



ELSEVIER

Available online at www.sciencedirect.com

Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks

Eytan Ruppin¹, Jason A Papin², Luis F de Figueiredo³ and Stefan Schuster³

With the advent of modern omics technologies, it has become feasible to reconstruct (quasi-) whole-cell metabolic networks and characterize them in more and more detail. Computer simulations of the dynamic behavior of such networks are difficult due to a lack of kinetic data and to computational limitations. In contrast, network analysis based on appropriate constraints such as the steady-state condition (constraint-based analysis) is feasible and allows one to derive conclusions about the system's metabolic capabilities. Here, we review methods for the reconstruction of metabolic networks, modeling techniques such as flux balance analysis and elementary flux modes and current progress in their development and applications. Game-theoretical methods for studying metabolic networks are discussed as well.

Addresses

¹ Department of Computer Science, Tel Aviv University, Tel Aviv, Israel

² Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA

³ Department of Bioinformatics, Friedrich Schiller University Ernst-Abbe-Pl. 2, 07743 Jena, Germany

Corresponding author: Schuster, Stefan (stefan.schu@uni-jena.de)

Current Opinion in Biotechnology 2010, **21**:1–9

This review comes from a themed issue on
Systems biology
Edited by Vitor Martins dos Santos and Jiri Damborsky

0958-1669/\$ – see front matter
© 2010 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.copbio.2010.07.002](https://doi.org/10.1016/j.copbio.2010.07.002)

Introduction

The study of metabolism has changed drastically during the last century. The concept of metabolic pathways was molded by the experimental methods available in the beginning of the 20th century resulting in a stepwise elucidation of metabolism (cf. [1[•]]). In the second half of the 20th century, the discovery of the structure and information coding of DNA laid the foundations for recombinant technology, making microorganisms more amenable to metabolic engineering [2].

The mathematical modeling and computer simulation of metabolic systems started with dynamic modeling [3[•]]. This is useful, for example, to simulate the occasional metabolic oscillations in bioreactors. Later, it was realized that the knowledge of kinetic parameters was insufficient in many cases and that a detailed dynamic simulation is

often unnecessary [4,5]. Many specific questions such as the effect of the activation or overexpression of an enzyme can be tackled by specially tailored techniques such as Metabolic Control Analysis (MCA) (cf. [2,3[•],4]). Moreover, the search for methods to analyze invariants of intracellular networks led to the development of the constraint-based modeling (CBM) approach, also called structural analysis or network analysis [4,5]. This requires even less input data than MCA. A subfield is called Metabolic Pathway Analysis [6,7,8^{••}], in which the structure of pathways (routes) going through the system is detected and/or optimal flux distributions are calculated based on the stoichiometry of the network and the directionality of reactions, the knowledge of which is often available. Elementary flux modes (EFMs) [9^{••},10^{••}] and extreme pathways [11^{••}] were established as unbiased mathematical representations of metabolic pathways.

CBM comprises metabolic flux analysis (MFA) (cf. [2]) by which flux distributions can be predicted using flux measurements in addition to network stoichiometry. The limitations in measurements led to the inclusion of additional constraints coming from the Darwinian theory of optimization in evolution. Notably, optimality principles such as maximizing growth rate or given reaction fluxes at normalized input rate are widely used, thus enabling one to predict fluxes by linear programming [12^{••},13^{••}]. This is the essence of Flux Balance Analysis (FBA), by which phenotypically relevant flux distributions in metabolic networks can be predicted (cf. [14,15]).

At the end of the 20th century, the development of new experimental technologies such as sequencing and chip technologies triggered the explosion of omics data. In metabolomics, several hundreds of metabolite concentrations can be measured simultaneously [16^{••},17]. In fluxomics, in contrast, it is difficult to measure more than a dozen fluxes simultaneously [18^{••}]. After DNA microarrays, transcriptomics has recently reached a higher level by RNA-Seq technologies [19]. A secondary source is bibliomics (text mining). High amounts of data enabled one to reconstruct more complex metabolic networks reaching a genome scale for a rapidly increasing number of species (cf. [20^{••},21^{••}]). In addition, recent efforts to integrate gene expression data [22,23^{••},24], sequencing data [25^{••}], proteomics [17], and other omics data have enabled the generation of reconstructions unique to particular life-cycle stages, environments, and genetic backgrounds.

These large models together with CBM methods represent a key foundational advance in Systems Biology [5,14] and

2 Systems biology

are essential for seeking comprehension of biological functioning throughout the integration of data with mathematical models. The modeling and computer simulation of metabolism in the genome and post-genome eras have been the subject of a number of reviews [6,7,20^{••},21^{••}]. In this paper we will review recent advancements in the methods for metabolic network reconstruction, the tools available for their analysis and several applications. A special focus will be on game-theoretical methods (cf. [26]). Finally, perspectives on further developments in Systems Biology will be outlined.

Reconstruction of metabolic networks

The process of metabolic network reconstruction begins with the annotation data of the relevant genes (cf. [20^{••},27]). This annotation provides the ‘parts list’ for the network. The metabolic reactions that the associated gene products catalyze are delineated by incorporating data on the metabolites and stoichiometry from databases (e.g. ExPasy and KEGG) and the literature. The stoichiometric coefficients of the metabolites or compounds in the associated reactions are typically represented in a stoichiometric matrix, N (sometimes denoted S) with its rows corresponding to the metabolites and the columns representing the chemical transformations that the gene products (enzymes) catalyze. The usefulness of CBM in the reconstruction process is outlined in [Box 1](#). Two key challenges are firstly, the integration of disparate high-throughput data and secondly, the inclusion of additional constraints to improve the predictive power. Recently, methods for the iterative refinement of the networks using high-throughput transcript data have been devel-

oped and used to significantly improve the reconstruction of *Chlamydomonas reinhardtii* and to identify key genes associated with biofuel production [25^{••}].

For a reconstructed network to be realistic, there must be a flux vector v fulfilling Eqn (1) in [Box 1](#) and covering practically all reactions. While there may be a few reactions that always subsist at thermodynamic equilibrium, the vast majority must be able to operate at non-equilibrium, that is, they must not be blocked due to missing reactions [28[•]] and have to fulfill mass conservation [29]. The coverage of a network by a flux distribution can be tested by a method called flux coupling analysis, which is based on linear programming [30^{••}].

Genome-scale reconstructions have been assembled for organisms from all kingdoms: archaea, eukaryota, and bacteria ([Figure 1](#)), and include single-cell and multi-cell organisms (for references to specific reconstructions, see [21^{••}]). Here, we mention only a few: *Escherichia coli* [31^{••}], *Saccharomyces cerevisiae* [32^{••}], *C. reinhardtii* [25^{••},33], *Arabidopsis thaliana* [34,35], mouse [36,37], and human [38^{••},39^{••}]. The goal is to account for all the enzymes encoded in the entire genome. However, the term ‘genome-scale’ is to express the dimension in which this is done, which does not necessarily imply that this difficult task would be accomplished with 100% completeness. Questions such as whether substrate A can be transformed into product B can only be tackled exhaustively in a whole-cell model, which provides a further motivation for their reconstruction [40^{••}].

Some recent efforts are focused on automating the reconstruction process (cf. [20^{••},41]), aided by the development of computational platforms to manage the data associated with gene–enzyme–reaction associations and reaction stoichiometry. This has had some success, although there is typically a recognized need for manual curation efforts [28[•],29,42]. A significant remaining challenge is the visualization of these networks, in spite of some recent efforts in this area [43].

Biotechnological applications of metabolic modeling

With the reconstructions that have been generated, the next important step is the development of analysis tools and frameworks to study functional properties of these networks [30^{••},40^{••},44[•],45[•]]. Various tools for metabolic modeling have been established and refined. These include dynamic modeling, optimization, game-theoretical methods, FBA, Metabolic Pathway Analysis and others (cf. Introduction).

FBA is based on optimality principles ([Box 1](#)). It is a matter of debate whether FBA always gives sufficiently correct results [46,47], see below. Various objective functions have been compared by [48[•]]. However, more

Box 1 Most metabolic systems subsist at stationary states. Even if they oscillate (only very few do), the average reaction rates (on a sufficiently long time scale), v , must fulfill the steady-state condition

$$N \cdot v = 0 \quad (1)$$

because, otherwise, the concentrations of metabolic intermediates would accumulate or be depleted in the long run. In addition, for some or all fluxes, inequality constraints can be written:

$$v_{i,\text{inf}} \leq v_i \leq v_{i,\text{sup}} \quad (2)$$

For example, $v_{i,\text{inf}}$ is zero for all irreversible reactions. For all reactions, $v_{i,\text{sup}}$ can be given by the respective maximal velocity if it is known.

Central to FBA is an optimization principle

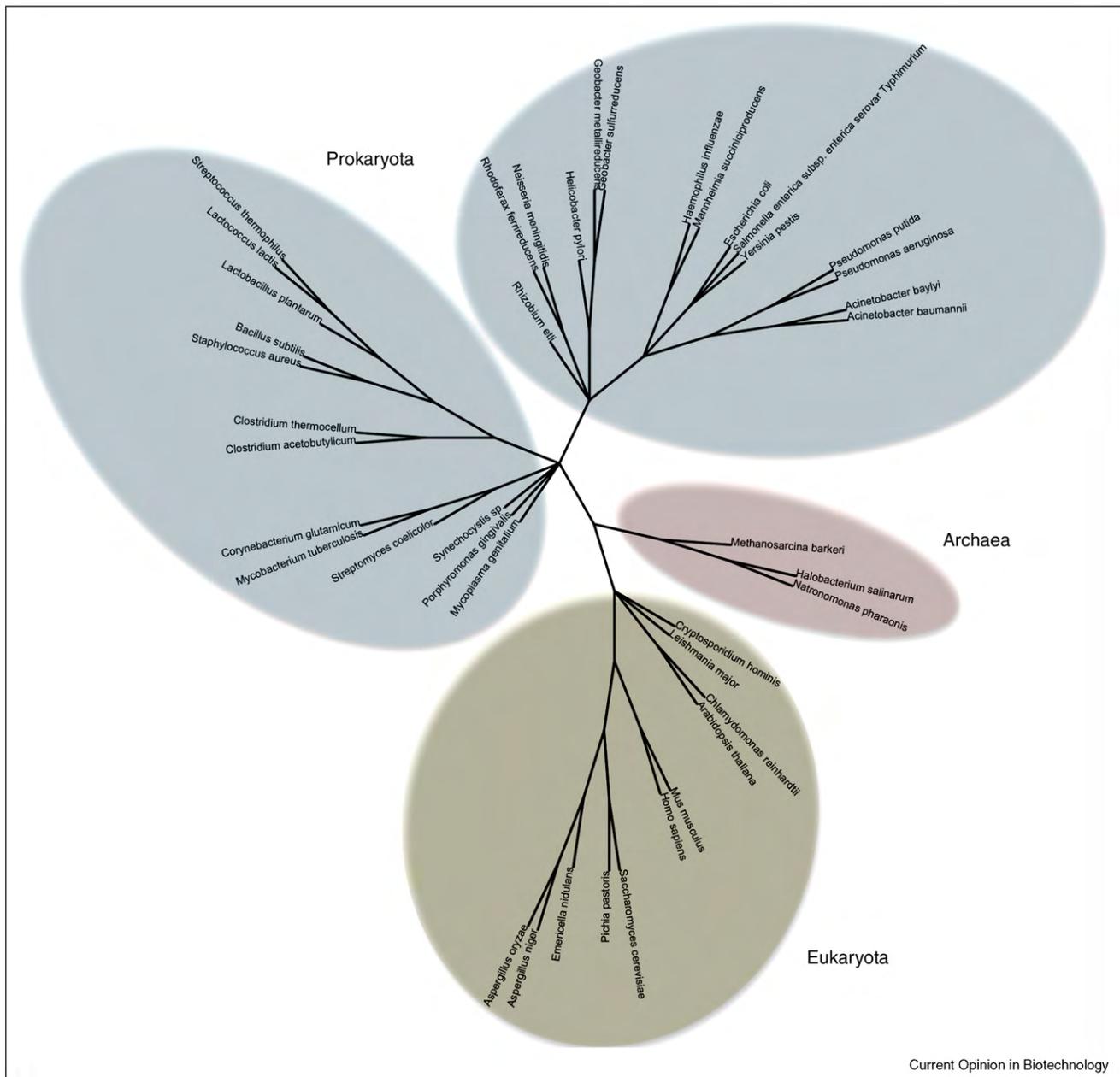
$$\text{maximize } \sum c_j v_j \quad (3)$$

subject to relations (1) and (2) [12^{••},13^{••},15]. The coefficients c_j denote the weights of the particular fluxes in the objective function, for example, the production of biomass. Typically, relation (2) includes one constraint that fixes or limits a relevant input flux, for example, glucose uptake.

The non-decomposability of elementary flux modes can be expressed by a constraint saying that the support of the flux vector is not a proper superset of the support of any other steady-state flux vector, v' :

$$S(v) = \{i : v_i \neq 0\} \text{ not proper superset of } S(v') \quad (4)$$

Figure 1



Phylogenetic relationship between the organisms for which metabolic network reconstructions exist, generated with iTOL [84].

sophisticated methods usually require kinetic data, with the above-mentioned problem of incomplete data availability. Some current approaches attempt to guess flux values from thermodynamics [49], purely structural properties [45^{*}] or RNA transcript data [23^{**}].

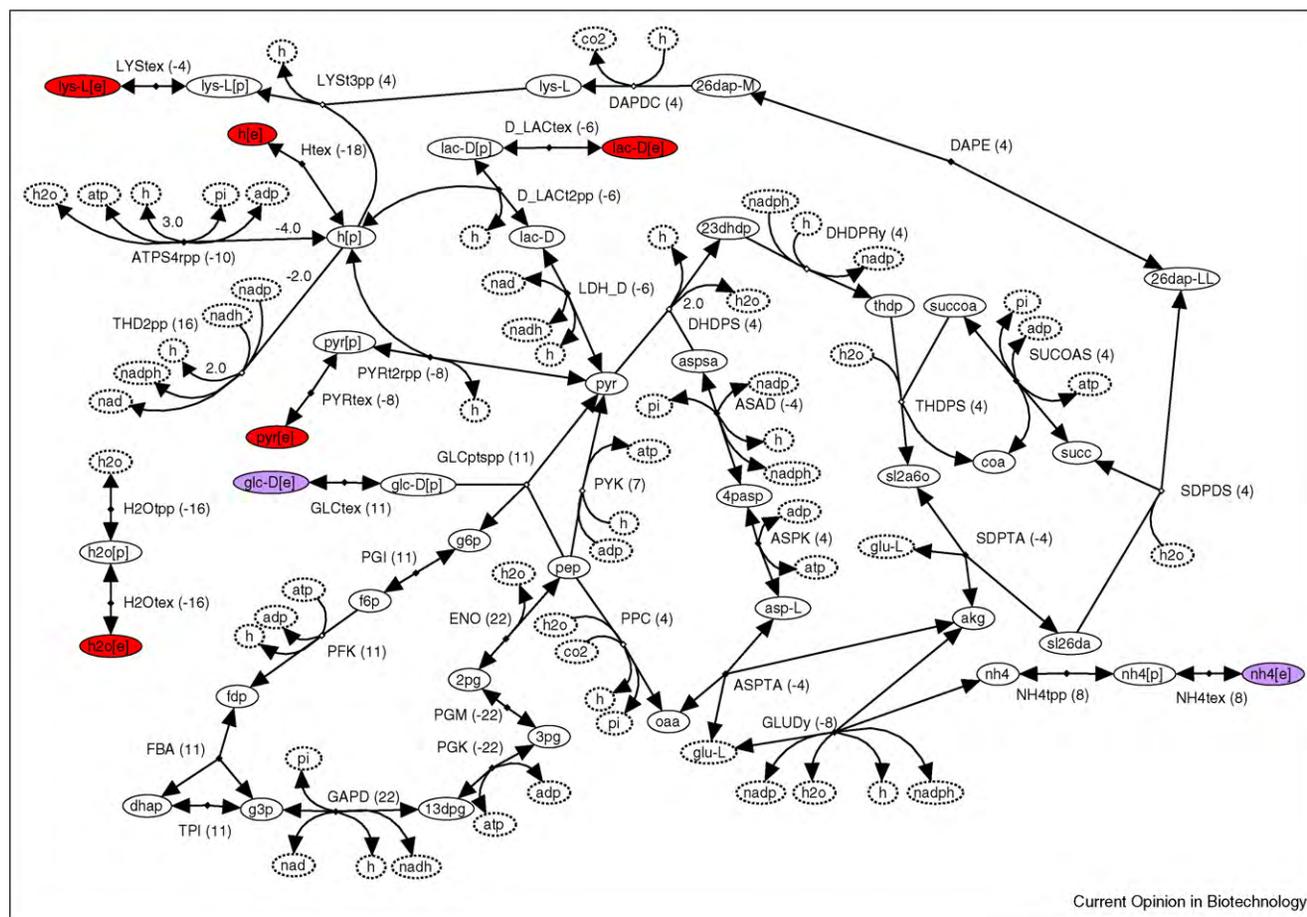
A useful tool in Metabolic Pathway Analysis is based on the concept of EFMs [9^{**}, 10^{**}]. An EFM is a minimal set of enzymes that can operate at steady-state such that all irreversible reactions involved proceed in the thermodynamically favored direction (Box 1). The related concept

of minimal T-invariants had been established earlier in Petri net theory (cf. [50]).

Maximal yields can be computed by EFMs or by FBA. The latter methodology is particularly suitable in large-scale networks, in which EFM analysis meets the problem of combinatorial explosion. In small-sized and moderate-sized networks, in contrast, the set of EFMs provides a more comprehensive overview of the network's metabolic capabilities because it also comprises suboptimal pathways and pathways optimal with respect to other

4 Systems biology

Figure 2



The shortest EFM producing lysine from glucose in *Escherichia coli* [53**]. The calculation was performed in the genome-scale network from [31**]. Violet nodes belong to the *in silico* growth medium, red nodes denote external metabolites. Duplicate nodes are dashed. Values in parentheses indicate reaction fluxes. For abbreviations, see [31**].

substrate-product pairs. Moreover, knockouts can easily be assessed by considering the remaining subset of EFMs. This sometimes leads to the counter-intuitive result that the average yield increases, as has been confirmed experimentally after the *in silico* analysis [8**].

By EFM analysis, previously unrecognized pathways can be detected (cf. [10**,55*], see also Box 2). Recently, we proposed two methods for pathway prediction in large-scale networks [40**,53**]. We computed the 10 shortest EFMs producing lysine in the genome-scale networks of *E. coli* and *Corynebacterium glutamicum* [53**] (Figure 2). Moreover, EFM analysis allows the quantification of robustness (see below).

There is a growing effort to use network models to identify drug targets and characterize mechanisms of disease. A recent study reconstructed and analyzed the metabolic networks of multiple strains of *Staphylococcus aureus* to identify novel drug targets [56]. A network-based pipeline

Box 2 The EFM method has manifold applications in biotechnology. First, it allows one to compute maximal molar yields (product-to-substrate ratios). For example, a previously undescribed pathway of efficient conversion of carbohydrate to oil in developing green plant seeds was detected [51**]. That pathway involves the pentose-phosphate pathway and the RUBISCO enzyme and provides 20% more acetyl-CoA for fatty acid synthesis than glycolysis. Trinh *et al.* [52*] designed, initially *in silico*, an *E. coli* strain with eight gene KO mutations. By EFM analysis, four pathways with non-growth-associated conversion of pentoses and hexoses into ethanol (important for biofuel production) at maximum theoretical yields and two pathways with tight coupling of growth with ethanol formation at high yields were obtained. Thereafter, they verified in experiment that the ethanol yields of the engineered strains closely matched the theoretical predictions. A third example is the EFM analysis of the synthesis of the commercially important amino acid, lysine (Figure 2) [7,53**,54*]. Depending on the bacterial species and on whether ATP was assumed to be sufficiently available, different maximum lysine-over-glucose yield values have been computed, for example 9:11 in *Corynebacterium glutamicum* when ATP must be regenerated by part of the glucose resource [54*]. There are many more studies in which EFMs were used, see [8**] for a recent review.

for identifying potential antimicrobials is being developed [57]. The human metabolic network reconstruction was analyzed to identify alternative enzyme targets for treating hyperlipidemia [39^{••}]. It has also been recently used to predict biomarker changes characterizing a large set of different genetically inherited metabolic disorders [58].

Using metabolic models to study basic biological questions

Metabolic games

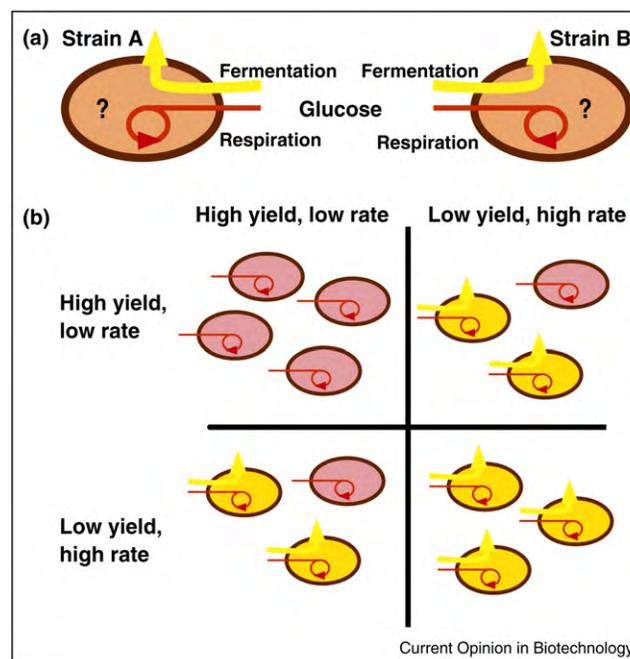
As has been seen above with FBA and EFMs, the concept of optimality has turned out to be extremely useful in understanding biological systems. Traditional optimization is, however, often insufficient for a deeper understanding of evolution. It usually neglects that the properties of the environment can change, and this in turn can change the optimal strategy. This is particularly important if the environment includes coevolving competitors that optimize their own metabolic capabilities. A mathematical framework to describe coevolution is provided by game theory (cf. [59]).

An illustrative example of the importance of competition in pathway evolution is the interplay between fermentation and respiration in ATP production [60^{••}]. Several organisms and cell types such as *S. cerevisiae* and *E. coli* use respiro-fermentation at high glucose levels for degrading glucose to produce ATP, that is, respiration at maximum rate and fermentation in addition (Figure 3a), while many wild yeasts such as *Kluyveromyces marxianus* do fully respire glucose under aerobic conditions [61]. Respiration has a higher ATP-over-glucose yield but a relatively low rate in comparison to fermentation. When two species or strains compete for the same substrate, a typical game-theoretical situation arises (Figure 3b). The fitness of either organism depends not only on its own strategy (pathway usage in this case) but also on that of the other because both strategies affect the common nutrient pool.

To apply game theory requires little kinetic information. Not much more than the constraints (1) and (2) in Box 1 are necessary. In the example of fermentation and respiration, the upper limits given by relation (2) have to be chosen appropriately. In addition, the different yields of the pathways must be considered.

In order to use external glucose as economically as possible, it would be best if all organisms in a given habitat opted for respiration. The evolutionary reason for the profligate utilization of glucose by baker's yeast is that it thus out-competes organisms that operate more economically. In the terminology of game theory (cf. [59]), yeast cells are trapped in a Nash equilibrium (stable solution of a game) of a Prisoner's Dilemma (cf. [62[•]]). In the light of these results, FBA should be critically

Figure 3



Game played by *S. cerevisiae* when growing at high glucose level under aerobic conditions. (a) The strains face the decision as to whether they use respiration (red arrow) or respiro-fermentation (red and yellow arrows). The higher rate of fermentation is indicated by a thicker arrow. (b) Schematic picture of the payoff matrix for this game. The best solution is to use respiration in order to have a higher yield. However, the respiro-fermenters would grow faster.

re-visited [46,47] although it was very successful in many cases [14,15,63^{••}]. If maximum yield were the relevant criterion for the choice of pathway, respiration would always be chosen by *S. cerevisiae*. Fermentation was predicted only by FBA when additional constraints or specially tailored optimization principles were used [64]. Game-theoretical approaches can help predict flux distributions without additional corrections.

Several other metabolic systems have been analyzed by game theory, such as distinct regimes of glycolysis [65] and metabolic strategies in biofilms [66]. *S. cerevisiae* is involved in yet another interesting game. It concerns the extracellular enzyme, invertase, which generates glucose [67[•]]. A cheating strategy is to take up glucose while saving the metabolic costs of production and secretion of invertase. Gore *et al.* [68^{••}] showed by experiments and a mathematical model that a stable coexistence between invertase-secreting and non-secreting yeast cells can be established (for an alternative mathematical model and the biotechnological relevance, see [69]). When the metabolic effort for exoenzyme production is low, all cells cooperate (harmony game); at intermediate costs, cooperators and cheaters coexist (hawk-dove game), while at high costs, all cells use the cheating strategy (Prisoner's Dilemma) [68^{••},69].

6 Systems biology

Several other biochemical examples including photosynthesis have been reviewed in [62[•]]. Moreover, the concept of Shapley value from game theory has been used in quantifying metabolic robustness [70], see below.

Robustness and its evolution

A general feature of living cells is their robustness to varying environmental conditions and genetic mutations. As metabolic network models provide an exciting opportunity to study genotype–phenotype relations on a genome scale, CBM models (and metabolic models in general) have been successfully used to study many facets of this fundamental relationship [71[•],72^{••},73,74]. These studies have mainly asked two basic questions: Firstly, how did genetic robustness evolve? Does it have a direct adaptive value, or is it a consequence of environmental robustness, or perhaps just a side effect of other network properties? Secondly, what network mechanisms underlie the observed robustness—is it primarily due to gene duplications, to alternative metabolic pathways, or related to untested environmental conditions?

Since employing FBA in an exhaustive search of all gene knockout combinations cannot proceed beyond combinations of four knocked out (KO) genes, Deutscher *et al.* [75^{••}] used a probabilistic approach. Thus, gene sets providing mutual functional backup until the depth of eight could be cataloged for *S. cerevisiae*. This has enabled them to characterize the '*k*-robustness' (the depth of backup interactions) of each gene, revealing that almost three quarters of yeast metabolic genes do participate in processes that are essential to growth in a standard laboratory environment, compared with only 13% previously found to be essential using single KOs. Optimization-based procedures for the exhaustive identification of multi-gene backup sets in genome-scale metabolic models have been recently developed [76], revealing new avenues available for redirecting metabolism, and uncovering complex patterns of gene interdependence. On the reverse side, genetic robustness may markedly hamper classical genetic studies using KOs to identify gene functionality, due to backups. Using the concept of Shapley value, Deutscher *et al.* [70] have shown that when assigning gene contributions for individual metabolic functions (such as the production of a given amino acid), the picture arising from single-perturbations is severely lacking and a multiple-perturbations approach turns out to be essential. Metabolic robustness under multiple KOs has also been studied in CBMs of several cell types by developing a robustness measure [73]. That measure is based on the ratio of the number of remaining EFMs after KO and the number of EFMs in the unperturbed situation.

Genetic interactions and network organization

By systematically generating double KOs of nonessential genes and assessing the resulting growth rate (fitness) of the organism, geneticists have traditionally identified

both positive (alleviating) and negative (aggravating) genetic interactions, which has been a traditional tool for discovering functional relationships between genes. A comprehensive experimental screening for this in a whole organism is currently underway for yeast [77^{••}]. Naturally, CBM models offer an opportunity to carry out such screens *in silico*. Segrè *et al.* [78] computed growth phenotypes of all single and double KOs of metabolic genes in *S. cerevisiae*, using FBA. The ensuing genetic interaction network could be clustered into modules composed of genes interacting with each other 'monochromatically' (i.e. with purely aggravating or purely alleviating interactions), emphasizing interactions between, rather than within, functional modules. Harrison *et al.* [79^{••}] investigated the extent to which the functional impact of single and double KOs in yeast changes across different environments, employing FBA across 53 different conditions. The synthetic lethal (SL) predictions of the model were then validated by an *in vivo* double gene KO experiment and by literature search. The strong context dependency of the pattern of SL interactions observed suggests that the environment plays an important role in shaping genetic robustness.

From unicellular to multi-cellular organisms

The vast majority of the work on metabolic CBM performed up until now has focused on unicellular organisms. Naturally, in recent years, there have been attempts at extending these methods to study the metabolism of multicellular and multi-tissue organisms, a considerably greater challenge. CBM reconstructions of human metabolic networks were performed up to 2007 only for cell types and organelles with a very limited scope of metabolism [80,81[•],82^{••}]. A fundamental step forward has been presented in recent reconstructions of the global, generic human metabolic network based on an extensive evaluation of genomic and bibliomic data [38^{••},39^{••}]. These networks include ~3000 reactions, ~2000 metabolites, and ~1500 genes mapped to the different reactions over 7 organelles. The generic model of [39^{••}] helped identify a set of functionally related reactions involving glutathione metabolism that were causally related to hemolytic anemia, and another set of functionally related enzymes containing HMG-CoA reductase, a common target for the cholesterol lowering statins. This model, however, is not tissue or cell specific. More recent efforts have presented methods for inferring context-specific networks [23^{••},24], which can be utilized to infer large-scale descriptions of the human tissues' metabolism. Accordingly, Shlomi *et al.* [23^{••}] have integrated tissue-specific gene and protein expression data to predict and validate versus publicly available data for the tissue-specific metabolic activity for 10 human tissues, identifying that post-transcriptional regulation plays a central role in shaping tissue-specific metabolic activity. Very recently, an extended approach of the latter has been used to build and study the first large-scale model of liver metabolism [83^{••}].

Concluding remarks and future directions/challenges

CBM methods are very useful for understanding the complex architecture of metabolism and for manifold biotechnological and medical applications. Even if kinetic parameters were to become available, an analysis of the network properties using tools of FBA and Metabolic Pathway Analysis often provides valuable insight before performing a dynamic simulation. As outlined above, the integration of omics data of different types into metabolic models has had much success. Nevertheless, its refinement and scaling-up certainly remains a challenge. Cell-specific and tissue-specific studies can now be performed for those multi-cellular organisms for which metabolic reconstructions are available, as is already being done for humans. Whole-cell modeling has also raised philosophical issues on what level of completeness can be reached in modeling.

More work in this field is also needed to study emergent properties, which is at the heart of Systems Biology, after the necessary assembly of the network constituents has been done. Another direction is Synthetic Biology. Specific goals are the design of minimal metabolisms (depending very much on the given set of nutrients) and minimal genomes. This could help design efficient microbes for biosyntheses. Game-theoretical methods, in particular, are helpful in assessing the impact of 'cheater' mutants in bioreactors, which may impair productivity.

Overall, one can safely maintain that the field of genome-scale metabolic modeling has undergone a tremendous development and growth in the last decade, in terms of the organisms spanned, the methodologies developed, and the themes covered. Certainly, if there is one specific field in systems biology where we have made significant strides towards the holy grail of generating a working cell *in silico*, this is the one.

Acknowledgements

Financial support from the following agencies is gratefully acknowledged: German-Israeli Foundation to ER and SS, Israeli Science Foundation (ISF) to ER, National Science Foundation CAREER award (0643548) to JP and German Ministry of Education and Research (BMBF) to SS. Financial support from Portuguese entities: Fundação Calouste Gulbenkian, Fundação para a Ciência e a Tecnologia (FCT) and Siemens SA Portugal (PhD grant SFRH/BD/32961/2006) to LF de F. The authors thank Christoph Kaleta and Steffen Klamt for helpful discussions.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Barnett JAG: **Glucose catabolism in yeast and muscle.** In **Selected Topics in the History of Biochemistry: Personal Recollections IX**, vol 44. Edited by Semenza G, Turner AJ. Elsevier; 2005:1-132.
This paper provides a historical perspective on glycolysis, being the first metabolic pathway formulated.

2. Stephanopoulos GN, Aristidou AA, Nielsen J (Eds): *Metabolic Engineering. Principles and Methodologies*. San Diego: Academic Press; 1998.
3. Heinrich R, Rapoport SM, Rapoport TA: **Metabolic regulation and mathematical models.** *Prog Biophys Mol Biol* 1977, **32**:1-82. An early review on mathematical modeling of metabolism.
4. Heinrich R, Schuster S: *The Regulation of Cellular Systems*. New York: Chapman and Hall; 1996.
5. Palsson BØ: *Systems Biology: Properties of Reconstructed Networks* Cambridge: Cambridge University Press; 2006.
6. Papin J, Stelling J, Price N, Klamt S, Schuster S, Palsson BØ: **Comparison of network-based pathway analysis methods.** *Trends Biotechnol* 2004, **22**:400-405.
7. Schuster S, von Kamp A, Pachkov M: **Understanding the roadmap of metabolism by pathway analysis.** *Methods Mol Biol* 2007, **358**:199-226.
8. Trinh CT, Wlaschin A, Sreenc F: **Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism.** *Appl Microbiol Biotechnol* 2009, **81**:813-826.
Excellent review on elementary-mode analysis focusing on its biotechnological applications.
9. Schuster S, Hilgetag C: **On elementary flux modes in biochemical reaction systems at steady state.** *J Biol Syst* 1994, **2**:165-182.
Pioneering paper, in which elementary flux modes were introduced.
10. Schuster S, Dandekar T, Fell DA: **Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering.** *Trends Biotechnol* 1999, **17**:53-60.
Clear introduction into EFM analysis, with illustrative examples. Moreover, in this paper, a catabolic pathway comprising part of the TCA cycle, glyoxylate shunt and PEP carboxykinase had been predicted for *E. coli*. This prediction was confirmed experimentally later [55*].
11. Schilling CH, Letscher D, Palsson BØ: **Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.** *J Theor Biol* 2000, **203**:229-248.
First paper on extreme pathways, to reduce the number of pathways necessary to describe the metabolic capabilities of a network.
12. Fell DA, Small JR: **Fat synthesis in adipose tissue. An examination of stoichiometric constraints.** *Biochem J* 1986, **238**:781-786.
Pioneering paper on FBA (although that term was not yet used then), including application to lipid metabolism.
13. Varma A, Palsson BØ: **Metabolic capabilities of Escherichia coli. I. Synthesis of biosynthetic precursors and cofactors.** *J Theor Biol* 1993, **165**:477-502.
An early paper in FBA, with mathematical foundations and illustrative examples.
14. Price ND, Reed JL, Palsson BØ: **Genome-scale models of microbial cells evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**:886-897.
15. Lee JM, Gianchandani EP, Papin JA: **Flux balance analysis in the era of metabolomics.** *Brief Bioinform* 2006, **7**:140-150.
16. Çakır T, Patil KR, Önsan Zİ, Ülgen KÖ, Kırdar B, Nielsen J: **Integration of metabolome data with metabolic networks reveals reporter reactions.** *Mol Syst Biol* 2006, **2**:50.
Addresses the important issue of metabolic versus hierarchical control of metabolic fluxes, in an integrative computational and experimental manner.
17. Yizhak K, Benyamini T, Liebermeister W, Ruppín E, Shlomi T: **Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model.** *ISMB 2010 Bioinform* 2010, **26**:i255-i260.
18. Sauer U: **Metabolic networks in motion: ¹³C-based flux analysis.** *Mol Syst Biol* 2006, **2**:62.
Excellent review on flux measurement and analysis.
19. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.

8 Systems biology

20. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ:
 •• **Reconstruction of biochemical networks in microorganisms.** *Nat Rev Microbiol* 2009, **7**:129-143.
 Excellent review on metabolic reconstructions.
21. Oberhardt MA, Palsson BØ, Papin JA: **Applications of genome-scale metabolic reconstructions.** *Mol Syst Biol* 2009, **5**:320.
 Excellent review on metabolic reconstructions, listing the coverage of species.
22. Lewis NE, Cho BK, Knight EM, Palsson BØ: **Gene expression profiling and the use of genome-scale in silico models of *Escherichia coli* for analysis: providing context for content.** *J Bacteriol* 2009, **191**:3437-3444.
23. Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E:
 •• **Network-based prediction of human tissue-specific metabolism.** *Nat Biotechnol* 2008, **26**:1003-1010.
 The first paper presenting a distinct set of specific metabolic descriptions of a variety of human tissues, which are validated via comparison to pertaining tissue-specific data on the activity of disease genes, among others.
24. Becker SA, Palsson BØ: **Context-specific metabolic networks are consistent with experiments.** *PLoS Comput Biol* 2008, **4**:e1000082.
25. Manichaikul A, Ghamsari L, Hom EFY, Lin C, Murray RR,
 •• Chang RL, Balaji S, Hao T, Shen Y, Chavali AK et al.: **Metabolic network analysis integrated with transcript verification for sequenced genomes.** *Nat Methods* 2009, **6**:589-592.
 Influential paper delineating a pipeline for the iterative refinement of the networks with high-throughput transcript verification.
26. Pfeiffer T, Schuster S: **Game-theoretical approaches to studying the evolution of biochemical systems.** *Trends Biochem Sci* 2005, **30**:20-25.
27. Thiele I, Palsson BØ: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nat Protoc* 2010, **5**:93-121.
28. Poolman MG, Bonde BK, Gevorgyan A, Patel HH, Fell DA:
 • **Challenges to be faced in the reconstruction of metabolic networks from public databases.** *IEE Proc Syst Biol* 2006, **153**:379-384.
 Databases are central to the reconstruction of genome-scale networks. This article points out important errors present in databases which render the reconstruction process difficult.
29. Gevorgyan A, Poolman MG, Fell DA: **Detection of stoichiometric inconsistencies in biomolecular models.** *Bioinformatics* 2008, **24**:2245-2251.
30. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**:301-312.
 Extending the concept of 'enzyme subsets', this paper introduces various refined concepts such as 'partially coupled reactions', taking into account irreversibility. This opened a new methodology — flux coupling analysis, and allowed the determination of blocked reactions.
31. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR,
 •• Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ: **A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Mol Syst Biol* 2007, **3**:121.
 Widely used genome-scale metabolic reconstruction of *E. coli*.
32. Duarte NC, Herrgård MJ, Palsson BØ: **Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model.** *Genome Res* 2004, **14**:1298-1309.
 Widely used genome-scale metabolic reconstruction of *S. cerevisiae*.
33. May P, Christian J-O, Kempa S, Walther D: **ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*.** *BMC Genomics* 2009, **10**:209.
34. de Oliveira Dal'Molin CG, Quek L-E, Palfreyman RW, Brumbley SM, Nielsen LK: **AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*.** *Plant Physiol* 2010, **152**:579-589.
35. Poolman MG, Miguet L, Sweetlove LJ, Fell DA: **A genome-scale metabolic model of *Arabidopsis thaliana* and some of its properties.** *Plant Physiol* 2009, **151**:1570-1581.
36. Sheikh K, Förster J, Nielsen LK: **Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*.** *Biotechnol Prog* 2005, **21**:112-121.
37. Selvarasu S, Karimi IA, Ghim GH, Lee D-Y: **Genome-scale modeling and in silico analysis of mouse cell metabolic network.** *Mol Biosyst* 2010, **6**:152-161.
38. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I:
 •• **The Edinburgh human metabolic network reconstruction and its functional analysis.** *Mol Syst Biol* 2007, **3**:135.
 Widely used genome-scale reconstruction of human metabolism.
39. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD,
 •• Srivas R, Palsson BØ: **Global reconstruction of the human metabolic network based on genomic and bibliomic data.** *Proc Natl Acad Sci U S A* 2007, **104**:1777-1782.
 A seminal, pioneering study charting out the first genome-scale, generic model of human metabolism, and showing its potential utility for disease modeling and drug target identification.
40. Kaleta C, de Figueiredo LF, Schuster S: **Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns.** *Genome Res* 2009, **19**:1872-1883.
 The new concept of elementary flux pattern can be used to find routes in subsystems that are compatible with flux distributions in genome-scale metabolic networks and, thus, to detect previously unknown pathways.
41. Reed JL, Famili I, Thiele I, Palsson BØ: **Towards multidimensional genome annotation.** *Nat Rev Genet* 2006, **7**:130-141.
42. Herrgård MJ, Fong SS, Palsson BØ: **Identification of genome-scale metabolic network models using experimentally measured flux profiles.** *PLoS Comput Biol* 2006, **2**:e72.
43. Schreiber F, Dwyer T, Marriott K, Wybrow M: **A generic algorithm for layout of biological networks.** *BMC Bioinformatics* 2009, **10**:375.
44. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP,
 • Nielsen J, Patil KR, Ferreira EC, Rocha M: **OptFlux: an open-source software platform for in silico metabolic engineering.** *BMC Syst Biol* 2010, **4**:45.
 Describes a software that combines various methods for metabolic network analysis. It also gives an overview of other software packages for Systems Biology.
45. Hädicke O, Klamt S: **CASOP: a computational approach for strain optimization aiming at high productivity.** *J Biotechnol* 2010, **147**:88-101.
 New method for designing genetic manipulation strategies, using elementary flux modes to weight the impact of blocking or increasing the flux through a given enzyme.
46. Teusink B, Wiersma A, Jacobs L, Notebaart RA, Smid EJ:
 • **Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation.** *PLoS Comput Biol* 2009, **5**:e1000410.
47. Schuster S, Pfeiffer T, Fell DA: **Is maximization of molar yield in metabolic networks favoured by evolution?** *J Theor Biol* 2008, **252**:497-504.
48. Schuetz R, Kuepfer L, Sauer U: **Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*.** *Mol Syst Biol* 2007, **3**:119.
 Eleven objective functions in FBA are compared, showing that no single objective describes the flux states under all conditions.
49. Boghigian BA, Shi H, Lee K, Pfeifer BA: **Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design.** *BMC Syst Biol* 2010, **4**:49.
50. Grafahrend-Belau E, Schreiber F, Heiner M, Sackmann A, Junker BH, Grunwald S, Speer A, Winder K, Koch I: **Modularization of biochemical networks based on classification of Petri net invariants.** *BMC Bioinform* 2008, **9**:90.
51. Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y: **Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds.** *Nature* 2004, **432**:779-782.
 Excellent application of elementary-mode analysis to plant metabolism, leading to discovery of previously undescribed, highly efficient pathway.

52. Trinh CT, Unrean P, Srienç F: **Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses.** *Appl Environ Microbiol* 2008, **74**:3634-3643.
Convincing application of EFM analysis to a system of high biotechnological relevance, by predicting and verifying the effect of KOs. The design of such streamlined metabolic functionality is interesting in view of Synthetic Biology.
53. de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ: **Computing the shortest elementary flux modes in genome-scale metabolic networks.** *Bioinformatics* 2009, **25**:3158-3165.
The first method computing elementary flux modes in genome-scale networks, with application to lysine synthesis in *E. coli* and *C. glutamicum*.
54. Wittmann C, Becker J: **The L-lysine story: from metabolic pathways to industrial production.** In *Amino Acid Biosynthesis – Pathways, Regulation and Metabolic Engineering*. Edited by Wendisch VF. Springer; 2007:39-70.
Nice overview of pathway analysis of lysine synthesis.
55. Fischer E, Sauer U: **A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*.** *J Biol Chem* 2003, **278**:46446-46451.
Experimental detection of the PEP-glyoxylate cycle predicted theoretically earlier, for example, by [10**].
56. Lee DS, Burd H, Liu J, Almaas E, Wiest O, Barabási A-L, Oltvai ZN, Kapatral V: **Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets.** *J Bacteriol* 2009, **191**:4015-4024.
57. Shen Y, Liu J, Estiu G, Isin B, Ahn Y-Y, Lee D-S, Barabási A-L, Kapatral V, Wiest O, Oltvai ZN: **Blueprint for antimicrobial hit discovery targeting metabolic networks.** *Proc Natl Acad Sci U S A* 2010, **107**:1082-1087.
58. Shlomi T, Cabili MN, Ruppín E: **Predicting metabolic biomarkers of human inborn errors of metabolism.** *Mol Syst Biol* 2009, **5**:263.
59. Gintis H: *Game Theory Evolving*. Princeton: Princeton University Press; 2009.
60. Pfeiffer T, Schuster S, Bonhoeffer S: **Cooperation and competition in the evolution of ATP-producing pathways.** *Science* 2001, **292**:504-507.
A pioneering paper on game-theoretical analysis of metabolic networks.
61. Veiga A, Arrabaça JD, Maria C, Loureiro-Dias MC: **Cyanide-resistant respiration is frequent, but confined to yeasts incapable of aerobic fermentation.** *FEMS Microbiol Lett* 2000, **190**:93-97.
62. Schuster S, Kreft J-U, Schroeter A, Pfeiffer T: **Use of game-theoretical methods in biochemistry and biophysics.** *J Biol Phys* 2008, **34**:1-17.
Overview of numerous applications of game theory.
63. Fong SS, Palsson BØ: **Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes.** *Nat Genet* 2004, **36**:1056-1058.
This paper provided significant experimental evidence for the principle of optimality in metabolic networks as implemented in Flux Balance Analysis.
64. Çakır T, Efe C, Dikicioglu D, Hortacsu A, Kirdar B, Oliver SG: **Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains.** *Biotechnol Prog* 2007, **23**:320-326.
65. Aledo JC, Pérez-Claros JA, Esteban del Valle A: **Switching between cooperation and competition in the use of extracellular glucose.** *J Mol Evol* 2007, **65**:328-339.
66. Kreft JU: **Biofilms promote altruism.** *Microbiology* 2004, **150**:2751-2760.
67. Greig D, Travisano M: **The Prisoner's Dilemma and polymorphism in yeast SUC genes.** *Proc R Soc B* 2004, **271**:S25-S26.
One of the first experimental studies on game-theoretical aspects of metabolism.
68. Gore J, Youk H, van Oudenaarden A: **Snowdrift game dynamics and facultative cheating in yeast.** *Nature* 2009, **459**:253-256.
Excellent combination of experimental and game-theoretical approaches to studying exoenzyme production.
69. Schuster S, Kreft J-U, Brenner N, Wessely F, Theissen G, Ruppín E, Schroeter A: **Cooperation and cheating in microbial exoenzyme production – theoretical analysis for biotechnological applications.** *Biotechnol J* 2010 doi: 10.1002/biot.200900303.
70. Deutscher D, Meilijson I, Schuster S, Ruppín E: **Can single knockouts accurately single out gene functions?** *BMC Systems Biol* 2008, **2**:50.
71. Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J: **Robustness of cellular functions.** *Cell* 2004, **118**:675-685.
Very good review on robustness of intracellular networks.
72. Blank LM, Kuepfer L, Sauer U: **Large-scale ¹³C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast.** *Genome Biol* 2005, **6**:R49.
A combined experimental and computational account aiming to decipher the underlying mechanisms of metabolic network robustness by comparing flux distributions in wild-type and knockout yeast strains.
73. Behre J, Wilhelm T, von Kamp A, Ruppín E, Schuster S: **Structural robustness of metabolic networks with respect to multiple knockouts.** *J Theor Biol* 2008, **252**:433-441.
74. Freilich S, Kreimer A, Borenstein E, Gophna U, Sharan R, Ruppín E: **Decoupling environment-dependent and independent genetic robustness across bacterial species.** *PLoS Comput Biol* 2010, **6**:e1000690.
75. Deutscher D, Meilijson I, Kupiec M, Ruppín E: **Multiple knockout analysis of genetic robustness in the yeast metabolic network.** *Nat Genet* 2006, **38**:993-998.
The first study to chart the backup architecture of genes using very large *in silico* knockout combinations and associate the latter with the genes' conservation and functional specificity.
76. Suthers PF, Zomorodi A, Maranas CD: **Genome-scale gene/reaction essentiality and synthetic lethality analysis.** *Mol Syst Biol* 2009, **5**:301.
77. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S et al.: **The genetic landscape of a cell.** *Science* 2010, **327**:425-431.
Comprehensive experimental study of double KOs and modularity in yeast.
78. Segrè D, DeLuna A, Church GM, Kishony R: **Modular epistasis in yeast metabolism.** *Nat Genet* 2005, **37**:77-83.
79. Harrison R, Papp B, Pál C, Oliver SG, Delneri D: **Plasticity of genetic interactions in metabolic networks of yeast.** *Proc Natl Acad Sci U S A* 2007, **104**:2307-2312.
A combined computational and experimental study showing the large extent of context-dependency of genetic interactions on the organism's environment.
80. Chan C, Berthiaume F, Lee K, Yarmush ML: **Metabolic flux analysis of cultured hepatocytes exposed to plasma.** *Biotechnol Bioeng* 2003, **81**:33-49.
81. Vo TD, Greenberg HJ, Palsson BØ: **Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data.** *J Biol Chem* 2004, **279**:39532-39540.
Detailed reconstruction of the metabolic network in a cell organelle.
82. Thiele I, Price ND, Vo TD, Palsson BØ: **Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet.** *J Biol Chem* 2005, **280**:11683-11695.
One of the first demonstrations of the potential use of CBM models for modeling and studying human disease conditions, employing sampling methods to track the flux profile changes after different perturbations.
83. Jerby L, Shlomi T, Ruppín E: **Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism.** *Mol Syst Biol* 2010:6 doi: 10.1038/msb.2010.56.
Differing from [23**], this study presents a general approach for building tissue models, that can be perturbed to study the metabolic implications of disorders and other context-dependent alterations.
84. Letunic I, Bork P: **Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.** *Bioinformatics* 2007, **23**:127-128.