



CHALMERS
UNIVERSITY OF TECHNOLOGY

The future of genome-scale modeling of yeast through integration of a transcriptional regulatory network

Downloaded from: <https://research.chalmers.se>, 2024-08-17 17:48 UTC

Citation for the original published paper (version of record):

Liu, G., Marras, A., Nielsen, J. (2014). The future of genome-scale modeling of yeast through integration of a transcriptional regulatory network. *Quantitative Biology*, 2(1): 30-46. <http://dx.doi.org/10.1007/s40484-014-0027-5>

N.B. When citing this work, cite the original published paper.

REVIEW

The future of genome-scale modeling of yeast through integration of a transcriptional regulatory network

Guodong Liu^{1,†}, Antonio Marras^{1,†} and Jens Nielsen^{1,2,*}

¹ Novo Nordisk Foundation Center for Biosustainability, Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg SE41296, Sweden

² Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm DK2970, Denmark

* Correspondence: nielsenj@chalmers.se

Received December 5, 2013; Revised January 29, 2014; Accepted February 5, 2014

Metabolism is regulated at multiple levels in response to the changes of internal or external conditions. Transcriptional regulation plays an important role in regulating many metabolic reactions by altering the concentrations of metabolic enzymes. Thus, integration of the transcriptional regulatory information is necessary to improve the accuracy and predictive ability of metabolic models. Here we review the strategies for the reconstruction of a transcriptional regulatory network (TRN) for yeast and the integration of such a reconstruction into a flux balance analysis-based metabolic model. While many large-scale TRN reconstructions have been reported for yeast, these reconstructions still need to be improved regarding the functionality and dynamic property of the regulatory interactions. In addition, mathematical modeling approaches need to be further developed to efficiently integrate transcriptional regulatory interactions to genome-scale metabolic models in a quantitative manner.

Keywords: transcriptional regulatory network; metabolic model; *Saccharomyces cerevisiae*; integration

INTRODUCTION

The application of high-throughput biological techniques has been increasing in the last decade. Annotated sequences, biochemical and physiological data are now available for many organisms, and the most exciting challenge for systems biology is to describe, integrate and analyze this huge amount of heterogeneous data in order to transform them into usable knowledge.

Metabolism is a highly coordinated system which provides energy and precursors of macromolecules for cellular functions. The metabolism of most cells is tightly regulated at multiple levels in response to external or internal stimuli. Constraint-based mathematical metabolic models have been extensively used as a tool to simulate the behavior of metabolic networks, guide metabolic engineering, and interpret biological data [1]. However, constraint-based metabolic modeling does not take

account of regulation. Consequently, these models cannot accurately carry out prediction when the regulatory mechanisms play a decisive role in defining the behavior of the system. So, in order to improve their prediction ability, it is useful to integrate regulatory information into constraint-based metabolic models. Integration of regulatory data into genome-scale metabolic models is a challenge for systems biology, but it could lead towards a better understanding of mechanisms governing the cell, and allows the analysis of the relationship between the environment and the metabolic capabilities of an organism.

Transcriptional regulation is one of the mechanisms controlling metabolic activities. The regulatory interactions between transcription factors (TFs) and their target genes form a transcriptional regulatory network (TRN) [2], which responds to the amount and kind of exogenous or endogenous metabolites (input) and adjusts the mRNA

[†] These authors contributed equally to this work.

synthesis of genes for specific metabolic pathways (output). Genome-scale TRNs have been reconstructed in many model organisms, such as *Escherichia coli* and *Saccharomyces cerevisiae*, due to the extensive knowledge and experimental data available for these organisms.

Here we will review studies on the reconstruction of the TRN of yeast, and how it can be integrated into constraint-based metabolic models. As there have been many thorough reviews on the methods for the reconstruction and analysis of TRNs [2–8], the progress that has been made in *S. cerevisiae* will be emphasized. For the integration of TRN reconstructions into metabolic models, many of the methods were first developed in *E. coli* and then extended to eukaryotes in a second stage. These methods have a general validity, and theoretically can be applied to any organisms. However, the results obtained in *S. cerevisiae* by the current integration techniques are often unsatisfactory, possibly due to the high complexity of gene regulation [9], the current limited knowledge available on such mechanisms, and the presence of many cellular compartments in eukaryotic cells. Thus, we are going to represent the current state of the art of these integration methods and emphasize the results obtained in *S. cerevisiae* when possible.

TRANSCRIPTIONAL REGULATION OF METABOLISM IN *S. CEREVISIAE*

The metabolic state is determined by multiple factors including substrate concentrations and the activities of metabolic enzymes [10]. Protein abundance, covalent modification and allosteric regulation all affect the activities of metabolic enzymes. Transcriptional regulation plays a role in controlling the abundances of some metabolic enzymes through regulating the levels of their corresponding mRNAs. The transcriptional regulation of metabolism in *S. cerevisiae* has been widely studied for the physiological effects and molecular mechanisms. From present results we can draw several lessons:

(1) Transcriptional regulation widely occurs in metabolic pathways. Enzymes in carbon (Figure 1), nitrogen, sulfur, phosphate and oxygen metabolisms are extensively regulated at the transcriptional level in *S. cerevisiae* [13,14], with about 100 TFs having been confirmed or suggested to be involved. In fact, about 40 TFs in *S. cerevisiae* are named after their functions in metabolism regulation, e.g., Adr1p (Alcohol Dehydrogenase Regulator), Rgt1p (Restores Glucose Transport), and Ino2p (INositol requiring).

(2) The contribution of transcriptional regulation in the regulation of metabolism is reaction- and condition-dependent. The abundances of metabolic enzymes are regulated at several other layers (e.g., mRNA degradation, translation, protein degradation) in addition to transcrip-

tion. Moreover, many metabolic enzymes are regulated at the activity level. Thus, it is quite natural that changes in the transcription levels of enzyme genes are often not consistent with those in corresponding metabolic fluxes in *S. cerevisiae* [15–17]. The correlation between gene transcription and flux is affected by the degree of other kinds of regulations (e.g., metabolite-enzyme interactions [18]), and is dynamic under different conditions even for the same reaction [19,20]. Some reactions for alternative carbon source (e.g., galactose and maltose) utilization [21], glyoxylate cycle, gluconeogenesis [16], ergosterol biosynthesis [19], amino acid metabolism, and nucleotide metabolism [14] are suggested to be transcriptionally-controlled. For example, a mutant with disruption of *PUT3*, which encodes a transcriptional activator of proline utilization genes, completely loses the ability to use proline as a sole nitrogen source [22]. On the other hand, transcriptional regulation plays only minor roles in controlling many other metabolic reactions, e.g., glycolysis [15,16].

(3) Diverse input (upstream signal to TFs) and output (TFs to targets) mechanisms are used by the TFs for regulation of metabolism. Similar with metabolic enzymes, the TFs themselves are regulated in abundance and/or activity levels by upstream signals. In addition, some TFs are regulated at the cellular location level (cytoplasm or nucleus). These regulations are achieved through diverse molecular mechanisms: (a) transcriptional control by other TFs, e.g., repression of gluconeogenesis activator gene *CAT8* by Mig1p [23]; (b) specific mRNA degradation, e.g., deadenylation of the mRNA of glucose repressor gene *NRG1* [24]; (c) translational control, e.g., translation initiation regulation of amino acid biosynthesis activator Gcn4p through upstream ORFs [25]; (d) regulated protein degradation, e.g., ubiquitin-mediated degradation of peptide transport repressor Cup9p [26]; (e) proteolytic processing, e.g., activation of amino acid utilization activator Stp1p and Stp2p by protease Ssy5p [27]; (f) covalent protein modification, e.g., phosphorylation of nitrogen source utilization activator Gln3p, which controls its cellular location [28]; (g) allosteric regulation by metabolites, e.g., activation of Put3p by proline [29]; (h) regulation by protein-protein interaction, e.g., inhibition of galactose utilization activator Gal4p by Gal80p [30]. Actually, many TFs are under the control of multiple regulatory mechanisms. For example, the above-mentioned Gal4p is also regulated at the transcriptional level [30], and Gcn4p is also regulated at the protein degradation level [31]. To exert their biological functions, some TFs can differently act on different targets. For example, when the medium is shifted from glucose to ethanol, Rds2p activates the transcription of gluconeogenic genes and represses that of genes suppressing gluconeogenesis [32]. Similarly, in the

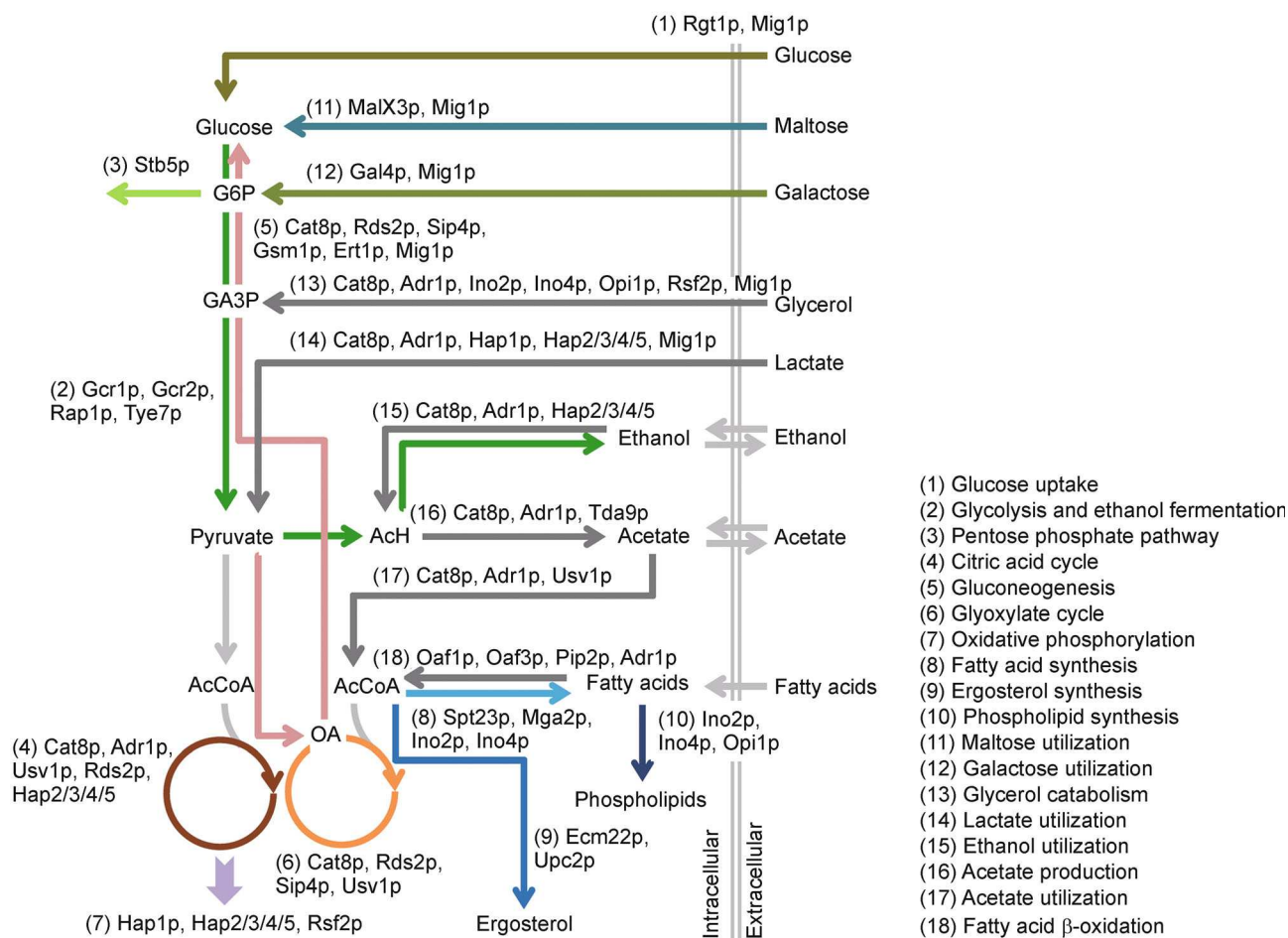


Figure 1. Transcription factors involved in the regulation of carbon metabolism in *S. cerevisiae*. Data were collected from YEASTRACT [11] and SGD [12] databases. Different metabolic pathways are shown in different colors, except those for non-fermentable carbon source utilization (13 to 18) sharing dark grey. Abbreviations: G6P, glucose 6-phosphate; GA3P, glyceraldehyde-3-phosphate; AcH, acetaldehyde; AcCoA, acetyl-CoA; OA, oxaloacetate

presence of serine, Cha4p activates the transcription of a serine catabolic gene and represses that of a serine biosynthetic gene, with the latter achieved by activating the expression of a neighboring intergenic non-coding transcript [33].

(4) Different types of associations exist between TFs with shared target genes. For example, phospholipid biosynthesis activators Ino2p and Ino4p form a complex on target gene promoters, and both TFs are necessary for inositol synthesis [34]. However, in the case of sulfur metabolism activators Met31p and Met32p, which are paralogs binding to the same site on the promoter, only disruption of both TFs led to methionine auxotrophy, suggesting redundant functions of them in methionine biosynthesis [35]. Four TFs involved in nitrogen source utilization, including repressors Dal80p and Gzf3p and activators Gat1p and Gln3p, influence each other by both transcriptional regulation and physical interaction [36].

This kind of associations between TFs needs to be carefully considered when converted to mathematical representations.

RECONSTRUCTION OF THE TRANSCRIPTIONAL REGULATORY NETWORK

Strategies for establishing TF-target interactions

Regulatory interactions linking TFs (regulatory proteins) and their targets (regulated genes) are the basic units of a TRN [37]. As many transcriptional regulatory mechanisms are less conserved among organisms [37,38], it is generally difficult to reconstruct TRNs starting from protein homology analysis, as can be done for metabolic network reconstructions [39]. Experimental studies, biochemical or genetic, bottom-up or top-down, are

usually needed to establish the transcriptional regulatory interactions in a specific organism. However, genomic annotation is also important because it provides useful information for experiment design (e.g., TF prediction) and data integration (e.g., sequence conservations) (Figure 2).

Biochemical approaches can be used to identify direct interactions between TFs and their target genes. Here direct interactions mean physical occupancies of TFs on the promoters of genes. The methods in this field (for reviews see Refs.[40,41]), especially electrophoretic mobility shift assay (EMSA) [42] and DNAase footprinting assay [43], have been frequently used to study *in vitro* interactions between specific TF-promoter pairs in *S. cerevisiae*. Chromatin immunoprecipitation combined with cDNA microarray analysis (ChIP-chip), which was first developed in yeast [44], is a powerful *in vivo* method

to find genome-wide direct targets of TFs. The first high-throughput ChIP-chip experiment in yeast identified the targets of 106 TFs in rich medium [9], and a later study provided interaction information for 203 TFs under more environmental conditions [45]. The data from these two studies are used in most TRN reconstruction studies in yeast (see section Reconstruction and Characterization of the TRN). The third high-throughput ChIP-chip experiment in yeast mainly focused on proteins involved in the transcriptional machinery [46]. Alternative ChIP-based techniques have been developed recently, including ChIP followed by high-throughput sequencing (ChIP-seq) for higher resolution, larger coverage, lower noise and more accurate quantitation [47], ChIP-exo achieving single-nucleotide resolution [48], and competition ChIP for determining the residence time of TFs on promoters [49]. As an alternative approach, affinity chromatography with

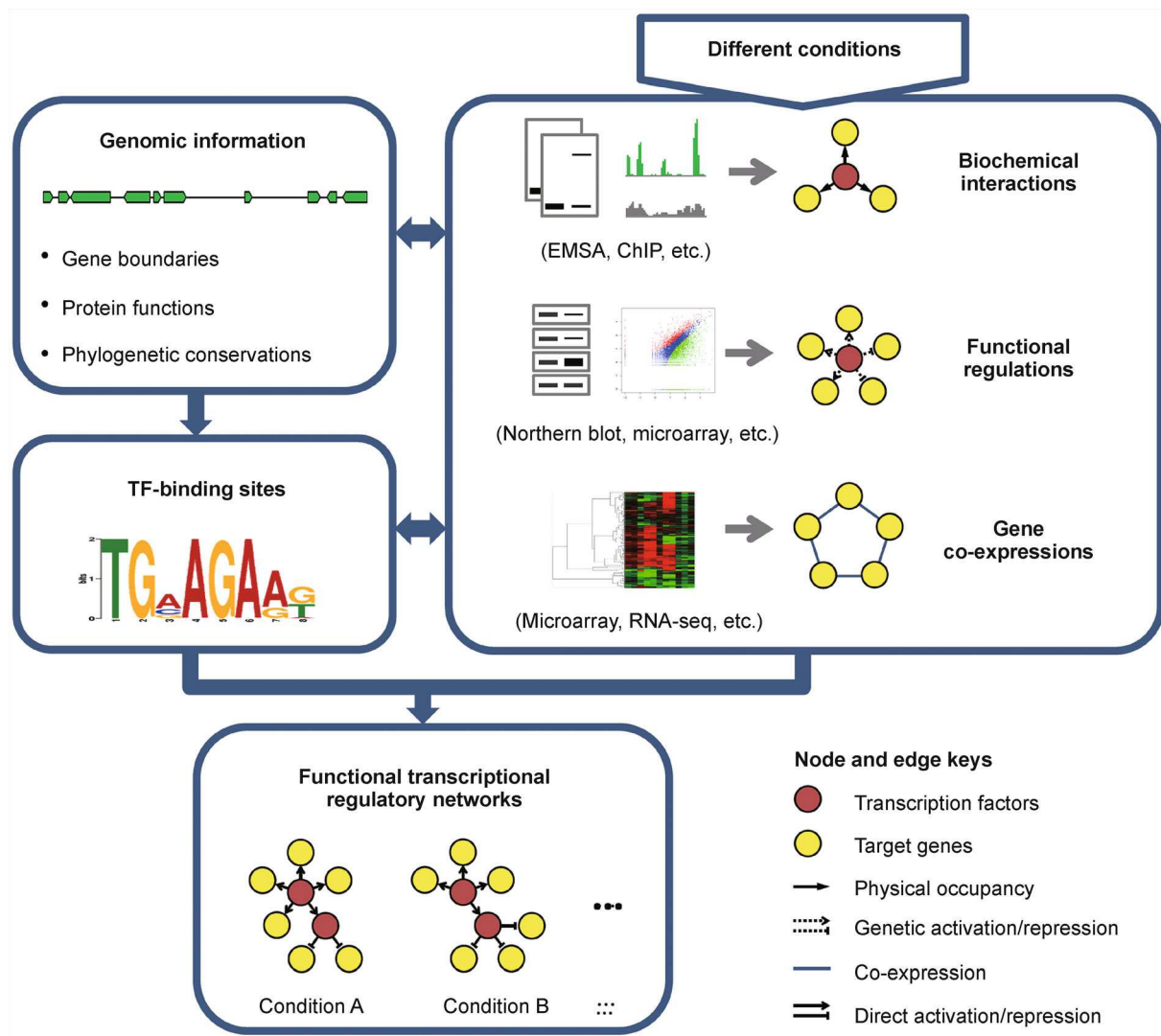


Figure 2. Strategies for the reconstruction of a functional TRN in *S. cerevisiae*

immobilization of DNA followed by proteomic profiling can be used to identify the TFs occupying on a specific promoter *in vivo* [50].

The results obtained using biochemical approaches have at least two drawbacks. First, TF occupancy does not necessarily represent functional regulation, e.g., Leu3p constitutively occupies its target promoters whether it functions or not [45]. In some cases, the functional activation of TFs (e.g., presence of cooperative factors) is needed in addition to their binding on promoters [51]. Second, the interaction information between TFs and target genes obtained by biochemical methods does not contain the directivity (repression or activation) of transcriptional regulations.

Genetic approaches cannot distinguish direct and indirect TF-target interactions, but they do provide information on the existence (*if*) and directivity (*how*) of regulatory functions of TFs. This function information is very important, because only functional regulations can be integrated to metabolic models as constraints. In genetic approaches, usually transcriptional regulations are perturbed by deletion, suppression, over-expression or mutation of TF genes first, and then target genes with affected transcription levels are discovered through comparative analysis of parent and mutant strains. The transcription levels of specific genes can be determined using Northern blotting, quantitative PCR or reporter gene assay, while genome-wide transcriptional changes can be measured using transcriptomic technologies such as microarrays or RNA-seq [52].

In *S. cerevisiae*, a genome-wide single-gene-deletion mutant library has been constructed [53], and genes with significant differential expressions in rich medium responding to individual deletions of 269 TFs were identified [54]. A reanalysis of these data estimated that about 98% of the differential expressions were mediated by indirect TF-target interactions [55]. These indirect regulations may be mediated by regulatory cascades (intermediate TFs) or physiological changes (e.g., deletion of one TF alters the concentration of some metabolite, which affects gene expression through other regulatory pathways). Transcriptomic responses to individual over-expressions of 55 growth-involved TFs were also reported in *S. cerevisiae* [56]. Compared with gene deletion, the gene over-expression strategy has the advantage in identifying condition-specific targets of some TFs even under standard growth conditions [57]. It should be noted that the function of some TFs may be masked by paralog compensation when genetic perturbation of one single TF was performed [58].

Transcriptome analysis has also been extensively used to study the responses of laboratory wild-type *S. cerevisiae* strains to environmental perturbations. Some of the co-expressed gene groups (modules) generated by

clustering analysis of these transcriptome data are assumed to be “co-regulated” genes under the same TF (s) [59]. TF-target gene interactions can be predicted from these co-expression modules, for some TFs themselves are regulated at the transcriptional level [60]. This method is particularly meaningful in finding condition-specific TF-target interactions, as there have been thousands of transcriptome datasets collected under various conditions (environments, developmental stages, time-series, etc.) in *S. cerevisiae* [61]. However, the correlations between the expression level of TF itself and those of their targets is weak for many TFs [62], which seriously limits the inference of the TRN solely from gene co-expression data.

TFs regulate the expression of target genes through occupying specific *cis*-acting transcription factor-binding sites (TFBSs). Thus, the “TF-target gene” relationships are actually the combination of “TF-TFBS” and “TFBS-target gene” relationships. Knowing the TFBS information is important for understanding the molecular mechanisms of TRNs. In addition, the presence, position, orientation and combination of TFBSs on promoters could be used to predict gene expression patterns in *S. cerevisiae* [63] and to infer the TRN (see section Reconstruction and Characterization of the TRN). The “TF-TFBS” relationships can be determined by *in vitro* biochemical methods such as EMSA and large-scale microarray-based assays. Protein-binding oligonucleotide microarrays (PBMs) were used to determine the TFBSs of 112 [64] and 89 [65] TFs in *S. cerevisiae* by different research groups. A microfluidics-based assay was shown to have higher sensitivity than PBM, and can measure the binding affinities of TFs [66]. On the other hand, an oligonucleotide-binding protein microarray comprising 282 *S. cerevisiae* TFs also succeeded in finding new interactions between TFs and oligonucleotides [67].

“TF-TFBS” relationships can also be computationally predicted based on the over-representation of TFBSs in the target gene sequences of TFs. Target gene sets identified by both ChIP experiments [48] and genetic perturbations [56] can be used for computational discovery of TFBSs, and obviously the former (direct target gene sets) are more likely to give significant results. These computational tools were also used to predict “TFBS-target gene” relationships from other kinds of gene sets, such as co-expressed [68] or functionally-related [69] genes in *S. cerevisiae*. In addition, as functional TFBSs are usually highly conserved, some of them can be identified by phylogenetic footprinting (i.e., looking for conserved regions after aligning promoter sequences of orthologous genes from closely related species) [70,71]. Integration of TFBSs determined by different approaches can give high-confidence data [45,72], and combination of fragmental “TF-TFBS” and

“TFBS-target gene” relationships successfully gave complete “TF-target gene” interactions, e.g., the discovery of two TFs controlling ribosome biogenesis in *S. cerevisiae* [65].

Many of the transcriptional regulatory interactions obtained using the above approaches in *S. cerevisiae* have been manually collected from original literature. The examples include a review summarizing the early results of individual TF-target interactions [65], databases YeTFaSCo [73] and ScerTF [74] extensively collecting and curating TFBSs, and databases YEASTRACT [11] and SCPD [75] containing both TFBSs and TF-target interactions. In the *Saccharomyces* Genome Database [12], regulatory relationships between TFs and target genes are being added for nearly all genes recently [76], which provide machine-readable data for the *S. cerevisiae* TRN research.

Reconstruction and characterizations of the TRN

Interactions between TFs, TFBSs and target genes obtained from different approaches are complementary due to their own advantages and drawbacks. For example, comparison of target genes from biochemical and genetic experiments found non-occupying regulation and non-regulatory occupancy of TFs on target genes [54]. Another example is that comparison of TFBSs obtained using *in vivo* and *in vitro* methods found combinational or “piggy-back” binding of TFs on promoters [72]. Thus, data of different resources need to be integrated to obtain high-quality TRN reconstructions [2].

Many algorithms for integrated TRN reconstruction and validation have been developed, such as those integrating ChIP data and gene expression data [9,77–82], TFBS information and gene expression data [83–87], or all the three types of data [88–90]. The algorithms are generally based on the assumption that the direct target genes of a TF should meet more than one of the following criteria: (1) being detected for the known TFBS of the TF on the promoter; (2) being physically occupied by the TF; (3) transcriptionally responding to the perturbation of the TF; (4) transcriptionally responding to the active condition(s) of the TF; (5) having related functional annotations. A good example in reconstructing integrative TRNs is a study on DNA damage response in *S. cerevisiae*, in which 30 TFs were first selected based on existing knowledge, and then gene expression profiling of wild-type and TF-deleted mutants and ChIP-chip were performed (both under normal and DNA-damaging conditions), providing multi-view data for integrative analysis [91]. For less-studied TFs whose “active” conditions are unknown, integration of *in vitro* TFBS information and co-expression data could be used to infer the possible conditions for further studies [84]. TF-target

interactions established using low- and high-throughput methods are also complementary regarding their coverage and signal-to-noise ratios. An integrative strategy for this is starting with well-studied small TRN reconstructions, followed by expanding it using interactions from high-throughput analyses [92].

The TRN reconstructions are usually represented as network graphs, in which nodes indicate TFs and target genes, and arrows indicate transcriptional regulations. In addition, TRNs were modeled using different mathematical formalisms such as ordinary differential equations, stochastic equations, Petri nets, or Bayesian networks (for reviews see Refs.[93–95]). Nevertheless, the lack of high-throughput experimental methods to measure biochemical kinetic parameters *in vivo*, necessary for the description of the system, limits the application of these approaches to only the description of small networks [96].

Currently, genome-scale TRN reconstructions were converted into mathematic models only based on the use of Boolean logic. This formalism only requires information about the structure of the network and the nature of interactions (activation/repression) between the network elements [97,98]. The state of elements in Boolean network is described by an on/off representation, so the state of the whole network is defined by an n -tuple of 1s and 0s according to the presence/absence of the particular elements. However, using the Boolean rules the level of detail of a system decreases and it is not possible to carry out quantitative prediction of the behavior of the system. Matrix based formalism has been used to study the properties of TRNs [99]. This formalism is based on a reformulation of the Boolean rules into a matrix representation. Through the reformulation, the TRN is described as a regulatory network matrix (R matrix) used to investigate its functional (expression) states in response to environmental conditions.

In *S. cerevisiae*, several large-scale TRN reconstructions have been reported during the past ten years. Figure 3 lists eleven of these reconstructions, which vary in terms of the number of TFs, target genes and TF-target interactions due to different data resources and different criteria used [9,11,45,46,100–106]. Most of these studies focused on the structural properties of TRNs, and the major conclusions from these studies are

- (1) The structure of the *S. cerevisiae* TRN is scale-free [107], with most target genes having a few TFs and most TFs regulating a few targets [6,105]. The favor of links between highly-connected nodes (hub TFs or genes) and low-connected nodes, and limiting links between hubs, divide the TRN into structural modules [104].
- (2) From the viewpoint of hierarchical organization, TFs in top, middle and bottom layers have significantly different sequence conservations,

transcript and protein abundances, numbers of direct targets, and effects on global gene expressions [102,108]. Interestingly, TFs in the middle layer were found to have the most direct targets.

- (3) The structure of the *S. cerevisiae* TRN is highly dynamic in responding to condition changes [45]. Local regulatory units of different topological structures are favored by different conditions, and most hub TFs are only active under certain conditions [101].
- (4) The TRN of *S. cerevisiae* is more complex compared with that of *E. coli*. A comparative analysis found that although the total numbers of TFs are similar, *S. cerevisiae* has ~5-fold more TF-target regulatory interactions, and ~12-fold more co-regulation links between TFs (i.e., two TFs share a common target), than those in *E. coli* [107].

INTEGRATION OF TRANSCRIPTIONAL REGULATORY NETWORK RECONSTRUCTIONS INTO METABOLIC MODELS

Constraint-based metabolic models

The concept of genome-scale metabolic reconstruction has been extensively researched and reviewed in the literature [5,109]. By using genomic and biochemical data available, it is possible to reconstruct genome-scale models of metabolism and analyze them to infer

information about the behavior of the cell. Such models can be used for simulations and integrative data analysis, and several toolboxes have been developed for handling these models, e.g., COBRA [110] and RAVEN [39].

A protocol to generate a genome-scale model includes different stages [111]. The first stage is the creation of a draft reconstruction by using a genome annotation of the organism of interest, and bioinformatics resources such as databases of reactions [112,113]. The second stage is the manual refining of the draft reconstruction to improve the quality of the model. This work is mainly based on the study of primary literature and includes the formulation of the biomass equation. The next stages are formed by the conversion of the reconstructed network into a mathematical model used to carry out simulations and by the evaluation of the quality of the reconstruction.

Many genome-scale reconstructions of the *S. cerevisiae* metabolic network have been developed during the past decade. A consensus network was reconstructed as a result of a common effort by researchers to produce a highly annotated network with a standard nomenclature (see Ref. [114] and visit <http://yeast.sourceforge.net/> for the consensus yeast metabolic project). Moreover, the development of new models is still in progress in order to improve our description of some parts of the yeast metabolic network such as the lipid metabolism [19,115].

Flux Balance Analysis (FBA) [116] is a mathematical method based on linear programming through which one can simulate the behavior of the reconstructed metabolic network under specific environmental conditions. The

| Data sources | | | | | No. of components | | | No. of conditions | Main concerns | References |
|--------------|----|----|----|----|-------------------|-------------------|--------------|-------------------|------------------------------|--------------------|
| PL | DB | HC | GE | PF | TFs | Targets | Interactions | | | |
| ■ | ■ | | | | 124 | 402 | 909 | Not considered | Topological structure | [105] |
| | ■ | | | | ..a) | 682 ^{b)} | 1289 | Not considered | Topological structure | [104] |
| | | ■ | | | 106 | 2343 | 4677 | 1 | Network units | [9] |
| ■ | | ■ | | | 180 | 3474 | 7419 | Not considered | Gene expression | [100] |
| ■ | | ■ | ■ | | 142 | 3420 | 7074 | 5 | Condition-specific structure | [101] |
| ■ | | ■ | | ■ | 102 | 1296 | 3553 | TF-dependent | Condition-specific structure | [45] |
| ■ | | ■ | ■ | | 157 | 4410 | 12873 | Not considered | Robustness | [106] |
| ■ | | ■ | ■ | | 55 | 320 | 599 | TF-dependent | Metabolism regulation | [103] |
| ■ | | ■ | | | 158 | 4369 | 13385 | Not considered | Hierarchical structure | [102] |
| ■ | | ■ | ■ | | 183 | 6403 | 48082 | TF-dependent | Data collection and curation | [11] ^{c)} |
| | | ■ | | | 202 | 5866 | 231704 | 2 | Transcription machineries | [46] |

Figure 3. Examples of TRN reconstructions in *S. cerevisiae*. Abbreviations: PL, primary literature; DB, databases; HC, high-throughput ChIP-based experiment data; GE, gene expression profiling data; PF, phylogenetic footprinting. a) data not reported; b) total number of TFs and target genes; c) data in October 2011

analysis is based on the fact that a reaction network is subject to a series of constraints that limit its behavior, such as environmental and physicochemical ones [117]. The constraints define a closed space which contains all the possible stationary states that the metabolic network can reach. The more constraints one adds, the smaller the solution space becomes.

In particular, FBA allows finding a particular solution (optimal solution) inside the solution space that maximizes or minimizes a particular objective function. One of the most used objective functions is the optimization of biomass formation, which is a reaction that describes the incorporation of all necessary precursors like amino acids, nucleotides and cofactors into cellular macromolecules that make up the functional cell [118]. Figure 4 summarizes the process of simulation of a metabolic network. Starting with a metabolic reconstruction and applying a series of constraints a solution space is defined. A particular solution inside the solution space is found using an optimization technique such as linear programming.

FBA is a powerful tool to study the properties of a metabolic network but it still has some limitations [119]. In particular, FBA can make incorrect predictions when the regulatory network plays a key role in the definition of the behavior of the cell [120]. For example, an FBA simulation of *E. coli* metabolic network during aerobic growth in presence of both glucose and lactose predicts an uptake of both carbon sources at the same time. This prediction is not in agreement with the experimental result that the cell prefers glucose over galactose, and misunderstandings in the final analysis of the process may arise [121].

Moreover, often the flux distribution obtained from linear programming is not unique but is part of a set of different flux distributions that generate the same value for the objective function. Every optimal flux distribution describes an alternative pathway usage that maximizes (or

minimizes) the objective function [122,123]. This fact can lead to several problems. For instance, if there are two possible metabolic pathways that lead to the same final product from the same substrate with an equivalent stoichiometry, FBA is not able to distinguish between them. As a consequence, they may both be used in the simulation. So, these miscalculations often occur given that, for example, isoenzymes are the simplest form of an alternative pathway involved in this process, and without addition of extra information FBA cannot predict which of the isoenzymes is active at a given moment of time. In a study of the central carbon metabolism in *S. cerevisiae*, the prediction of fluxes using different objective functions was shown to be able to catch the essential feature of the behavior of a system perturbed by gene knockouts, but with some shortcomings likely due to the presence of alternative pathways and regulatory mechanisms [124].

The process of integration of transcriptional regulation into metabolic models

Two kinds of constraints can be applied to a reaction network: nonadjustable and adjustable constraints [125]. The first kind is invariable because it is related to characteristics of the process that can hardly change such as network topology, thermodynamics, and maximal flux capacities. In contrast, the second kind (of constraints) depends on regulatory events so they can vary over time as well as under different environmental conditions. One can consider these constraints to be temporal constraints applied to the metabolic network.

In the first generation of constraints-based models only nonadjustable constraints were used, while the addition of adjustable constraints led to the second generation of constraints-based models [126]. Incorporation of regulation into genome-scale metabolic models is, however, still a challenge and requires more work.

A small example is useful to clarify the application of

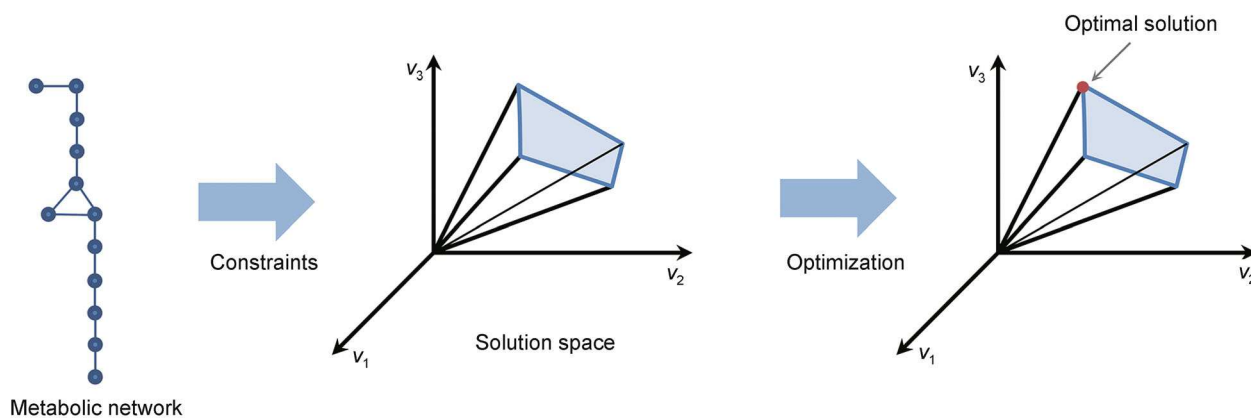


Figure 4. The constraint-based approach for metabolism modeling

adjustable constraints. In the network represented in Figure 5A, the metabolism is represented by a single reaction that carries out the transformation of the substrate (M-1) into the product (M-2) using an enzyme (Enz). The product of the reaction suppresses the activity of the transcription factor which activates the expression of the enzyme.

The behavior of this small network can be described by some simple Boolean rules (Figure 5B). According to these rules the reaction is active when the substrate M-1 and the enzyme are both present. The enzyme is present when both the transcription factor and the gene are present and the product M-2 is absent. This small network contains five elements, thus all the possible states of the network are $2^5 = 32$. The reaction can be turned on or turned off depending on which state the network finds itself in. It is quite clear that even a simple network can show a very complex behavior. The most important consequence of the imposition of regulatory constraints is that it reduces the solution space of the metabolic network (Figure 6) and this fact can also be examined using topological techniques [126].

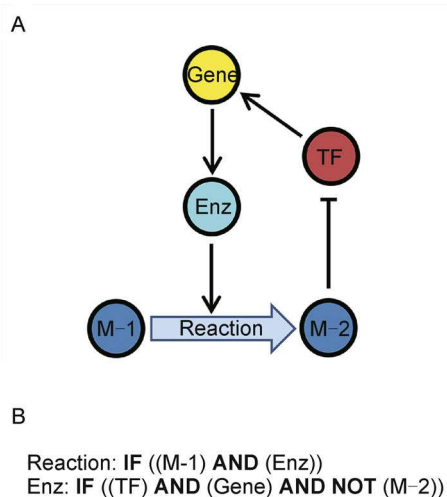


Figure 5. A simple integrated metabolic model (A) and its representation with a Boolean formalism (B)

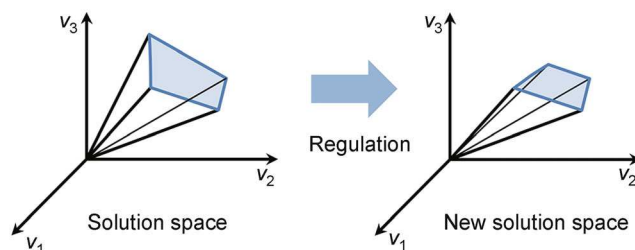


Figure 6. The reduction of the solution space of the metabolic network using regulatory constraints

The rFBA method and its current development

Historically, first attempts to integrate FBA with regulation started with the observation that time constants characterizing metabolism and regulation differ by one order of magnitude at least [120]. In particular, the time constants that describe the behavior of the metabolism are faster (order of milliseconds) than the time constants that describe the transcription regulation (order of minutes). As a consequence, a dynamic process can be simulated as a sequence of steady states if one looks at the behavior of metabolic network during a time series as a sequence of short intervals of time in which the metabolic network reaches a quasi-steady state.

So the whole time of the simulation is split in several time intervals in which a quasi-steady state for the metabolic network is assumed. For the first time interval, the knowledge of the environmental condition is available, thus the FBA method can be used in order to simulate the response of the metabolic network to the environment. Additionally one can determine the status of the regulatory network by solving the Boolean equations for the time interval based on the presence or absence of particular metabolites, regulatory proteins or fluxes in the internal environment.

Finally, the output of the simulation is used to update the environmental condition for the next time interval by using a numerical integration in order to calculate the new concentrations of the external species [127]. For each of the following time steps the state of the regulatory network is calculated and used in addition to the environmental condition calculated from the preceding time step in order to constrain an FBA simulation.

The regulatory constraint process is accomplished by checking the presence of the enzyme in the specific time interval based on the results of the regulatory rules. If the enzyme is not present, the flux of the reaction at the given interval is set equal to zero. If the enzyme is present, the flux is determined with FBA simulation in that given time interval.

As a consequence, the linear optimization for the metabolic network and the calculation of the state of the

regulatory network are made in two separate moments, and the regulation is represented as a time dependent (adjustable) set of constraints applied to an FBA simulation. This method is called regulatory FBA (rFBA) and allows for the prediction of dynamic profiles of the cell growth [120,121,128].

A common way in order to validate genome-scale models is by confronting simulated growth and exchanging fluxes with experimental data. Although this is the common accepted practice, some efforts have been done to try to validate models studying the correlations between the *in silico* calculated and *in vivo* measured internal fluxes. This is also important in order to analyze the ability of the models to predict internal fluxes [124,129]. One of the limitations of this comparison is that the number of measured fluxes is limited compared with the number of fluxes present in a genome-scale reconstruction. Furthermore, large-scale datasets usually cover mainly the central carbon metabolism and less other pathways, e.g., biosynthesis of biomass components like lipids and nucleotides [130].

Moreover, it is not yet clear what the best methodology is in order to select reactions which allow the reconstruction of the whole system state, although recent results about the observability of complex systems, such as metabolic network, suggest that in yeast the minimum number of sensor nodes should be about 10% of the number of reactions of the genome-scale reconstruction [131].

An integrated FBA such as rFBA can be used to overcome this limitation by improving the ability to make predictions and validate the model by using large gene transcription datasets available today. On one hand, the prediction of metabolic flux in a given environmental condition is possible using integrated FBA and this can suggest possible interpretations for up or down regulation of genes. On the other hand, this approach can also allow for the creation of *in silico* gene expression arrays and comparison with experimental data to validate the integrated model. For instance, Covert and coworkers in a large-scale simulation of *E. coli* [128] used their model to predict differential expression of genes. Afterwards they compared the *in silico* gene expression with *in vivo* gene expression to update their model by rewriting, relaxing or removing regulatory rules in order to improve the prediction ability of their reconstruction in an iterative fashion.

A large-scale study of an integrated yeast rFBA model was performed by Herrgård et al. [103]. They reconstructed a yeast TRN formed by 55 transcription factors related with the metabolism. Afterwards they integrated their TRN reconstruction with a genome-scale metabolic reconstruction and used this integrated yeast model to

study the interactions between the regulatory and the metabolic networks. They also predicted growth phenotypes and gene expression changes. This work clearly showed us how the identification of new interactions and analysis of different kinds of data in yeast are possible by using only one model framework.

Herrgård et al. used an iterative approach conceptual similar to the one used by Covert et al., which was based on the update of regulatory rules. In fact, they expanded the regulatory network of the model by using experimental data taken from the literature. This approach showed that discrepancies between the model predictions and the experimental data can be used in order to drive experiments to investigate and identify potential new target genes for each TF.

Later other efforts for extending rFBA to simulate the transcriptional regulation of the whole system led towards the integration of different types of model formalisms to describe the different process inside the cell [132]. This means using ordinary differential equation (ODE) when kinetic parameters are available, in order to simulate the signaling network and keep using Boolean rules to simulate the transcription regulatory system. This approach gave birth to hybrid methods used for *E. coli* central metabolism (iFBA [133]) and for a portion of the *S. cerevisiae* system (idFBA [134]). The final goal of this kind of approaches is to build a whole-cell system model as has been done for *Mycoplasma genitalium* [135].

The application of this hybrid models showed improvement over the rFBA approach. In particular, the ability of the model to simulate the dead ends [133] was enhanced. Dead ends are reactions present in the reconstructed network but not used in the optimal solution because their products are not involved in the growth objective function. However, the metabolic products of the dead ends are often very important in the cell for other reasons, for example, they can be mediators of signaling.

In particular, the method called idFBA [134] has been applied to a module of *S. cerevisiae* composed by a small portion of signaling, metabolic and transcriptional regulatory networks related with high-osmolarity glycerol response mitogen-activated protein kinase pathway. This work showed that an integrated system spanning different time scales can be studied with an integrated FBA based approach.

The MILP reformulation and the data driven approaches

In the rFBA method, the TRN and the metabolic network are analyzed separately and simulated separately. The presence of alternative solutions for both networks makes it difficult to integrate them and imposes a requirement for

selecting which of the steady state solutions have to be selected for simulating the next time interval. Nevertheless, Boolean rules can be converted into integer inequalities, and these inequalities can be used to formulate a mixed integer linear program problem (MILP). Thus, an integrated model reformulated in terms of an MILP problem can be obtained by changing the formalism used in order to describe the TRN. This MILP reformulation integrates the regulatory structure with the metabolic model in a single model.

The steady-state regulatory flux balance analysis (SR-FBA) [136] method is based on the reformulation of the integrated model into an MILP problem and it was used in order to find steady state solutions for the metabolic network and for the regulation network which are consistent between them. The solutions of the TRN were mapped on the metabolic network and the coherence of the integration was studied by checking the consistency between the states of the TRN and the metabolic network, in an effort to improve the evaluation of the optima alternative solutions by classifying each gene based on its expression and flux activity states.

The SR-FBA approach gave birth to important development such as the OptORF [137] and the GeneForce [138] algorithm. The limited and incomplete information available about the regulatory mechanisms and the inherent complexity of the system can lead to incomplete integrated models which can be inconsistent with experimental evidences. The GeneForce algorithm was developed to refine and solve inconsistency in large-scale integrated models by suggesting corrections for the model to improve its prediction ability. This algorithm tries to identify the minimum number of rules that can be violated so as to allow the model to perform the correct growth prediction, preserving as much as possible the regulatory rules already applied. Thus, by relaxing the regulation rules the solution space is remodeled to include solutions previously eliminated by the first shaping of the integration process. Moreover, the procedure of correction of these inconsistencies can suggest hypotheses which can drive future experiments and lead to new discoveries.

The GeneForce algorithm follows an approach based on using the experimental data available to lead modifications for a model in order to perform a best fit of experimental data. This data driven approach is currently applied in a more extended way by other methods. They are based on using the omics data to constrain an FBA simulation in trying to integrate directly gene expression into FBA without using any TRN reconstruction. Indeed, the process of reconstruction of a TRN in terms of Boolean rules is time consuming and difficult, because the literature available have to be

searched manually. Consequently, this process has been done only for few organisms [103,128,139].

For this reason, the data driven approaches have been used extensively recently. Several algorithms have been developed and many examples of this approach are present in the literature, such as the E-flux [140] and the GIMME [141] method and sometimes implemented in one single software [142]. Some of them use thresholds in order to reformulate the gene expression levels into on/off description; others try to overcome the use of arbitrary thresholds (for a recent review see Ref. [143]). Nevertheless, the results performed by data driven approaches can be limited because of the lack of correlation between the gene expression and the protein level [144]. However, this correlation can be improved by using a network based model [18] under certain circumstances.

Furthermore, the separation between a data driven method and a regulatory reconstruction driven method is not clearly defined sometimes. For instance, the probabilistic regulation of metabolism (PROM) [145] algorithm starts from a regulatory network structure and gene expression data in order to model the transcription regulation by applying a probabilistic approach. The PROM algorithm was initially applied to *E. coli* and *Mycobacterium tuberculosis*, and was later applied to *S. cerevisiae* to predict the structure of its TRN reconstruction [146]. Moreover, some methods use the information available from gene expression and regulatory networks together to improve the prediction ability of their models. One relevant example of this approach is the MADE [147] algorithm which has been applied to a *S. cerevisiae* metabolic reconstruction. The MADE algorithm uses experimental gene expression data by setting a binary approximation of gene expression changes in different conditions after a statistical test to evaluate the most probable approximation. In this way MADE algorithm is able to formulate an MILP problem and avoid using arbitrary thresholds. The MADE method has been integrated during a second stage with the yeast TRN reconstruction, and this step improved and refined the prediction ability of the method.

FUTURE PERSPECTIVES

Although many large-scale TRNs have been reconstructed, and some of them have been integrated into metabolic models for *S. cerevisiae*, there are still several important problems to be solved in this area. Many of the transcriptional regulatory interactions obtained by biochemical approaches need to be studied for the efficiency of their functions. In addition, the dynamics of the TRN need to be described in more detail by examining the transcriptional regulatory interactions under more envir-

omental conditions. For the integration of TRN reconstructions into metabolic models, only a small number of approaches used for *E. coli* have been also applied to *S. cerevisiae* so far. The transition from the modeling of prokaryotic organisms to eukaryotic organisms is challenging due to the increase in complexity of such organisms and the scarcity of our knowledge about the regulatory mechanisms.

The merging process of the Boolean logic qualitative description with the FBA quantitative results limits the outcome of the integrated system. Even when this integration is possible, the use of an on/off logic to describe the regulatory network reduces the ability of the model to make predictions. The Boolean on/off description of regulatory interactions is not sufficient to catch all the regulatory events of the cell, because it is not able to describe the intermediate levels of regulation present in a continuous system. For instance, when this binary representation is applied to describe the role of essential genes, a deletion of such genes due to a down-regulation event has the consequence that the model is not feasible. As a result, the development of methods allow a low flux even when an essential gene is down-regulated, and new large-scale methods based on non-Boolean formalisms are advantageous for a better integration of TRN reconstructions with metabolic models. Some steps toward this direction have been made with the introduction of probabilistic and data driven approaches using the MILP formulation, but other steps can be made by the extension of rFBA based techniques. Nevertheless, these are not simple tasks because the improvements of the experimental methodologies as well as the improvements of the modeling techniques are necessary in order to gain more quantitative data about the TRN in the future.

Recently, the efforts to reconstruct the metabolic network have been extended to create reconstructions of other systems present inside the cell. This new kind of modelling is based on an explicit representation of the elements which carry out the cellular tasks like the transcription and the translation process. This kind of reconstruction dramatically improves the dimension and the level of detail of the model. The reconstruction of *E. coli*'s transcriptional and translational machinery [148] allowed the formulation of the next generation genome-scale metabolic model (ME-model) and ME-models have been recently developed for *E. coli* [149,150] and *Thermotoga maritima* [151]. This is a promising approach for building a scaffold that can be used for the integration of omics data at several layers of the model. Nevertheless, if this approach can be carried out for a eukaryotic organism is an open question. There is a lack of information about several processes inside the yeast cell, so it is not clear if it will be possible to create a ME-model for yeast in the immediate future, although a step

toward this objective has been made with the reconstruction of the yeast secretory machinery [152].

ACKNOWLEDGEMENTS

This work has been funded by the European Research Council (INSYSBIO, Grant No. 247013) and the Novo Nordisk Foundation.

CONFLICT OF INTERESTS

The authors Guodong Liu, Antonio Marras and Jens Nielsen declare that they have no conflict of interests.

REFERENCES

- Osterlund, T., Nookaew, I. and Nielsen, J. (2012) Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol. Adv.*, 30, 979–988.
- Herrgård, M. J., Covert, M. W. and Palsson, B. O. (2004) Reconstruction of microbial transcriptional regulatory networks. *Curr. Opin. Biotechnol.*, 15, 70–77.
- Blais, A. and Dynlacht, B. D. (2005) Constructing transcriptional regulatory networks. *Genes Dev.*, 19, 1499–1511.
- Chua, G., Robinson, M. D., Morris, Q. and Hughes, T. R. (2004) Transcriptional networks: reverse-engineering gene regulation on a global scale. *Curr. Opin. Microbiol.*, 7, 638–646.
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. and Palsson, B. O. (2009) Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.*, 7, 129–143.
- Hughes, T. R. and de Boer, C. G. (2013) Mapping yeast transcriptional networks. *Genetics*, 195, 9–36.
- Kim, T. M. and Park, P. J. (2011) Advances in analysis of transcriptional regulatory networks. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 3, 21–35.
- Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J. and Stolovitzky, G., and the DREAM5 Consortium. (2012) Wisdom of crowds for robust gene network inference. *Nat. Methods*, 9, 796–804.
- Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298, 799–804.
- Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R. J., Jules, M., Link, H., Aymerich, S., Stelling, J. and Sauer, U. (2013) Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Mol. Syst. Biol.*, 9, 709.
- Abdulrehman, D., Monteiro, P. T., Teixeira, M. C., Mira, N. P., Lourenço, A. B., dos Santos, S. C., Cabrito, T. R., Francisco, A. P., Madeira, S. C., Aires, R. S., et al. (2011) YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in *Saccharomyces cerevisiae* through a web services interface. *Nucleic Acids Res.*, 39, D136–D140.
- Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., et al. (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, 40, D700–D705.
- Broach, J. R. (2012) Nutritional control of growth and development in yeast. *Genetics*, 192, 73–105.

14. Ljungdahl, P. O. and Daignan-Fornier, B. (2012) Regulation of amino acid, nucleotide, and phosphate metabolism in *Saccharomyces cerevisiae*. *Genetics*, 190, 885–929.
15. Daran-Lapujade, P., Rossell, S., van Gulik, W. M., Luttk, M. A., de Groot, M. J., Slijper, M., Heck, A. J., Daran, J. M., de Winde, J. H., Westerhoff, H. V., et al. (2007) The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at post-transcriptional levels. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 15753–15758.
16. Daran-Lapujade, P., Jansen, M. L., Daran, J. M., van Gulik, W., de Winde, J. H. and Pronk, J. T. (2004) Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*. A chemostat culture study. *J. Biol. Chem.*, 279, 9125–9138.
17. Fendt, S. M., Oliveira, A. P., Christen, S., Picotti, P., Dechant, R. C. and Sauer, U. (2010) Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Mol. Syst. Biol.*, 6, 432.
18. Moxley, J. F., Jewett, M. C., Antoniewicz, M. R., Villas-Boas, S. G., Alper, H., Wheeler, R. T., Tong, L., Hinnebusch, A. G., Ideker, T., Nielsen, J., et al. (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc. Natl. Acad. Sci. U.S.A.*, 106, 6477–6482.
19. Österlund, T., Nookaew, I., Bordel, S. and Nielsen, J. (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst. Biol.*, 7, 36.
20. Bordel, S., Agren, R. and Nielsen, J. (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLOS Comput. Biol.*, 6, e1000859.
21. Klein, C. J., Rasmussen, J. J., Rønnow, B., Olsson, L. and Nielsen, J. (1999) Investigation of the impact of *MIG1* and *MIG2* on the physiology of *Saccharomyces cerevisiae*. *J. Biotechnol.*, 68, 197–212.
22. Marczak, J. E. and Brandriss, M. C. (1989) Isolation of constitutive mutations affecting the proline utilization pathway in *Saccharomyces cerevisiae* and molecular analysis of the *PUT3* transcriptional activator. *Mol. Cell. Biol.*, 9, 4696–4705.
23. Hedges, D., Proft, M. and Entian, K. D. (1995) CAT8, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, 15, 1915–1922.
24. Lo, T. L., Qu, Y., Uwamahoro, N., Quenault, T., Beilharz, T. H. and Traven, A. (2012) The mRNA decay pathway regulates the expression of the Flo11 adhesin and biofilm formation in *Saccharomyces cerevisiae*. *Genetics*, 191, 1387–1391.
25. Hinnebusch, A. G. (1984) Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, 81, 6442–6446.
26. Turner, G. C., Du, F. and Varshavsky, A. (2000) Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. *Nature*, 405, 579–583.
27. Andréasson, C. and Ljungdahl, P. O. (2002) Receptor-mediated endoproteolytic activation of two transcription factors in yeast. *Genes Dev.*, 16, 3158–3172.
28. Cox, K. H., Tate, J. J. and Cooper, T. G. (2002) Cytoplasmic compartmentation of Gln3 during nitrogen catabolite repression and the mechanism of its nuclear localization during carbon starvation in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 277, 37559–37566.
29. Sellick, C. A. and Reece, R. J. (2003) Modulation of transcription factor function by an amino acid: activation of Put3p by proline. *EMBO J.*, 22, 5147–5153.
30. Pannala, V. R., Bhat, P. J., Bhartiya, S. and Venkatesh, K. V. (2010) Systems biology of *GAL* regulon in *Saccharomyces cerevisiae*. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 2, 98–106.
31. Kornitzer, D., Raboy, B., Kulka, R. G. and Fink, G. R. (1994) Regulated degradation of the transcription factor Gcn4. *EMBO J.*, 13, 6021–6030.
32. Soontornngun, N., Laroche, M., Drouin, S., Robert, F. and Turcotte, B. (2007) Regulation of gluconeogenesis in *Saccharomyces cerevisiae* is mediated by activator and repressor functions of Rds2. *Mol. Cell. Biol.*, 27, 7895–7905.
33. Martens, J. A., Wu, P. Y. and Winston, F. (2005) Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev.*, 19, 2695–2704.
34. Ambroziak, J. and Henry, S. A. (1994) *INO2* and *INO4* gene products, positive regulators of phospholipid biosynthesis in *Saccharomyces cerevisiae*, form a complex that binds to the *INO1* promoter. *J. Biol. Chem.*, 269, 15344–15349.
35. Blaiseau, P. L., Isnard, A. D., Surdin-Kerjan, Y. and Thomas, D. (1997) Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. *Mol. Cell. Biol.*, 17, 3640–3648.
36. Georis, I., Feller, A., Vierendeels, F. and Dubois, E. (2009) The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. *Mol. Cell. Biol.*, 29, 3803–3815.
37. Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. and Teichmann, S. A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.*, 14, 283–291.
38. Lavoie, H., Hogues, H. and Whiteway, M. (2009) Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. *Curr. Opin. Microbiol.*, 12, 655–663.
39. Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I. and Nielsen, J. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLOS Comput. Biol.*, 9, e1002980.
40. Geertz, M. and Maerkl, S. J. (2010) Experimental strategies for studying transcription factor-DNA binding specificities. *Brief Funct. Genomics*, 9, 362–373.
41. Xie, Z., Hu, S., Qian, J., Blackshaw, S. and Zhu, H. (2011) Systematic characterization of protein-DNA interactions. *Cell. Mol. Life Sci.*, 68, 1657–1668.
42. Garner, M. M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.*, 9, 3047–3060.
43. Galas, D. J. and Schmitz, A. (1978) DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.*, 5, 3157–3170.
44. Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000) Genome-wide location and function of DNA binding proteins. *Science*, 290, 2306–2309.
45. Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., et al. (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, 431, 99–104.

46. Venters, B. J., Wachi, S., Mavrich, T. N., Andersen, B. E., Jena, P., Sinnamon, A. J., Jain, P., Roller, N. S., Jiang, C., Hemeryck-Walsh, C., et al. (2011) A comprehensive genomic binding map of gene and chromatin regulatory proteins in *Saccharomyces*. *Mol. Cell*, 41, 480–492.
47. Park, P. J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, 10, 669–680.
48. Rhee, H. S. and Pugh, B. F. (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, 147, 1408–1419.
49. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. and Lieb, J. D. (2012) Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*, 484, 251–255.
50. Mirzaei, H., Knijnenburg, T. A., Kim, B., Robinson, M., Picotti, P., Carter, G. W., Li, S., Dilworth, D. J., Eng, J. K., Aitchison, J. D., et al. (2013) Systematic measurement of transcription factor-DNA interactions by targeted mass spectrometry identifies candidate gene regulatory proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 110, 3645–3650.
51. MacQuarrie, K. L., Fong, A. P., Morse, R. H. and Tapscott, S. J. (2011) Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.*, 27, 141–148.
52. Nookaew, I., Papini, M., Pornputtpong, N., Scalcinati, G., Fagerberg, L., Uhlén, M. and Nielsen, J. (2012) A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, 40, 10084–10097.
53. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucan-Danila, A., Anderson, K., André, B., et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418, 387–391.
54. Hu, Z., Killion, P. J. and Iyer, V. R. (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.*, 39, 683–687.
55. Reimand, J., Vaquerizas, J. M., Todd, A. E., Vilo, J. and Luscombe, N. M. (2010) Comprehensive reanalysis of transcription factor knockout expression data in *Saccharomyces cerevisiae* reveals many new targets. *Nucleic Acids Res.*, 38, 4768–4777.
56. Chua, G., Morris, Q. D., Sopko, R., Robinson, M. D., Ryan, O., Chan, E. T., Frey, B. J., Andrews, B. J., Boone, C. and Hughes, T. R. (2006) Identifying transcription factor functions and targets by phenotypic activation. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 12045–12050.
57. Tang, L., Liu, X. and Clarke, N. D. (2006) Inferring direct regulatory targets from expression and genome location analyses: a comparison of transcription factor deletion and overexpression. *BMC Genomics*, 7, 215.
58. Gitter, A., Siegfried, Z., Klutstein, M., Fornes, O., Oliva, B., Simon, I. and Bar-Joseph, Z. (2009) Backup in gene regulatory networks explains differences between binding and knockout results. *Mol. Syst. Biol.*, 5, 276.
59. Reiss, D. J., Baliga, N. S. and Bonneau, R. (2006) Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics*, 7, 280.
60. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.*, 34, 166–176.
61. Hibbs, M. A., Hess, D. C., Myers, C. L., Huttenhower, C., Li, K. and Troyanskaya, O. G. (2007) Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics*, 23, 2692–2699.
62. Wu, M. and Chan, C. (2012) Learning transcriptional regulation on a genome scale: a theoretical analysis based on gene expression data. *Brief Bioinformatics*, 13, 150–161.
63. Beer, M. A. and Tavazoie, S. (2004) Predicting gene expression from sequence. *Cell*, 117, 185–198.
64. Badis, G., Chan, E. T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C. D., Gossett, A. J., Hasinoff, M. J., Warren, C. L., et al. (2008) A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell*, 32, 878–887.
65. Zhu, C., Byers, K. J., McCord, R. P., Shi, Z., Berger, M. F., Newburger, D. E., Saulrieta, K., Smith, Z., Shah, M. V., Radhakrishnan, M., et al. (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.*, 19, 556–566.
66. Fordyce, P. M., Gerber, D., Tran, D., Zheng, J., Li, H., DeRisi, J. L. and Quake, S. R. (2010) De novo identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nat. Biotechnol.*, 28, 970–975.
67. Ho, S. W., Jona, G., Chen, C. T., Johnston, M. and Snyder, M. (2006) Linking DNA-binding proteins to their recognition sequences by using protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 9940–9945.
68. Roth, F. P., Hughes, J. D., Estep, P. W. and Church, G. M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.*, 16, 939–945.
69. Hughes, J. D., Estep, P. W., Tavazoie, S. and Church, G. M. (2000) Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.*, 296, 1205–1214.
70. Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A. and Johnston, M. (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science*, 301, 71–76.
71. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. and Lander, E. S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, 423, 241–254.
72. Gordán, R., Murphy, K. F., McCord, R. P., Zhu, C., Vedenko, A. and Bulyk, M. L. (2011) Curated collection of yeast transcription factor DNA binding specificity data reveals novel structural and gene regulatory insights. *Genome Biol.*, 12, R125.
73. de Boer, C. G. and Hughes, T. R. (2012) YeTFaSCO: a database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res.*, 40, D169–D179.
74. Spivak, A. T. and Stormo, G. D. (2012) ScerTF: a comprehensive database of benchmarked position weight matrices for *Saccharomyces* species. *Nucleic Acids Res.*, 40, D162–D168.
75. Zhu, J. and Zhang, M. Q. (1999) SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics*, 15, 607–611.
76. Costanzo, M. C., Engel, S. R., Wong, E. D., Lloyd, P., Karra, K., Chan, E. T., Weng, S., Paskov, K. M., Roe, G. R., Binkley, G., et al. (2014) *Saccharomyces* genome database provides new regulation data. *Nucleic Acids Res.*, 42, D717–D725.
77. Liu, X., Jessen, W. J., Sivaganesan, S., Aronow, B. J. and Medvedovic, M. (2007) Bayesian hierarchical model for transcrip-

- tional module discovery by jointly modeling gene expression and ChIP-chip data. *BMC Bioinformatics*, 8, 283.
78. Bar-Joseph, Z., Gerber, G. K., Lee, T. I., Rinaldi, N. J., Yoo, J. Y., Robert, F., Gordon, D. B., Fraenkel, E., Jaakkola, T. S., Young, R. A., et al. (2003) Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.*, 21, 1337–1342.
 79. Gao, F., Foat, B. C. and Bussemaker, H. J. (2004) Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics*, 5, 31.
 80. Kato, M., Hata, N., Banerjee, N., Futcher, B. and Zhang, M. Q. (2004) Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol.*, 5, R56.
 81. Youn, A., Reiss, D. J. and Stuetzle, W. (2010) Learning transcriptional networks from the integration of ChIP-chip and expression data in a non-parametric model. *Bioinformatics*, 26, 1879–1886.
 82. Novershtern, N., Regev, A. and Friedman, N. (2011) Physical Module Networks: an integrative approach for reconstructing transcription regulation. *Bioinformatics*, 27, i177–i185.
 83. Haverty, P. M., Hansen, U. and Weng, Z. (2004) Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification. *Nucleic Acids Res.*, 32, 179–188.
 84. McCord, R. P., Berger, M. F., Philippakis, A. A. and Bulyk, M. L. (2007) Inferring condition-specific transcription factor function from DNA binding and gene expression data. *Mol. Syst. Biol.*, 3, 100.
 85. Xing, B. and van der Laan, M. J. (2005) A statistical method for constructing transcriptional regulatory networks using gene expression and sequence data. *J. Comput. Biol.*, 12, 229–246.
 86. Wang, W., Cherry, J. M., Botstein, D. and Li, H. (2002) A systematic approach to reconstructing transcription networks in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.*, 99, 16893–16898.
 87. Chang, L. W., Nagarajan, R., Magee, J. A., Milbrandt, J. and Stormo, G. D. (2006) A systematic model to predict transcriptional regulatory mechanisms based on overrepresentation of transcription factor binding profiles. *Genome Res.*, 16, 405–413.
 88. Qin, J., Li, M. J., Wang, P., Zhang, M. Q. and Wang, J. (2011) ChIP-Array: combinatory analysis of ChIP-seq/chip and microarray gene expression data to discover direct/indirect targets of a transcription factor. *Nucleic Acids Res.*, 39, W430–436.
 89. Chen, G., Jensen, S. T. and Stoekert, C. J. Jr. (2007) Clustering of genes into regulons using integrated modeling-COGRIM. *Genome Biol.*, 8, R4.
 90. Lemmens, K., Dhollander, T., De Bie, T., Monsieurs, P., Engelen, K., Smets, B., Winderickx, J., De Moor, B. and Marchal, K. (2006) Inferring transcriptional modules from ChIP-chip, motif and microarray data. *Genome Biol.*, 7, R37.
 91. Workman, C. T., Mak, H. C., McCuine, S., Tagne, J. B., Agarwal, M., Ozier, O., Begley, T. J., Samson, L. D. and Ideker, T. (2006) A systems approach to mapping DNA damage response pathways. *Science*, 312, 1054–1059.
 92. Herrgård, M. J., Covert, M. W. and Palsson, B. O. (2003) Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res.*, 13, 2423–2434.
 93. Machado, D., Costa, R. S., Rocha, M., Ferreira, E. C., Tidor, B. and Rocha, I. (2011) Modeling formalisms in Systems Biology. *AMB Express*, 1, 45.
 94. Schlitt, T. and Brazma, A. (2007) Current approaches to gene regulatory network modelling. *BMC Bioinformatics*, 8, S9.
 95. Karlebach, G. and Shamir, R. (2008) Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.*, 9, 770–780.
 96. de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, 9, 67–103.
 97. Glass, L. and Kauffman, S. A. (1973) The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.*, 39, 103–129.
 98. Kauffman, S., Peterson, C., Samuelsson, B. and Troein, C. (2003) Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. U.S.A.*, 100, 14796–14799.
 99. Gianchandani, E. P., Papin, J. A., Price, N. D., Joyce, A. R. and Palsson, B. O. (2006) Matrix formalism to describe functional states of transcriptional regulatory systems. *PLOS Comput. Biol.*, 2, e101.
 100. Yu, H., Luscombe, N. M., Qian, J. and Gerstein, M. (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.*, 19, 422–427.
 101. Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. and Gerstein, M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431, 308–312.
 102. Jothi, R., Balaji, S., Wuster, A., Grochow, J. A., Gsponer, J., Przytycka, T. M., Aravind, L. and Babu, M. M. (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol. Syst. Biol.*, 5, 294.
 103. Herrgård, M. J., Lee, B. S., Portnoy, V. and Palsson, B. O. (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.*, 16, 627–635.
 104. Maslov, S. and Sneppen, K. (2002) Specificity and stability in topology of protein networks. *Science*, 296, 910–913.
 105. Guelzim, N., Bottani, S., Bourguin, P. and Képès, F. (2002) Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.*, 31, 60–63.
 106. Balaji, S., Iyer, L. M., Aravind, L. and Babu, M. M. (2006) Uncovering a hidden distributed architecture behind scale-free transcriptional regulatory networks. *J. Mol. Biol.*, 360, 204–212.
 107. Guzmán-Vargas, L. and Santillán, M. (2008) Comparative analysis of the transcription-factor gene regulatory networks of *E. coli* and *S. cerevisiae*. *BMC Syst. Biol.*, 2, 13.
 108. Yu, H. and Gerstein, M. (2006) Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.*, 103, 14724–14731.
 109. 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.
 110. Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.*, 6, 1290–1307.
 111. Thiele, I. and Palsson, B. O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, 5, 93–121.
 112. Chang, A., Scheer, M., Grote, A., Schomburg, I. and Schomburg, D. (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.*, 37, D588–D592.
 113. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M.,

- Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, 34, D354–D357.
114. Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.*, 26, 1155–1160.
115. Nookaew, I., Jewett, M. C., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J. and Bhumiratana, S. (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst. Biol.*, 2, 71.
116. Orth, J. D., Thiele, I. and Palsson, B. O. (2010) What is flux balance analysis? *Nat. Biotechnol.*, 28, 245–248.
117. Price, N. D., Reed, J. L. and Palsson, B. O. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, 2, 886–897.
118. Feist, A. M. and Palsson, B. O. (2010) The biomass objective function. *Curr. Opin. Microbiol.*, 13, 344–349.
119. Edwards, J. S. and Palsson, B. O. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U.S.A.*, 97, 5528–5533.
120. Covert, M. W., Schilling, C. H. and Palsson, B. (2001) Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, 213, 73–88.
121. Covert, M. W. and Palsson, B. O. (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.*, 277, 28058–28064.
122. Lee, S., Phalakornkule, C., Domach, M. M. and Grossmann, I. E. (2000) Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.*, 24, 711–716.
123. 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.
124. Snitkin, E. S. and Segrè, D. (2008) Optimality criteria for the prediction of metabolic fluxes in yeast mutants. *Genome Inform.*, 20, 123–134.
125. Llaneras, F. and Picó, J. (2008) Stoichiometric modelling of cell metabolism. *J. Biosci. Bioeng.*, 105, 1–11.
126. Covert, M. W. and Palsson, B. O. (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.*, 221, 309–325.
127. Varma, A. and Palsson, B. O. (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.*, 60, 3724–3731.
128. Covert, M. W., Knight, E. M., Reed, J. L., Herrgård, M. J. and Palsson, B. O. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429, 92–96.
129. 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.
130. Zamboni, N. (2011) ¹³C metabolic flux analysis in complex systems. *Curr. Opin. Biotechnol.*, 22, 103–108.
131. Liu, Y. Y., Slotine, J. J. and Barabási, A. L. (2013) Observability of complex systems. *Proc. Natl. Acad. Sci. U.S.A.*, 110, 2460–2465.
132. Smallbone, K., Simeonidis, E., Broomhead, D. S. and Kell, D. B. (2007) Something from nothing: bridging the gap between constraint-based and kinetic modelling. *FEBS J.*, 274, 5576–5585.
133. Covert, M. W., Xiao, N., Chen, T. J. and Karr, J. R. (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 24, 2044–2050.
134. Lee, J. M., Gianchandani, E. P., Eddy, J. A. and Papin, J. A. (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLOS Comput. Biol.*, 4, e1000086.
135. Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B. Jr, Assad-Garcia, N., Glass, J. I. and Covert, M. W. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell*, 150, 389–401.
136. Shlomi, T., Eisenberg, Y., Sharan, R. and Ruppin, E. (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.*, 3, 101.
137. Kim, J. and Reed, J. L. (2010) OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.*, 4, 53.
138. Barua, D., Kim, J. and Reed, J. L. (2010) An automated phenotype-driven approach (*GeneForce*) for refining metabolic and regulatory models. *PLOS Comput. Biol.*, 6, e1000970.
139. Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessièrès, P., Aymerich, S. and Fromion, V. (2008) Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst. Biol.*, 2, 20.
140. Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., Cheng, T. Y., Moody, D. B., Murray, M. and Galagan, J. E. (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLOS Comput. Biol.*, 5, e1000489.
141. Becker, S. A. and Palsson, B. O. (2008) Context-specific metabolic networks are consistent with experiments. *PLOS Comput. Biol.*, 4, e1000082.
142. Jensen, P. A., Lutz, K. A. and Papin, J. A. (2011) TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst. Biol.*, 5, 147.
143. Blazier, A. S. and Papin, J. A. (2012) Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.*, 3, 299.
144. Gygi, S. P., Rochon, Y., Franz, B. R. and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.*, 19, 1720–1730.
145. Chandrasekaran, S. and Price, N. D. (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 17845–17850.
146. Chandrasekaran, S. and Price, N. D. (2013) Metabolic constraint-based refinement of transcriptional regulatory networks. *PLOS Comput. Biol.*, 9, e1003370.
147. 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.
148. Thiele, I., Jamshidi, N., Fleming, R. M. and Palsson, B. O. (2009) Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLOS Comput. Biol.*, 5, e1000312.

149. Thiele, I., Fleming, R. M., Que, R., Bordbar, A., Diep, D. and Palsson, B. O. (2012) Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS ONE*, 7, e45635.
150. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. and Palsson, B. O. (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.*, 9, 693.
151. Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., Schrimpe-Rutledge, A. C., Smith, R. D., Adkins, J. N., Zengler, K., et al. (2012) In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.*, 3, 929.
152. Feizi, A., Österlund, T., Petranovic, D., Bordel, S. and Nielsen, J. (2013) Genome-scale modeling of the protein secretory machinery in yeast. *PLoS ONE*, 8, e63284.