ISSN Online: 2349-7785

Application of Computer Assisted Bimolecular Interaction Modelling in Predictive Microbiology, Current State and Future Prospects

Saikat Bandyopadhyay [1], Lopamudra Ray [1&2]*

[1] School of Biotechnology, KIIT University, Bhubaneswar, Odisha, India-751024, India
[2] School of Law, KIIT University, Bhubaneswar, Odisha, India-751024, India

Received 10 January 2020; accepted in revised form 27 February 2020

Abstract: With the rapid evolution of Biotechnology and Green Technology, Microbial synthesis of commercially relevant biological compounds has been emerging for the past few decades. These economically valuable bio-compounds that have numerous applications to food, agriculture, chemical, and pharmaceutical industries. These low yield high-value products include several antibacterial and anticancer drugs, organic acids, amino acids, vitamins, industrial chemicals, and even biofuels. Biological synthesis of these extremely complex products often employs complex biochemical pathways performed under controlled culture environments, exploiting live cells; Often followed by appropriate downstream extraction and purification unit operations depending upon the nature and type of the product. In the last few decades, the latest innovations enabling constant improvement of nucleotide sequencing and computational methods for downstream analysis of sequence data have attracted a great number of biologists, mathematicians, and programmers across the world, in form of a tool potentially capable of drawing important biological conclusions. The introduction of these novel approaches has greatly contributed to the abundance of publicly available sequence data and analytical algorithms through a plentiful and ever-increasing number of studies turning towards the big data approaches. The inception of 'Multiomics' gave birth to *in silco* methods for identification of vital bimolecular interactions accountable for major phenotypes including those which account for, biosynthesis of several expensive bio-compounds. Most unpretentious application of these constraint-based models is to prioritize target pathways for "knock-in" or "knock out" approaches, and also to identify important pathways that are to be built into industrially relevant production organism(s) in synthetic biology. However, a more fanciful application of this mechanistic predictions could be drawing an inference about the environment in which the organism is living or was grown to express a certain physiological state. However, multiple major questions need to be addressed before one starts predicting optimal culture conditions using Omics level information. Here we focus on the major technological and scientific concepts that make up the core of, major scientific questions that need to be answered to improve the predictive power of such technologies, future prospects, and challenges associated with such integrative technology and their potential effect on the global economy.

Introduction

Before the insurgence of genomics, strain development, strain improvement and optimization of culture condition variables were generally sustained through random chemical mutagenesis followed by resource-intensive screening, followed by selection procedures. Later, in the 1980s, the introduction of Recombinant DNA technology gave birth to ''Industrial Biotechnology'. In these early days of Genetic engineering methods for sustainable production of antibiotics (Penicillin from *Penicillium chrysogenum*; annual market size surpassing US$ 1.5 billion), vitamins (L-ascorbic acid production through the 'Reichstein

*Corresponding author (Lopamudra Ray)
E-mail: < lray@kiitbiotech.ac.in >

process' and biocatalysis employing *Gluconobacter oxydans*; annual market size exceeding US$ 600 million), organic acids (citric acid production by *Aspergillus* sp.; annual market size exceeding US$ 1.5 billion), and amino acids (L-glutamate and L-lysine production by *Corynebacterium glutamicum*; annual production exceeding 600,000 tons) were the primary focus [1]. Since the beginning of industrial microbiology, Yeast (*Saccharomyces cerevisiae*) was the most commonly used microbial cell factory for industrial applications and many of the modern methods are equally inclined towards using yeast, as it is well studied under several controlled environments. Over the last few decades, Yeast has been used to produce alcohol, yoghurt, Insulin, Bioethanol, Vitamins and even Amino acid. Thus yeast-derived product is generally accepted easily in the global market.

In the earliest days Industrial Biotechnology, process designing was solely dependent on the concepts of bioprocess engineering to enhance production-yield and robustness. However, optimization of the process variables for attaining commercial feasibility consumed a large amount of time and resources affecting the overall cost of the product. The mechanistic association between genotype, physiological traits and environmental parameters were often beyond the scope of these methods. In the meantime, vast applications of these classical techniques, improvement of the tools in 'Molecular Biology' especially DNA sequencing and Gene editing paved the path for relatively newer concepts and technologies for site-directed mutagenesis and delivered biologist the scopes of metabolic engineering by direct alteration of the gene sequence or gene-modules. These methods were largely associated with the production of, 1,3-propanediol [2], Isobutanol [3], Succinic [4], 1,4-butanediol [5], Artemisinin [6], and Omega-3 Eicosapentaenoic acid [7]. Despite all of these successful examples, efficient metabolic engineering was still limited by the complexity of cellular metabolism and the associated regulatory networks. Moreover, the synthesis of exogenous and non-natural compounds requires heterologous expression of novel genes and pathways [8] which often lead to metabolic competition and metabolic imbalance and possibly [9]. An aspect of substrate utilization diversity often also needs to be considered for the commercial utilization of cell factories.

These challenges drove biologists towards the need for holistic approaches for acquiring knowledge about the mechanistic principles that govern the physiological properties, that laid the foundation of Systems biology and Synthetic biology [10]. Contrary to the conventional methods, System biology deals with the totality of a system and focuses on elucidating the interactions between each constitutive components to predict cellular behaviour. These methods are often driven by the High-Throughput technologies and computational methods [11,12]. With the rapid decrease in the cost of 'per-base' sequence, High throughput techniques (i.e. genomics, epigenomics and transcriptomics) has promoted mathematical modelling to help biological science attain an unprecedented mechanistic resolution and identification of causal associations [13]. Since then, the evolution of the 'Industrial systems biology' has continuously fuelled big data-driven integration of experimental and computational methods. These modern methods aim to predict growth kinetics and the phenotypical behaviours of microorganisms by mapping genotype to the protein-protein interaction (PPI) network. A directed graph is generated where each node in the graph represent proteins and edges represent interactions between the proteins and small molecules identifies the EC numbers and metabolic reactions from relevant databases, formally known as Genome-Scale Metabolic model. These models are peculiarly valuable for decoding mechanisms underlying improved phenotypes in strains derived through mutagenesis and screening or adaptive laboratory evolution [14]. As these methods allow integration of multiple High Throughput Omics (HTO) techniques [15] that enables prediction of biological events. Synthetic biology is fundamentally driven by these predictive algorithms. With the arrival of Synthetic Biology, biologists today have also gained the privilege of reconfiguring the existing operational systems. With the ever-increasing number of concrete evidences where methods include Big data-driven engineering, minimize the resource-consuming experiment-driven hypothesis generation. The unprecedented

predictive power promotes designing of interactions between macromolecules, signalling in the regulatory networks, and the metabolic networks. However, these advanced techniques [16] are often capable of producing unexpected outcomes to be studied through experimental methods. Hence, Synthetic biology has its own excitement. Never the less, growing ethical issues and the global governance of synthetic biology is a growing concern. With the advancement of synthetic biology, the need for establishing accountability external to the system along with the fragmentation of social authority is also a pressing issue [17].

A more challenging application of these predictive models could also be a prediction of optimal growth condition for an organism in question. One could easily predict the optimal growth condition cell factories with specific genetic capabilities, therefore reducing the number of experimental evaluation. These predictive methods could also complement Game theory model and Lotka Volterra model for community analysis. Current microbial community modelling often fails to establish a comprehensive correlation between genetics and community features, this issue can potentially be addressed through the integration of Multiomics. These Big data-driven methods are also likely to be useful where novel wild-type organisms are to be used in a commercial process (Bio-fertilizer industry/ Bioremediation/synthetic community designing etc.). However, several questions need to be addressed before we develop efficient mathematical models for such predictions.

Through this review, we aim to outline major concepts and current application of Multiomics driven metabolic models in 'Industrial Systems Biology'. We also aim to draw attention towards a few other potential areas of microbiological research where system-wide metabolic modelling could significantly improve current understanding. Finally, we move on to point out how Predictive microbiology could impact the global economy.

Prediction of biological capabilities using genome-scale models of metabolism (GEM)

Prediction of biological capabilities of an organism in the bottom-up approach of systems biology is vastly dependent upon the reconstruction of the 'Reactome'. A Reactome is an assembly of the biochemical reactions that are supported by the organism's genetic content. This assembly in principle is very similar to the whole-genome assembly from Shotgun sequencing data, however, in a typical Reactome reconstruction different levels of biological information are incorporated in a directed graph to visualize and map each reaction. These reactions infer to biological capabilities. There are several tools (i.e. KEGG, MetaCyc, REAVEN Toolbox etc.) which aim to automate the reconstruction process. Earlier reconstruction was mostly dependent upon manual curation, a challenging task that requires advanced knowledge of computational techniques along with expertise in biological systems to identify interacting compounds, reactions acting on each compound and finally the protein (enzymes) that catalyses these reactions along with their corresponding open reading frames. To avoid the manual curation modern automatic reconstruction methods often utilize a globally curated database as a template for the reconstruction. Finally, both the automate or manual reconstruction generates a directed graph representing the bimolecular interaction network. This network facilitates mathematical modelling and thus enable computational analysis to predict the mechanistic basis of biological capabilities. In the automated process, Shotgun Whole Genome Sequencing (WGS) usually serves as the starting point for Genome-Scale metabolic reconstruction. Next, the genes present in the organism (system) are easily be predicted using available gene prediction algorithms. These gene prediction methods often employ sequence similarly search, GC-content matching using the available databases. A few machine learning-based method also exits alongside the conventional approaches. However, all of these gene prediction algorithms produce output of the Enzyme Commission Numbers. The biochemical/kinetic information about these reactions and associated bio-compounds are then established using the biochemical knowledge base (otherwise known as the 'Bibliome'). With the increasing popularity of genomics, GEMs has grown to enable biologi-

Serial no.	Biological questions to be answered for thereconstruction of a Reactome.
1	What are the substrates and products involved in a particular reaction?
2	How to quantify the stoichiometric coefficients for metabolites that participates in a reaction(s)?
3	Are the reactions in question are reversible or not?
4	In what cellular compartment does a particular reaction occur?
5	What are the key gene(s) that are required for a particular reaction to occur, and how to identify their genomic locations?

cal science to attain predictive attribute where Constraint-based metabolic reconstruction proved to be a major force. Principally, the reconstruction process fundamentally treats each reaction as the basic elements of the network and then aims to answer basic biological questions about each reaction by combining information from multiple sources (i.e. databases and primary literature). These associations articulate the gene-protein-reaction relationship (GPR) and organize reactions in different compartments of the cell in a format of organised subsystems [18]. Nevertheless, before computing the network properties from these topological models, a crucial step is the conversion of topology into a mathematically quantifiable format.

Mathematical representation of GEM(s) enables computational prediction of physiological states: The mathematical representation of GEM is a tabular format commonly known as the 'Stoichiometric Matrix'. The paradigm of Fluxomics is based on the fundamental aspects of stoichiometry. Fluxomics studies flux maps/flux-distributions (imposition of systemic constraints on the possible flow patterns of metabolites through a metabolic network). Knowledge-driven imposition of these constraints upon the metabolic network (COBRA approach), primarily differentiates this Genomics-driven approach from the biophysical approach driven by measurement of kinetics in the system. The mathematical constraints are equations represent balances or imposed restrictions. The constraints are often based on fundamental biological knowledge. The stoichiometric matrix, on the other hand, imposes constraints on the flux-distribution. This matrix ensures the steady-state equilibrium of the system, where the aggregate amount of any compound being produced is equal to the total amount being consumed. The flux distribution of every reaction in the matrix may also have an upper and lower bound that define the maximum and minimum allowable fluxes, which consecutively links the turnover number (abundance) of enzymes to the constructed Reactome. All these factors interactively define a space of allowable flux distributions that reflects rates of consumption or production for each metabolite. The flux vector aims to quantify the state of the network to infer the physiological property like uptake rates and secretion rates [19]. Computed network states that are coherent across all the enforced constraints are isolated as candidate physiological state. These candidate physiological states serve as the basis of constraint-based predictions.

The oldest known COBRA method, the Flux Balance Analysis (FBA) is an approach that predicts the flow of metabolites in the Reactome [20]. In a typical FBA, several different network states are possible under the given constraints, that satisfy the governing equations. The solutions are originated using linear programming. However, these methods are only useful if the optimal solution lies at the boundaries of the solution space impinging up against governing constraints. The function of FBA utilizes the nutrient availability data to precondition the output metabolite and enables *in silico* tracing of balanced paths across the directed graph of the Reactome. The traced reactions are often correlated with objective functions that describes the removal of the target metabolite from the network, never the less this approach is one of the many approaches of all the molecules flowing and interacting through the reactome.

While simulating bacterial metabolism, GEMs account for both the environmental and genetic parameters. The environmental parameters can be

altered by introducing changes in the growth media. On the other hand, as GEMs are essentially assembled bibliomic data, thus any number of genes and reactions can easily be removed from the Reactome, which in turn makes GEM(s) favourable for identification of knock-out/knock-in candidate reactions/genes. Together, the assumption of steady-state for the internal metabolite, the stoichiometry, and reversibility of each reaction, allow articulation of a region for the allowed flux distributions. The steady-state assumption and stoichiometric model-based quantitation and prediction of growth phenotype and has also been applied to multiple facets of industrial synthetic biology. Famous applications even include- (1) estimation of the optimal state for growth under different cultivation conditions [21], (2) maximization of ATP or NADH production [22], (3) Optimization and evaluation of industrial processes for the production of high-value target metabolites [23], (4) prediction of global quantitative relationships between the input rates of nutrients, the output rates of by-products and growth rates [24].

Computational frameworks for analysis and functional prediction with GEM

Although FBA by far is the most popular method for GEM analysis and predictions, several other sophisticated computational frameworks exits that aim to identify candidate genes or pathways for metabolic engineering using genome-scale models. These methods aim to explore the metabolic potential of a given microbial cell factory. Implementation of these algorithms are often achieved using (1) Linear programming; (2) Quadratic programming; (3) Mixed Integer linear programming; and (4) Evolutionary programming.

FBA; Flux Variability Analysis (FVA); and Flux Coupling Analysis (FCA) are all based on linear programming. FVA ranks the possibilities of variation in each reaction rate when the environmental factors are altered. On the other hand, FCA utilizes thermodynamic modelling that elucidates the correlation between different reaction rates to enable FBA with molecular crowding. A quadratic programming (QP) based method for Minimization of Metabolic Adjustment (MOMA) identifies unique flux distributions that are the closest to the observed flux distribution in a wild-type strain. MOMA assumes that metabolic operations in a knockout mutant or engineered cell are very similar to the wild-type strain and for this reason, MOMA often outperforms the conventional FBA [25].

Metabolic engineering often also includes the optimization of the process to attain sustainability. A widespread computational framework to achieve optimization of the product yields often aim to integrate the production of the craved product to the growth kinetics. OptKnock, a bi-level optimization framework aims to identify optimal gene knockout strategies to attain phenotypes with higher production capability for the desired metabolite [26,27]. Other similar algorithms such as OptReg [28], OptStrain [29], OMNI (optimal metabolic network identification) [30,31], and regulatory on/off minimization (ROOM) [32] has increased the predictive capacity of metabolic modelling by quite a fold. OptReg combines glucose uptake rate, minimum ATP production and 13C experimental flux data and to determine flux distributions in the wild-type strain and finally uses a bi-level optimization algorithm that finds deletion targets. This method can also be used for successful identification of over-expression and down-regulation targets. OptStrain uses a collection of databases to guide the addition of non-native reactions to the wild-type strain that, maximize the yield of the desired product. On the other OMNI identifies the bottleneck reactions that cause the discrepancies between the experimental and measured fluxes. Finally, MOMA and ROOM aims to determine putative flux distributions after gene deletions by minimizing the number of significant flux changes [33,34].

The evolutionary algorithm OptGene identifies deletion targets in microorganisms at lower computational costs by limiting the number of solutions to be found using simulated annealing (SA) [35] and Set-based Evolutionary Algorithms (SEA) [36]. Nevertheless, each of the mentioned algorithms has its own sets of advantages as well as limitations. Thus, choosing the right tool often depends upon for specific application. Detailed description of the constraint-based prediction alongside

the relevant methods can be found in the review by Park *et al*. [37].

Current applications of GEMs

With the recent advent of computational biology, manually curated high-quality GEM(s) of many model organisms and other well-known organisms are already complete. These reconstructions are often publically available and often serve as a scaffold for automated recognition. However, construction of GEMS for novel organisms are often limited by the lack of relevant experimental data and particularly absenteeism of large-scale physiological and omics data. High quality published GEMs of *E. coli* and *S. cerevisae*, have already been utilized in a broad range of applications. Over the last decade including the production of small molecules, metabolites, antibiotics, and even bioethanol. A major study in 2016 involving on *E. coli* GEM (iJO1366) [38], predicted the biological capability of *E. coli* for industrial production of 279 non-native chemicals. On the other hand, the recent interest in GEMs of well-known specific chemical producers *Clostridium aceto-butylicum* and *Streptomyces coelicolor* has generated several high quality qurated GEM(s) for industrial application. The updated GEM of *C. acetobutylicum* (iCac967) [39], Streptomyces and other actinomycetes strains possess large industrial interest due to their capabilities of secondary metabolites production, which resulted in a comprehensive *S. coelicolor* A3(2) GEM, and the algorithm FSEOF (flux scanning based on enforced objective flux) that was successfully used for commercial to overproduce actinorhodin. Other manually curated GEM of actinomycetes include GEM of *Actinoplanes* sp. SE50/110 [40], *Salinispora tropica* [41] and *Saccharopolyspora spinosa* [42]. These GEMs are particularly popular for their application in the antibiotic industry. Available GEMs of *Chlamydomonas reinhardtii* [43] and *Phaeodactylum tricornutum* [44], and *Synechococcus* sp. PCC 207002 [45] and *Synecho-cystis* sp. PCC 6803 [46] is a point of major recent for elucidating electron flows during photosynthesis. These GEMs have contributed greatly towards several biological discoveries allowing mechanistic knowledge base [47].

Community-wide metabolic modelling and functional associations

With rapid developments in both metabolic modelling resources and sequencing technologies modelling of microbial communities that involve either natural [48] or artificial metabolic interaction [49] has rapidly gained massive popularity in applied microbiology. Automatic metabolic modelling tools have largely promoted reconstruction of the GEMs for pan-genome analyses, which in turn paved the path for modelling cross-feeding interactions arising in a complex microbial community (i.e. human gut microbiome, marine microbiome, earth microbiome etc). Sophisticated computational frameworks that aim to articulate the dynamic interspecies interactions in the complex microbial communities. These methods aim to quantify levels of metabolites exchange within the community as a function of biomass.

A popular resource for community-wide metabolic modelling?

The CASINO (Community and Systems-level Interactive Optimization) was recently used to characterize interactions between gut microbiota and the effects of human diets on cross-feeding interactions [50]. A similar multi-level objective function optimization method, d-OptCom additionally includes kinetic parameters for identifying nutrient uptake. This method predicts concentrations of community members', biomass and concentration of shared metabolites over time. This method was used to model metabolic reactions in a uranium-reducing community consisting of *Geobacter sulfurreducens*, *Rhodoferax ferrireducens*, and *Shewanella oneidensis* [51]. These two methods serve as the basis for Dynamic FBA. Over the time many similar methods have been developed, nevertheless, modern methods introduce additional (differential) equations to optimize model parameters. Optimized model parameters often enhance the predictive capacity when studding specific aspects of microbial community. Additionally, FVA & FCA is often used to predict co-culturing dependent exclusive secretion events. A novel method, COMETS (Computation of Microbial Ecosystems in Time and Space) employs additional diffusion equations

along with the Dynamic FBA. The diffusion equations enable conceptualization of species ratios and spatial configuration of the community members. Finally, a similar method MCM (Microbial Community Modeler) predicts the dynamics of relative cell densities and pathway activities of community members. In spite of this massive development in computational methods, efficient prediction through these algorithms is limited to smaller communities consisting of relatively well-studied community members. To elaborate on the reconstruction and use of community-wide metabolic interactions, integrated omics data holds great potential towards a better understanding of metabolic interactions among more diverse community members. Gene-identification driven implementation of evolutionary game theory and Lotka Volterra model assists to understand the evolutionary rise of metabolic interdependencies and functional evolution in microbial communities [52].

Challenges and future prospects

While the community-scale metabolic modelling continues to gain massive popularity, several major computational and experimental challenges need to be overcome to attain accuracy in predicting qualitative and quantitative behaviours of a microbial community. The primary challenge in this regard is a lack of experimental inter- and interspecies flux measurements, secondly, measurement of analogous co-culture flux measurements is often quite difficult to acquire. Also, the scope of studding extracellular flux of individual strain and cross-feeding reactions associated with a community member is often limited, as metabolites are often simultaneously produced and consumed by multiple community members. With the recent advances in carbon-13 labelling, modern experimental methods have been able to resolve intracellular fluxes in two-species communities [53]. This technique can be improved to elucidate flux distributions among genetically diverse community members.

Another major challenge in this area is building metabolic models from metagenomics DNA sequence, as current methods of community profiling lack species-level resolution, thus metabolic genes predicted through the shotgun metagenomics approach lacks details on which community members these genes belong to, making compartmentalization of these metabolic reactions specifically difficult. Also, the transporter mechanisms predicted based on sequence information, fail to resolve, which specific metabolites are being taken up or excreted by these transporters. Therefore, improved transporter annotation and their data-based characterization will potentially amend the anticipations of nutrient uptake, product secretion and metabolite exchange in microbial communities.

The spatial configuration of community members is often beyond the scope of current microbiome modelling; however spatial chemical gradients are a common feature of microbial communities where agitation is absent. Moreover, the cellular behaviour depends largely on chemical concentration in the local environment. Thus, future microbiome models should also include accesses for predicting concentration gradients in response to flow, diffusion, and microbial metabolism.

A major application of the constraint-based modelling of bio-chemical interactions is to study a various range of organisms and microbial communities, including synthetic and natural communities associated with the ocean, marine, and human environments. These models can be further used for designing efficient microbial consortium to be used in bio-fertilizer and bioremediation. Mathematical constrains can also be used to identify underlying mechanisms behind antibiotic resistance and pathogenicity.

Today large scale studies of flux distribution also include the integration of proteomic/ transcriptomic data into the metabolic network using parsimonious flux balance analysis (pFBA) and most recently Linear Bound Flux Balance Analysis (LBFBA) [54] also transcriptional regulated flux balance analysis (TRFBA) [55] algorithm enables integration of transcriptional regulatory and metabolic models using a set of expression data. This method introduces two new linear constraints by considering the expression levels of genes as a new continuous variable. The first constrain limits the rate of a particular reaction and

the second constrain correlates expression level of each target gene to its regulatory genes.

Conclusion

The modern era of predictive microbiology is based upon the assumption that the reactions of populations of microorganisms to environmental factors are consistent. Topological modelling and systems biology offer great potential for the establishment of a comprehensive agreement of microorganisms, their interaction with other species in a community. The flow of information in the community drives the community-wide dynamic population behaviours. Studding the emergent patterns of information flow and accurate functional association is a fundamental question in microbiology that can be addressed through Big data-driven multi-omic modelling as interactions between different microorganisms regulate their metabolism, thus empowering identification of the most important biomolecular interactions along with the associated microbial guilds.

The chemical factors of biological processes include carbon source availability, nutrient availability, electron donor/acceptor availability, pH, and chemical stressors and the physical factors enforced by the micro/macro geography include, humidity, conductivity, temperature, pressure, diffusion, and texture and density of the extracellular matrix. As this complication increases, there is also an increased need for developing a new set of fundamental principles, concepts and algorithms that will further reveal the secrets of microbial and cellular communities [56].

The substantial advantage of these metabolic models is that they allow seamless incorporation of the data generated with the new High Throughput Omics (HTO) techniques where multi-dimensional data integration can give rise to a more accurate and naturalistic quantification of microbial processes. Hence, elevated metabolic models can assist as a bridge between molecular/biochemical research and environmental engineering practices, effectively functioning as a tool that can better link the work of microbiologists and engineers.

In the near future, modelling approaches will help to decrypt patterns of compounds and energy flow in environmental systems. This unprecedented predictive capability must be applied for the sustainable and integral development of human socio-economic activities amongst nature.

References

1. **Gavrilescu, M. and Chisti, Y. (2005).** Biotechnology-a sustainable alternative for chemical industry. Biotechnology Advances. 23(7-8): 471-499.
2. **Nakamura, C.E., Whited, G.M. (2003).** Metabolic engineering for the microbial production of 1,3-propanediol. Curr. Opin. Biotechnol. 14: 454-459.
3. **Feldman, R.M.R., Gunawardena, U., Urano, J., Meinhold, P., Aristidou, A., Dundon, C.A., Smith, C. (2013).** Yeast organism producing isobutanol at a high yield. Edited by: Google Patents.
4. **Zhu, X., Tan, Z., Xu, H., Chen, J., Tang, J., Zhang, X. (2014).** Metabolic evolution of two reducing equivalent-conserving pathways for high-yield succinate production in *Escherichia coli*. Metab Eng. 24: 87-96.
5. **Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R. et al. (2011).** Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat. Chem. Biol. 7: 445-452.
6. **Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D. et al. (2013).** High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 496: 528-532.
7. **Xue, Z., Sharpe, P.L., Hong, S.P., Yadav, N.S., Xie, D., Short, D.R., Damude, H.G., Rupert, R.A., Seip, J.E., Wang, J. et al. (2013).** Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. Nat. Biotechnol. 31: 734-740.
8. **Lee, J.W., Na, D., Park, J.M., Lee, J., Choi, S., Lee, S.Y. (2012).** Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat. Chem. Biol. 8: 536-546.

9. **Biggs, B.W., De Paepe, B., Santos, C.N.S., De Mey, M., Ajikumar, P.K. (2014).** Multivariate modular metabolic engineering for pathway and strain optimization. Curr. Opin. Biotechnol. 29: 156-162.
10. **Otero, J.M., Nielsen, J. (2010).** Industrial systems biology. Biotechnol. Bioeng. 105: 439-460.
11. **Kitano, H. (2002).** Systems biology: a brief over-view. Science. 295: 1662-1664.
12. **Dias, O., Rocha, I. (2015).** Systems biology in fungi. In: Paterson R (ed) Mol. Biol. Food water borne mycotoxigenic mycotic fungi. CRC Press, Boca Raton, FL, pp 69-92.
13. **Nielsen, J., Jewett, M.C., (2008).** Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. FEMS Yeast Res. 8: 122-131.
14. **Hong, K.K., Vongsangnak, W., Vemuri, G.N., Nielsen, J. (2011).** Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. Proc. Natl. Acad. Sci. U S A, 108: 12179-12184.
15. **Caspeta, L., Chen, Y., Ghiaci, P., Feizi, A., Buskov, S., Hallstrom, B.M., Petranovic, D., Nielsen, J. (2014).** Altered sterol composition. renders yeast thermotolerant. Science. 346: 75-78.
16. **Heinemann, M., Panke, S. (2006).** Synthetic biology-putting engineering into biology. Bioinformatics. 22: 2790-2799.
17. **Zhang, J.Y., Marris, C., Rose, N. (2011).** BIOS working paper no: 4.
18. **Price, N.D., Papin, J. A., Schilling, C.H. and Palsson, B.O. (2003).** Genome-scale microbial in silico models: the constraints-based approach. Trends in Biotechnology. 21(4): 162-169.
19. **Reed, J.L. (2012).** Shrinking the Metabolic Solution Space Using Experimental Datasets. PLoS Comput. Biol. 8(8): e1002662.
20. **Orth, J.D., Thiele., I., Palsson, B.O. (2010).** What is flux balance analysis? Nat. Biotechnol. 28: 245-248.
21. **Karthik, R., Nagasuma, C. (2009).** Flux balance analysis of biological systems: applications and challenges, Briefings in Bioinformatics. 10(4): 435-449.
22. **Ramakrishna., R., Edwards., J.S., McCulloch., A., Palsson, B.O. (2001).** Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am. J. Physiol. Reg. I, 280: R695-R704.
23. **Lee., D.S. Burd., H., Liu., J.X., Almaas., E., Wiest., O., Barabasi., A.L., Oltvai., Z.N., Kapatral, V. (2009). C**omparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. J. Bacteriol. 191: 4015-4024.
24. **Cakir., T., Efe., C., Dikicioglu., D., Hortacsu., A., Kirdar., B., Oliver, S.G. (2007).** Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. Biotechnol. Prog. 23: 320-326.
25. **Rodrigues, J.F.M., Wagner, A. (2009).** Evolutionary plasticity and innovations in complex metabolic reaction networks. PLoS Computational Biology. 5(12): e1000613.
26. **Pharkya., P., Maranas, C.D. (2006).** An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab. Eng. 8: 1-13.
27. **Fischer., E., Zamboni., N., Sauer, U. (2004).** High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived C-13 constraints. Anal. Biochem. 325: 308-316.
28. **Pharkya., P., Burgard., A.P., Maranas, C.D. (2004).** OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14: 2367-2376.
29. **Patil., K.R., Rocha., I., Forster., J. Nielsen, J. (2005).** Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinform. 6: 12.
30. **Herrgard., M.J., Fong., S.S., Palsson, B.O. (2006).** Identification of genome-scale metabolic

network models using experimentally measured flux profiles. Plos Comput. Biol. 2: 676-686.

31. **Fleming., R.M.T. Thiele., I., Nasheuer, H.P. (2009).** Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. Biophys. Chem. 145: 47-56.
32. **Shlomi., T., Berkman., O., Ruppin, E., (2005).** Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. USA. 102: 7695-7700.
33. **Nakahigashi, K. et al., (2009).** Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol. Syst. Biol. 5: 14.
34. **Zhao., Q.Y., Kurata, H. (2009).** Genetic modification of flux for flux prediction of mutants. Bioinformatics. 25: 1702-1708.
35. **Rocha., M., Maia., P., Mendes., R., Pinto., J.P., Ferreira., E.C., Nielsen., J., Patil., K.R., Rocha, I. (2008).** Natural computation meta-heuristics for the in silico optimization of microbial strains. BMC Bioinform. 9: 16.
36. **Gonzalez., O.R., Kuper., C., Jung., K., Naval., P.C., Mendoza, E. (2007).** Parameter estimation using Simulated Annealing for S-system models of biochemical networks. Bioinformatics. 23: 480-486.
37. **Park., J.M., Kim., T.Y., Lee, S.Y. (2009).** Constraints-based genome-scale metabolic simulation for systems metabolic engineering Biotechnol. Adv. 27: 979-988.
38. **Zhang, X., Tervo, C.J., Reed, J.L. (2016).** Metabolic assessment of E. coli as a Biofactory for commercial products. Metab Eng. 35: 64-74.
39. **Yoo, M., Bestel-Corre, G., Croux, C., Riviere, A., Meynial-Salles, I., Soucaille, P. (2015).** A quantitative system-scale characterization of the metabolism of Clostridium acetobutylicum. MBio, 6: e01808-01815.
40. **Wang, Y., Xu, N., Ye, C., Liu, L., Shi, Z., Wu, J. (2015).** Reconstruction and in silico analysis of an Actinoplanes sp.SE50/110 genome-scale metabolic model for acarbose production. Front Microbiol. 6: 632.
41. **Contador, CA., Rodriguez, V., Andrews, B.A., Asenjo, J.A. (2015).** Genome-scale reconstruction of *Salinispora tropica* CNB-440 metabolism to study strain-specific adaptation. Antonie Van Leeuwenhoek. 108: 1075-1090.
42. **Wang, X., Zhang, C., Wang, M., Lu, W. (2014).** Genome-scale metabolic network reconstruction of *Saccharopolyspora spinose* for spinosad production improvement. Microb. Cell. 42 Fact, 13: 41.
43. **Imam, S., Schauble, S., Valenzuela, J., Lopez, Garcia, de Lomana, A., Carter, W., Price, N.D., Baliga, NS. (2015).** A refined genome scale reconstruction of Chlamydomonas metabolism provides a platform for systems-level analyses. Plant J. 84: 1239-1256.
44. **Levering, J., Broddrick, J., Dupont, C.L., Peers, G., Beeri, K., Mayers, J., Gallina, A.A., Allen, A.E., Palsson, B.O., Zengler, K. (2016).** Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PLoS One. 11: e0155038.
45. **Hendry, J.I., Prasannan, C.B., Joshi, A., Dasgupta, S., Wangikar, P.P. (2016).** Metabolic model of Synechococcus sp. PCC 7002: Prediction of flux distribution and network modification for enhanced biofuel production. Bioresour. Technol. 213: 190-197.
46. **Mohammadi, R., Fallah-Mehrabadi, J., Bidkhori, G., Zahiri, J., Javad, Niroomand, M., Masoudi-Nejad, A., (2016).** A systems biology approach to reconcile metabolic network models with application to Synechocystis sp. PCC 6803 for biofuel production. Mol. Biosyst. 23(12): 2552-2561.
47. **Baroukh, C., Munoz-Tamayo, R., Steyer, J.P., Bernard, O. (2015).** A state of the art of metabolic networks of unicellular microalgae and cyanobacteria for biofuel production. Metab Eng. 30: 49-60.

48. **Embree, M., Liu, J.K., A.l-Bassam, M.M., Zengler, K. (2015).** Networks of energetic and metabolic interactions define dynamics in microbial communities. Proc. Natl. Acad. Sci. USA. 112: 15450-15455.
49. **Louca. S., Doebeli, M. (2015).** Calibration and analysis of genome-based models for microbialecology. Elife. 4: e08208.
50. **Shoaie, S., Ghaffari, P., Kovatcheva-Datchary, P., Mardinoglu, A., Sen, P., Pujos-Guillot, E., de, Wouters, T., Juste, C., Rizkalla, S., Chilloux J. *et al.*, (2015).** Quantifying diet-induced metabolic changes of the human gut microbiome. Cell Metab. 22:320-331.
51. **Zomorrodi, A.R., Islam, M.M., Maranas, C.D. (2014).** Dynamic multi-level and multi objective metabolic modeling of microbial communities. ACS Synth. Biol. 3:247-257.
52. **Zomorrodi, A.R., and Segrè, D. (2017).** Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. Nature Communications. doi:10.1038/s41467-017 01407-5.
53. **Gebreselassie, N.A. and Antoniewicz, M.R. (2015).** 13 C-metabolic flux analysis of co-cultures: A novel approach. Metabolic Engineering. 31: 132-139.
54. **Tian, M. and Reed, J.L. (2018).** Integrating Proteomic or Transcriptomic Data into Metabolic Models Using Linear Bound Flux Balance Analysis. Bioinformatics. 34(22): 3882-3888.
55. **Motamedian, E., Mohammadi, M., Shojaosadati, S.A., and Heydari, M. (2017).** TRFBA: an algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. Bioinformatics. 33(7): 1057-1063.
56. **Zengler, K., and Palsson, B. O. (2012).** A road map for the development of community systems (CoSy) biology. Nat. Rev. Microbiol. 10: 366-372.