak et al., 2014). They could show that changes in metabolite concentrations relative to a reference steady state can be predicted by their formulation that includes information on network connectivity in addition to differential mRNA expression levels. All those works utilizing kinetic information demonstrate the necessity of

dynamic models for a more comprehensive analysis of metabolic networks.

Kinetic models of biochemical reactions not only provide a rational platform for omics data – especially metabolomics – to be incorporated in the estimation of metabolic fluxes but also they enable the prediction and study of the dynamics of metabolic networks far beyond the steady state (**Box 1**). Such models were only possible for small-scale metabolic networks until recently (Teusink et al., 2000; Chassagnole et al., 2002), since, they require detailed information on the enzyme kinetics of each individual reaction. Estimation of kinetic parameters is a major obstacle in the applicability of dynamic modeling of metabolic networks. New platforms and algorithms were established to circumvent this problem so that the estimation of explicit kinetic parameters is not a prerequisite to study the dynamic capacity and behavior of the system (Link et al., 2014). Approximative kinetic models (lin-log, power-law, mass-action) on the other hand, try to fit a standard rate expression formula to all reactions of the network to increase the range of their applicability to larger networks (Visser et al., 2004; Sorribas et al., 2007). Thanks to approximative kinetics, attempts to reconstruct large-scale kinetic metabolic models with more than 100 reactions were recently presented (Smallbone et al., 2010; Chakrabarti et al., 2013; Stanford et al., 2013), but their prediction power is limited to the conditions adequately close to the corresponding steady state.

As a better alternative to approximative kinetics, an approach was established and utilized based on the concept of parametric Jacobian, which covers the behavior of all possible kinetic models that are consistent with an experimentally observed operating point (Steuer et al., 2006). This approach provides an opportunity to detect and analyze bifurcation characteristics of the metabolic network without the need for explicit determination of kinetic parameters. Ensemble modeling of metabolic networks (Tran et al., 2008) is an elegant idea for large-scale kinetic modeling of biochemical reaction networks. In this method, each enzymatic reaction is broken down to its elementary reactions that all follow mass-action kinetics. An ensemble of thermodynamically consistent kinetic models with different dynamic behavior that all converge to a reference steady state is collected with the help of intracellular metabolome data. This ensemble is then filtered by the results of perturbation experiments to filter out inconsistent models from the ensemble and to increase the predictability of remaining models. The approach was successfully applied, among others, to construct kinetic models of *E. coli* (Khodayari et al., 2014) and cancer metabolisms (Khazaei et al., 2012), leading to promising flux predictions.

TOP-DOWN APPROACHES TO DISCOVER CONDITION-SPECIFIC METABOLIC NETWORKS

Time series of metabolite concentrations in response to a perturbation, and also replicates of metabolome data at a specific steady state, both implicitly contain information on the structure of active metabolic network. Reverse engineering of these data to infer the condition-specific metabolic network without necessarily prior knowledge on the genome of the organism and its static metabolic network is an alternative to all bottom-up approaches that are based on the availability of a large-scale stoichiometric model of the organism. Although promising, less attention has

been paid to these top-down approaches compared to bottom-ups mainly because of the technical obstacles in gathering reliable metabolome data in large scale. This limitation will be removed with future advancements in the detection and quantification of intracellular metabolites such as higher coverage and temporal resolution. At this stage, however, several research groups have established algorithms and methods for reverse engineering of metabolic networks by using either time series or steady-state replicates of metabolite concentrations (Crampin et al., 2004; Chou and Voit, 2009; Hendrickx et al., 2011; Lecca and Priami, 2013).

NETWORK DISCOVERY BASED ON TIME-SERIES DATA

The use of time-series metabolite concentration data to predict the underlying network connectivity information first appeared in the literature about two decades ago. Time-lagged correlations combined with a projection technique called multidimensional scaling were shown to construct the structure of generic biochemical networks with few nodes (Arkin and Ross, 1995). Correlation between time-series profiles of metabolites, with the consideration of the delay in the influence of one metabolite on the next, is the basis of the time-lagged correlation method for the inference of metabolic networks. The approach, called correlation metric construction, was later experimentally verified *in vitro* by inferring the first steps of glycolytic pathway in a 14-metabolite system (Arkin et al., 1997). Modified versions of the approach appeared later (Samoilov et al., 2001; Lecca et al., 2012). In the latter, metabolic pathway of an anticancer drug was deduced from the time-lagged correlations of corresponding metabolite concentration measurements. The modification introduced by the former work was recently improved by using mutual information similarity score rather than simple linear correlation (Villaverde et al., 2014). The authors compared their method, called MIDER, with several other methods by applying it to different types of cellular networks, including *in vitro* glycolytic pathway data. The approach outperformed the other methods.

Another method to reconstitute a network using time-series data is based on perturbation experiments around steady state. The initial curve of concentration changes of metabolites in response to a pulse change on the concentration of a metabolite is processed with the method of zero initial slopes (Vance et al., 2002). The method successfully inferred the structure of glycolysis based on *in vitro* experimental data (Torralba et al., 2003). Performance comparison of the method with the correlation metric construction approach was later provided based on *in silico* data of *S. cerevisiae* and *E. coli* central metabolic networks (Hendrickx et al., 2011). An approach based also on perturbation experiments, but with a different formulation aiming to calculate Jacobian matrix from time derivatives of concentration data, was first applied to gene networks (Schmidt et al., 2005). A modified version of the approach recently used *in vivo* metabolite concentration measurements from tomato seedlings to reconstruct quercetin glycosylation pathway (Astola et al., 2011).

Apart from such model-free structure identification methods, model-based methods use time-series metabolite concentration data not only to identify network structure but also to estimate proper model parameters such as rate constants of kinetic

expressions (Chou and Voit, 2009). Majority of these approaches use power-law (also called S-system) formulation (Savageau and Voit, 1987) to approximate reaction kinetics. An approach, for example, used S-system modeling with a multi-objective optimization by simultaneously minimizing the number of interactions and the error in the fitting (Liu and Wang, 2008). They applied their method to major metabolites involved in ethanol fermentation. An earlier work analyzed a small three-metabolite network of phospholipid metabolism by combining S-system modeling and an evolutionary modeling method, genetic programming (Ando et al., 2002). Later, a new representation of S-system approach, called S-trees, was combined with genetic programming to reverse-engineer yeast fermentation pathway in a more efficient manner by using *in silico* time-series concentration data of five metabolites (Cho et al., 2006). In a sophisticated approach, others used symbolic regression based on genetic programming to infer both the structure and the model of yeast glycolytic oscillations from *in silico* data (Schmidt et al., 2011). Their use of acyclic graph encoding rather than tree-based encoding together with symbolic regression approach ensured the identification of parsimonious (sparse) models. Rather than S-system formulation, mass-action kinetics can also be used to infer pathway connectivity and reaction mechanism (Srividhya et al., 2007). This minimizes the computational burden on the algorithm since only rate constants are to be estimated as parameters in the mass-action formulation. The authors tested their method with real time course experimental metabolome data of *Lactococcus lactis* glycolysis. A graphical user interface was later made available by the same group to ease the inference of kinetics and network architecture from dynamic data of biochemical pathways (Mourão et al., 2011). Genetic programming was also combined with mass-action kinetics in an algorithm, which ensures the estimation of biochemically more plausible models (Gormley et al., 2013). The small phospholipid network of (Ando et al., 2002) was inferred in a more compact way by this algorithm.

NETWORK DISCOVERY BASED ON STEADY-STATE DATA

The use of steady-state metabolome data to infer metabolic network structure has also drawn attention in the last decade. The biological variability in the metabolism of the organisms at around steady state is a known phenomenon due to slight variations in the enzyme levels or due to slight natural or environment-induced fluctuations within cellular processes. Slight variations in the steady-state measurements of metabolite levels can be informative on the network structure (Steuer et al., 2003; Camacho et al., 2005; Çakır et al., 2009). The most common approach here is to use the similarity measures such as Pearson correlation to assign edges between metabolites. One should note that such correlations are not necessarily strong among neighboring metabolites whereas there could be strong correlations among distant metabolites in the network (Camacho et al., 2005). In a comprehensive study, different alternative similarity measures (linear vs. non-linear, and full vs. partial) were applied to *in silico* metabolome data belonging to two microorganisms to systematically analyze method performances (Çakır et al., 2009). The results revealed no clear superiority between linear (Pearson correlation) and non-linear (mutual information) similarity measures. The

best performing method was identified as n th order partial Pearson correlation, known also as graphical Gaussian modeling. Graphical Gaussian modeling was also applied to metabolome data from blood serum samples to reconstruct human fatty acid metabolism (Krumstiek et al., 2011). Others (Nemenman et al., 2007) analyzed *in silico* metabolome data of red blood cell metabolism by ARACNE approach (Margolin et al., 2006), which is based on pruning mutual information scores. An elegant improvement on ARACNE based reverse engineering of metabolic profiling data was suggested later (Bandaru et al., 2011). The approach puts a constraint on the possible metabolic transformations to satisfy the mass conservation between the connected metabolites. Synthetic data covering up to about 200 metabolites were generated to test the approach. One issue in such similarity-based approaches is that only pairwise interactions are aimed to be found. However, a metabolic reaction can involve more than two metabolites. Based on this reasoning, an attempt to also deduce triple interactions by using ternary mutual information was suggested (Diệp et al., 2011). Analysis of synthetic yeast glycolysis data and red blood cell data showed the success of this approach in capturing higher order interactions.

A different approach to discover active metabolic networks from steady-state data is based on Lyapunov equation. In Eq. 1, the rate vector, \mathbf{v} , is a complex non-linear function of concentrations, \mathbf{C} . For systems around steady state, the equation can be expressed in terms of Jacobian matrix, \mathbf{J} , by the help of linear approximation:

$$\frac{d\mathbf{X}}{dt} \approx \mathbf{J}\mathbf{X} \quad (3)$$

with $\mathbf{X} = \mathbf{C} - \mathbf{C}_s$, and \mathbf{C}_s shows the steady-state metabolite concentrations. Jacobian matrix stores detailed information on the structure of the underlying network; such as the directionality of interaction, strength of interaction, and regulation type of interaction. For small fluctuations around steady state, the right-hand side of Eq. 3 becomes zero, and the left-hand side can be expressed in such a way that a link between the covariance matrix of metabolome data, $\mathbf{\Gamma}$, and Jacobian matrix is provided. The details of the derivation are given elsewhere (Van Kampen, 1992; Steuer et al., 2003).

$$\mathbf{J}\mathbf{\Gamma} + \mathbf{\Gamma}\mathbf{J}^T = -2\mathbf{D} \quad (4)$$

\mathbf{D} in the equation shows the extent of fluctuations. Eq. 4, known as Lyapunov equation, can be used to infer metabolic network structure since it provides a link between the data-based covariance matrix and network connectivity stored in \mathbf{J} . Reverse-engineering metabolome data by using the Lyapunov equation was first discussed via a hypothetical three-metabolite system (Steuer et al., 2003). A recent work provided a theoretical analysis on the use of the Lyapunov equation to infer network structure from steady-state metabolome data (Öksüz et al., 2013). The authors used a rearranged version of the Lyapunov equation:

$$\mathbf{A}\mathbf{j} = 2\mathbf{d} \quad (5)$$

Here, \mathbf{j} and \mathbf{d} are vectorized versions of \mathbf{J} and \mathbf{D} matrices. \mathbf{A} is a matrix based on the covariance of data. In that work, directed networks were inferred from *in silico* metabolome data of *S. cerevisiae*

glycolysis, *E. coli* central carbon metabolism, and brain glycolysis by solving Eq. 5 for \mathbf{j} using a genetic-algorithm based formulation. In the optimization formulation, the dual objective function was simultaneous maximization of the sparse structure and minimization of the residual norm of the equation. When compared to the inference results based on n th order partial Pearson correlation, a much higher prediction accuracy was reported. One other advantage of the optimization-based approach is the fact that Eq. 5 infers a directed network whereas correlation-based approaches cannot predict directions of interactions. The Lyapunov equation was recently used to infer differential changes in Jacobian matrix rather than the inference of network structure by predicting Jacobian matrix itself (Sun and Weckwerth, 2012; Kügler and Yang, 2014; Nägele et al., 2014).

PATHS TO RECONCILE BOTTOM-UP AND TOP-DOWN METABOLIC NETWORK DISCOVERY APPROACHES

Previous sections reviewed bottom-up and top-down metabolic network discovery approaches from literature. Top-down approaches are dependent on intracellular metabolome data, and there are bottom-up approaches, which aim to use omics data as additional constraints. The simultaneous use of both approaches to discover better condition-specific networks has not been a focus in the scientific community. Here, we will elaborate on the ways to reconcile these two approaches when intracellular metabolome data of a condition in question are available.

All model-based top-down approaches using time-series data also infer a Jacobian matrix of the model. Many other top-down approaches are based on correlations between metabolites. There is a significant relationship between the correlation strengths and the strengths of interactions implied by Jacobian entries (Çakır et al., 2009). Therefore, correlation strengths or Jacobian-interaction strengths of the inferred edges can be used as edge scores in the bottom-up constraint-based modeling approaches as additional constraints for a better identification of the active metabolic network as follows: all inferred edges in a top-down approach based on metabolome data are ranked with respect to their edge scores. Afterward, cut-off values for high- and low-scores are determined. If a high-score edge also appears in the corresponding static genome-scale stoichiometric model, that reaction is assigned a high weight. If a high edge-score does not have a corresponding connection in the genome-scale model, this could imply a novel or a regulatory interaction. As it is known, genome-scale metabolic models do not account for regulatory interactions of metabolites with enzymes, however, top-down approaches do not have this limitation since they are purely data-based. If the edge-score is low, the corresponding reaction in the stoichiometric model is assigned a low weight. Similarly, if the top-down approach assigns no edge between two metabolites, which are linked with a reaction in the stoichiometric model, such reactions are also assigned low weight. All other reactions can be assigned with a medium-weight. Then, a mixed-integer programming based optimization framework can be used with Eq. 2 such that the resulting condition-specific flux distribution is as consistent as possible with the edge scores, including maximum possible number of high-weight reactions and minimum possible number of low-weight reactions as active. Thereby, the strength of

top-down predictions can be used for better bottom-up flux predictions.

Use of transcriptome or proteome data as constraints in metabolic-flux calculations resulted in several alternative methods such as GIMME, iMAT, and INIT. These approaches remove reactions from the static metabolic reaction set if the controlling gene or protein is not active. However, a recent work comparing all these methods could not identify a method with clear superiority over the parsimonious FBA (Machado and Herrgård, 2014). This approach can be combined with edge scores (inferred Jacobian-interaction strength or calculated correlation strength) information to yield better network identification. GIMME-like approaches remove reactions from the model, this means also removal of metabolites. Two different approaches can be used: (i) removed reactions whose main substrates and products show high edge scores must be retained in the reaction set, implying an active edge (ii) reactions whose main substrates and products show very low and insignificant correlations must be candidates to be removed from the reaction set, implying an inactive edge if their removal does not hamper the objective function. Such a flux calculation powered by the top-down inference of network edges can lead to a more refined network.

One reconciliation approach will be the integrative use of flux-balance equation (Eq. 2) and rearranged Lyapunov equation (Eq. 5). Flux-balance equation was widely used in the last two decades because of its simplicity, requiring only the stoichiometric coefficients of reactions, and few measurement constraints. The rearranged Lyapunov equation bears a similar simplicity since it is only based on the covariances of metabolome measurements. The only major issue, as it is the case in flux-balance equation, is a proper choice of objective function to solve the equation. Since both \mathbf{J} and \mathbf{v} , the unknowns in both equations, represent the active network structure, the coupled use of these two equations can be beneficial from two different aspects: (i) a better flux distribution can be found thanks to the metabolome-based constraint provided by Eq. 5, (ii) the information stored in stoichiometric matrix, since it will reveal all possible non-interacting pairs, will provide a constraint to get a better estimate of Jacobian matrix by setting edge scores of some pairs to zero.

An approach getting popular to construct genome-scale kinetic models is ensemble modeling. This modeling approach constructs kinetic models from an ensemble of models, and filters the inconsistent models out by using the results of perturbation experiments (Tran et al., 2008; Khodayari et al., 2014). On the other hand, a number of methods infer networks from time-series data by using a model-based approach. The output of such methods is both the network structure and the dynamic kinetic model with estimated parameters (Srividhya et al., 2007; Liu and Wang, 2008). A number of alternative models are scanned in these methods to infer the most suitable one. Therefore, the strengths of model-based network inference and ensemble-based kinetic model reconstruction can be combined to yield better frameworks.

In summary, both bottom-up and top-down discovery of metabolic networks have come a long way in the last 20 years, providing the scientific community with a number of computational methods, as reviewed in this review. Considering the improvements that are being experienced both on the coverage and precision of

metabolome data, the coming decade will witness an exponential increase in the number of metabolome datasets, similar to what was experienced with transcriptome data in the last decade. This review aimed at drawing attention to this point, as ways to reconcile the two major metabolic network discovery approaches will gain increasing importance.

ACKNOWLEDGMENTS

The financial support by TUBITAK, The Scientific and Technological Research Council of Turkey, through a career grant (Project Code: 110M464) is gratefully acknowledged.

REFERENCES

- Agren, R., Bordel, S., Mardinoglu, A., Pornputtpong, N., Nookaew, I., and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* 8:e1002518. doi:10.1371/journal.pcbi.1002518
- Åkesson, M., Förster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* 6, 285–293. doi:10.1016/j.ymben.2003.12.002
- Alcántara, R., Axelsen, K. B., Morgat, A., Belda, E., Coudert, E., Bridge, A., et al. (2012). Rhea – a manually curated resource of biochemical reactions. *Nucleic Acids Res.* 40, D754–D760. doi:10.1093/nar/gkr1126
- Altman, T., Travers, M., Kothari, A., Caspi, R., and Karp, P. D. (2013). A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics* 14:112. doi:10.1186/1471-2105-14-112
- Ando, S., Sakamoto, E., and Iba, H. (2002). Evolutionary modeling and inference of gene network. *Inf. Sci.* 145, 237–259. doi:10.1016/S0020-0255(02)00235-9
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007). Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86. doi:10.1016/j.ymben.2006.09.001
- Arkin, A., and Ross, J. (1995). Statistical construction of chemical reaction mechanisms from measured time-series. *J. Phys. Chem.* 99, 970–979. doi:10.1021/j100003a020
- Arkin, A., Shen, P., and Ross, J. (1997). A test case of correlation metric construction of a reaction pathway from measurements. *Science* 277, 1275–1279. doi:10.1126/science.277.5330.1275
- Astola, L., Groenenboom, M., Roldan, V. G., Van Eeuwijk, F., Hall, R. D., Bovy, A., et al. (2011). “Metabolic pathway inference from time series data: a non iterative approach,” in *Pattern Recognition in Bioinformatics*, eds M. Loog, L. Wessels, M. J. T. Reinders, and D. de Ridder (Berlin: Springer), 97–108.
- Bandaru, P., Bansal, M., and Nemenman, I. (2011). Mass conservation and inference of metabolic networks from high-throughput mass spectrometry data. *J. Comput. Biol.* 18, 147–154. doi:10.1089/cmb.2010.0222
- Becker, S. A., and Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* 4:e1000082. doi:10.1371/journal.pcbi.1000082
- 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. doi:10.1038/nchembio.186
- Blazier, A. S., and Papin, J. A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* 3:299. doi:10.3389/fphys.2012.00299
- Bordbar, A., Mo, M. L., Nakayasu, E. S., Schrimpe-Rutledge, A. C., Kim, Y.-M., Metz, T. O., et al. (2012). Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Mol. Syst. Biol.* 8, 558. doi:10.1038/msb.2012.21
- Bruggeman, F. J., and Westerhoff, H. V. (2007). The nature of systems biology. *Trends Microbiol.* 15, 45–50. doi:10.1016/j.tim.2006.11.003
- Çakır, T., Efe, Ç., Dikicioglu, D., Hortacısu, A., Kırdar, B., and Oliver, S. G. (2007). Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. *Biotechnol. Prog.* 23, 320–326. doi:10.1021/bp060272r
- Çakır, T., Hendriks, M. M., Westerhuis, J. A., and Smilde, A. K. (2009). Metabolic network discovery through reverse engineering of metabolome data. *Metabolomics* 5, 318–329. doi:10.1007/s11306-009-0156-4

- Camacho, D., de la Fuente, A., and Mendes, P. (2005). The origin of correlations in metabolomics data. *Metabolomics* 1, 53–63. doi:10.1007/s11306-005-1107-3
- Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, D459–D471. doi:10.1093/nar/gkt1103
- Chakrabarti, A., Miskovic, L., Soh, K. C., and Hatzimanikatis, V. (2013). Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8, 1043–1057. doi:10.1002/biot.201300091
- Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* 79, 53–73. doi:10.1002/bit.10288
- Cho, D.-Y., Cho, K.-H., and Zhang, B.-T. (2006). Identification of biochemical networks by S-tree based genetic programming. *Bioinformatics* 22, 1631–1640. doi:10.1093/bioinformatics/btl122
- Chou, I.-C., and Voit, E. O. (2009). Recent developments in parameter estimation and structure identification of biochemical and genomic systems. *Math. Biosci.* 219, 57–83. doi:10.1016/j.mbs.2009.03.002
- 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. doi:10.1038/msb.2013.66
- Crampin, E. J., Schnell, S., and McSharry, P. E. (2004). Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog. Biophys. Mol. Biol.* 86, 77–112. doi:10.1016/j.pbiomolbio.2004.04.002
- Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., et al. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Res.* 42, D472–D477. doi:10.1093/nar/gkt1102
- Daran-Lapujade, P., Rossell, S., van Gulik, W. M., Luttk, M. A. H., de Groot, M. J. L., Slijper, M., et al. (2007). The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15753–15758. doi:10.1073/pnas.0707476104
- Devantier, R., Scheithauer, B., Villas-Bôas, S. G., Pedersen, S., and Olsson, L. (2005). Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations. *Biotechnol. Bioeng.* 90, 703–714. doi:10.1002/bit.20457
- Diệp, N. Q., Hoan, P. T., Báo, H. T., Hùng, T. Đ., and Thắng, P. Q. (2011). Computational reconstruction of metabolic networks from high-throughput profiling data. *J. Comput. Sci. Cybern* 27, 23–35. doi:10.15625/1813-9663/27/1/460
- Dunn, W. B., Bailey, N. J., and Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst* 130, 606–625. doi:10.1039/b418288j
- Feist, A. M., and Palsson, B. O. (2010). The biomass objective function. *Curr. Opin. Microbiol.* 13, 344–349. doi:10.1016/j.mib.2010.03.003
- Gormley, P., Li, K., Wolkenhauer, O., Irwin, G. W., and Du, D. (2013). Reverse engineering of biochemical reaction networks using co-evolution with eng-genes. *Cogn. Comput.* 5, 106–118. doi:10.1007/s12559-012-9159-y
- Hamilton, J. J., Dwivedi, V., and Reed, J. L. (2013). Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophys. J.* 105, 512–522. doi:10.1016/j.bpj.2013.06.011
- Hendrickx, D. M., Hendriks, M. M. W. B., Eilers, P. H. C., Smilde, A. K., and Hoefsloot, H. C. J. (2011). Reverse engineering of metabolic networks, a critical assessment. *Mol. Biosyst.* 7, 511–520. doi:10.1039/c0mb00083c
- Henry, C. S., Jankowski, M. D., Broadbelt, L. J., and Hatzimanikatis, V. (2006). Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophys. J.* 90, 1453–1461. doi:10.1529/biophysj.105.071720
- Hoppe, A. (2012). What mRNA abundances can tell us about metabolism. *Metabolites* 2, 614–631. doi:10.3390/metabo2030614
- Hoppe, A., Hoffmann, S., and Holzhütter, 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. doi:10.1186/1752-0509-1-23
- Jensen, P. A., and Papin, J. A. (2011). Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* 27, 541–547. doi:10.1093/bioinformatics/btq702
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi:10.1093/nar/gkt1076
- Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* 7, 296–307. doi:10.1016/j.mib.2004.04.012
- Khazaei, T., McGuigan, A., and Mahadevan, R. (2012). Ensemble modeling of cancer metabolism. *Front. Physiol.* 3:135. doi:10.3389/fphys.2012.00135
- Khodayari, A., Zomorodi, A. R., Liao, J. C., and Maranas, C. D. (2014). A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab. Eng.* 25, 50–62. doi:10.1016/j.ymben.2014.05.014
- Kim, T. Y., Sohn, S. B., Kim, Y. B., Kim, W. J., and Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr. Opin. Biotechnol.* 23, 617–623. doi:10.1016/j.copbio.2011.10.007
- Krumsiek, J., Suhre, K., Illig, T., Adamski, J., and Theis, F. J. (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* 5:21. doi:10.1186/1752-0509-5-21
- Kügler, P., and Yang, W. (2014). Identification of alterations in the Jacobian of biochemical reaction networks from steady state covariance data at two conditions. *J. Math. Biol.* 68, 1757–1783. doi:10.1007/s00285-013-0685-3
- Kumar, A., Suthers, P. E., and Maranas, C. D. (2012). MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics* 13:6. doi:10.1186/1471-2105-13-6
- Lecca, P., Morigio, D., Fantaccini, G., Casagrande, A., and Priami, C. (2012). Inferring biochemical reaction pathways: the case of the gemcitabine pharmacokinetics. *BMC Syst. Biol.* 6:51. doi:10.1186/1752-0509-6-51
- Lecca, P., and Priami, C. (2013). Biological network inference for drug discovery. *Drug Discov. Today* 18, 256–264. doi:10.1016/j.drudis.2012.11.001
- Lee, D., Smallbone, K., Dunn, W. B., Murabito, E., Winder, C. L., Kell, D. B., et al. (2012). Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.* 6:73. doi:10.1186/1752-0509-6-73
- 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. doi:10.1038/msb.2010.47
- 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. doi:10.1038/nrmicro2737
- Link, H., Christodoulou, D., and Sauer, U. (2014). Advancing metabolic models with kinetic information. *Curr. Opin. Biotechnol.* 29, 8–14. doi:10.1016/j.copbio.2014.01.015
- Liu, P.-K., and Wang, F.-S. (2008). Inference of biochemical network models in S-system using multiobjective optimization approach. *Bioinformatics* 24, 1085–1092. doi:10.1093/bioinformatics/btn075
- Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* 10:e1003989. doi:10.1371/journal.pcbi.1003989
- Mahadevan, R., and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5, 264–276. doi:10.1016/j.ymben.2003.09.002
- Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi:10.1186/1471-2105-7-S1-S7
- Millard, P., Massou, S., Wittmann, C., Portais, J.-C., and Létisse, F. (2014). Sampling of intracellular metabolites for stationary and non-stationary (13)C metabolic flux analysis in *Escherichia coli*. *Anal. Biochem.* 465C, 38–49. doi:10.1016/j.ab.2014.07.026
- Mo, M. L., Palsson, B. O., and Herrgård, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* 3:37. doi:10.1186/1752-0509-3-37
- Mourão, M. A., Srividhya, J., McSharry, P. E., Crampin, E. J., and Schnell, S. (2011). A graphical user interface for a method to infer kinetics and network architecture (MIKANA). *PLoS ONE* 6:e27534. doi:10.1371/journal.pone.0027534
- Mueller, D., and Heinze, E. (2013). Stable isotope-assisted metabolomics to detect metabolic flux changes in mammalian cell cultures. *Curr. Opin. Biotechnol.* 24, 54–59. doi:10.1016/j.copbio.2012.10.015
- Müller, A., and Bockmayr, A. (2013). Fast thermodynamically constrained flux variability analysis. *Bioinformatics* 29, 903–909. doi:10.1093/bioinformatics/btt059
- Nägele, T., Mair, A., Sun, X., Fragner, L., Teige, M., and Weckwerth, W. (2014). Solving the differential biochemical Jacobian from metabolomics covariance data. *PLoS ONE* 9:e92299. doi:10.1371/journal.pone.0092299

- Navid, A., and Almaas, E. (2012). Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst. Biol.* 6:150. doi:10.1186/1752-0509-6-150
- Nemenman, I., Escala, G. S., Hlavacek, W. S., Unkefer, P. J., Unkefer, C. J., and Wall, M. E. (2007). Reconstruction of metabolic networks from high-throughput metabolite profiling data: in silico analysis of red blood cell metabolism. *Ann. N. Y. Acad. Sci.* 1115, 102–115. doi:10.1196/annals.1407.013
- Nikereel, E., Berkhout, J., Hu, F., Teusink, B., Reinders, M. J. T., and de Ridder, D. (2012). Understanding regulation of metabolism through feasibility analysis. *PLoS ONE* 7:e39396. doi:10.1371/journal.pone.0039396
- Oberhardt, M. A., Palsson, B. Ø., and Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5, 320. doi:10.1038/msb.2009.77
- Öksüz, M., Sadikoglu, H., and Çakır, T. (2013). Sparsity as cellular objective to infer directed metabolic networks from steady-state metabolome data: a theoretical analysis. *PLoS ONE* 8:e84505. doi:10.1371/journal.pone.0084505
- Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi:10.1038/nbt.1614
- Petranovic, D., and Nielsen, J. (2008). Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol.* 26, 584–590. doi:10.1016/j.tibtech.2008.07.008
- Postmus, J., Canelas, A. B., Bouwman, J., Bakker, B. M., van Gulik, W., de Mattos, M. J. T., et al. (2008). Quantitative analysis of the high temperature-induced glycolytic flux increase in *Saccharomyces cerevisiae* reveals dominant metabolic regulation. *J. Biol. Chem.* 283, 23524–23532. doi:10.1074/jbc.M802908200
- Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., et al. (2011). The human serum metabolome. *PLoS ONE* 6:e16957. doi:10.1371/journal.pone.0016957
- Quek, L.-E., Wittmann, C., Nielsen, L. K., and Krömer, J. O. (2009). OpenFLUX: efficient modelling software for ¹³C-based metabolic flux analysis. *Microb. Cell Fact.* 8, 25. doi:10.1186/1475-2859-8-25
- Samoilov, M., Arkin, A., and Ross, J. (2001). On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos* 11, 108–114. doi:10.1063/1.1336499
- Sauer, U. (2006). Metabolic networks in motion: ¹³C-based flux analysis. *Mol. Syst. Biol.* 2, 62. doi:10.1038/msb4100109
- Savageau, M. A., and Voit, E. O. (1987). Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Math. Biosci.* 87, 83–115. doi:10.1016/0025-5564(87)90035-6
- Schaub, J., Mauch, K., and Reuss, M. (2008). Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ¹³C labeling data. *Biotechnol. Bioeng.* 99, 1170–1185. doi:10.1002/bit.21675
- Schmidt, H., Cho, K.-H., and Jacobsen, E. W. (2005). Identification of small scale biochemical networks based on general type system perturbations. *FEBS J.* 272, 2141–2151. doi:10.1111/j.1742-4658.2005.04605.x
- Schmidt, M. D., Vallabhajosyula, R. R., Jenkins, J. W., Hood, J. E., Soni, A. S., Wikswa, J. P., et al. (2011). Automated refinement and inference of analytical models for metabolic networks. *Phys. Biol.* 8, 055011. doi:10.1088/1478-3975/8/5/055011
- Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* 3, 119. doi:10.1038/msb4100162
- Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., and Rupp, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* 26, 1003–1010. doi:10.1038/nbt.1487
- Smallbone, K., Simeonidis, E., Swainston, N., and Mendes, P. (2010). Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst. Biol.* 4:6. doi:10.1186/1752-0509-4-6
- Soh, K. C., and Hatzimanikatis, V. (2010). Network thermodynamics in the post-genomic era. *Curr. Opin. Microbiol.* 13, 350–357. doi:10.1016/j.mib.2010.03.001
- Sorribas, A., Hernández-Bermejo, B., Vilaprinyo, E., and Alves, R. (2007). Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnol. Bioeng.* 97, 1259–1277. doi:10.1002/bit.21316
- Srividhya, J., Crampin, E. J., McSharry, P. E., and Schnell, S. (2007). Reconstructing biochemical pathways from time course data. *Proteomics* 7, 828–838. doi:10.1002/pmic.200600428
- Stanford, N. J., Lubitz, T., Smallbone, K., Klipp, E., Mendes, P., and Liebermeister, W. (2013). Systematic construction of kinetic models from genome-scale metabolic networks. *PLoS ONE* 8:e79195. doi:10.1371/journal.pone.0079195
- Steuer, R., Gross, T., Selbig, J., and Blasius, B. (2006). Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11868–11873. doi:10.1073/pnas.0600013103
- Steuer, R., Kurths, J., Fiehn, O., and Weckwerth, W. (2003). Observing and interpreting correlations in metabolomic networks. *Bioinformatics* 19, 1019–1026. doi:10.1093/bioinformatics/btg120
- Sun, X., and Weckwerth, W. (2012). COVAIN: a toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* 8, 81–93. doi:10.1007/s11306-012-0399-3
- Tarlak, F., Sadikoglu, H., and Çakır, T. (2014). The role of flexibility and optimality in the prediction of intracellular fluxes of microbial central carbon metabolism. *Mol. Biosyst.* 10, 2459–2465. doi:10.1039/c4mb00117f
- Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., van der Weijden, C. C., Schepers, M., et al. (2000). Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* 267, 5313–5329. doi:10.1046/j.1432-1327.2000.01527.x
- Thiele, I., and Palsson, B. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi:10.1038/nprot.2009.203
- Torralba, A. S., Yu, K., Shen, P., Oefner, P. J., and Ross, J. (2003). Experimental test of a method for determining causal connectivities of species in reactions. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1494–1498. doi:10.1073/pnas.262790699
- Toya, Y., Ishii, N., Hirasawa, T., Naba, M., Hirai, K., Sugawara, K., et al. (2007). Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. *J. Chromatogr.* 1159, 134–141. doi:10.1016/j.chroma.2007.04.011
- Tran, L. M., Rizk, M. L., and Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophys. J.* 95, 5606–5617. doi:10.1529/biophysj.108.135442
- Van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. Amsterdam: Elsevier Science.
- Van Winden, W. A., van Dam, J. C., Ras, C., Kleijn, R. J., Vinke, J. L., van Gulik, W. M., et al. (2005). Metabolic-flux analysis of *Saccharomyces cerevisiae* CEN.PK113-7D based on mass isotopomer measurements of (¹³C)-labeled primary metabolites. *FEMS Yeast Res.* 5, 559–568. doi:10.1016/j.femsyr.2004.10.007
- Vance, W., Arkin, A., and Ross, J. (2002). Determination of causal connectivities of species in reaction networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5816–5821. doi:10.1073/pnas.022049699
- Varma, A., and Palsson, B. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110: basic concepts, scientific and practical use. *Appl. Environ. Microbiol.* 60, 3724–3731.
- Villaverde, A. F., Ross, J., Morán, F., and Banga, J. R. (2014). MIDER: network inference with mutual information distance and entropy reduction. *PLoS ONE* 9:e96732. doi:10.1371/journal.pone.0096732
- Visser, D., Schmid, J. W., Mauch, K., Reuss, M., and Heijnen, J. J. (2004). Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metab. Eng.* 6, 378–390. doi:10.1016/j.ymben.2004.07.001
- Weckwerth, W., Loureiro, M. E., Wenzel, K., and Fiehn, O. (2004). Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7809–7814. doi:10.1073/pnas.0303415101
- Weitzel, M., Nöh, K., Dalman, T., Niedenfür, S., Stute, B., and Wiechert, W. (2013). ¹³CFLUX2 – high-performance software suite for (¹³C)-metabolic flux analysis. *Bioinformatics* 29, 143–145. doi:10.1093/bioinformatics/bts646
- Wiechert, W., Möllney, M., Petersen, S., and de Graaf, A. A. (2001). A universal framework for ¹³C metabolic flux analysis. *Metab. Eng.* 3, 265–283. doi:10.1006/mbe.2001.0187
- Wiechert, W., and Nöh, K. (2013). Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Curr. Opin. Biotechnol.* 24, 979–986. doi:10.1016/j.copbio.2013.03.024
- Yizhak, K., Benyamini, T., Liebermeister, W., Rupp, E., and Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* 26, i255–i260. doi:10.1093/bioinformatics/btq183
- Young, J. D., Walther, J. L., Antoniewicz, M. R., Yoo, H., and Stephanopoulos, G. (2008). An elementary metabolite unit (EMU) based method of isotopically non-stationary flux analysis. *Biotechnol. Bioeng.* 99, 686–699. doi:10.1002/bit.21632

- Zamboni, N., Fendt, S.-M., Rühl, M., and Sauer, U. (2009). ^{13}C -based metabolic flux analysis. *Nat. Protoc.* 4, 878–892. doi:10.1038/nprot.2009.58
- Zamboni, N., Fischer, E., and Sauer, U. (2005). FiatFlux – a software for metabolic flux analysis from ^{13}C -glucose experiments. *BMC Bioinformatics* 6:209. doi:10.1186/1471-2105-6-209
- Zelezniak, A., Sheridan, S., and Patil, K. R. (2014). Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. *PLoS Comput. Biol.* 10:e1003572. doi:10.1371/journal.pcbi.1003572

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 September 2014; accepted: 14 November 2014; published online: 03 December 2014.

Citation: Çakır T and Khatibipour MJ (2014) Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation. *Front. Bioeng. Biotechnol.* 2:62. doi: 10.3389/fbioe.2014.00062

This article was submitted to *Systems Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2014 Çakır and Khatibipour. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.