Metabolic Pathway Analysis 2017 Conference

Bozeman, Montana USA

24-28 July 2017

Oral Presentation Abstracts (in order of presentation)

T-1

Title: Exploring the combinatorial space of complete pathways to chemicals

Author: Costas D. Maranas

Affiliation(s): Department of Chemical Engineering, Penn State, University Park, PA, USA

Abstract: Computational pathway design tools provide a systematic way to traverse production routes to high-value chemicals. Important considerations such as reaction rules, network size, the complexity of pathway topology, co-substrate and co-product choices, mass-conservation, cofactor balance, thermodynamic feasibility, chassis selection, yield, and cost are not generally placed within the same decision framework and are largely dealt with in a posteriori fashion. We recently developed two computational tools: (i) optStoic/minFlux and (ii) rePrime/novoStoic. The optStoic/minFlux is a two-stage MILP-based computational procedure wherein the first step, optStoic, explores the maximum extent of converting carbon substrate(s) to desired product(s) through a non-intuitive combination of co-reactants and co-products while maintaining overall thermodynamic feasibility and mass balances [1]. In the subsequent step, the minRxn/minFlux algorithm identifies the minimal network of reactions to perform the overall conversion. This formalism was applied to identify alternate non-oxidative glycolysis and methane fixation pathways [1]. We recently updated the optStoic/minflux procedure to eliminate the occurrence of thermodynamically infeasible subnetworks in the identified pathways. The updated algorithm was employed to prospect for over 37,000 pathways that are capable of converting glucose to pyruvate while generating pre-determined numbers of ATP. These pathways were filtered based on their thermodynamic feasibility under physiological metabolites concentration bounds and ranked based on the minimal protein cost required to operate them. Pareto analysis of these glycolytic pathways revealed interesting tradeoffs between pathway energy efficiency and protein cost. In particular, the canonical ED and EMP pathways were found to be optimized through natural evolution for energy efficiency. Improved synthetic pathways with lower protein cost were also identified. To further expand our pathway designing capabilities beyond the known repertoire of enzymatic reactions, we incorporate hypothetical reactions predicted using reaction rules. Reaction rules expand our solution space and allow us to explore enzymatic capabilities which are yet to be identified (i.e., due to substrate promiscuity) or those that could be designed through protein engineering. The computational procedure rePrime/novoStoic we present here designs routes that extend beyond the currently known bioconversion space while simultaneously considering all aforementioned design criteria. First, we track and codify as rules, all reaction centers using a novel prime factorization based encoding technique (rePrime). The biochemical tenets, encoded within (reAxiom) reaction rules guide the pathway-designing algorithm (novoStoic) to devise mass balanced bio-conversion strategies. rePrime is a recursive procedure and novoStoic is posed as linear mixed-integer optimization formulation. We demonstrate the use of novoStoic in pathway elucidation towards first predicting intermediates of ill-defined degradation pathways for polycylic aromatic hydrocarbons (PAHs), and by designing novel synthetic routes for small molecules such as 1,4 butanediol and non-natural products such as phenylephrine, naproxen, epinine, and N-methyl-aspartate from aromatic precursors.

1. Chowdhury, A. and C.D. Maranas, *Designing overall stoichiometric conversions and intervening metabolic reactions.* Sci Rep, 2015. **5**: p. 16009.

T-2

Title: Design of Multi-Biosynthetic Paths

Author: Elmar Heinzle, Lisa Katharina Blass, Christian Weyler

Primary affiliation: Saarland University, Germany

Abstract: Complex smaller molecules are of major interest as pharmaceutically active compounds or precursors. Their synthesis is mostly observed in organisms whose physiology is largely unknown and that are usually difficult to cultivate. Modern genetic engineering methods are increasingly powerful in transferring biosynthesis gene clusters into robust, well-known heterologous hosts as E. coli or yeast strains. We investigate production using permeabilized cells or designed enzyme cocktails to improve selectivity, yield and titer for one single molecular variant of e.g. polyketides or non-ribosomally synthesized peptides.

We developed a method that using available whole-genome information and builds on existing knowledge about metabolic networks. In a first step the metabolic network of a super organism is created using all data available in KEGG about metabolic reactions. The 7405 metabolites are separated in different groups, e.g. potential start metabolites and cofactors. The network data are combined with thermodynamic data about equilibria using eQuilibrator. We modified a method earlier presented by Pey et al. (2011) by introducing path-finding, stoichiometric and other constraints to generate a set of pathway candidates using MILP. These are ranked using a set of criteria, e.g. length of pathway, identified starting molecules, thermodynamic feasibility, number of heterologous enzymes required, number of cofactors requiring regeneration. The ranked list provides overall balances, the thermodynamic profile of the biosynthesis path, lists of potential side reactions. The pathways are visualized using Cytoscape. This output provides a most useful basis for a following expert assessment.

Concluding, we will present some examples, where we successfully applied this method.

T-3

Title: MODCELL: A Multiobjective Strain Design Platform for Modular Cell Engineering

Authors: Cong Trinh, Sergio Garcia

Primary affiliation: University of Tennessee, Knoxville

Abstract: Metabolic engineering has enabled the use of microbial cell factories for industrial production of biochemicals. However, developing an optimal strain for synthesis of one product with the conventional strategy is laborious and expensive. To accelerate and reduce the cost of strain engineering, we develop the modular cell (MODCELL) design principle by exploiting the modular organization of metabolic networks and combinatorial possibilities of metabolic modules that enable the synthesis of a large space of biochemicals. Using the multiobjective optimization methods, we develop novel algorithms to implement the MODCELL design for genome-scale metabolic networks. We demonstrate MODCELL to design the biomass-degrading *Clostridium thermocellum* modular cell for combinatorial synthesis of biochemicals, e.g., alcohols and bioesters from lignocellulosic biomass. We envision the MODCELL will provide a useful platform for modular cell engineering.

T-4

Title: Managing uncertainty in metabolic network structures

Author: Jason Papin

Primary affiliation: University of Virginia

Abstract: A major barrier preventing more widespread use of genome-scale metabolic network reconstructions (GENREs), particularly to study non-model organisms, is the extensive time required to produce a high-quality GENRE. Many automated approaches have been developed which reduce this time requirement, but automatically-reconstructed draft GENREs still require curation before useful predictions can be made. We present a novel approach to the analysis of GENREs which improves the predictive capabilities of draft GENREs by representing many alternative network structures, all equally consistent with available data, and generating predictions from this ensemble. This ensemble approach is compatible with many reconstruction methods. We refer to this new approach as Ensemble Flux Balance Analysis (EnsembleFBA). We validate EnsembleFBA by predicting growth and gene essentiality in the model organism *Pseudomonas aeruginosa*. We demonstrate how EnsembleFBA can be included in a systems biology workflow by predicting essential genes in six *Streptococcus* species and mapping the essential genes to small molecule ligands from DrugBank. We found that some metabolic subsystems contributed disproportionately to the set of predicted essential reactions in a way that was unique to each *Streptococcus* species, leading to species-specific outcomes from small molecule interactions. Through our analyses of *P. aeruginosa* and six *Streptococci*, we show that ensembles increase the quality of predictions without drastically increasing reconstruction time, thus making GENRE approaches more practical for applications which require predictions for many non-model organisms.

T-5

Title: MENTOS: a thermodynamic approach for estimating metabolites and fluxes

Authors: Jeremy Zucker, Neeraj Kumar, Bill Cannon

Primary affiliation: Pacific Northwest National Laboratory

Abstract: Constraint-based methods such as flux balance analysis (FBA) have been successfully used to predict the steady-state metabolic fluxes that maximize the growth rate, but FBA provides no information about the metabolite concentrations. Genome-scale thermodynamic-based methods have been developed that constrain the reaction direction of metabolic fluxes using measured metabolite concentrations. We presen a new thermodynamic optimization method that can be used to predict both the metabolite concentrations and the fluxes that extract ener from the environment as quickly and efficiently as possible, given the constraints of the network and a suitable set of boundary conditions. The method can also be used to estimate the concentrations of unmeasured internal metabolites, given a set of measured internal metabolites.

T-6

Title: Computing EFMs consistent with equilibrium constants

Authors: Sabine Peres, P. Dague, M. Jolicoeur, S. Schuster

Primary affiliation(s): University Paris-Sud, INRA, University of Jena

Abstract: The notion of elementary flux mode (EFM) is a key concept in the analysis of metabolic networks from a pathway-oriented perspective. The set of EFMs represents the potential pathways in a metabolic network but the biologically feasible pathways are limited by various biological constraints: thermodynamic constraints, kinetics and regulations. We present a method for computing thermodynamically feasible elementary flux modes (tEFMs) using equilibrium constants without need of internal metabolite concentrations. The method is compared with the method based on a binary distinction between reversible and irreversible reactions. When all reactions are reversible, adding the constraints based on equilibrium constants reduces the number of elementary flux modes (EFMs) by a factor of two. Declaring in advance some reactions as irreversible, based on reliable biochemical expertise, can in general reduce the number of EFMs by a greater factor. But, even in this case, computing tEFMs can rule out some EFMs which are biochemically irrelevant. We applied our method to a published model described with binary distinction: the central carbon metabolism of Chinese hamster ovary cells. The suppression of the EFMs that are not consistent with the equilibrium constants appears to be biologically relevant.

T-7

Title: Systems Analysis of Intracellular pH Vulnerabilities for Cancer Therapy

Authors: Erez Persi[1, *, #], Miquel Duran-Frigola[2, *], Mehdi Damaghi[3, *], William R. Roush[4], Patrick Aloy[2, 5], John L. Cleveland[6], Robert J. Gillies[3], Eytan Ruppin[1, #]

1 Center for Bioinformatics and Computational Biology, Institute of Advanced Computer Studies, Department of Computer Science, University of Maryland College Park, Maryland 20742, USA

2 Joint IRB-BSC-CRG Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain

3 Department of Cancer Imaging and Metabolism, Moffitt Cancer Center and Research Institute, Tampa, Florida 33612, USA

4 Department of Chemistry, The Scripps Research Institute, 110 Scripps Way, Jupiter, USA

5 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

6 Department of Tumor Biology, Moffitt Cancer Center & Research Institute, Tampa, Florida 33612, USA

Abstract: A reverse pH gradient is a hallmark of cancer metabolism, where tumor cells acidify the extracellular milieu yet alkalinize their cytoplasm. While consequences of extracellular acidosis are known, the roles of intracellular alkalization are poorly understood. By reconstructing and integrating enzymatic pH-dependent activity profiles into cell-specific genome-scale metabolic models, we developed a computational methodology that defines how intracellular pH (pHi) modulates metabolism. The model reveals that alkaline pHi maximizes proliferation, glycolysis and hypoxia adaptation, whereas acidic pHi disables metabolic adaptations and compromises tumor cell growth and survival. The modeling also predicted vulnerable metabolic targets, whose experimental knockdown compromised breast cancer cell growth and survival in a selective and pHi dependent manner. This systems analysis establishes the essential roles of alkaline pHi in cancer metabolism and provides a framework for exploring pHi roles across biology.

T-8

Title: The Hidden costs of enzymatic catalysis

Authors: Elad Noor, Meike Wortel, Wolfram Liebermeister

Primary affiliation: ETH Zurich

Abstract: The existence of a trade-off between the biomass yield and growth rate of cells has been used to explain aerobic fermentation in cancer cells (Warburg effect), yeast cells (Crabtree effect) and in bacteria such as *E. coli.* This trade-off relies on the assumption that even though fermentation pathways produce 5-10 times less ATP per glucose, respiration requires so much more resources and is therefore inefficient when carbon is not limiting. Is this trade-off a universal constraint imposed by thermodynamics, or a coincidental feature of the specific enzyme kinetic parameters that evolved in these organisms? To answer this question we developed a new method called Flux-balance Enzyme Cost Minimization (fECM) to model the costs of both respiration and fermentation (along with ~1000 other flux combinations called elementary flux modes). We find that the trade-off in E. coli is not universal and depends strongly on the availability of oxygen. This framework successfully predicts in vivo enzyme concentrations, and has applications in metabolic engineering where similar candidate pathways can be compared not just by their yields, but also by their costs.

T-9

Title: Evolution explains the universality and simplicity of microbial metabolism

Authors: Daan de Groot, Systems Bioinformatics, Vrije Universiteit, The Netherlands

Robert Planqué, Mathematics, Vrije Universiteit, Amsterdam, The Netherlands

Frank J. Bruggeman, Systems Bioinformatics, Vrije Universiteit, The Netherlands

Bas Teusink, Systems Bioinformatics, Vrije Universiteit, The Netherlands

Abstract: Many evolutionary distant microbial species show highly similar metabolic behaviour, such as overflow metabolism, diauxic growth and catabolite repression, and mixed substrate usage. Overflow metabolism is also shown by autonomously proliferating higher eukaryotic cells, such as cancer cells (known as the Warburg effect). This raises the question whether and how natural selection drives cells towards those common metabolic strategies. We identify an evolutionary extremum principle that determines metabolic behaviour under growth-promoting conditions: fitness maximisation requires minimisation of metabolic complexity. The principle is a mathematical consequence of maximising metabolic fluxes under (enzymatic) constraints. We prove that the number of active constraints determines the number of active minimal metabolic pathways (Elementary Flux Modes) that carry flux in the optimal solution. These constraints are low in number and universal across species, which explains why different species display common fitness effects of protein expression and similar metabolic behaviour. The theory is used to derive three necessary conditions for a micro-organism to show a mixed strategy. The fundamental principle is presented as a (graphical) mathematical framework that provides a deeper, unifying understanding of (mixed) metabolic strategies. We use it to re-interpret experimental results on overflow metabolism and the co-consumption of substrates, demonstrating that the theory has important operational use in the interpretation and design of experiments that can identify the most likely active constraints.

T-10

Title: Multi-constraint approach for the design of lean-proteome strains

Authors: Egils Stalidzans, I. Morell, A. Seiman, A. Pentjuss, R. Vilu

Primary affiliation: Latvia University of Agriculture

Abstract: Modeling aproach is used in metabolic engineering to predict organism behavior after implementation of designed changes. Each model of biological system is a simplification of reality taking into account just part of real life constraints. The available modeling approaches mostly concentrate on metabolism and mass conservation issues leaving out of scope some other cellular processes and parameters that can be matematically calculated or estimated. In case of protein as a target product it becomes important to take into account the potential of engineered organism after removal of some groups of proteins (lean-proteome strains) that are not essential in particular biotechological process.

The proposed SILICON modeling platform integrates constraint based modelling approach for genome scale model and Single Cell Model, where thermodynamic, kinetic constraints, measured input-output fluxes and media composition, proteomics data, and central metabolic network data is taken into account along with cell geometry, cell cycle, necessary trancriptional and translational molecular resources and energetic costs and other parameters. Single Cell Model calculated flux distributions (about five hundred reactions) are integrated in genome scale modeling part and shows possible metabolic flux pattern in a genome scale model. Increased scope of biological and biochemical constraints enables better assessment of design feasibility, limiting factors and other peculiarities of organism behavior under conditions of increased target protein production.

T-11

Title: Enzymes and Substrates Are Balanced at Minimal Combined Mass Concentration

Author: Martin Lercher, Hugo Dourado, Veronica Maurino

Affiliation: Heinrich Heine University

Abstract: A fundamental problem in biology is how cells organize their resource investment. Cellular metabolism, for example, typically involves hundreds of enzymes and metabolites, but it is unclear according to which principles their concentrations are set. Reasoning that natural selection will drive cells towards achieving a given physiological state at minimal cost, we derive a general equation that predicts the concentration of a metabolite from the concentration of the most abundant and costly enzyme consuming it. Simulations of cellular growth as well as experimental data demonstrate that costs are approximately proportional to molecular masses. For effectively irreversible reactions, the cell maximizes its metabolic efficiency by investing equally into substrate and unbound enzyme molecules. Without fitting any free parameters, the resulting model predicts in vivo substrate concentrations from enzyme concentrations and substrate affinities with high accuracy across data from *E. coli* and diverse eukaryotes (R^2=0.79, geometric mean fold-error 1.74). The corresponding organizing principle – the minimization of the summed mass concentrations of solutes – may facilitate reducing the complexity of kinetic models and will contribute to the design of more efficient synthetic cellular systems.

T-12

Title: Spatiotemporal dynamics of gut microbiota from *in vitro* and *in silico* models

Author: Terrance Hwa

Primary affliation: University of California, San Diego

Abstract: The human gut harbors a dynamic microbial community whose composition bears great importance for the health of the host. Here, we investigate how colonic physiology impacts bacterial growth behaviors, which ultimately dictate the gut microbiota composition. Combining measurements of bacterial growth physiology with analysis of published data on human physiology into a quantitative modeling framework, we show how hydrodynamic forces in the colon, in concert with other physiological factors, determine the abundances of the major bacterial phyla in the gut. Our model quantitatively explains the observed variation of microbiota composition among healthy adults, and predicts colonic water absorption (manifested as stool consistency) and nutrient intake to be two key factors determining this composition. The model further reveals that both factors, which have been identified in recent correlative studies, exert their effects through the same mechanism: changes in colonic pH that differentially affect the growth of different bacteria. Our findings show that a predictive and mechanistic understanding of microbial ecology in the human gut is possible, and offer the hope for the rational design of intervention strategies to actively control the microbiota.

T-13

Title: How a few tolerant individuals can save a population under stress: a model of growth, death, and phenotype switching during formaldehyde exposure

Authors: Jessica Lee, Siavash Riazi, Christopher H. Remien, Christopher J. Marx

Primary affiliation: University of Idaho

Abstract: Apparent lags in population behavior occasionally have been revealed to actually be caused by population heterogenetiy. Rare "persister" cells are one such example, where a small minority of non-growing cells can tolerate lethal doses of antibiotics. These dynamics are well-modeled as two discrete populations that phenotypically switch states and have different growth and death characteristics. We have recently discovered the critical role of phenotypic heterogeneity for survival of *Methylobacterium extorquens* to threshold levels of formaldehyde. Although formaldehyde is an intermediate of methanol metabolism in these organisms, concentrations above 2 mM lead to loss of viability. We have found that there are is a wide, continuous distribution of cellular tolerance to formaldehyde, such that an ever-smaller fraction of cells can survive and grow at ever-higher concentrations. We have developed a novel mathematical model of tolerance using a PDE to capture this continuum of tolerance levels, such that state-switching is best expressed as an advection-diffusion process. Our growth/death/switching model captures the critical features of the system, such as the rate and formaldehyde tolerance of switching, that are nearly impossible to measure experimentally; and to predict the conditions under which a minority of tolerant cells can rescue the whole population through removal of a communal stressor.

T-14

Title: Connecting Flux Balance at the Environmental and Organismal Levels

Author: Isaac Klapper

Primary affiliation: Temple University

Abstract: Metabolic pathway analysis, and flux balance in particular, has been extensively studied over the past decades at the organismal level. More recently, attention is also being directed to community level balances and, as a natural extension, to the larger scale environment in which that community functions. Such extensions are warranted as flux at the organismal level is ultimately subject to constraint by flux and transport at the environmental level. We discuss mathematical (space and time multiscale issues) and physical/chemical (transport and energetic issues) aspects that arise in the context of specific examples of phototrophic and medically-related microbial communties. A key issue is how to efficiently connect organismal level processes (e.g. metabolic fluxes) with environmental level processes (e.g. diffusive transport) while respecting differences in spatial and temporal scales.

T-15

Title: Exploring the metabolic potential of human gut microbiota

Author: Ines Thiele

Primary affiliation: Luxembourg Centre for Systems Biomedicine, University of Luxembourg, http://thielelab.eu

Abstract: Computational modeling of microbial metabolism has gained increasing attention for phenotypic characterization. Such modeling is achieved by assembling in a bottom-up manner a high-fidelity computational representation of a microbe's metabolic network based on genomic, biochemical, and physiological data. Various computational tools exist to characterize its phenotypic properties (e.g., amino acid or vitamin production capabilities, potential carbon and energy sources). Until recently, only few refined metabolic models of human gut microbes had been published. We have developed a semi-automated pipeline for the assembly of high-fidelity microbial metabolic networks and applied this pipeline to generate a collection of more than 770 gut microbial metabolic networks. We then phenotypically characterized these microbes *in silico* and validated predictions obtained for two microbes *in vitro*. I will show that we can now combine these metabolic networks, e.g., based on metagenomic data, to predict emergent metabolic capabilities of the microbial community and their potential effect on the human host. As such, computational modeling can expand and accelerate the insight gained from metagenomic studies in health and disease.

T-16

Title: *in silico* and *in vitro* analysis of resource allocation in biofilm consortia

Authors: Ross Carlson, Mathew Fields, Tomas Gedeon, Luke Hanley, Michael Henson, Jeffrey Heys

Primary affiliations: MSU Bozeman, University of Illinois, Chicago, University of Massachusetts, Amherst

Abstract: Multispecies biofilms are ubiquitous in medical, environmental, and engineered microbial systems. In fact, most naturally occurring microorganisms exist in biofilm consortia. These complicated, self-assembling communities compete for limited resources via interactions between cells and their environment. While foundational to the majority of microbial life, the basic design principles including resource allocation strategies of consortia biofilms are poorly understood. Design principles for community resource allocation have been extracted from three experimental systems including a medical chronic wound consortium, an environmental extremophile consortium and a synthetic, engineered bacterial consortium. The three systems are all being analyzed through a combination of in silico and in vitro studies. Predictive multiscale modeling frameworks including flux balance analysis-based and agent-based models are generating quantitatively accurate predictions of biofilm dynamics, species distributions and responses to perturbations. The in silico research is complemented with in vitro studies including spatially resolved biofilm measurements including dissolved oxygen fluxes and species distributions to quantify the physiologies and resource budgets of consortia members. These three distinct, yet tractable, systems share basic behaviors that can be explained using two powerful ecological theories, namely the resource ratio theory and the maximum power principle. The shared system properties including improved stress tolerance, metabolite cross feeding and enhanced productivity are important design criteria for the rational control of medical, environmental and applied biosystems.

T-17

Title: Elementary Flux Vectors: Closing the Gap between Elementary Flux Modes and Flux Balance Analysis in Metabolic Networks

Authors: Steffen Klamt, Regensburger G, Gerstl MP, Jungreuthmayer C, Schuster S, Mahadevan R, Zanghellini J, Müller S.

Primary affiliation(s): Max Planck Institute, Magdeburg, University of Toronto, Jena University

Abstract: Elementary flux modes (EFMs) provide a valuable approach to explore the space of feasible steady-state flux distributions in metabolic networks. However, EFMs cannot account for inhomogeneous constraints such as known lower or upper flux bounds or allocation constraints frequently used in the context of flux balance analysis. These constraints turn the solution space from a flux cone to a flux polyhedron. In order to generalize EFMs from flux cones to flux polyhedra, the concept of elementary flux vectors (EFVs) was proposed by Urbanczik one decade ago [IET Systems Biology (2007) 1:274-279]. So far it has attracted much less attention than EFMs, possibly because the concept seems, at a first glance, to be more involved. Moreover, apart from some specific uses, the whole spectrum of potential applications of EFVs has not been clearly communicated so far.

We here revisit the concept of EFVs, emphasize the close relationships between EFMs and EFVs, and highlight that almost all applications of EFMs are, in an analogous manner, possible with EFVs in flux polyhedra. In fact, certain properties can only be studied with EFVs. We conclude that EFVs provide a powerful and unifying framework for constraint-based modeling of metabolic networks because they close the gap between EFM analyses (operating on the flux cone) and flux balance analysis and related optimization techniques (usually operating on a flux polyhedron due to inhomogeneous constraints). We also show that EFVs can be calculated by well-established algorithms developed for computation of EFMs, which should boost the applicability of the approach.

Reference: Klamt S, Regensburger G, Gerstl MP, Jungreuthmayer C, Schuster S, Mahadevan R, Zanghellini J, Müller S. (2017) From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. PLoS Comput Biol 13:e1005409.

T-18

Title: Extremum Principles in Metabolic Network

Authors: John Barrett, Friedrich Srienc

Primary affiliation: University of Minnesota, Chem. Eng. / BioTechnology Institute

Abstract: Living cells represent open systems that convert inflowing nutrients via their metabolism into products for excretion. Statements of mass conservation, in the form of balanced stoichiometric reactions, can be used to define metabolic models that describe the fundamental pathways of functioning cells. In addition, the operation of these fundamental pathways is constrained by the principles of thermodynamics. To investigate the relationships between reaction kinetics, material conversions and thermodynamic constraints, we have studied the behavior of a continuous stirred tank reactor as a well-defined representation of an open system. Specifically, we are interested in the operational constraints that result from extremum conditions on thermodynamic quantities such as entropy, Gibbs free energy, and their rates of production. The results of this study provide insight into the operation of metabolic reaction networks and their evolution in growing cell populations.

T-19

Title: Identifying optimal metabolic nodes and intervention strategies for dynamically controlled microorganisms using two stage minimal cut sets

Authors: Naveen Venayak, Axel von Kamp, Steffen Klamt, Radhakrishnan Mahadevan

Primary affiliation: University of Toronto, Max Planck Institute, Magdeburg

Abstract: Microorganisms are capable of producing a large variety of valuable chemicals. Due to stoichiometric constraints, metabolic optimization of these cell factories for chemical production can come at the expense of native functionality, such as cell growth. This trade-off is of immense importance due to its significant impact on overall productivity. To overcome such limitations, there has been recent interest in implementing sensor-actuator circuits to dynamically control metabolism, and gain temporal control of the microbial phenotype. These systems can be used to implement a two-stage fermentation, where the first stage is dedicated to cell growth and the second stage to production. Here, we present the first algorithm to determine optimal interventions to dynamically control metabolism for chemical production, by eliminating unwanted (low yield) phenotypes. This algorithm is based on the minimal cut set (MCS) algorithm and searches for a suitable minimal combination of static and dynamic (valve) interventions. The static interventions allow for a high growth rate (first stage), while additionally switching off the valves in the production stage enables high-yield production. We exploit the computational efficiency of this algorithm to determine metabolic valves which can be applied to a broad spectrum of products, leading to targets for the creation of platform strains. Furthermore, we explore the efficacy of controlling the phenotype based on oxygen availability, which has historically been a preferred method to implement two-stage fermentation. This algorithm presents for the first time a direct and efficient route to determine optimal interventions for dynamically controlled microbes.

T-20

Title: Model-Guided Engineering of Microbial Biocatalysts

Author: Jennifer L. Reed

Primary affiliation: University of Wisconsin, Madison

Abstract: Microbes have been engineered to produce a variety of chemicals, including biofuels, commodity chemicals, specialty chemicals, and therapeutics. Chemical production can be enhanced by connecting synthesis pathways to host metabolism, re-wiring regulatory networks, improving precursor production, and optimizing gene expression. A number of computational systems biology approaches have been developed to facilitate metabolic engineering efforts by suggesting which combination of genetic changes would improve chemical production. Network analysis methods can be used to identify paths from inexpensive substrates (e.g., sugars) to high-value chemical products and to identify central metabolic precursors that can be converted into a variety of chemical products. Genome-scale metabolic models can be used to predict how gene deletions, gene additions, and gene expression changes would impact chemical product yields, growth rates, and/or productivities. Additionally, machine learning and active learning algorithms can be used to optimize gene expression constructs to efficiently convert metabolic precursors into desired products. Case studies will be presented that show how computational tools can guide development of strains with enhanced sugar utilization, precursor production, and chemical production. Together this work illustrates how integrating computational and experimental efforts can lead to the rapid development of microbial biocatalysts.

T-21

Title: Optimizing the production of bulk chemicals from carbon monoxide using a genome-scale model of *Clostridium autoethanogenum*

Authors: Rupert Norman, Millat T, Schatschneider S, Henstra AM, Hartman HB, Poolman MG, Fell DA, Winzer K, Minton NP, Hodgman C

Primary affiliation(s): The University of Nottingham, Oxford Brookes University

Abstact: Recent international directives promoting the reduced consumption of fossil fuels have warranted methods for effective carbon recycling. Subsequently, *Clostridium autoethanogenum* has attracted academic and industrial interest due to its ability to convert syngas components (CO, CO_2 & H_2) into valuable platform chemicals, including ethanol and 2,3-butanediol - a jet fuel additive. Developing the metabolic conversions catalysed by *C. autoethanogenum* into an efficient bioprocess requires the accurate prediction of optimal metabolic steady states, which in turn necessitates the construction of a genome-scale model (GSM).

We have successfully constructed a predictive model, suitable for the integration of omics data sets and prediction of gene knock-out targets, consisting of 795 reactions and 786 metabolites. Our model-simulated growth yields agree well with experimentally observed specific growth rates, while elementary modes analysis (EMA) confirms the availability of metabolic routes for acetate, ethanol, lactate and butanediol production. Elevated ethanol production is predicted to result from a reduction in pH levels. Similarly, we found that the switch from acetate to ethanol production occurs with increasing CO uptake rates under non-carbon limited conditions, finally leading to lactate production as a consequence of electron stress. Our results are consistent with trends observed in continuous cultures.

Our interdisciplinary approach for the construction, analysis and application of a genome-scale model provides insight into biological and biochemical principles which govern experimentally observed metabolic behaviour. Our results offer a rationale to aid the optimization of commodity chemicals from waste gases on an industrial scale.

T-22

Title: Metabolic modeling in food biotechnology

Authors: Ahmad Zeidan, Ana Rute Neves

Primary affiliation: Chr. Hansen A/S Denmark

Abstract: Genome-scale metabolic network reconstruction and constraint-based modeling (CBM) are increasingly important tools in microbial systems biology. The majority of CBM applications are focused on rationally identifying metabolic engineering strategies for strain improvement. In the food industry, however, strict legislations and the negative perception by consumers of genetically modified foods render the use of recombinant DNA technologies inapplicable.

Here, we illustrate the primary non-GMO-based applications of CBM at Chr. Hansen, a global supplier of food cultures and enzymes. First, the use of strain-specific genome-scale models (GEMs) for unraveling the metabolic potential of industrial strains and optimizing their biomass yield will be presented, using the probiotic bacterium *Bifidobacterium animalis* ssp. lactis BB-12® as an example. An extensively curated and validated GEM for the strain was developed and used to identify its essential nutritional requirements, design a chemically defined medium supporting its growth, and gain new insights into its fructose and sulfur metabolism as well as vitamin and amino acid biosynthesis. Secondly, a framework for CBM-based identification of metabolite analogues for rational strain improvement via non-GMO approaches will be described. The framework provides an alternative approach to strain design by searching for metabolite targets, which when 'knocked-out' in presence of their analogues can result in a desirable phenotype following dominant selection or adaptive laboratory evolution. Finally, the use of CBM for the mechanistic understanding of biological processes in microbial communities in food products will be illustrated. Current limitations and challenges for the full exploitation of CBM in food biotechnology will also be discussed.

T-23

Title: Explaining the asymmetric label incorporation during photosynthesis

Author: Oliver Ebenhöh

Primary affiliation: Heinrich Heine University Düsseldorf

Abstract: Sixty years ago in 1957, Martin Gibbs discovered that radioactively labelled carbon dioxide is asymmetrically incorporated into sugars during photosynthesis. This observation, later termed 'Photosynthetic Gibbs Effect', was puzzling and appeared counter-intuitive, because RuBisCO, the enzyme fixing carbon dioxide to a five-carbon sugar,releases two identical three-carbon molecules, from which sugarars e symetrically formed. Many different explanations have been proposed to explain the observed asymmetries, and as usual the simplest were also the most plausible. Already in 1964, James Bassham explained the appearance of asymmetries by different pool sizes of intermediates and argued that other reproducible patterns result from a 'quirk' of carbons by reversible reactions catalysed by transketolase. Despite such plausible qualitative arguments, a quantitative explanation of the observed labelling dynamics has never been given.

Here, we propose a simple model of the Calvin-Benson-Bassham cycle, which is based on thermodynamic considerations of the cycle and focusses on the paths of carbon atoms. We demonstrate that the observed patterns of label incorporation are an emergent property of the cycle's dynamics and do not require any further assumptions beyond the cycle's stoichiometry and thermodynamics. The observed patterns are a result of the particular thermodynamic properties, which clearly separate the enzymatic steps into close-to-equilibrium and far-from-equilibrium reactions.

With our model, we can quantify the effect of single enzymatic steps on the label incorporation and thus we provide the first fully quantitative explanation of the Photosynthetic Gibbs Effect six decades after its discovery.

T-24

Title: Elementary modes analysis of photorespiration

Authors: David Fell, Benazir Huma, Mark G Poolman, Sudip Kun

Primary affiliation: Oxford Brookes University

Abstract: Photorespiration is the metabolism associated with recovery of Calvin cycle metabolites after Rubisco catalyses the addition of oxygen to ribulose bisphosphate instead of carbon dioxide. Starting from a genome scale metabolic model of C3 plant metabolism, we extracted a smaller submodel of the reactions associated with photorespiration in the chloroplast, peroxisome and mitochondrion, as well as the transport reactions of the associated metabolites between compartments. This model had 28 elementary modes of photorespiratory metabolism. Amongst these is the classic photorespiratory cycle, but there are also previously undescribed pathways that involve photorespiration coupled with mitochondrial metabolism and ATP production, the glutathione-ascorbate (GSH--ASC) cycle and nitrate reduction to ammonia. Thus these modes demonstrate the underlying basis of the metabolic linkages with photorespiration that have been inferred experimentally, but especially in the case of nitrate assimilation, have not previously been satisfactorily explained. The set of reactions common to all the elementary modes shows good agreement with those of the gene products of mutants that have been reported to have a photorespiratory phenotype. Finally, the set of modes provides a demonstration that photorespiration itself has no intrinsic impact on the assimilation quotient (CO_2 fixed per O_2 released), except in those modes associated with concomitant nitrate reduction, contrary to widespread belief.

T-25

Title: Flux analysis of the plant MEP pathway

Authors: Johann Rohwer, Erica Perreca, Bettina Raguschke, Diego González Cabanelas, Jonathan Gershenzon, Louwrance Wright

Primary affiliation(s): Stellenbosch University & MPI for Chemical Ecology

Abstract: The methylerythritol phosphate (MEP) pathway produces the common precursors for all terpenoids in most bacteria and the chloroplasts of higher plants, playing a central role in the synthesis of plant photosynthetic pigments and defence compounds. We present two studies on the regulation of the MEP pathway. First, we investigated in Arabidopsis whether a futile cycle exists in the transport of the intermediate, methylerythritol cyclodiphosphate (MEcDP), between the chloroplast and cytosol. From isotopic labelling data, plastidic and extraplastidic pool sizes of MEcDP were calculated throughout a 10-h daylight period, which enabled the calculation of an MEcDP export rate. Strikingly, the export flux was orders of magnitude smaller (0.4%) than the main MEP pathway flux, and largely irreversible. Significant futile cycling of MEcDP was thus absent.

In a second study, the effect of drought stress on MEP pathway flux was analysed in spruce. The rate of isoprene emission (the pathway product) was measured directly with mass spectrometry. In parallel, the metabolic pool sizes of three MEP pathway intermediates were measured under the same conditions. MEP pathway flux was calculated from label incorporation into these intermediates from $^{13}CO_2$. While the MEP pathway flux decreased significantly under drought stress, the directly measured rate of isoprene emission was largely unaffected. Under normal conditions, MEP pathway flux was 4–5 times higher than the isoprene emission flux, showing that photosynthetic pigments were the major products. Under drought-stressed conditions, this factor reduced to three-fold, suggesting a change in carbon allocation to pathway products.

T-26

Title: Dynamic modelling and flux analysis

Author: Mario Jolicoeur

Primary affiliation: École Polytechnique de Montréal

Abstract: The use of dynamic models to describe a cell behaviour is gaining in interest. Besides its capacity to "animate" a biosystem metabolic network, such models also allow following the time-evolution of metabolic fluxes performing a dynamic flux analysis. Although in silico, such simulation tool enables observations that are tedious or impossible to perform experimentally. The starting cost for developing a model with each flux kinetics described relies on acquiring experimental concentration data of extracellular and intracellular volumes, as well as literature data on kinetic parameters, when available. Expressomic data is obviously a plus. From this knowledge, the process of building the model structure can be initiated. Various optimization algorithms can be applied to estimate parameters value, based on a sensitivity analysis on model behaviour. This process is highly instructive while refining model structure testing hypothesis on the network map and flux regulation mechanisms. Starting with a minimal metabolic network can be a wise idea to accelerate the model development process. Therefore, once anchored on an experimental reality, model simulations, which only require initial conditions and kinetic parameters value, will allow questioning the cell metabolic behaviour and thus visualise flux dynamics with time as well as the specific contribution of each regulation mechanisms. Examples on the use of such dynamic models with bacteria, microalgae, plant and animal cells will be shown, for biomedical and bioprocess applications, under fed-batch and perfusion culture modes. Current limitations and future trends will be discussed.

T-27

Title: Towards modeling dynamic regulations in ecosystems

Authors: Antonella Succurro, Oliver Ebenhoeh, Daniel Segrè

Primary affiliation(s): Cologne University, Heinrich Heine University Düsseldorf, Boston University

Abstract: Ecosystems can be defined as different organisms interacting with each other as well as with the environment. The mechanisms regulating the interplay between each unit are still to be understood, but it is known that metabolic interactions play a central role. To capture the ecosystem dynamics in a quantitative and predictive theory is today a compelling and challenging task.

Constraint based methods like Flux Balance Analysis (FBA) are powerful ways to investigate reaction flux distributions of Metabolic Network Models (MNM) at the steady state. By construction, however, information on metabolite concentrations or reaction kinetics is lost, making it impossible to capture the dynamics of ecosystem interactions. Another aspect lost is the action of cofactors (typically micronutrients, like Iron or vitamins), which are known to be often exchanged for organic carbon sources in mutualistic consortia between eukaryotes and bacteria.

To overcome these limitation we developed an integrated modeling framework coupling dynamic FBA (dFBA) with Ordinary Differential Equations (ODEs) systems in order to interface the metabolic steady state (FBA solutions) with biochemical processes and regulatory mechanisms (pure ODEs). In this way we bridge metabolic adjustements with ecosystem regulation mechanisms and can include the effect of micronutrients availability. We present a case study on a small toy network where we introduce a storage system that allows for the accumulation of metabolites and accounts for regulation of the pool level in a way inspired by control theory and two competing networks are simulate

T-28

Title: Progressing Towards a Deep Integration of Chemistry and Biology to Discover New Protein Functions, Pathways, and Ecological Principles

Authors: Christopher Henry, Jose Faria, James Jeffreyes, Janaka Edirisinghe, Pam Weisenhorn, Nidhi Gupta, Sam Seaver, Andrew Hanson, Keith Tyo, Ronald Taylor, Hyun-Seob Song, Hans Bernstein

Primary affiliation(s): Argonne National Laboratory, Pacific Northwest National Laboratory

Abstract: We are currently exploring how metabolic modeling, cheminformatics, comparative genomics, and omics data integration can be combined to discover new biology in individual microbes and microbial communities. We have applied this approach to discover new protein functions, annotate pathways, and predict interactions between species within a microbiome.

Metabolic models lie at the apex of our approach, serving as the abstraction that is used to integrate all other approaches together. As such, the quality of these models is paramount. Thus we recently invested significant effort improving the quality of our model reconstruction pipeline, ensuring that models represent annotated metabolic functions more comprehensively, and ensuring that the solutions selected by gapfilling algorithms do not result in biologically unreasonable behavior.

Next, we explore how cheminformatics may be merged with metabolic models to discover new protein functions. Specifically, by applying chemical rules to predict potential spontaneous chemistry that may take place inside the cell, we identified regions of metabolism where proteins are required to mitigate the effects of spontaneous metabolite damage. This led to the identification and characterization of new damage control enzymes in the metabolic network, resulting in a more comprehensive understanding of metabolism, including extensive spontaneous reactions that are likely taking place. Our broad application of reactions rules representing both spontaneous and enzymatic chemistry has also led to the identification of numerous unknown peaks in metabolomics datasets.

We then combine cheminformatics with comparative genomics and omics data to identify genes responsible for catalyzing a pathway known to occur in a specific organism. While no single component of this approach is sufficient to completely narrow the list of gene candidates, but applying all of these approaches together, we can arrive at a top set of candidate genes with a high level of confidence.

Finally, we apply all of these approaches to study interactions between multiple species within a microbiome, identifying the metabolites exchanged between species, and filling the gaps required to permit the microbiome system to function in the environment where it is observed to grow. These studies reveal key ecological principles that govern the behavior and structure of the species interactions within these systems.

All of these tools and capabilities have been fully integrated and released within the DOE Systems Biology Knowledge. We will discuss how the tools in KBase may now be applied to perform similar studies on new datasets uploaded by external users. We will also point to KBase Narratives that include all described studies in detail.

T-29

Title: Grow or store? Exploring metabolic decision making under feast/famine conditions using dynamic ^{13}C flux analysis

Authors: Leonor Guedes da Silva, Koen Verhagen, Andy Wiranata Wijaya, Robbert Kleerebezem, Mark C.M. van Loosdrecht, Aljoscha S. Wahl

Primary affiliation: TU Delft

Abstract: Natural habitats of microorganisms are dynamic environments with non-continuous supply of carbon and energy sources, in which intermediate storage of substrates can increase competitiveness. *Plasticicumulans acidivorans* are polyhydroxybutyrate (PHB) accumulating bacteria enriched from activated sludge using carbon feast-famine cycles as selective pressure. Despite growing slowly, *P. acidivorans* outcompetes other bacteria by quickly taking up acetate and storing it intracellularly as PHB to later use it for growth. As soon as acetate is depleted, these bacteria immediately 'switch' their metabolism from PHB production to consumption entailing a very interesting regulatory challenge as parallel activity could lead to significant losses (futile cycling). While the stoichiometry for both feast and famine phases has been extensively described in literature, the switch regulation is not yet fully understood.

To elucidate the responsible regulatory processes, an enrichment of *P. acidivorans* was studied using targeted intracellular metabolite analysis over time, with emphasis on the feast to famine switch. In combination with extracellular rates, the measured intracellular metabolite pools are used to design a labelling experiment to obtain actual intracellular fluxes (dynamic ^{13}C flux analysis). Here the challenge is to create an isotopically non-stationary state (usually mediated by changing the substrate's isotopic composition) to study the metabolic response in the transition from presence-to-absence of substrate.

In this way, we aim to unravel the responsible regulatory mechanism governing the metabolic switch from storage-to-consumption and use this knowledge not only to understand its ecological relevance, but to also propose novel metabolic strategies for microbial cell factory design.

T-30

Title: Modeling cyanobacterial growth

Authors: Ralf Steuer, A-M Reimers, H Knoop, A Bockmayr

Primary affiliation: Humboldt-University Berlin

Abstract: Photoautotrophic growth requires a highly coordinated distribution of cellular resources to different intracellular processes, including the de novo synthesis of proteins, ribosomes, lipids, and other cellular components. In our contribution, we present a computational framework to investigate the optimal allocation of cellular resources during diurnal phototrophic growth using a genome-scale metabolic reconstruction of the cyanobacterium *Synechococcus elongatus* PCC 7942. Specifically, we formulate phototrophic growth as an autocatalytic process and solve the resulting time-dependent resource allocation problem using constraint-based analysis. Based on a narrow and well-defined set of parameters, our approach results in the prediction of growth properties over a full diurnal cycle. The computational model allows us to study the optimality of metabolite partitioning during diurnal growth. The cyclic pattern of glycogen accumulation is an emergent property of the model and has timing characteristics that are in excellent agreement with experimental findings. Our approach provides insight into the time-dependent resource allocation problem of phototrophic diurnal growth and may serve as a general framework to assess the optimality of metabolic strategies that evolved in phototrophic organisms under diurnal conditions.

T-31

Title: Dynamic Metabolic Flux Analysis of Oil Biosynthesis in *Camelina sativa* Seeds

Authors: Teresa J Clark, Mike Pollard, and Yair Shachar-Hill

Primary affiliation: Michigan State University

Abstract: Seed oil is of great economic importance for food, animal feed, and industrial applications. Triacylglycerol (TAG), the major seed storage lipid, is composed of a glycerol backbone and fatty acid hydrocarbon chains. Fatty acid modification and addition to glycerol can occur through alternative routes within a metabolic reaction network, but the relative importance of fluxes through these alternative routes is unclear. In *Camelina sativa*, a promising oilseed crop, we delineate these fluxes and demonstrate the existence of two distinct metabolically active pools of phosphatidylcholine (PC), which is a key intermediate in oilseed TAG biosynthesis. The first PC pool is used in *de novo* synthesis of diacylglycerol (DAG), the immediate precursor of TAG. This is evidenced by initial DAG products containing higher levels of sn2 labeling compared to sn1 labeling (~9:1). This pool comprises ~60% of total PC. The second PC pool is involved in acyl editing and its formation is catalyzed by the Rod1 enzyme. This is evidenced by ~40% of total PC containing similar labeling in sn1 and sn2 (~1:1). We present a quantitative flux map for TAG biosynthesis, which will facilitate rational engineering of seed oils by providing a means for predicting the effectiveness of altering the expression of genes associated with this pathway.

T-32

Title: Modelling the phosphorus pools of *Chlorella vulgaris*

Authors: Dipali Singh, Ines Hotopp, Oliver Ebenhöh

Primary affiliation: Heinrich Heine University

Abstract: In the interdisciplinary collaborative project “AlgalFertilizer” we investigate the capability of *Chlorella vulgaris* to uptake P from waste water and the molecular mechanisms underlying P uptake and storage in algae. For this, we develop mathematical models to understand the uptake of P and the underlying dynamics of the conversion of P pools in algal cells under different environmental conditions.??

The model developed in our study consists of four phosphate pools : the external inorganic phosphate (Pex) pool which provides the phosphate source, and three internal algal phosphate pools. The external phosphate is taken up as free inorganic phosphate (Ortho-P), polyphosphate (Poly-P) is used as short term storage, and organic phosphate (Organic-P) is essential for the algal growth. In addition, the model also takes into account the influence of light on algal growth through an increasing biomass density and the dependence of growth on the availability of growth machinery (proteins). The model is able to simulate the characteristic lag phage at the beginning of the algal growth, the uptake of external phosphorus into an algal cell and the dynamic distribution between the Poly-P and the Ortho-P pools within the cell. It provides an insight on the underlying dynamics of different P pools and supports the hypothesis that the lag phase results from the 'legacy' of the starvation phase, in which many important proteins for reproduction have been degraded, and therefore the growth machinery needs to be re-established. This also lead to quantitative predictions, that can be experimentally tested in growth experiments.?

M-1

Title: The COBRA Toolbox 3.0 and beyond

Authors: Ronan Fleming, The COBRA Toolbox developers consortium

Primary affiliation: University of Luxembourg

Abstract: The COnstraint-Based Reconstruction and Analysis Toolbox (COBRA) is a comprehensive software suite for quantitative prediction of cellular and multicellular biochemical networks with constraint-based modelling. It implements a comprehensive collection of basic and advanced modelling methods, including reconstruction and model generation as well as biased and unbiased model-driven analysis methods. Since its release in 2011, The COBRA Toolbox 2.0 has been widely used for modelling, analysing and predicting a variety of metabolic phenotypes using genome-scale biochemical networks. Here we present the main developments in The COBRA Toolbox 3.0, including expansion of functionality to cover existing and new modelling methods, implementation of support for input and output of new standards for model sharing formats, implementation of support for an extended suite of optimisation solvers, multi-lingual integration with C, FORTRAN, Julia, Perl and Python code for enhanced functionality in the areas of computational efficiency, high performance computing and integration with omics data. This enhanced functionality now enables the modelling of communities of biochemical networks, from microbial communities to multicellular mammalian systems. In particular, the integration with a new set of novel high numerical precision optimisation solvers enables robust and efficient modelling of multi-scale biochemical networks, such as those obtained from integration of metabolism and macromolecular synthesis. With an increasing number of contributions from developers around the world, the code base is evolving at a fast pace. In order to guarantee consistent, stable, and high quality code, a continuous integration approach has been implemented. A semi-automated continuous integration environment ensure that every change in code, as minor as it may be, is fully tested before being released as part of a stable version. Each contribution is verified automatically, and for each code submission, a comprehensive test suite is run in order to detect bugs and integration errors before release. Consequently, the automated testing environment ensures higher quality code and the release of well tested computer code. All documentation and code released as part of the openCOBRA project on https://github.com/opencobra/cobratoolbox and https://github.com/opencobra/cobratoolbox respectively. We conclude with a description of future plans for The COBRA Toolbox.

M-2

Title: Improving automated model reconstruction

Authors: José Faria, Janaka N. Edirisinghe, Filipe Liu, Christopher S. Henry

Primary affiliation: Argonne National Laboratory

Abstract: The Department of Energy Systems Biology Knowledgebase (KBase) is a platform designed to solve the grand challenges of Systems Biology. KBase has implemented bioinformatics tools that allow for multiple workflows including genome annotation, comparative genomics, and metabolic modeling. KBase now also includes a comprehensive database of over 80K reference genomes (approximately 5K complete genomes) from NCBI's RefSeq. We have selected a phylogenetically diverse set of approximately 1000 genomes and built draft genome-scale metabolic models constructed using the ModelSEED pipeline. In constructing these models, we were interested in improving the gene functional role – reaction templates and accuracy of biomass compositions used by ModelSEED when producing a draft model. We have curated the existing ModelSEED templates by both fixing and introducing new gene – reaction mappings. Previous ModelSEED biomass compositions often excluded metabolites that are essential for the viability of many organisms, while simultaneously including some non-essential metabolites. These errors in the biomass compositions lead to errors in the growth conditions and gene knockout phenotypes predicted by the models. To correct these problems, we conducted a sensitivity analysis on our full set of models, using a biomass composition that includes all possible essential biomass precursors. We also improved the ModelSEED gap filling database with new rules to restrict the addition of non-biological significant reactions. We then used this data to improve all of our models, validating our improved models with a diverse set of growth and knockout phenotype data. Our improved models are now available for download, viewing, and comparison in the KBase.

M-3

Title: Unification of Genome Scale Models

Authors: Filipe Liu, Xavier J, Ramalho F, Rocha M, Rocha I

Primary affiliation: University of Minho

Abstract: Genome scale metabolic models (GEMs), along with constraint based analysis techniques, are capable of predicting key metabolic characteristics of organisms, revealing their limitations and strengths.

Up until today, hundreds of curated GEMs have been published. These models provide rich curated information for future reconstructions and studies. However, a common drawback of combining several GEMs is the lack of standardization between them. Recently, genome scale model repositories have been developed, however unable to include most of the published GEMs or performing standardization.

An integration pipeline was thus developed to standardize GEMs, and tested against a set of nearly one hundred published models of prokaryotes. These models contain a high diversity of annotation strategies and were built from early 2000 to the present date.

The pipeline follows a cyclic semi-automated integration process. User interaction is necessary only for profiling and report analysis, which was essential to detect faults found in a few GEMs, allowing an easy curation. Identification of the metabolites is fully automated using several methods, such as taking the annotation found within the model, pattern decomposition of the identifiers, name similarity against compound databases, and user provided input (e.g., curation, supplementary materials). The integration scored an average of 70% integration in the model metabolites.

Standardization of drain reactions is essential to enable compatibility between models, and flux balance analysis was used for validation purposes. A total of 75% of the GEMs shown positive growth for default uptake conditions, in 20% no bounded uptake constraints were detected, while the remaining 5% had no growth in all conditions tested.

M-4

Title: Memote - A testing suite for constraints-based metabolic models

Authors: Christian Lieven, Moritz E. Beber, Nikolaus Sonnenschein

Primary affiliation: Novo Nordisk Foundation Center for Biosustainability

Abstract: Constraints-based metabolic models have become fundamental and trusted tools in systems biology. Several layers of biological information are combined in a compact format in order to describe a metabolic model. A richly annotated model is required for its various areas of application and represents a veritable knowledge base about an organism's metabolism. However, coherently describing a complex interlinked system such as metabolism is a challenge in and of itself that is only aggravated by the current lack of cohesive, widely-accepted, testable, and modern standards.

Here, we introduce memote (Metabolic Model Tests {https://github.com/biosustain/memote}), a Python package designed to run a given model through a set of hard and soft tests and generate a report that reflects model integrity. Soft tests focus on aspects that do not influence the performance of the model, such as syntactic conventions whereas hard tests determine whether a model is fully functional.

While memote can be run locally as a stand-alone testing suite, it shows its full potential when combined with web-based version controlling (Github) and continuous integration tools (Travis CI). Every tracked edit of a model automatically triggers the memote test suite, and generates a corresponding report that facilitates factual debate of model changes.

Thus, memote not only allows researchers to more quickly iterate through the design-build-test cycle but also provides the scientific community with a measure of quality that is consistent across setups, as well as an opportunity to interact and collaborate by establishing workflows for publicly hosted and version controlled models.

M-5

Title: MFAPipe: Open source software for parallel labeling, steady state metabolic flux analysis.

Author: Mark Borkum

Primary affiliation: Pacific Northwest National Laboratory

Abstract: Modern fluxomics studies are enabled by robust, high-performance software packages that empower their users to validate and verify their hypotheses with the support of both experimentally obtained and computationally simulated, multi-instrumental data. Recent advances [1] in the mathematical formulation of stable isotopic labeling (SIL) models have enabled the development of new scientific capabilities; most pertinently, the new ability to fit model parameters to smooth functions of arbitrary isotopic labeling states of heteronuclear moieties; enabling the new ability to fit model parameters to experimentally obtained data from both low- and high-resolution mass spectrometry (MS) experiments, and both low- and high-dimension nuclear magnetic resonance (NMR) experiments.

In this presentation and corresponding workshop, we motivate and discuss the development of MFAPipe; a new, open source software package for flux balance analysis (FBA) and single- and parallel-labeling, steady state metabolic flux analysis (MFA). MFAPipe has no proprietary dependencies and is distributed under an open source software license that permits both academic and non-academic use in both commercial and non-commercial settings.

Workshop participants will be guided, step-by-step, through the MFAPipe installation process, and will be given the opportunity to perform an MFAPipe-supported fluxomics investigation of the citric acid cycle for a model strain of *E. coli* using different combinations of SIL data from MS and NMR instrumentations.

The prerequisites for this workshop are Git (see https://git-scm.org for more information) and The Haskell Tool Stack (see https://www.haskellstack.org for more information). This workshop is tested on macOS and Unix-like operating systems.

References

1. Borkum, Mark I., et al. "Modeling framework for isotopic labeling of heteronuclear moieties." Journal of Cheminformatics 9.1 (2017): 14.