Hatzimanikatis LCSB articles

# pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis.

2018

Bioinformatics (Oxford, England)

Salvy, Pierre

Fengos, Georgios

Ataman, Meric

Pathier, Thomas

Soh, Keng C

Hatzimanikatis, Vassily

Summary: pyTFA and matTFA are the first published implementations of the original TFA paper. Specifically, they include explicit formulation of Gibbs energies and metabolite concentrations, which enables straightforward integration of metabolite concentration measurements. Motivation: High-throughput analytic technologies provide a wealth of omics data that can be used to perform thorough analyses for a multitude of studies in the areas of Systems Biology and Biotechnology. Nevertheless, most studies are still limited to constraint-based Flux Balance Analyses (FBA), neglecting an important physicochemical constraint: thermodynamics. Thermodynamics-based Flux Analysis (TFA) in metabolic models enables the integration of quantitative metabolomics data to study their effects on the net-flux directionality of reactions in the network. In addition, it allows us to estimate how far each reaction operates from thermodynamic equilibrium, which provides critical information for guiding metabolic engineering decisions. Results: We present a Python package (pyTFA) and a Matlab toolbox (matTFA) that implement TFA. We show an example of application on both a reduced and a genome-scale model of E. coli., and demonstrate TFA and data integration through TFA reduce the feasible flux space with respect to FBA. Availability and implementation: Documented implementation of TFA framework both in Python (pyTFA) and Matlab (matTFA) are available on www.github.com/EPFL-LCSB/. Supplementary information: Supplementary data are available at Bioinformatics online.

# Kinetic models of metabolism that consider alternative steady-state solutions of intracellular fluxes and concentrations.

2018

Metabolic engineering

Hameri, Tuure

Fengos, Georgios

Ataman, Meric

Miskovic, Ljubisa

Hatzimanikatis, Vassily

Large-scale kinetic models are used for designing, predicting, and understanding the metabolic responses of living cells. Kinetic models are particularly attractive for the biosynthesis of target molecules in cells as they are typically better than other types of models at capturing the complex cellular biochemistry. Using simpler stoichiometric models as scaffolds, kinetic models are built around a steady-state flux profile and a metabolite concentration vector that are typically determined via optimization. However, as the underlying optimization problem is underdetermined, even after incorporating available experimental omics data, one cannot uniquely determine the operational configuration in terms of metabolic fluxes and metabolite concentrations. As a result, some reactions can operate in either the forward or reverse direction while still agreeing with the observed physiology. Here, we analyze how the underlying uncertainty in intracellular fluxes and concentrations affects predictions of constructed kinetic models and their design in metabolic engineering and systems biology studies. To this end, we integrated the omics data of optimally grown Escherichia coli into a stoichiometric model and constructed populations of non-linear large-scale kinetic models of alternative steady-state solutions consistent with the physiology of the E. coli aerobic metabolism. We performed metabolic control analysis (MCA) on these models, highlighting that MCA-based metabolic engineering decisions are strongly affected by the selected steady state and appear to be more sensitive to concentration values rather than flux values. To incorporate this into future studies, we propose a workflow for moving towards more reliable and robust predictions that are consistent with all alternative steady-state solutions. This workflow can be applied to all kinetic models to improve the consistency and accuracy of their predictions. Additionally, we show that, irrespective of the alternative steady-state solution, increased activity of phosphofructokinase and decreased ATP maintenance requirements would improve cellular growth of optimally grown E. coli.

# Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors.

2018

ACS synthetic biology

Tokic, Milenko

Hadadi, Noushin

Ataman, Meric

Neves, Dario

Ebert, Birgitta E

Blank, Lars M

Miskovic, Ljubisa

Hatzimanikatis, Vassily

The limited supply of fossil fuels and the establishment of new environmental policies shifted research in industry and academia toward sustainable production of the second generation of biofuels, with methyl ethyl ketone (MEK) being one promising fuel candidate. MEK is a commercially valuable petrochemical with an extensive application as a solvent. However, as of today, a sustainable and economically viable production of MEK has not yet been achieved despite several attempts of introducing biosynthetic pathways in industrial microorganisms. We used BNICE.ch as a retrobiosynthesis tool to discover all novel pathways around MEK. Out of 1325 identified compounds connecting to MEK with one reaction step, we selected 3-oxopentanoate, but-3-en-2-one, but-1-en-2-olate, butylamine, and 2-hydroxy-2-methylbutanenitrile for further study. We reconstructed 3679610 novel biosynthetic pathways toward these 5 compounds. We then embedded these pathways into the genome-scale model of E. coli, and a set of 18622 were found to be the most biologically feasible ones on the basis of thermodynamics and their yields. For each novel reaction in the viable pathways, we proposed the most similar KEGG reactions, with their gene and protein sequences, as candidates for either a direct experimental implementation or as a basis for enzyme engineering. Through pathway similarity analysis we classified the pathways and identified the enzymes and precursors that were indispensable for the production of the target molecules. These retrobiosynthesis studies demonstrate the potential of BNICE.ch for discovery, systematic evaluation, and analysis of novel pathways in synthetic biology and metabolic engineering studies.

# Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions.

2017

Nucleic acids research

Bryan, Louise C

Weilandt, Daniel R

Bachmann, Andreas L

Kilic, Sinan

Lechner, Carolin C

Odermatt, Pascal D

Fantner, Georg E

Georgeon, Sandrine

Hantschel, Oliver

Hatzimanikatis, Vassily

Fierz, Beat

Chromatin recruitment of effector proteins involved in gene regulation depends on multivalent interaction with histone post-translational modifications (PTMs) and structural features of the chromatin fiber. Due to the complex interactions involved, it is currently not understood how effectors dynamically sample the chromatin landscape. Here, we dissect the dynamic chromatin interactions of a family of multivalent effectors, heterochromatin protein 1 (HP1) proteins, using single-molecule fluorescence imaging and computational modeling. We show that the three human HP1 isoforms are recruited and retained on chromatin by a dynamic exchange between histone PTM and DNA bound states. These interactions depend on local chromatin structure, the HP1 isoforms as well as on PTMs on HP1 itself. Of the HP1 isoforms, HP1alpha exhibits the longest residence times and fastest binding rates due to DNA interactions in addition to PTM binding. HP1alpha phosphorylation further increases chromatin retention through strengthening of multivalency while reducing DNA binding. As DNA binding in combination with specific PTM recognition is found in many chromatin effectors, we propose a general dynamic capture mechanism for effector recruitment. Multiple weak protein and DNA interactions result in a multivalent interaction network that targets effectors to a specific chromatin modification state, where their activity is required.

# Identification and dynamics of the human ZDHHC16-ZDHHC6 palmitoylation cascade.

2017

eLife

Abrami, Laurence

Dallavilla, Tiziano

Sandoz, Patrick A

Demir, Mustafa

Kunz, Beatrice

Savoglidis, Georgios

Hatzimanikatis, Vassily

van der Goot, F Gisou

S-Palmitoylation is the only reversible post-translational lipid modification. Knowledge about the DHHC palmitoyltransferase family is still limited. Here we show that human ZDHHC6, which modifies key proteins of the endoplasmic reticulum, is controlled by an upstream palmitoyltransferase, ZDHHC16, revealing the first palmitoylation cascade. The combination of site specific mutagenesis of the three ZDHHC6 palmitoylation sites, experimental determination of kinetic parameters and data-driven mathematical modelling allowed us to obtain detailed information on the eight differentially palmitoylated ZDHHC6 species. We found that species rapidly interconvert through the action of ZDHHC16 and the Acyl Protein Thioesterase APT2, that each species varies in terms of turnover rate and activity, altogether allowing the cell to robustly tune its ZDHHC6 activity.

# lumpGEM: Systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites.

2017

PLoS computational biology

Ataman, Meric

Hatzimanikatis, Vassily

In the post-genomic era, Genome-scale metabolic networks (GEMs) have emerged as invaluable tools to understand metabolic capabilities of organisms. Different parts of these metabolic networks are defined as subsystems/pathways, which are sets of functional roles to implement a specific biological process or structural complex, such as glycolysis and TCA cycle. Subsystem/pathway definition is also employed to delineate the biosynthetic routes that produce biomass building blocks. In databases, such as MetaCyc and SEED, these representations are composed of linear routes from precursors to target biomass building blocks. However, this approach cannot capture the nested, complex nature of GEMs. Here we implemented an algorithm, lumpGEM, which generates biosynthetic subnetworks composed of reactions that can synthesize a target metabolite from a set of defined core precursor metabolites. lumpGEM captures balanced subnetworks, which account for the fate of all metabolites along the synthesis routes, thus encapsulating reactions from various subsystems/pathways to balance these metabolites in the metabolic network. Moreover, lumpGEM collapses these subnetworks into elementally balanced lumped reactions that specify the cost of all precursor metabolites and cofactors. It also generates alternative subnetworks and lumped reactions for the same metabolite, accounting for the flexibility of organisms. lumpGEM is applicable to any GEM and any target metabolite defined in the network. Lumped reactions generated by lumpGEM can be also used to generate properly balanced reduced core metabolic models.

# redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models.

2017

PLoS computational biology

Ataman, Meric

Hernandez Gardiol, Daniel F

Fengos, Georgios

Hatzimanikatis, Vassily

Genome-scale metabolic reconstructions have proven to be valuable resources in enhancing our understanding of metabolic networks as they encapsulate all known metabolic capabilities of the organisms from genes to proteins to their functions. However the complexity of these large metabolic networks often hinders their utility in various practical applications. Although reduced models are commonly used for modeling and in integrating experimental data, they are often inconsistent across different studies and laboratories due to different criteria and detail, which can compromise transferability of the findings and also integration of experimental data from different groups. In this study, we have developed a systematic semi-automatic approach to reduce genome-scale models into core models in a consistent and logical manner focusing on the central metabolism or subsystems of interest. The method minimizes the loss of information using an approach that combines graph-based search and optimization methods. The resulting core models are shown to be able to capture key properties of the genome-scale models and preserve consistency in terms of biomass and by-product yields, flux and concentration variability and gene essentiality. The development of these "consistently-reduced" models will help to clarify and facilitate integration of different experimental data to draw new understanding that can be directly extendable to genome-scale models.

# Bioenergetics-based modeling of Plasmodium falciparum metabolism reveals its essential genes, nutritional requirements, and thermodynamic bottlenecks.

2017

PLoS computational biology

Chiappino-Pepe, Anush

Tymoshenko, Stepan

Ataman, Meric

Soldati-Favre, Dominique

Hatzimanikatis, Vassily

Novel antimalarial therapies are urgently needed for the fight against drug-resistant parasites. The metabolism of malaria parasites in infected cells is an attractive source of drug targets but is rather complex. Computational methods can handle this complexity and allow integrative analyses of cell metabolism. In this study, we present a genome-scale metabolic model (iPfa) of the deadliest malaria parasite, Plasmodium falciparum, and its thermodynamics-based flux analysis (TFA). Using previous absolute concentration data of the intraerythrocytic parasite, we applied TFA to iPfa and predicted up to 63 essential genes and 26 essential pairs of genes. Of the 63 genes, 35 have been experimentally validated and reported in the literature, and 28 have not been experimentally tested and include previously hypothesized or novel predictions of essential metabolic capabilities. Without metabolomics data, four of the genes would have been incorrectly predicted to be non-essential. TFA also indicated that substrate channeling should exist in two metabolic pathways to ensure the thermodynamic feasibility of the flux. Finally, analysis of the metabolic capabilities of P. falciparum led to the identification of both the minimal nutritional requirements and the genes that can become indispensable upon substrate inaccessibility. This model provides novel insight into the metabolic needs and capabilities of the malaria parasite and highlights metabolites and pathways that should be measured and characterized to identify potential thermodynamic bottlenecks and substrate channeling. The hypotheses presented seek to guide experimental studies to facilitate a better understanding of the parasite metabolism and the identification of targets for more efficient intervention.

# Mechanistic Modeling of Genetic Circuits for ArsR Arsenic Regulation.

2017

ACS synthetic biology

Berset, Yves

Merulla, Davide

Joublin, Aurelie

Hatzimanikatis, Vassily

van der Meer, Jan R

Bioreporters are living cells that generate an easily measurable signal in the presence of a chemical compound. They acquire their functionality from synthetic gene circuits, the configuration of which defines the response signal and signal-to-noise ratio. Bioreporters based on the Escherichia coli ArsR system have raised significant interest for quantifying arsenic pollution, but they need to be carefully optimized to accurately work in the required low concentration range (1-10 mug arsenite L(-1)). To better understand the general functioning of ArsR-based genetic circuits, we developed a comprehensive mechanistic model that was empirically tested and validated in E. coli carrying different circuit configurations. The model accounts for the different elements in the circuits (proteins, DNA, chemical species), and their detailed affinities and interactions, and predicts the (fluorescent) output from the bioreporter cell as a function of arsenite concentration. The model was parametrized using existing ArsR biochemical data, and then complemented by parameter estimations from the accompanying experimental data using a scatter search algorithm. Model predictions and experimental data were largely coherent for feedback and uncoupled circuit configurations, different ArsR alleles, promoter strengths, and presence or absence of arsenic efflux in the bioreporters. Interestingly, the model predicted a particular useful circuit variant having steeper response at low arsenite concentrations, which was experimentally confirmed and may be useful as arsenic bioreporter in the field. From the extensive validation we expect the mechanistic model to further be a useful framework for detailed modeling of other synthetic circuits.

# Reconstruction of biological pathways and metabolic networks from in silico labeled metabolites.

2016

Biotechnology journal

Hadadi, Noushin

Hafner, Jasmin

Soh, Keng Cher

Hatzimanikatis, Vassily

Reaction atom mappings track the positional changes of all of the atoms between the substrates and the products as they undergo the biochemical transformation. However, information on atom transitions in the context of metabolic pathways is not widely available in the literature. The understanding of metabolic pathways at the atomic level is of great importance as it can deconvolute the overlapping catabolic/anabolic pathways resulting in the observed metabolic phenotype. The automated identification of atom transitions within a metabolic network is a very challenging task since the degree of complexity of metabolic networks dramatically increases when we transit from metabolite-level studies to atom-level studies. Despite being studied extensively in various approaches, the field of atom mapping of metabolic networks is lacking an automated approach, which (i) accounts for the information of reaction mechanism for atom mapping and (ii) is extendable from individual atom-mapped reactions to atom-mapped reaction networks. Hereby, we introduce a computational framework, iAM.NICE (in silico Atom Mapped Network Integrated Computational Explorer), for the systematic atom-level reconstruction of metabolic networks from in silico labelled substrates. iAM.NICE is to our knowledge the first automated atom-mapping algorithm that is based on the underlying enzymatic biotransformation mechanisms, and its application goes beyond individual reactions and it can be used for the reconstruction of atom-mapped metabolic networks. We illustrate the applicability of our method through the reconstruction of atom-mapped reactions of the KEGG database and we provide an example of an atom-level representation of the core metabolic network of E. coli.

# ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies.

2016

ACS synthetic biology

Hadadi, Noushin

Hafner, Jasmin

Shajkofci, Adrian

Zisaki, Aikaterini

Hatzimanikatis, Vassily

Because the complexity of metabolism cannot be intuitively understood or analyzed, computational methods are indispensable for studying biochemistry and deepening our understanding of cellular metabolism to promote new discoveries. We used the computational framework BNICE.ch along with cheminformatic tools to assemble the whole theoretical reactome from the known metabolome through expansion of the known biochemistry presented in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We constructed the ATLAS of Biochemistry, a database of all theoretical biochemical reactions based on known biochemical principles and compounds. ATLAS includes more than 130000 hypothetical enzymatic reactions that connect two or more KEGG metabolites through novel enzymatic reactions that have never been reported to occur in living organisms. Moreover, ATLAS reactions integrate 42% of KEGG metabolites that are not currently present in any KEGG reaction into one or more novel enzymatic reactions. The generated repository of information is organized in a Web-based database ( http://lcsb-databases.epfl.ch/atlas/ ) that allows the user to search for all possible routes from any substrate compound to any product. The resulting pathways involve known and novel enzymatic steps that may indicate unidentified enzymatic activities and provide potential targets for protein engineering. Our approach of introducing novel biochemistry into pathway design and associated databases will be important for synthetic biology and metabolic engineering.

# Analysis of Translation Elongation Dynamics in the Context of an Escherichia coli Cell.

2016

Biophysical journal

Vieira, Joana Pinto

Racle, Julien

Hatzimanikatis, Vassily

Understanding the mechanisms behind translation and its rate-limiting steps is crucial for both the development of drug targets and improvement of heterologous protein production with many biotechnological applications, such as in pharmaceutical and biofuel industries. Despite many advances in the knowledge of the ribosome structure and function, there is still much discussion around the determinants of translation elongation with experiments and computational studies pointing in different directions. Here, we use a stochastic framework to simulate the process of translation in the context of an Escherichia coli cell by gathering the available biochemical data into a ribosome kinetics description. Our results from the study of translation in E. coli at different growth rates contradict the increase of mean elongation rate with growth rate established in the literature. We show that both the level of tRNA competition and the type of cognate binding interaction contribute to the modulation of elongation rate, and that optimization of a heterologous transcript for faster elongation rate is achieved by combining the two. We derive an equation that can accurately predict codon elongation rates based on the abundances of free tRNA in the cell, and can be used to assist transcript design. Finally, we show that non-cognate tRNA-ribosome binding has an important weight in translation, and plays an active role in the modulation of mean elongation rate as shown by our amino-acid starvation/surplus studies.

# Model-Driven Understanding of Palmitoylation Dynamics: Regulated Acylation of the Endoplasmic Reticulum Chaperone Calnexin.

2016

PLoS computational biology

Dallavilla, Tiziano

Abrami, Laurence

Sandoz, Patrick A

Savoglidis, Georgios

Hatzimanikatis, Vassily

van der Goot, F Gisou

Cellular functions are largely regulated by reversible post-translational modifications of proteins which act as switches. Amongst these, S-palmitoylation is unique in that it confers hydrophobicity. Due to technical difficulties, the understanding of this modification has lagged behind. To investigate principles underlying dynamics and regulation of palmitoylation, we have here studied a key cellular protein, the ER chaperone calnexin, which requires dual palmitoylation for function. Apprehending the complex inter-conversion between single-, double- and non-palmitoylated species required combining experimental determination of kinetic parameters with extensive mathematical modelling. We found that calnexin, due to the presence of two cooperative sites, becomes stably acylated, which not only confers function but also a remarkable increase in stability. Unexpectedly, stochastic simulations revealed that palmitoylation does not occur soon after synthesis, but many hours later. This prediction guided us to find that phosphorylation actively delays calnexin palmitoylation in resting cells. Altogether this study reveals that cells synthesize 5 times more calnexin than needed under resting condition, most of which is degraded. This unused pool can be mobilized by preventing phosphorylation or increasing the activity of the palmitoyltransferase DHHC6.

# Identification of metabolic engineering targets for the enhancement of 1,4-butanediol production in recombinant E. coli using large-scale kinetic models.

2016

Metabolic engineering

Andreozzi, Stefano

Chakrabarti, Anirikh

Soh, Keng Cher

Burgard, Anthony

Yang, Tae Hoon

Van Dien, Stephen

Miskovic, Ljubisa, Hatzimanikatis, Vassily

Rational metabolic engineering methods are increasingly employed in designing the commercially viable processes for the production of chemicals relevant to pharmaceutical, biotechnology, and food and beverage industries. With the growing availability of omics data and of methodologies capable to integrate the available data into models, mathematical modeling and computational analysis are becoming important in designing recombinant cellular organisms and optimizing cell performance with respect to desired criteria. In this contribution, we used the computational framework ORACLE (Optimization and Risk Analysis of Complex Living Entities) to analyze the physiology of recombinant Escherichia coli producing 1,4-butanediol (BDO) and to identify potential strategies for improved production of BDO. The framework allowed us to integrate data across multiple levels and to construct a population of large-scale kinetic models despite the lack of available information about kinetic properties of every enzyme in the metabolic pathways. We analyzed these models and we found that the enzymes that primarily control the fluxes leading to BDO production are part of central glycolysis, the lower branch of tricarboxylic acid (TCA) cycle and the novel BDO production route. Interestingly, among the enzymes between the glucose uptake and the BDO pathway, the enzymes belonging to the lower branch of TCA cycle have been identified as the most important for improving BDO production and yield. We also quantified the effects of changes of the target enzymes on other intracellular states like energy charge, cofactor levels, redox state, cellular growth, and byproduct formation. Independent earlier experiments on this strain confirmed that the computationally obtained conclusions are consistent with the experimentally tested designs, and the findings of the present studies can provide guidance for future work on strain improvement. Overall, these studies demonstrate the potential and effectiveness of ORACLE for the accelerated design of microbial cell factories.

# iSCHRUNK--In Silico Approach to Characterization and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks.

2015

Metabolic engineering

Andreozzi, Stefano

Miskovic, Ljubisa

Hatzimanikatis, Vassily

Accurate determination of physiological states of cellular metabolism requires detailed information about metabolic fluxes, metabolite concentrations and distribution of enzyme states. Integration of fluxomics and metabolomics data, and thermodynamics-based metabolic flux analysis contribute to improved understanding of steady-state properties of metabolism. However, knowledge about kinetics and enzyme activities though essential for quantitative understanding of metabolic dynamics remains scarce and involves uncertainty. Here, we present a computational methodology that allow us to determine and quantify the kinetic parameters that correspond to a certain physiology as it is described by a given metabolic flux profile and a given metabolite concentration vector. Though we initially determine kinetic parameters that involve a high degree of uncertainty, through the use of kinetic modeling and machine learning principles we are able to obtain more accurate ranges of kinetic parameters, and hence we are able to reduce the uncertainty in the model analysis. We computed the distribution of kinetic parameters for glucose-fed E. coli producing 1,4-butanediol and we discovered that the observed physiological state corresponds to a narrow range of kinetic parameters of only a few enzymes, whereas the kinetic parameters of other enzymes can vary widely. Furthermore, this analysis suggests which are the enzymes that should be manipulated in order to engineer the reference state of the cell in a desired way. The proposed approach also sets up the foundations of a novel type of approaches for efficient, non-asymptotic, uniform sampling of solution spaces.

# Heading in the right direction: thermodynamics-based network analysis and pathway engineering.

2015

Current opinion in biotechnology

Ataman, Meric

Hatzimanikatis, Vassily

Thermodynamics-based network analysis through the introduction of thermodynamic constraints in metabolic models allows a deeper analysis of metabolism and guides pathway engineering. The number and the areas of applications of thermodynamics-based network analysis methods have been increasing in the last ten years. We review recent applications of these methods and we identify the areas that such analysis can contribute significantly, and the needs for future developments. We find that organisms with multiple compartments and extremophiles present challenges for modeling and thermodynamics-based flux analysis. The evolution of current and new methods must also address the issues of the multiple alternatives in flux directionalities and the uncertainties and partial information from analytical methods.

# Rites of passage: requirements and standards for building kinetic models of metabolic phenotypes.

2015

Current opinion in biotechnology

Miskovic, Ljubisa

Tokic, Milenko

Fengos, Georgios

Hatzimanikatis, Vassily

The overarching ambition of kinetic metabolic modeling is to capture the dynamic behavior of metabolism to such an extent that systems and synthetic biology strategies can reliably be tested in silico. The lack of kinetic data hampers the development of kinetic models, and most of the current models use ad hoc reduced stoichiometry or oversimplified kinetic rate expressions, which may limit their predictive strength. There is a need to introduce the community-level standards that will organize and accelerate the future developments in this area. We introduce here a set of requirements that will ensure the model quality, we examine the current kinetic models with respect to these requirements, and we propose a general workflow for constructing models that satisfy these requirements.

# Noise analysis of genome-scale protein synthesis using a discrete computational model of translation.

2015

The Journal of chemical physics

Racle, Julien

Stefaniuk, Adam Jan

Hatzimanikatis, Vassily

Noise in genetic networks has been the subject of extensive experimental and computational studies. However, very few of these studies have considered noise properties using mechanistic models that account for the discrete movement of ribosomes and RNA polymerases along their corresponding templates (messenger RNA (mRNA) and DNA). The large size of these systems, which scales with the number of genes, mRNA copies, codons per mRNA, and ribosomes, is responsible for some of the challenges. Additionally, one should be able to describe the dynamics of ribosome exchange between the free ribosome pool and those bound to mRNAs, as well as how mRNA species compete for ribosomes. We developed an efficient algorithm for stochastic simulations that addresses these issues and used it to study the contribution and trade-offs of noise to translation properties (rates, time delays, and rate-limiting steps). The algorithm scales linearly with the number of mRNA copies, which allowed us to study the importance of genome-scale competition between mRNAs for the same ribosomes. We determined that noise is minimized under conditions maximizing the specific synthesis rate. Moreover, sensitivity analysis of the stochastic system revealed the importance of the elongation rate in the resultant noise, whereas the translation initiation rate constant was more closely related to the average protein synthesis rate. We observed significant differences between our results and the noise properties of the most commonly used translation models. Overall, our studies demonstrate that the use of full mechanistic models is essential for the study of noise in translation and transcription.

# Integrative approaches for signalling and metabolic networks.

2015

Integrative biology : quantitative biosciences from nano to macro

Hatzimanikatis, Vassily

Saez-Rodriguez, Julio

The study and analysis of the organization of biochemical reactions into complex networks is central to integrative biology. This themed issue on networks has assembled reviews and original research papers that demonstrate how biological chemistry from the basic mechanistic design of individual reactions to hundreds of reactions organized into networks determine cellular function. The collection spans from metabolism to signal transduction, and from detailed biochemical formalisms to coarse graph-based approaches.

We can appreciate the model development process, the methods for the analysis and the applications of genome-scale models of metabolic networks through the thorough review of the genome-scale models of yeast by Sánchez and Nielsen (DOI: 10.1039/C5IB00083A). The authors present the history of yeast metabolic models and they discuss how omics data from different levels can be used in the development of such models and how in turn these models can be used for the integration, analysis and interpretation of different omics data. In a similar direction Sang Yup Lee and colleagues (DOI: 10.1039/C5IB00002E) offer a review of genome-scale human metabolic modeling, and they discuss how state-of-the-art high-throughput techniques and data are analyzed by advanced computational methods to formulate tissue/cell type-specific human metabolic models. Such context-specific metabolic models are used for studying metabolic diseases, the role of metabolism in other diseases such as cancer and infection, and the capabilities of biological chemistry to produce fuels and chemicals.

Mahadevan, Stephanopoulos and colleagues (DOI: 10.1039/C5IB00095E) assembled the biochemical reactions in the metabolic network of Moorella thermoacetica, a versatile acetogenic bacterium that is able to fix CO2 and transform syngas (CO + H2) into acetyl-CoA. These biosynthetic capabilities make this organism an attractive host for many biotechnology applications. The authors demonstrate how stoichiometric modeling and flux balance analysis can help us understand better the chemistry that leads from carbon dioxide and carbon monoxide to life and to every-day useful chemicals. E. coli is the benchmark organism for the analysis and design of metabolism in the development of almost every industrial chemical. In another study that combined metabolic modeling and experiments, Mahadevan and colleagues (DOI: 10.1039/C5IB00096C) have investigated the role of redundant reductions in the physiology of E. coli through the analysis of double-knockout mutants. They found that these mutants are not agnostic to the order in which their genes are deleted, and the order in which genes are deleted determines the phenotype of the mutants during the sub-optimal growth phase. However, these observations could not be explained using stoichiometric models alone and they would require the consideration of kinetic models that describe the regulatory effects at the level of gene expression and the regulation of enzyme kinetics.

The integration of enzyme kinetics for the analysis of the dynamic properties of these networks is one of the major challenges in the analysis of large- to genome-scale metabolic networks. The contribution by Klipp et al. (DOI: 10.1039/C5IB00050E) proposes a method for contextualizing existing dynamic metabolic models through the formulation and addition of reactions that account in a simplified but consistent manner for some of the pathways that are missing from the dynamic metabolic model. Such approaches can help increase significantly the accuracy of dynamic metabolic models. The nonlinear enzyme kinetics in the formulation of dynamic metabolic models introduce many challenges common in nonlinear dynamics. One of these challenges is the robustness of the system, which is the ability of the system to maintain a stable, physiological steady state. Liao and colleagues (DOI: 10.1039/C4IB00257A) propose a metric that can allow characterization of the robustness of a nonlinear dynamic system. Such a metric can be used for the analysis and design of complex biochemical networks and for the improvement of mathematical models that describe these systems.

Gunawardena and colleagues (DOI: 10.1039/C5IB00009B) further study the issue of robustness in the specific context of bifunctional enzymes. These enzymes control robustly different biochemical processes such as metabolic branch points and osmoregulatory networks. The authors use tools of computational algebraic geometry to find relationships (invariants) between the steady state concentrations of the modified and unmodified substrate. This allows them to characterize robust behavior in these systems, and illustrate the power of these invariants to study biochemical networks.

Similarly to metabolism, formulating and analyzing models that describe the underlying biochemical reactions is the most common way to study signaling networks. Analytical methods that rely on the biochemical structure, such as those presented by Gunawardena et al., can be applied to identify key properties. Since signaling is largely a dynamic process, the reactions are often converted into differential equations for simulation and analysis. However, the size and complexity of the models increase exponentially with the scope of the model, and require parameter values often unknown.

As an alternative, qualitative models and in particular logic-based models have become popular recently to analyze properties of large signaling networks. In this issue, Zinovyev and colleagues (DOI: 10.1039/C5IB00029G) use Boolean formalism (the simplest variant of a logic model, where each variable is either ON or OFF) to characterize genetic interactions (i.e. epistasis) in signaling and regulatory networks. “Genetic interactions” refers here to cases where the combined effect of two genes cannot be predicted from the effect of both of them alone. Genetic interactions are useful to characterize functional relationships between genes, and can provide novel insight to treat diseases. Because Boolean logic models are so simple, they can be built for many systems, even if our knowledge is not very refined, and hence their approach can be broadly applicable.

While the knowledge required to build logic models is significantly less than for biochemical models, there are many pathways for which there is information about the proteins involved and the potential connections (directed and often signed), but not enough granularity to build logic functions. In addition, while fairly efficient and scalable, the methods to simulate and analyze these models can generally not handle the size needed to ultimately cover genome-wide networks. Melas et al. (DOI: 10.1039/C4IB00294F) propose an approach to identify, in directed and signed networks of thousands of proteins, the pathways that causally link a given perturbation with an altered profile of gene expression. The authors apply their method to study the mode of action of drugs in the context of drug induced lung injury, but it can also be used to dissect the effect of other perturbations or alterations (e.g. mutations) on downstream processes.

Through the articles described above, this collection provides a glance at the variety and power of network biology approaches. We hope to see more papers on this area in the future in Integrative Biology. An area of active research is around models that bridge the different layers of molecular processes within cells (Gonçalves et al., Molecular BioSystems, 2013, 9, 1576, DOI: 10.1039/C3MB25489E).

We can foresee further development of the computational methods, refinement of the models and integration of additional accumulated data from multiple levels of the genome sequence to metabolite levels. Thereby these models should become increasingly useful across many areas of fundamental and integrative biology, as well as in applications ranging from biotechnology to analysis of disease and eventually the design of personalized therapies.

# Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways.

2015

Current opinion in chemical biology

Hadadi, Noushin

Hatzimanikatis, Vassily

Designing putative metabolic pathways is of great interest in synthetic biology. Retrobiosynthesis is a discipline that involves the design, evaluation, and optimization of de novo biosynthetic pathways for the production of high-value compounds and drugs from renewable resources and natural or engineered enzymes. The best candidate pathways are then engineered within a metabolic network of microorganisms that serve as synthetic platforms for synthetic biology. The complexity of biological chemistry and metabolism requires computational approaches to explore the full possibilities of engineering synthetic pathways towards target compounds. Herein, we discuss recent developments in the design of computational tools for retrosynthetic biochemistry and outline the workflow and design elements for such tools.

# Metabolic Needs and Capabilities of Toxoplasma gondii through Combined Computational and Experimental Analysis.

2015

PLoS computational biology

Tymoshenko, Stepan

Oppenheim, Rebecca D

Agren, Rasmus

Nielsen, Jens

Soldati-Favre, Dominique

Hatzimanikatis, Vassily

Toxoplasma gondii is a human pathogen prevalent worldwide that poses a challenging and unmet need for novel treatment of toxoplasmosis. Using a semi-automated reconstruction algorithm, we reconstructed a genome-scale metabolic model, ToxoNet1. The reconstruction process and flux-balance analysis of the model offer a systematic overview of the metabolic capabilities of this parasite. Using ToxoNet1 we have identified significant gaps in the current knowledge of Toxoplasma metabolic pathways and have clarified its minimal nutritional requirements for replication. By probing the model via metabolic tasks, we have further defined sets of alternative precursors necessary for parasite growth. Within a human host cell environment, ToxoNet1 predicts a minimal set of 53 enzyme-coding genes and 76 reactions to be essential for parasite replication. Double-gene-essentiality analysis identified 20 pairs of genes for which simultaneous deletion is deleterious. To validate several predictions of ToxoNet1 we have performed experimental analyses of cytosolic acetyl-CoA biosynthesis. ATP-citrate lyase and acetyl-CoA synthase were localised and their corresponding genes disrupted, establishing that each of these enzymes is dispensable for the growth of T. gondii, however together they make a synthetic lethal pair.

# Antihypertensive drugs metabolism: an update to pharmacokinetic profiles and computational approaches.

2014

Current pharmaceutical design

Zisaki, Aikaterini

Miskovic, Ljubisa

Hatzimanikatis, Vassily

Drug discovery and development is a high-risk enterprise that requires significant investments in capital, time and scientific expertise. The studies of xenobiotic metabolism remain as one of the main topics in the research and development of drugs, cosmetics and nutritional supplements. Antihypertensive drugs are used for the treatment of high blood pressure, which is one the most frequent symptoms of the patients that undergo cardiovascular diseases such as myocardial infraction and strokes. In current cardiovascular disease pharmacology, four drug clusters - Angiotensin Converting Enzyme Inhibitors, Beta-Blockers, Calcium Channel Blockers and Diuretics - cover the major therapeutic characteristics of the most antihypertensive drugs. The pharmacokinetic and specifically the metabolic profile of the antihypertensive agents are intensively studied because of the broad inter-individual variability on plasma concentrations and the diversity on the efficacy response especially due to the P450 dependent metabolic status they present. Several computational methods have been developed with the aim to: (i) model and better understand the human drug metabolism; and (ii) enhance the experimental investigation of the metabolism of small xenobiotic molecules. The main predictive tools these methods employ are rule-based approaches, quantitative structure metabolism/activity relationships and docking approaches. This review paper provides detailed metabolic profiles of the major clusters of antihypertensive agents, including their metabolites and their metabolizing enzymes, and it also provides specific information concerning the computational approaches that have been used to predict the metabolic profile of several antihypertensive drugs.

# Kinetic models in industrial biotechnology - Improving cell factory performance.

2014

Metabolic engineering

Almquist, Joachim

Cvijovic, Marija

Hatzimanikatis, Vassily

Nielsen, Jens

Jirstrand, Mats

An increasing number of industrial bioprocesses capitalize on living cells by using them as cell factories that convert sugars into chemicals. These processes range from the production of bulk chemicals in yeasts and bacteria to the synthesis of therapeutic proteins in mammalian cell lines. One of the tools in the continuous search for improved performance of such production systems is the development and application of mathematical models. To be of value for industrial biotechnology, mathematical models should be able to assist in the rational design of cell factory properties or in the production processes in which they are utilized. Kinetic models are particularly suitable towards this end because they are capable of representing the complex biochemistry of cells in a more complete way compared to most other types of models. They can, at least in principle, be used to in detail understand, predict, and evaluate the effects of adding, removing, or modifying molecular components of a cell factory and for supporting the design of the bioreactor or fermentation process. However, several challenges still remain before kinetic modeling will reach the degree of maturity required for routine application in industry. Here we review the current status of kinetic cell factory modeling. Emphasis is on modeling methodology concepts, including model network structure, kinetic rate expressions, parameter estimation, optimization methods, identifiability analysis, model reduction, and model validation, but several applications of kinetic models for the improvement of cell factories are also discussed.

# A computational framework for integration of lipidomics data into metabolic pathways.

2014

Metabolic engineering

Hadadi, Noushin

Cher Soh, Keng

Seijo, Marianne

Zisaki, Aikaterini

Guan, Xueli

Wenk, Markus R

Hatzimanikatis, Vassily

Lipids are important compounds for human physiology and as renewable resources for fuels and chemicals. In lipid research, there is a big gap between the currently available pathway-level representations of lipids and lipid structure databases in which the number of compounds is expanding rapidly with high-throughput mass spectrometry methods. In this work, we introduce a computational approach to bridge this gap by making associations between metabolic pathways and the lipid structures discovered increasingly thorough lipidomics studies. Our approach, called NICELips (Network Integrated Computational Explorer for Lipidomics), is based on the formulation of generalized enzymatic reaction rules for lipid metabolism, and it employs the generalized rules to postulate novel pathways of lipid metabolism. It further integrates all discovered lipids in biological networks of enzymatic reactions that consist their biosynthesis and biodegradation pathways. We illustrate the utility of our approach through a case study of bis(monoacylglycero)phosphate (BMP), a biologically important glycerophospholipid with immature synthesis and catabolic route(s). Using NICELips, we were able to propose various synthesis and degradation pathways for this compound and several other lipids with unknown metabolism like BMP, and in addition several alternative novel biosynthesis and biodegradation pathways for lipids with known metabolism. NICELips has potential applications in designing therapeutic interventions for lipid-associated disorders and in the metabolic engineering of model organisms for improving the biobased production of lipid-derived fuels and chemicals.

# A genome-scale integration and analysis of Lactococcus lactis translation data.

2013

PLoS computational biology

Racle, Julien

Picard, Flora

Girbal, Laurence

Cocaign-Bousquet, Muriel

Hatzimanikatis, Vassily

Protein synthesis is a template polymerization process composed by three main steps: initiation, elongation, and termination. During translation, ribosomes are engaged into polysomes whose size is used for the quantitative characterization of translatome. However, simultaneous transcription and translation in the bacterial cytosol complicates the analysis of translatome data. We established a procedure for robust estimation of the ribosomal density in hundreds of genes from Lactococcus lactis polysome size measurements. We used a mechanistic model of translation to integrate the information about the ribosomal density and for the first time we estimated the protein synthesis rate for each gene and identified the rate limiting steps. Contrary to conventional considerations, we find significant number of genes to be elongation limited. This number increases during stress conditions compared to optimal growth and proteins synthesized at maximum rate are predominantly elongation limited. Consistent with bacterial physiology, we found proteins with similar rate and control characteristics belonging to the same functional categories. Under stress conditions, we found that synthesis rate of regulatory proteins is becoming comparable to proteins favored under optimal growth. These findings suggest that the coupling of metabolic states and protein synthesis is more important than previously thought.

# Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints.

2013

Biotechnology journal

Chakrabarti, Anirikh

Miskovic, Ljubisa

Soh, Keng Cher

Hatzimanikatis, Vassily

Mathematical modeling is an essential tool for the comprehensive understanding of cell metabolism and its interactions with the environmental and process conditions. Recent developments in the construction and analysis of stoichiometric models made it possible to define limits on steady-state metabolic behavior using flux balance analysis. However, detailed information on enzyme kinetics and enzyme regulation is needed to formulate kinetic models that can accurately capture the dynamic metabolic responses. The use of mechanistic enzyme kinetics is a difficult task due to uncertainty in the kinetic properties of enzymes. Therefore, the majority of recent works considered only mass action kinetics for reactions in metabolic networks. Herein, we applied the optimization and risk analysis of complex living entities (ORACLE) framework and constructed a large-scale mechanistic kinetic model of optimally grown Escherichia coli. We investigated the complex interplay between stoichiometry, thermodynamics, and kinetics in determining the flexibility and capabilities of metabolism. Our results indicate that enzyme saturation is a necessary consideration in modeling metabolic networks and it extends the feasible ranges of metabolic fluxes and metabolite concentrations. Our results further suggest that enzymes in metabolic networks have evolved to function at different saturation states to ensure greater flexibility and robustness of cellular metabolism.

# Functional genomics of Plasmodium falciparum using metabolic modelling and analysis.

2013

Briefings in functional genomics

Tymoshenko, Stepan

Oppenheim, Rebecca D

Soldati-Favre, Dominique

Hatzimanikatis, Vassily

Plasmodium falciparum is an obligate intracellular parasite and the leading cause of severe malaria responsible for tremendous morbidity and mortality particularly in sub-Saharan Africa. Successful completion of the P. falciparum genome sequencing project in 2002 provided a comprehensive foundation for functional genomic studies on this pathogen in the following decade. Over this period, a large spectrum of experimental approaches has been deployed to improve and expand the scope of functionally annotated genes. Meanwhile, rapidly evolving methods of systems biology have also begun to contribute to a more global understanding of various aspects of the biology and pathogenesis of malaria. Herein we provide an overview on metabolic modelling, which has the capability to integrate information from functional genomics studies in P. falciparum and guide future malaria research efforts towards the identification of novel candidate drug targets.

# A computational framework for the design of optimal protein synthesis.

2012

Biotechnology and bioengineering

Racle, Julien

Overney, Jan

Hatzimanikatis, Vassily

Despite the establishment of design principles to optimize codon choice for heterologous expression vector design, the relationship between codon sequence and final protein yield remains poorly understood. In this work, we present a computational framework for the identification of a set of mutant codon sequences for optimized heterologous protein production, which uses a codon-sequence mechanistic model of protein synthesis. Through a sensitivity analysis on the optimal steady state configuration of protein synthesis we are able to identify the set of codons, that are the most rate limiting with respect to steady state protein synthesis rate, and we replace them with synonymous codons recognized by charged tRNAs more efficient for translation, so that the resulting codon-elongation rate is higher. Repeating this procedure, we iteratively optimize the codon sequence for higher protein synthesis rate taking into account multiple constraints of various types. We determine a small set of optimized synonymous codon sequences that are very close to each other in sequence space, but they have an impact on properties such as ribosomal utilization or secondary structure. This limited number of sequences can then be offered for further experimental study. Overall, the proposed method is very valuable in understanding the effects of the different properties of mRNA sequences on the final protein yield in heterologous protein production and it can find applications in synthetic biology and biotechnology.

# From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks.

2011

FEMS yeast research

Soh, Keng Cher

Miskovic, Ljubisa

Hatzimanikatis, Vassily

Many important problems in cell biology arise from the dense nonlinear interactions between functional modules. The importance of mathematical modelling and computer simulation in understanding cellular processes is now indisputable and widely appreciated. Genome-scale metabolic models have gained much popularity and utility in helping us to understand and test hypotheses about these complex networks. However, there are some caveats that come with the use and interpretation of different types of metabolic models, which we aim to highlight here. We discuss and illustrate how the integration of thermodynamic and kinetic properties of the yeast metabolic networks in network analyses can help in understanding and utilizing this organism more successfully in the areas of metabolic engineering, synthetic biology and disease treatment.

# DREAMS of metabolism.

2010

Trends in biotechnology

Soh, Keng Cher

Hatzimanikatis, Vassily

Metabolic networks have been studied for several decades, and sophisticated computational frameworks are needed to augment experimental approaches to harness these complex networks. BNICE (Biochemical Network Integrated Computational Explorer), a computational approach for the discovery of novel biochemical pathways that is based on biochemical transformations, overcomes many of the current limitations. BNICE and similar frameworks can be used in several different areas: (i) 'Design' of novel pathways for metabolic engineering; (ii) 'Retrosynthesis' of metabolic compounds; (iii) 'Evolution' analysis between metabolic pathways of different organisms; (iv) 'Analysis' of metabolic pathways; (v) 'Mining' of omics data; and (vi) 'Selection' of targets for enzyme engineering. Here, we discuss the issues and challenges in building such frameworks as well as the gamut of applications in biotechnology, metabolic engineering and synthetic biology.

# Production of biofuels and biochemicals: in need of an ORACLE.

2010

Trends in biotechnology

Miskovic, Ljubisa

Hatzimanikatis, Vassily

The engineering of cells for the production of fuels and chemicals involves simultaneous optimization of multiple objectives, such as specific productivity, extended substrate range and improved tolerance - all under a great degree of uncertainty. The achievement of these objectives under physiological and process constraints will be impossible without the use of mathematical modeling. However, the limited information and the uncertainty in the available information require new methods for modeling and simulation that will characterize the uncertainty and will quantify, in a statistical sense, the expectations of success of alternative metabolic engineering strategies. We discuss these considerations toward developing a framework for the Optimization and Risk Analysis of Complex Living Entities (ORACLE) - a computational method that integrates available information into a mathematical structure to calculate control coefficients.

# Network thermodynamics in the post-genomic era.

2010

Current opinion in microbiology

Soh, Keng Cher

Hatzimanikatis, Vassily

Network models have been used to study the underlying processes and principles of biological systems for decades, providing many insights into the complexity of life. Biological systems require a constant flow of free energy to drive these processes that operate away from thermodynamic equilibrium. With the advent of high-throughput omics technologies, more and more thermodynamic knowledge about the biological components, processes and their interactions are surfacing that we can integrate using large-scale biological network models. This allows us to ask many fundamental questions about these networks, such as, how far away from equilibrium must the reactions in a network be displaced in order to allow growth, or what are the possible thermodynamic objectives of the cell.