The 2021

Metabolic Pathway Analysis: The Frontiers

Conference
University of Tennessee, Knoxville

August 2–6, 2021 • Knoxville, Tennessee

Table *of* Contents

Introduction

The fundamental study of biology has undergone a significant transformation from qualitative description to quantitative prediction of biological systems. Metabolic Pathway Analysis (MPA) is a leading meeting for the scientific community analyzing topological structure-function relationships of metabolic networks to study fundamentals of biology and biotechnological applications.

MPA 2021 is the eighth in the biennial series that integrates a highly interdisciplinary group of international biologists, computer scientists, engineers, and mathematicians for five days of discussions focused on state-of-the-art advances in network analysis methodology, theory, and application.

The conference will focus on the frontiers of theoretical and computational approaches to decode biological principles behind metabolic networks, helping better understand rules of life and enabling biosystems design to tackle challenging problems related to health, food, energy, and the environment.

The presentations will be organized into five themes including:

- Fundamentals of Network Structure and Metabolism
- Modular and Biosystems Design
- Resource Allocation and Metabolism
- Advances in Methods, Algorithms, and Tools Development
- Applications in Systems/Synthetic Biology/Metabolic Engineering.

Unique to MPA2021 will be also a series of tutorials organized on the first day of the conference that aims to cover the latest tool development from skilled professionals.

Committee

Cong Trinh (University of Tennessee, Knoxville, USA)

Egils Stalidzans (University of Latvia, Riga)

Herbert Sauro (University of Washington, Seattle, USA)

Hyun-Seob Song (Pacific Northwest National Laboratory, USA)

Isabel Rocha (ITQB, NOVA University Lisbon, Portugal)

Mark Poolman (Oxford Brookes University, United Kingdom)

Oliver Ebenhoeh (Henrich Heine University, Germany)

Ross P. Carlson (Montana State University, Bozeman, USA)

Sabine Peres (University of Paris-Sud Paris-Saclay, France)

Stefan Schuster (Friedrich-Schiller-University Jena, Germany)

Zita Soons (Uniklinikum RWTH Aachen, Germany)

Letter of Welcome

Dear MPA2021 Attendees:

Greetings from Knoxville! On behalf of the Organizing Committee, I would like to welcome you to "The 2021 Metabolic Pathway Analysis: The Frontiers". MPA2021 is the 8th international biennial conference that integrates a highly interdisciplinary group of biologists, computer scientists, physicists, engineers and mathematicians for five days of discussions focused on state-of-the-art advances in network analysis methodology, theory and application.

The in-person attendance will be held in Strong Hall on the University of Tennessee (UT) main campus in Knoxville. Knoxville is the third-largest city in Tennessee, located at the foothills of the Smoky Mountains. It offers small town charm with the amenities you expect to find in a big city. Knoxville's downtown and Old City areas feature museums, shopping, restaurants, a movie cinema, parks, and nighttime entertainment. We are excited to connect with colleagues and friends both virtually and in person at MPA2021.

MPA2021 focuses on the frontiers of theoretical and computational approaches to decode biological principles behind metabolic networks, helping better understand rules of life and enabling biosystems design to tackle challenging problems related to health, food, energy, and the environment. The presentations feature into the traditional themes on i) Fundamentals of Network Structure and Metabolism, ii) Resource Allocation and Metabolism, iii) Advances in Methods, Algorithms, and Tools Development, and iv) Applications in Systems/Synthetic Biology/Metabolic Engineering together with a new topic on v) Modular and Biosystems Design. MPA 2021 also offers three tutorials organized on the first day of the conference that aims to cover the latest tool development from skilled professionals.

MPA2021 features two keynote speakers including Prof. Costas Maranas from Penn State University and Prof. Terence Hwa from University of California at San Diego that feature the frontiers on metabolic pathway analysis tool development and investigation into resource allocation and cellular metabolism. We also have an excellent collection of 12 invited talks from international leading experts as well as contributed talks from professionals and junior scientists with broad diversity.

Finally, we are very grateful for the financial support from the US Department of Energy, Eastman Chemical Company as well UT's Office of Research and Engagement, Tickle College of Engineering, and Department of Chemical and Biomolecular Engineering.

I hope you all enjoy the MPA2021 conference that aims to promote rigorous scientific discussion and viewpoint diversity in a collegial and welcoming environment. Stay safe and healthy!

Sincerely,

Cong T. Trinh (Chair)

Invited Speakers

Steven Abel
University of Tennessee, Knoxville, United States

Daniel Amador-Noguez
University of Wisconsin, United States

Nanette Boyle
Colorado School of Mines, United States

Anne Goelzer
Institut National de la Recherche Agronomique, France *virtual only

William Harcombe
University of Minnesota, United States

Soha Hassoun
Tufts University, United States

Christopher Henry
Argonne National Laboratory, United States

Terence Hwa
University of California, San Diego, United States *virtual only

Steffen Klamt
Max Planck Institute, Magdeburg, Germany

Costas Maranas
Pennsylvania State University, United States

Balazs Papp
Biological Research Center, Hungary

Ines Thiele
National University of Ireland, Ireland *virtual only

Sponsors

EASTMAN

The results of insight™

U.S. DEPARTMENT OF ENERGY

THE UNIVERSITY OF TENNESSEE KNOXVILLE
DEPARTMENT OF CHEMICAL & BIOMOLECULAR ENGINEERING

THE UNIVERSITY OF TENNESSEE KNOXVILLE
RESEARCH, INNOVATION & ECONOMIC DEVELOPMENT

THE UNIVERSITY OF TENNESSEE KNOXVILLE
TICKLE COLLEGE OF ENGINEERING

Tutorials

On the first day of MPA, we will be offering tutorials that aim to cover the latest tool development from skilled professors. Attendees were asked to pre-register for each tutorial. Tutorials will happen in Rooms 101, 103, and 104.

09:30–12:30 Registrations/In-parallel Tutorial Sessions

Workshop 1: Analyzing and designing metabolic networks with CellNetAnalyzer – Steffen Klamt, Max Planck Institute, Magdeburg, Germany [virtual]

CellNetAnalyzer (CNA) is a widely used MATLAB package for analyzing biological (metabolic, signaling and regulatory) networks. CNA and supports both command-line based operations as well as a graphical user interface with embedded network visualizations. In the first part of this workshop (1,5h) we will demonstrate key features of CNA for stoichiometric and constraint-based modeling of metabolic networks including flux (balance) analysis, flux and yield optimization, phase plane analysis, elementary mode analysis, computational (re)design of metabolic networks based on minimal cut sets, and many more. The second part of the tutorial (1,5h) will consist of hands-on exercises with example networks where the participants will learn how to use *CellNetAnalyzer* in practice.

Workshop 2: Working with Metabolic Kinetic Models: Deployment and Analysis – Herbert Sauro, University of Washington, Washington, United States [virtual]

In this workshop, we will introduce the tellurium python package for building kinetic models. You will learn the best strategies for publishing models so that they are reproducible. You will learn to build models using the Antimony language which is the fastest and most straight forward way to build metabolic models, as well as learn to leverage existing models from Biomodels, BiGG and other resources. We will introduce you to two new tools that you should find very useful: an easy-to-use model fitting package, SBstoat and an indispensable static analysis package SBviper for identifying bugs in models. Finally, we will illustrate some of the core features including time course simulation, steady state analysis, structural analysis, and metabolic control analysis.

Workshop 3: Metabolic Modeling in the DOE Systems Biology Knowledgebase (KBase) – Christopher Henry, Argonne National Laboratory, Illinois, United States [virtual]

The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is freely available and a developer extensible platform enabling scientists to analyze their own data within the context of public data and share findings across the system. The workshop will focus on the genome-scale modeling tools available in KBase. Participants will learn how to reconstruct and analyze metabolic models for microbes, plants, and communities.

Contacts

Conference Email:

MPA2021UT@gmail.com

Organizer Contacts:

Jeremy K. Easterday

Tel: (+1) 865-974-0250

Email: jeasterd@utk.edu

Jessica Swett

Tel: (+1) 865-974-0269

Email: jswett@utk.edu

Cong T. Trinh

Tel: (+1) 865-974-8121

Email: ctrinh@utk.edu

In any emergency situation, you should call 911.

Schedule *Monday, August 2*

09:30–12:30 **Registrations/In-parallel Tutorial Sessions**

Attendees should register for one of the below workshops.

• **Workshop 1,** Strong Hall 101: Analyzing and designing metabolic networks with CellNetAnalyzer – Steffen Klamt, Max Planck Institute, Magdeburg, Germany

• **Workshop 2,** Strong Hall 103 Working with Metabolic Kinetic Models: Deployment and Analysis – Herbert Sauro, University of Washington, Washington, United States

• **Workshop 3,** Strong Hall 104: Metabolic Modeling in the DOE Systems Biology Knowledgebase (KBase) – Christopher Henry, Argonne National Laboratory, Illinois, United States

13:00–13:10 **Introduction**

• **13:00–13:05:** Conference Chair: Cong T. Trinh, University of Tennessee, Knoxville, Tennessee, United States

• **13:05–13:10:** Conference organizer: Logistics

13:10–13:55 **Keynote Lecture**

Session chairs: Cong T. Trinh and Oliver Ebenhoeh

Bringing to bear retrosynthesis tools and enzyme engineering for pathway synthesis – Costas Maranas, Pennsylvania State University, Pennsylvania, United States

13:55–14:00 **Coffee Break**

14:00–15:30 **Applications in Systems/Synthetic Biology/Metabolic Engineering (1)**

Session chairs: Steve Abel and Herbert Sauro

• **14:00–14:20:** Identification of Metabolic Pairs Allow a Reliable and Quantitative Analysis of the Finger Sweat Metabolome – Jurgen Zanghellini, University of Vienna, Austria

• **14:20–14:40:** Metabolic Modeling of Cellular Senescence through Multi-Omics Integration – Marian Breuer, Maastricht University, Netherlands

• **14:40–15:00:** Metabolic complementarity applied to the screening of microbiota and the identification of key species – Clemence Frioux, Inria Bordeaux Sud-Ouest, France

• **15:00–15:30:** Using Machine Learning to Characterize Metabolic Pathway Activities – Soha Hassoun, Tufts University, Massachusetts, United States

15:30–15:35 **Coffee Break**

15:35–17:30 **Poster Session 1**

Block 1*

18:30–20:00 **Reception/Dinner at the Hilton** [in person only]

Schedule *Tuesday, August 3*

07:30–08:30 **Registration/Opening**

08:30–10:15 **Fundamentals of Network Structure and Metabolism (1)**

Session chairs: Herbert Sauro and Stefan Schuster

• **08:30–09:00:** Principles of metabolome conservation in mammals – Balázs Papp, Biological Research Center ELKH, Hungary

• **09:00–09:20:** GC-MS-based 13C Metabolic Flux Analysis in Pseudomonads and related species – Michael Kohlstedt, Institute of Systems Biotechnology, Saarland University, Germany

• **09:20–09:40:** Stoichiometric Modeling of String Chemistries – Devlin Moyer, Boston University, United States

• **09:40–10:00:** Modeling Lignin Biosynthesis using Proteomics and Isotopic Labeling Data in Brachpodium Distachyon – Jaime Barros-Rios, University of North Texas, United States

10:15–10:45 **Coffee Break/Networking**

10:45–12:30 **Applications in Systems/Synthetic Biology/Metabolic Engineering (2)**

Session chairs: Ross Carlson and Zita Soons

• **10:45–11:15:** Data driven advanced metabolic modeling of diurnal growth in the model green alga *Chlamydomonas reinhardtii* – Nanette Boyle, Colorado School of Mines, Colorado, United States

• **11:15–11:35:** Metabolic Network Modeling of Degrader-Cheater Interactions in a Chitin-Decomposing Model Soil Community – Aimee Kessell, University of Nebraska-Lincoln, Nebraska, United States

• **11:35-11:55:** Flux balance network expansion predicts stage-specific human peri-implantation embryo metabolism – Andisheh Dadashi, University of New Mexico, United States

• **11:55–12:00:** Dynamic flux balance analysis reveals the distribution of biochemical subtypes in CAM photosynthesis – Antonio Rigueiro-Masejo, Heinrich Heine University Düsseldorf, Germany

• **12:00–12:05:** Metabolic profiling of cultured erythroblasts for the large scale production of red blood cells – Joan S. Gallego Murillo, Delft University, Netherlands

• **12:05–12:10:** Predicting the Effect of Carbon Substrate on Bioplastic Production in *Rhodopseudomonas palustris* – Adil Alsiyabi, University of Nebraska-Lincoln, Nebraska, United States

• **12:10–12:15:** Metabolism of *Campylobacter jejuni:* an integrated *in silico* and *in vitro* study – Dipali Singh, Quadram Institute, United Kingdom

12:30–13:30 **Lunch/Networking**

12:30-13:30 **Committee Meeting** [virtual]

13:30–14:30 **Fundamentals of Network Structure and Metabolism (2)**

Session chairs: Herbert Sauro and Stefan Schuster

13:30–13:50: Thermodynamics of Evolving Metabolic Reaction Networks – Friedrich Srienc, University of Minnesota-Twin Cities, Minnesota, United States

13:50–14:10: Linear energy converter inspired metabolic analysis couples rate and yield in metabolic models – St. Elmo Wilken, Heinrich-Heine-Universität, Düsseldorf, Germany

14:10–14:30: What Is a Stoichiometric Matrix? – Ronan Fleming, Leiden University & National University of Ireland, Galway, Ireland

14:30–15:00 **Coffee Break/Networking**

15:00–17:30 **Poster Session 2**

Block 2*

18:30–20:00 **Group Dinner at Barley's Taproom and Pizzeria [in person only]**

Schedule *Wednesday, August 4*

07:30–08:30 **Registration/Opening**

08:30–10:15 **Modular and Biosystems Design (1)**

Session chairs: Oliver Ebenhoeh and Cong T. Trinh

• **08:30–09:00:** New Conceptual and Algorithmic Developments Extending the Framework of Minimal Cut Sets for Metabolic Network Design – Steffen Klamt, Max Planck Institute, Magdeburg, Germany

• **09:00–09:25:** *In Vivo* Thermodynamic Analysis Of Metabolic Networks – Daniel Amador-Noguez, University of Wisconsin-Madison, Wisconsin, United States

•**09:25–09:50:** Modular Design Principles In Biological Systems: Theory, Computation, And Experimental Validation – Cong T. Trinh, University of Tennessee, Knoxville, Tennessee, United States

• **09:50–10:10:** Stress-induced cross-feeding of internal metabolites provides a dynamic mechanism of microbial cooperation – Kapil Amarnath, University of California at San Diego, California, United States

10:15–10:45 **Coffee Break/Networking**

10:45–12:30 **Advances in Methods, Algorithms, and Tools Development (1)**

Session chairs: Sabine Peres and Hyun-Seob Song

• **10:45–11:05:** Metabolic Pathway Analysis in presence of biological constraints – Phillippe Dague, Université Paris-Saclay, France

• **11:05–11:25:** ModelSeed Release 2: High Throughput Genome-Scale Metabolic Model Reconstruction and Analysis – Jose Faria, Argonne National Laboratory, Illinois, United States

• **11:25–11:45:** Pickaxe, JN1224min, & the MINE: Unearthing Enzyme Promiscuity for New Applications – Keith Tyo, Northwestern University, Illinois, United States

• **11:45–12:05:** Integrated Knowledge Mining, Genome-Scale Modeling, and Machine Learning for Predicting *Yarrowia lipolytica* Bioproduction – YinJie Tang, Washington University in St. Louis, Missouri, United States

• **12:05–12:10:** ModelSEEDpy: A python library for automated model reconstruction and analysis – Filipe Liu, Argonne National Laboratory, Illinois, United States

• **12:10-12:15:** Systematizing the different notions of growth-coupled strain design and a single framework for their computation - Phillip Schneider, Max Planck Institute, Magdeburg, Germany

12:30–13:00 **Lunch box** [in person only]

12:30–14:30 **Virtual Networking**

13:00–20:00 **Trip to Smoky Mountains** [in person only]

Schedule *Thursday, August 5*

07:30–08:30 **Registration/Opening**

08:30–10:15 **Resource Allocation and Metabolism (1)**

Session chairs: Egils Stalidzans and Ross Carlson

• **08:30–09:00:** Towards Resource Allocation Models for Multicellular Organisms – Anne Goelzer, Université Paris-Saclay, France

• **09:00–09:20:** Efficient Dry Mass Utilization May Explain Bacterial Growth Laws That Relate Catalyst Concentrations With Growth Rate – Martin Lercher, Heinrich Heine University Düsseldorf, Germany

• **09:20–09:40:** Integrative Modeling of Enzymatic Coordination of Saccharification And Fermentation In Consolidated Bioprocessing of Cellulose – Firnaaz Ahamed, University of Nebraska-Lincoln, Nebraska, United States

• **09:40–10:00:** Growth Mechanics: The Economy, Control, and Optimality of Self-Replication – Hugo Dourado, Heinrich Heine University, Germany

• **10:00–10:20:** Functional analysis of metabolism quantifies protein costs for the synthesis of biomass components in *E. coli* – Matteo Mori, University of California at San Diego, California, United States

10:30–11:00 **Coffee Break/Networking**

11:00–11:40 **Keynote Lecture**

Session chairs: Egils Stalidzans and Ross Carlson

Dynamic Allocation of Proteomic Resources by *E. coli* in Rich and Minimal Media – Terence Hwa, University of California at San Diego, United States

11:40–12:30 **Resource Allocation and Metabolism (2)**

Session chairs: Egils Stalidzans and Ross Carlson

• **11:40–12:00:** Pseudomonas aeruginosa Reverse Diauxie Is a Multidimensional, Optimized, Resource Utilization Strategy – Ross Carlson, Montana State University, Montana, United States

• **12:00–12:20:** Comparing Glutamine and Glucose as Fuels of Cancer and Yeast Cells: A Linear Optimization Model – Stefan Schuster, University of Jena, Germany

• **12:20–12:25:** Proteome efficiency of metabolic pathways in *Escherichia coli* – Xiao-Pan Hu, Heinrich Heine University Düsseldorf, Germany

12:30–13:30 **Lunch/Networking**

13:30–15:00 **Advances in Methods, Algorithms, and Tools Development (2)**

Session chairs: Sabine Peres and Hyun-Seob Song

• **13:30–13:50:** EFMlrs: Massive Parallel and Scalable Enumeration of Elementary Flux Modes – Bianca Buchner, University of Vienna, Austria

• **13:50–14:10:** Applications of Answer Set Programming to the analysis of constraint-based models – Maxime Mahout, Université Paris Saclay, France

• **14:10–14:30:** Introducing New Approaches to Gapfilling and Dynamic Flux Balance Analysis for Genome-Scale Models – Rajib Saha, University of Nebraska-Lincoln, Nebraska, United States

• **14:30–15:00:** Advances in Metabolic Modeling Tools in KBase and ModelSEED Enable Integration of Multiomics Data for Pathway Discovery – Chris Henry, Argonne National Laboratory, Illinois, United States

15:00–15:30 **Coffee Break/Networking**

Schedule *Thursday, August 5*

15:30–17:00 **Advances in Methods, Algorithms, and Tools Development (3)**

Session chairs: Sabine Peres and Hyun-Seob Song

- **15:30–15:50:** ASTHERISC: A method to design microbial communities with maximal thermodynamic driving force for the production of chemicals – Pavlos Stephanos Bekiaris, Max Planck Institute of Magdeburg, Germany

- **15:50–16:10:** Integration Of Metabolic, Ecological, And Kinetic Models To Predict The Dynamics In Context-Dependent Microbial Interactions – Hyun-Seob Song, University of Nebraska-Lincoln, Nebraska, United States

- **16:10–16:40:** A Metabolic Approach To Microbial Community Robustness – William Harcombe, University of Minnesota Twin Cities, Minnesota, United States

18:30–21:00 **Reception/Dinner at Calhoun's on the River** [in person only]

Schedule *Friday, August 6*

07:30–08:30 **Registration/Opening**

08:30–10:20 **Applications in Systems/Synthetic Biology/Metabolic Engineering (3)**

Session chairs: Zita Soons and Isabel Rocha

• **08:30–09:00:** Whole-body metabolic modelling provides novel insight into host-microbiome crosstalk – Ines Thiele, National University of Ireland, Galway, Ireland

• **09:00–09:20:** Determination of metformin transport parameters between plasma and red blood cells of humans in different scale models – Egils Stalidzans, Latvian Biomedical Research and Study Centre, Riga, Latvia

• **09:20–09:40:** Improving Quantitative Genome-Scale Metabolic Modeling of CHO – Diana Széliová, Austrian Centre of Industrial Biotechnology, Vienna, Austria

• **09:40–10:00:** Modelling of glycogen metabolism and glycogen-related disorders – Yvan Rousset, Heinrich Heine University, Germany

• **10:00–10:20:** The Role of RDCVFL in a Mathematical Model of Photoreceptor Interactions – Erika Camacho, Arizona State University, Arizona, United States

10:20–10:45 **Coffee Break/Networking**

10:45–12:15 **Applications in Systems/Synthetic Biology/Metabolic Engineering (4)**

Session chairs: Oliver Ebenhoeh and Cong T. Trinh

• **10:45–11:15:** Macromolecular crowding regulates spatial organization and gene expression in cell-sized vesicles – Steve Abel, University of Tennessee, Knoxville, Tennessee, United States

• **11:15–11:35:** Is the cytosolic density of a prokaryotic cell optimized for metabolic efficiency? – Tin Yau Pang, Heinrich Heine University Düsseldorf, Germany

• **11:35–11:55:** An Improved Kinetic Model to Explain the Physiological Response of Escherichia coli Under High ATP Demand – Giulia Slaviero, Max Planck Institute, Magdeburg, Germany

• **11:55-12:15:** Dissecting the Metabolic Reprogramming of Maize Root under Nitrogren Limiting Stress Condition – Niaz Chowdhury, University of Nebraska-Lincoln, Nebraska, United States

12:15–12:45 **Poster Awards/MPA2023 Venue/Closing Remarks**

12:45–14:00 **Lunch at the Hilton** [in person only]

Schedule *Poster Sessions*

Block 1

Monday 15:35–17:3

First Name	Last Name	Affiliation	Poster Title
Francesco	Balzerani	University of Navarra, Spain	Prediction of Phenolic Compound Metabolism in the Human Gut Microbiota using Machine Learning Methods
Naroa	Barrena	University of Navarra, Tecnun School of Engineering, San Sebastian	Integration of regulatory and signaling networks with genome-scale metabolic networks for the prediction of drug targets
Oliver	Bodeit	Heinrich-Heine-Universitaet Duesseldorf	A Programming Interface to Resource Balance Analysis, Enabling Development of Workflows on Cellular Resource Allocation
Cailean	Carter	Quadram Institute	Tuatara - a tool for constructing and handling multiple genome-scale metabolic models
Janaka	Edirisinghe	Argonne National Laboratory	Automated Genome-Scale Fungal Model construction in KBase: Towards eliminating redundancies in Fungal Biochemistry Template
Fatima	Foflonker	Argonne National Laboratory	A protein-enabled genome-scale metabolic model of maize in plantSEED
André	Fonseca	ITQB, Universidade Nova de Lisboa	Dynamic modeling of the shikimate pathway and central carbon metabolism of E. coli
Almut	Heinken	National University of Ireland Galway	Large-scale metabolic reconstruction of the human microbiome accounting for strain-specific drug metabolism
Clémence	Joseph	KU Leuven	CellScanner, a user-friendly tool to identify and enumerate cells in flow cytometry data
Filipe	Liu	Computing, Environment, and Life Sciences Division, Argonne National Laboratory, Lemont, Illinois	ModelSEEDpy: A python library for automated model reconstruction and analysis
Tim	Nies	Heinrich Heine University Düsseldorf	A computational and Experimental Analysis: information about PAM instrument parameters may affect our research
Tim Daniel	Rose	LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich	Analysis of Global Changes of the Lipid Metabolism on Lipid Metabolic Networks with the Lipid Network Explorer
Peter	Schubert	Heinrich-Heine-University Duesseldorf	A Python tool making SBML kinetic modelling more accessible

Schedule *Poster Sessions*

Block 1, continued

Monday 15:35–17:30

First Name	Last Name	Affiliation	Poster Title
Dipali	Singh	Quadram Institute	Metabolism of Campylobacter jejuni: an integrated in silico and in vitro study
Cyrille	Thinnes	National University of Ireland Galway	MicroMaps: Microbiome Metabolic Network Visualization using Systematic Team Empowerment
Vikas	Upadhyay	Penn State University	dGPredictor: Automated fragmentation method for Gibbs energy change prediction of metabolic reaction and de novo pathway design
Luis	Valcárcel	Tecnun School of Engineering, University of Navarra, San Sebastian, Spain	Linear model for the integration of labeling data with genome-scale metabolic networks

Schedule *Poster Sessions*

Block 2

Tuesday 15:00–17:30

First Name	Last Name	Affiliation	Poster Title
Firnaaz	Ahamed	University of Nebraska-Lincoln	A cybernetic trait-based model identifies microbial regulation as a key driver of priming effects
Adil	Alsiyabi	University of Nebraska-Lincoln	Predicting the Effect of Carbon Substrate on Bioplastic Production in Rhodopseudomonas Palustris
Kapil	Amarnath	UC San Diego	Stress-induced cross-feeding of internal metabolites provides a dynamic mechanism of microbial cooperation
Adrienne	Arnold	Montana State University	Methanotroph acclimation to cultivation stresses and the role of byproduct synthesis
Ross	Carlson	Montana State University	Multiscale Analysis of Autotroph-Heterotroph Interactions in a High-Temperature Microbial Community
Wassili	Dimitriew	Friedrich Schiller University Jena	Modeling and deciphering the regulation of microbial central metabolism by dynamic optimization
Joan Sebastián	Gallego Murillo	Delft University of Technology	Metabolic profiling of cultured erythroblasts for the large scale production of red blood cells
Xiao-Pan	Hu	Heinrich Heine University Düsseldorf, Germany	Proteome efficiency of metabolic pathways in Escherichia coli
Aimee	Kessell	University of Nebraska-Lincoln	Metabolic Network Modeling to Determine Microbially-Driven Biogeochemical Functions in Globally-Distributed River Corridors
Lukas	Korn	University of Jena	Elementary-mode analysis of the methylcitrate cycle
Maxime	Lecomte	INRAE UMR STLO	Metabolic Modeling Deciphers Interactions in a Cheese Bacterial Community
Naeun	Lee	University of Nebraska-Lincoln	Omics and Metabolic Network Integration to Predict the Impacts of Copper Dysregulation on the Liver Functions and Energy Metabolism

Schedule *Poster Sessions*

Block 2, continued

Tuesday 15:00–17:30

First Name	Last Name	Affiliation	Poster Title
Tobias	Pfennig	Heinrich-Heine-University Düsseldorf	Emergence of Metabolic Cooperation Between Cells in Schematic Balanced Growth Models
Antonio	Rigueiro-Mesejo	Heinrich Heine University Duesseldorf	Dynamic flux balance analysis reveals the distribution of biochemical subtypes in CAM photosynthesis
Zahmeeth	Sakkaff	Argonne National Laboratory	Information- and Communication-Centric Approach in Cell Metabolism
Philipp	Schneider	Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg	Systematizing the different notions of growth-coupled strain design and a single framework for their computation
Hyeongmin	Seo	University of Tennessee, Knoxville	Engineering Modularity of Ester Biosynthesis Across Biological Scales
Ralf	Steuer	Humboldt-University of Berlin	Optimal resource allocation and the limits of phototrophic productivity
Anna	Stikane	University of Latvia	Exploring valorisation of fermentation by-products through metabolic modelling
Cong	Trinh	University of Tennessee, Knoxville	Computational Design and Biological Insights into Modular Cell Design for Large Libraries of Exchangeable Product Synthesis Modules in Escherichia Coli
Luis	Valcárcel	Tecnun School of Engineering, University of Navarra, San Sebastian, Spain	Computational models for the identification of metabolic vulnerabilities in Multiple Myeloma
Koen	Verhagen	Delft University of Technology	Modelling of glucose recycling through storage metabolism in Saccharomyces cerevisiae under dynamic conditions

Title: Macromolecular Crowding Regulates Spatial Organization and Gene Expression in Cell-Sized Vesicles

Author: Steven M. Abel

Primary Affiliation: University of Tennessee, Knoxville, USA

Abstract

The spatial organization of cells can impact the emergent behavior of biochemical reaction networks governing gene expression, signaling, and metabolism. For example, bacterial cells commonly exhibit membraneless organization of gene expression. In contrast, cell-free expression systems have been limited in their ability to mimic the cellular environment because a means to control spatial organization is lacking. Here, using a combination of computer simulations and experiments, we show that macromolecular crowding can differentially modulate the organization of DNA plasmids and ribosomes in cell-sized vesicles, leading to changes in gene expression and providing a route toward more cell-like synthetic systems. We further examine cell-free protein synthesis in vesicles using a coupled mRNA/protein reporter technique. With an increase in crowding, the mRNA abundance remains approximately constant while the protein abundance decreases. Computer simulations reveal that the decrease in translational efficiency arises due to slower diffusion and the spatial segregation of transcription and translation. Our work demonstrates a flexible platform to understand mechanisms of self-organization of membraneless structures in cells and the spatial control of gene expression. We conclude by discussing the role of spatial organization in biochemical reaction networks more broadly, including implications for metabolic pathways.

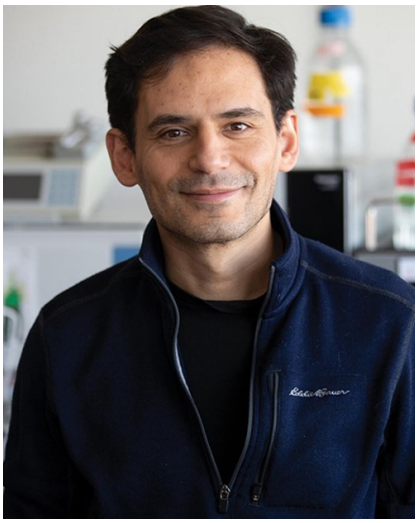
Bio

Steve Abel is an Associate Professor of Chemical and Biomolecular Engineering at the University of Tennessee, Knoxville. He received undergraduate degrees in Chemical Physics, Mathematics, and Physics from Rice University and a PhD in Chemical Engineering from Stanford University, where his research focused on theoretical statistical mechanics and glassy materials. He was a postdoctoral associate at the Massachusetts Institute of Technology, where he used theory and computation to study signal transduction in T cells. Abel's current interests encompass problems in cell biology, immunology, and soft biological materials, which his group addresses using theoretical and computational methods. Specific interests include the physical regulation of immune-cell activation, stochastic and spatial effects in biochemical reaction networks, membrane and cytoskeletal biophysics, and intracellular transport. Much of this work is carried out in close collaboration with experimentalists. Abel is a recipient of the National Science Foundation CAREER award and has won numerous teaching awards.

Title: *In vivo* Thermodynamic Analysis of Metabolic Networks

Author: Daniel Amador-Noguez

Primary Affiliation: Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA



Abstract

Recent breakthroughs in genome editing and metabolic engineering expand the range of microorganisms and synthesis routes that may be used to produce biofuels and valuable bioproducts from renewable biomass resources. Thus, there is an increasing need for new approaches to characterize the metabolic capabilities of nascent industrial organisms and improve the efficiency of their biosynthetic pathways.

Thermodynamics constitutes a key determinant of flux and enzyme efficiency in metabolic networks. A biochemical reaction with a strong thermodynamic driving force (i.e. with a large negative ΔG) will achieve a higher net flux given a fixed amount enzyme than one closer to equilibrium. Within a pathway, steps closer to equilibrium will be the least enzyme efficient. Thermodynamic analysis can therefore provide unique insights in synthetic pathway design by identifying bottlenecks, pinpointing the enzymes for which changes in activity will have the largest effect on flux, and predicting the most efficient route for product synthesis.

During this talk, I will discuss the development of experimental-computational approaches for in vivo determination of Gibbs free energies (ΔG) in metabolic networks. In our first iteration of this approach, we combined quantitative metabolomics with ^{2}H and ^{13}C metabolic flux analysis to measure step-by-step ΔG of central metabolic pathways in different microbes. This showed that the Entner-Doudoroff glycolytic pathway in Zymomonas mobilis, a prolific ethanol producer, is nearly twice as thermodynamically favorable as the classical glycolytic pathway in well-studied microbes such as E. coli or yeast, helping explain its rapid sugar catabolism. In a subsequent study, we found that the glycolytic pathway of Clostridium thermocellum, a cellulolytic biofuel producer, operates surprisingly close to thermodynamic equilibrium, allowing increased ATP yield per glucose molecule but at a slower glycolytic rate. Taken together, these results illustrate the tradeoff between energy yield, catabolic rate, and thermodynamic driving force across the glycolytic pathways of diverse organisms, reflecting evolutionary adaptations to distinct lifestyles and habitats. Our long-term goal is the construction of accurate metabolic models that incorporate thermodynamic constrains and guide rational engineering of microbial networks.

Bio

Daniel Amador-Noguez is an Associate Professor in the Department of Bacteriology at the University of Wisconsin-Madison. Dr. Amador-Noguez received his B.S. in Chemistry from Monterrey Institute of Technology (ITESM), Monterrey, Mexico. He earned his Ph.D. in Molecular Genetics from Baylor College of Medicine, Houston TX. He then worked as a post-doctoral associate with Joshua Rabinowitz at Princeton University. Daniel joined the faculty of the Department of Bacteriology at UW-Madison in 2013. His research program seeks to generate a quantitative and holistic understanding of how metabolic networks are regulated in microbes. The Amador-Noguez laboratory integrates systems-level approaches, especially LC-MS-based metabolomics, with computational modeling and genetic engineering to understand how metabolic fluxes are controlled and how microbes adapt their metabolism in response to environmental challenges and during developmental processes. His laboratory has three main research areas: 1) metabolic regulation in biofuel producers, and 2) metabolic remodeling during biofilm development, and 3) biochemical activities of the gut microbiome.

Title: Data Driven Advanced Metabolic Modeling of Diurnal Growth in the Model Green Alga *Chlamydomonas reinhardtii*

Author: Nanette R. Boyle

Primary Affiliation: Colorado School of Mines, United States

Abstract

Photosynthetic microorganisms have the potential to become economical and sustainable sources of fuels, as the energy required for the cell to grow can be sourced from natural sunlight alone; however, we have yet to harness their full power due to a general lack of tools for engineering their metabolism. Metabolic models have been shown to drastically reduce the development time for commercial production strains of heterotrophic bacteria; however, these models are less applicable to photosynthetic systems due to the transient nature of diurnal (day/night) growth. Current metabolic models are not capable of accurately predicting growth rates in day/night growth cycles, let alone genetic changes which would lead to increased yields. Our work is focused on constructing an approach to diurnal modeling that allows for extension of current metabolic models into a transient space, using organism specific circadian information. We have used circadian gene expression data from Chlamydomonas reinhardtii to cluster gene expression and convert discrete data into continuous functions. We then implemented these functions as additional constraints on our metabolic model and are currently integrating this constrained model into an agent-based framework. We will present the result of this work and demonstrate how these modeling techniques are able to further improve the model and better predict growth in diurnal light. Ultimately, the availability of such models will introduce a new frontier in the ability to use in silico tools to investigate the metabolism, growth and phenotype of photosynthetic microorganisms. It will enable us to gain insight into why photosynthetic organisms have drastically different productivities when grown in continuous light compared to diurnal cycles and how to circumvent this.

Bio

Nanette Boyle is an associate professor in chemical & biological engineering at Colorado School of Mines. She received her PhD in chemical engineering from Purdue University in in 2009, where she was an NSF graduate research fellow. Nanette joined the faculty at Mines in 2013 after completing two postdoctoral fellowships: with Prof. Ryan Gill at CU Boulder (where she was awarded the NIH NRSA postdoctoral fellowship) and Prof. Sabeeha Merchant at UCLA in biochemistry. She is a DOE early career program awardee and a Scialog Fellow for Negative Emission Science. The Boyle Laboratory focuses on utilizing photosynthetic microorganisms, such as cyanobacteria and algae, to develop more sustainable sources of fuels, chemical feedstocks and value added chemicals. To do this, the lab uses a wide variety of tools including metabolic modeling, systems biology and synthetic biology.

Title: Towards Resource Allocation Models for Multicellular Organisms

Author: Anne Goelzer

Primary Affiliation: Université Paris-Saclay, INRAE, MaIAGE, Jouy-en-Josas, France

Abstract

Predicting quantitatively the behavior of living organisms from the finest scales to the individual scale in normal and complex environmental conditions remains highly challenging for the system biology community. Part of the difficulty consists in integrating the scales where the decisions of the adaptation to the environment take place, i.e. the cellular and infra-cellular scales, in the context of the individual [1]. Since 2009, a significant step has been achieved in the modeling of cellular and infra-cellular scales in microbial cells [2]. Considering the cell as a set of functional constraints (e.g. biophysical, stoichiometric or osmotic, etc.) falls within the so-called modeling field of constraint-based modeling (CBM) that calls for the resolution of linear optimization problem [3]. We developed in [4] and validated experimentally in [5] a new CBM method, named Resource Balance Analysis (RBA). RBA predicts for a specific environment, the set of possible cellular configurations (growth rate, metabolic fluxes, abundances of molecular machines, including ribosomes, enzymes, transporters) compatible with the available external resources, and has the potential to predict the cell response to a large set of complex environmental conditions [6]. Recently, the RBA theoretical framework was extended to eukaryotic cells [7] and used to generate RBA models for yeast and for the mesophyll cell of the plant *Arabidopsis thaliana*. In this talk, we present the RBA framework and the first results of RBA simulations for eukaryotic cells, and finally discuss the next steps towards the modeling of multicellular organisms.

References

[1]. Goelzer, A., & Fromion, V. (2017). Resource allocation in living organisms. *Biochemical Society Transactions*, 45(4), 945-952.

[2]. Goelzer, A., & Fromion, V. (2011). Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta* (BBA)-General Subjects, 1810(10), 978-988.

[3]. Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis?. *Nature biotechnology*, 28(3), 245-248.

[4]. Goelzer, A., Fromion, V., & Scorletti, G. (2011). Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6), 1210-1218.

[5]. Goelzer, A., et al. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic engineering*, 32, 232-243.

[6]. Tournier, L., Goelzer, A., & Fromion, V. (2017). Optimal resource allocation enables mathematical exploration of microbial metabolic configurations. *Journal of mathematical biology*, 75(6), 1349-1380.

[7]. Goelzer, A., & Fromion, V. (2019). RBA for eukaryotic cells: foundations and theoretical developments. *bioRxiv*, 750182.

Bio

Anne Goelzer received the engineering degree from Ensicaen in 2001 and obtained a Ph.D. degree in control theory/computer science from the Ecole Centrale de Lyon (France) in 2010. From 2001 to 2003, she was an engineer, consulting in control theory at Renault (France). Since 2003, she is with the Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). Her main research interest is to develop mathematical methods and tools to understand the design principles of living systems, and predict their behavior in complex environmental conditions. She has a strong expertise in the multi-scale modeling of biological systems, control theory, applied mathematics and systems biology.

Title: A Metabolic Approach to Microbial Community Robustness

Author: William R. Harcombe

Primary Affiliation: University of Minnesota, USA

Abstract

As we strive to manage critical microbial communities we must understand how these systems respond to perturbation. We use a combination of dynamic flux balance analysis [1] and laboratory co-cultures to investigate how the physiology of cells determines the content, function and dynamics of microbial communities. We specifically investigate a cross-feeding system in which *Escherichia coli* and *Salmonella enterica* rely on each other for essential metabolites. We find that cross-feeding tends to stabilize species ratios [1,2]. However, cross-feeding tends to make populations less robust to both mutations and antibiotics [2-4]. Our work suggests that metabolic interactions between species can have predictable impacts on both ecological and evolutionary responses to perturbation in microbial communities.

References

[1] Harcombe et al., "Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics", *Cell Reports*, 2014, 7:1104-1115.

[2] Chubiz et al., " Species interactions differ in their genetic robustness", *Frontiers in Microbiology*, 2015, 6: 271

[3] Adamowicz et al., "Cross-feeding modulates antibiotic tolerance in bacterial communities", *ISME*, 2018 12:2723-2735

[4] Adamowicz et al., "Cross-feeding modulates the rate and mechanisms of antibiotic resistance evolution in a model microbial community of *Escherichia coli* and *Salmonella enterica*. *PLoS Pathogens*. 2020 16:e1008700

Bio

Will Harcombe is an associate professor in Ecology, Evolution and Behavior and the BioTechnology Institute at the University of Minnesota. He is also part of the Microbiology, Immunology and Cancer Biology, and the Plant and Microbial Biology graduate programs. Dr. Harcombe trained as an evolutionary biologist with Dr. Jim Bull at the University of Texas investigating the predictability of evolution in bacteriophage and bacteria. He then gained experience with computational systems biology as an NIH postdoctoral fellow with Dr. Chris Marx at Harvard University. He joined the University of Minnesota in 2014 and is now the Wardle Chair of Microbial Ecology. His lab focuses on using microbial physiology to quantitatively understand, predict and manage the eco-evolutionary dynamics of microbial communities.

Title: Using Machine Learning to Characterize Metabolic Pathway Activities

Author: Soha Hassoun

Primary Affiliation: Department of Computer Science, Department of Chemical and Biological Engineering, Tufts University, Medford, MA

Abstract

Advances in machine learning present unique opportunities to fundamentally enhance our ability to analyze and understand metabolism. In this talk, we propose two machine learning techniques to address two important problems related to interpreting metablomics data within the context of the biological network of the underlying sample.

First, we address the problem of determining pathway activities based on targeted and untargeted metablomics measurements. We present an inference-based approach, termed Probabilistic modeling for Untargeted Metabolomics Analysis (PUMA). Capturing metabolomics measurements and the biological network for the biological sample under study in a generative model, PUMA predicts the likelihood of pathways being active, and then derives probabilistic annotations. Unlike prior pathway analysis tools that analyze differentially active pathways, PUMA defines a pathway as active if the likelihood that the path generated the observed measurements is above a particular (user-defined) threshold. Due to the lack of "ground truth" metabolomics datasets, where all measurements are annotated and pathway activities are known, PUMA is first validated on synthetic datasets that are designed to mimic cellular processes. PUMA, on average, outperforms pathway enrichment analysis by 8%. When applied to case studies, PUMA annotation results were in agreement to those obtained using other tools that utilize additional information in the form of spectral signatures. Importantly, For an experimentally validated 50-compound dataset, annotations using PUMA yielded 0.833 precision and 0.676 recall.

Next, we address the problem of analyzing time-series metabolomics data. Analyzing longitudinal data sheds light on dynamic metabolic shifts and can elucidate disease progression and assess the impact of nutritional and drug interventions. Pairing this data with the sample's underlying network model allows for identifying important time dynamics in metabolic pathways. Here, we adapt a machine learning technique called dynamic stochastic block models to uncover dynamic shifts in metabolomics data. Stochastic block models (SBM) are used for detecting groups (or blocks or clusters) within a network. Dynamic SBMs infer the evolution of such blocks through time while aiming to stabilize within-group connectivity. To adapt Dynamic SBM to analyze metabolomics data, we represent relationships among measured metabolites by their correlation similarity to other metabolites. Dynamic SMB thus identifies groups of measured metabolites that have similar correlation patterns. Using the underlying biological network for the sample under study, we weight correlation similarity by the corresponding distance in the biological network. We adopt dynSMB, the Dynamic Stochastic Block Model R package https://cran.r-project.org/web/packages/dynsbm/index.html, to analyze metabolomics data. We demonstrate this method on metabolomics data from Chinese hamster ovary cells culture (Sumit et al, 2019). A custom visualization tool allows the exploration of model parameters that yield different groups and the migration of metabolites among groups. Results reveal that dynamic SBM can detect metabolic temporal shifts associated within several metabolic pathways. We compare our findings against PCA-based methods and metabolite enrichment set analysis that were reported earlier for this dataset.

This work is in collaboration with Professors Liping Liu and Micheal Hughes, and graduate students Ramtin Hosseini, Neda Hassanpour, BJ Stubbs, and Margaret Martin. The research is supported by the NIGMS of the National Institutes of Health, Award R01GM132391. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Bio

Soha Hassoun is Professor in the Department of Computer Science at Tufts University with a secondary appointment in the Department of Chemical and Biological Engineering at Tufts. She has her Master's degree in course EECS from MIT and her PhD in CSE at the University of Washington, Seattle, WA. Her research is at the intersection of systems biology and machine learning. Visit her lab's web page for more information.

Title: Advances in Metabolic Modeling Tools in KBase and ModelSEED Enable Integration of Multiomics Data for Pathway Discovery

Author: Christopher Henry

Primary Affiliation: Argonne National Laboratory, United States

Abstract

Increasingly multi-omics data is becoming more accessible for the study of a wide range of complex biological systems. Today, large-scale metagenomes can be readily obtained from soil microbiome systems, while the instruments and protocols surrounding the collection of metabolomic and proteomic data are constantly improving. Yet analysis methods still struggle to annotate these individual datasets, let alone combine them to discover new biological principles. For example, one of the great challenges associated with the use and interpretation of metabolomics data is the large portion of observed peaks that cannot be readily associated with known biochemical compounds. With the lack of clear identities for peaks, and with many identified peaks lacking known pathways, analysis is often limited to correlations alone. In this talk, we will discuss recent advances in tools and workflows in KBase and ModelSEED that are expanding the possibilities and opportunities for the use of metabolic models to integrate multi-omics data for the discovery of novel biochemical pathways.

Specifically, we have made significant improvements to our pipeline for the rapid reconstruction of metabolic models from sequence data, including isolate genomes and metagenomes. Now models have hundreds of additional genes and reactions, produce energy in biologically relevant ways, and include tailored templates for archaea, bacteria, plants, fungi, and cyanobacteria. We also offer a fully integrated pipeline for the prediction of novel biochemical compounds and reactions using cheminformatics approaches, including prediction of novel promiscuous enzymatic reactions and spontaneous chemical reactions. Finally, we have flux balance analysis workflows for combining genomic-based and novel chemical networks together to predict pathways to explain metabolomics data.

Scientifically, we will explore how these improved tools permit us to study pathway variation across the microbial tree of life, learn insights about microbial diversity and variation from microbiome data, and study evolutionary implications over how potential spontaneous reactions occur across the known metabolic pathways. We'll demonstrate our multi-omics integration tools to discover new pathways in the JCVI minimal genome and to mechanistically map metabolites to microbes within the human microbiome. Our exploration of microbiome data demonstrates organizing principles for the assembly and function of microbiome systems.

Finally, we will discuss an exciting new way of accessing ModelSEED reconstruction capabilities and analyses through a new ModelSEEDpy module, designed to be completely compatible with COBRApy.

While still in development, we will have a prototype release of this module at MPA.

Bio

Christopher Henry is a scientist in the Data Science and Learning division of Argonne National Laboratory. He obtained his PhD in Chemical Engineering from Northwestern University, where he conducted research in biochemical thermodynamics, cheminformatics, and metabolic modeling. He went on to develop the ModelSEED pipeline for automated metabolic model reconstruction during his postdoc at Argonne. Today Dr. Henry is co-PI of the DOE Systems-biology Knowledgebase (www.kbase.us) project, leading development of a web-based collaborative platform for conducting large scale analysis and integration of genomic and multi-omics data. He continues to perform research exploring how metabolic models may be applied to integrate omics data to improve our understanding of gene function and microbial behavior, particularly in complex microbiome systems.

Title: Dynamic Allocation of Proteomic Resources by E. coli in Rich and Minimal Media

Author: Terence Hwa

Primary Affiliation: Department of Physics, University of California at San Diego, La Jolla, CA, USA

Abstract

Bacteria can grow in many different nutrient environments, and the rapidity of their adaptation is an important component of the overall fitness. For enteric bacteria, a common nutrient environment is the “rich medium” containing ample amino acids, sugars, and other substances. Here we perform quantitative studies on the allocation of proteomic resources by Escherichia coli in various rich media and during transition between rich and minimal media. Faster growth in rich media is attributed primarily to the allocation of the proteome from amino acid biosynthesis in minimal media to protein synthesis in rich media. However, this allocation results in a substantial growth lag during downshift from rich to minimal media. Quantitative analysis reveals a continuous spectrum of growth bottlenecks in amino acid synthesis. A striking linear relation is established between the onset time of each pathway and the pre-shift abundances of the corresponding enzymes relative to their post-shift abundances. To describe the downshift kinetics, we developed a coarse-grained kinetic model based on two key ingredients: flux-controlled regulation of proteome allocation for ribosome vs AA synthesis, and end-product inhibition of individual pathways. Our model quantitatively captures the growth recovery dynamics, including the onset time of each AA synthesis pathway, based only on steady-state proteome compositions, without the need to know any kinetic parameters or the underlying signaling pathways, and without invoking any ad hoc fitting parameters. The results will be discussed in terms of dynamic optimization theories.

Bio

Terry Hwa is the Presidential Chair and Distinguished Professor of Physics at UC San Diego, with joint appointment in the Division of Biological Sciences, and is the founding director of the quantitative biology program at UC San Diego. He is a fellow of the American Academy of Microbiology, the American Physical Society, and a member of the National Academy of Sciences. Hwa develops theoretical and experimental approaches to gain quantitative, predictive understanding of living systems. He is particularly known for quantitative studies of bacterial physiology, including the establishment of a number of bacterial growth laws which led to a principle governing proteome allocation. The Hwa lab is continuing to extend their physiological approach to characterize bacterial species singly and in consortium, to uncover underlying principles governing the spatiotemporal dynamics of microbial communities, and to apply these principles to synthetic biology applications.

Title: New Conceptual and Algorithmic Developments Extending the Framework of Minimal Cut Sets for Metabolic Network Design

Author: Steffen Klamt

Primary Affiliation: Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Abstract

Several constraint-based methods have been developed to guide the targeted engineering of microbial cell factories. One particular framework is based on the concept of Minimal Cut Sets (MCS), which allows the computation of various strain designs based on the specification of protected/desired (e.g. minimum growth rate and high product yield) and undesired (e.g. low product yield) metabolic behaviors. Here we present several new conceptual and algorithmic developments, which, on the one hand, significantly extend the classes of metabolic design problems that can be handled with MCS and, on the other hand, improve the efficiency of MCS computations in genome-scale networks. These latest developments include [1-4]:

- Use of multiple target and desired regions to precisely tailor the solution space of a (mutant's) metabolic network [1]. For example, this extension allows one to enforce specific behaviors under different conditions (e.g., in two-stage production processes).
- Target and desired regions may now also include optimality constraints allowing the direct computation of bi-level-like (e.g. OptKnock) strain designs as MCS [4].
- Individual cost factors can be assigned to each intervention and reaction/gene deletions and additions can be combined [1].
- Based on the MCS2 approach [5], a new MILP formulation for the (dual) calculation of MCS is presented [2]. Compared to the classical (Farkas'-lemma-based) MILP approach for MCS computation, this more concise formulation makes use of the kernel matrix leading to significantly shorter runtimes for most metabolic design problems tested.
- Novel compression rules for gene-protein-reaction (GPR) associations are introduced that accelerate the computation of gene-based MCS in large networks (with an average factor of 5) [1].
- Computed MCS can be automatically ranked via different criteria [3].

References

[1] Schneider P, von Kamp A, Klamt S (2020) An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets. PLOS Computational Biology 16(7): e1008110.

[2] Klamt S, Mahadevan R, von Kamp A (2020) Speeding up the core algorithm for the dual calculation of minimal cut sets in large metabolic networks. BMC Bioinformatics 21: 510

[3] Schneider P, Klamt S (2019) Characterizing and ranking computed metabolic engineering strategies. Bioinformatics35(17): 3063–3072.

[4] Schneider et al. (2021) submitted.

[5] Miraskarshahi R, Zabeti H, Stephen T, Chindelevitch L. MCS2: minimal coordinated supports for fast enumeration of minimal cut sets in metabolic networks. Bioinformatics. 2019;35: i615–i623

Bio

Steffen Klamt heads the group "Analysis and Redesign of Biological Networks" at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg, Germany. He received his Diploma in Systems Science from the Osnabrück University in 1998 and his Ph.D. from the Stuttgart University in 2005.

His group develops methods for systems and computational biology and combines them with experimental investigations. One central research focus is modeling and computational design of metabolic networks with applications in metabolic and biosystems engineering. His group developed CellNetAnalyzer, a comprehensive package for the *in silico* analysis and rational design of biochemical networks. Experimental work in his group focuses on Escherichia coli as model organism and tight integration of wet-lab and dry-lab investigations allows prompt verification of model-based predictions and metabolic engineering strategies.

Title: Bringing to Bear Retrosynthesis Tools and Enzyme Engineering for Pathway Synthesis

Author: Costas D. Maranas

Primary Affiliation: Department of Chemical Engineering, The Pennsylvania State University

Abstract

Metabolic pathway design is a cross-cutting task present in a wide range of endeavors, from biofuel and bio-renewables design to the biodegradation or funneling of complex aromatics, and chemicals of industrial interest. It is also a multi-disciplinary task which requires knowledge and encoding of the relevant biochemistry of pathways, understanding of the thermodynamic feasibility of individual enzymatic steps, biophysical understanding of activity and specificity of the enzymes involved, and subsequent host selection and metabolic engineering. In this talk, we will discuss recent progress from our group on a variety of fronts starting with pathway retrosynthesis allowing for novel steps (i.e., novoStoic), prediction of Gibbs free energy of change for novel and uncharacterized reaction steps (i.e., DGPredictor) for ensuring thermodynamic feasibility of pathways, rank-ordering of enzyme candidates in terms of their potential for adaptability for the targeted novel conversion, and computational enzyme redesign (i.e., IPRO +/-) for tuning substrate specificity. Both algorithmic developments and software implementations will be discussed along with applications and future perspectives.

Bio

Costas D. Maranas (b. 1967) is the Donald B. Broughton Professor in the Department of Chemical Engineering at The Pennsylvania State University. He received a Diploma in Chemical Engineering at the Aristotle University, Greece in 1990 and a Ph.D. in Chemical Engineering from Princeton University in 1995. He has been in the faculty of the department of Chemical Engineering at Penn State since 1995. He is the recipient of the 2021 Food, Pharmaceutical, and Bioengineering AIChE Division Award, 2020 Biotechnology Progress Award for Excellence in Biological Engineering Publication and the 2020 International Metabolic Engineering Award. Earlier awards include the Allan P. Colburn Award for Excellence in Publications by a Young Member of AIChE (2002) and the Outstanding Young Investigator Award of the Computing and Systems Technology AIChE Division (2006). Penn State recognitions include the Engineering Alumni Society (PSEAS) Premier (2016) and Outstanding (2012) Research Award. He is a member of a number of journal Editorial Boards including PLOS Computational Biology, BMC Systems Biology, Biotechnology Journal and Metabolic Engineering. He is a Fellow of the American Institute of Chemical Engineers (AIChE) and the American Institute of Medical and Biological Engineering (AIMBE). He is the Lead for the "Use Inspired Research" in the Center for Bioenergy Innovation (CBI) DOE center, a participant in the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) DOE center, and a member of the leadership team of the NSF Molecular Maker Lab Institute (MMLI).

The C. Maranas group develops and deploys computational framework informed by systems engineering and mathematical optimization to understand, analyze and redesign metabolism and proteins. Research interests include: Computational protein design; enzyme and antibody engineering; design of protein pores for bioseparations; reconstruction, curation and analysis of metabolic networks; computational strain design and synthetic biology; metabolism of photosynthetic organisms; metabolism of obligatory anaerobes; modeling of microbial communities; optimization theory and algorithms. He has co-authored over 200 refereed journal publications and proceedings including a textbook on "Optimization Methods in Metabolic Networks" (2016). He has supervised 36 PhD theses with many group alumni occupying leading positions in industry and academia. He lives in Sate College, PA with his wife and children.

Title: Principles of Metabolome Conservation in Mammals

Author: Balázs Papp

Primary Affiliation: Biological Research Centre, Institute of Biochemistry, Synthetic and Systems Biology Unit, Eötvös Loránd Research Network (ELKH), Szeged, Hungary

Abstract

Intracellular metabolite concentrations reflect the physiological state of the cell and their changes underlie various diseases. Recent works showed that metabolite concentrations are broadly conserved between distantly related organisms, indicating functional constraints to preserve them. Yet, it remains unexplored to what extent and why metabolites differ in their degree of evolutionary conservation. Here we applied phylogenetic methods to a metabolome dataset of 26 mammalian species and found extensive differences in the rates of concentration changes across metabolites. We identify several functional properties that shape the conservation of metabolite levels. Remarkably, absolute metabolite abundance emerges as the main determinant of conservation, with highly abundant metabolites evolving slowly. Systems modelling revealed that abundant metabolites are more critical to maintain key fluxes in the metabolic network, explaining their elevated conservation. Finally, we show that evolutionary conservation predicts metabolites involved in inborn errors of metabolism. Overall, our findings demonstrate that evolutionary comparisons of metabolite levels inform on the functional organization of metabolism and on their pathogenic alterations.

Bio

Balázs Papp is a principal investigator at the Biological Research Centre and at the Hungarian Centre of Excellence for Molecular Medicine (HCEMM) in Szeged, Hungary. He obtained his Ph.D. in evolutionary genetics in 2004 from the Eötvös Loránd University, Budapest, Hungary. Before establishing a research group in Szeged, he worked at the University of Bath, the University of Manchester and the University of Cambridge, all of which are in the UK. His main interest lies in studying the evolution of molecular networks by combining computational and high-throughput experimental approaches. In particular, his lab integrates metabolomics and phylogenetic approaches to understand the evolutionary forces driving metabolic diversity.

Title: Whole-Body Metabolic Modelling Provides Novel Insight into Host-Microbiome Crosstalk

Author: Ines Thiele

Primary Affiliation: Division of Microbiology, School of Medicine, National University of Ireland, Galway

Abstract

Precision medicine relies on the availability of realistic, mechanistic models that capture the complexity of the human body. Comprehensive computational models of human metabolism have been assembled by the systems biology community, which summarise known metabolic processes occurring in at least one human cell or organ. However, these models have not yet been expanded to connect with whole-body level processes. To address this shortcoming, we have built whole-body metabolic models of a male (deemed Harvey) and a female (deemed Harvetta) starting from the existing human metabolic models, physiological and anatomic information, comprehensive proteomic and metabolomic data, as well as biochemical data obtained from an extensive manual literature review. We tested the predictive capabilities of the resulting whole-body metabolic models against the current knowledge of organ-specific and inter-organ metabolism. The final models contain 28 organs. Importantly, these whole-body models can be expanded to include the strain-resolved metabolic models of gut microbes. By parameterising the whole-body metabolic models with physiological and metabolomic data, we connected physiology with molecular-level processes through networks of genes, proteins, and biochemical reactions. As a sample application of the whole-body metabolic models, I will demonstrate how different microbial composition leads to differences in host metabolism, such as the capability to produce important neurotransmitters in the brain and flux through liver enzymes, with implications for the gut-brain axis as well as for microbiome-mediated liver toxicity. The predictions were consistent with our current understanding but also highlighted that different microbiota composition can lead to high inter-person variability. I envisage the microbiome-associated whole-body metabolic models will usher in a new era for research into causal host-microbiome relationships and greatly accelerate the development of targeted dietary and microbial intervention strategies.

Bio

Ines Thiele is the principal investigator of the Molecular Systems Physiology group at the National University of Ireland, Galway.

Her research aims to improve the understanding of how diet influences human health. Therefore, she uses a computational modelling approach, termed constraint-based modelling, which has gained increasing importance in systems biology. Her group builds comprehensive models of human cells and human-associated microbes; then employs them together with experimental data to investigate how nutrition and genetic predisposition can affect one's health. In particular, she is interested in applying her computational modelling approach for better understanding of inherited and neurodegenerative diseases. Ines Thiele has been pioneering models and methods allowing large-scale computational modelling of the human gut microbiome and its metabolic effect on human metabolism.

Ines Thiele earned her PhD in bioinformatics from the University of California, San Diego, in 2009. Ines Thiele was an Assistant and Associate Professor at the University of Iceland (2009 - 2013), and Associate Professor at the University of Luxembourg (2013-2019).

In 2013, Ines Thiele received the ATTRACT fellowship from the Fonds National de la Recherche (Luxembourg). In 2015, she was elected as EMBO Young Investigator. In 2017, she was awarded the prestigious ERC starting grant. In 2020, Ines Thiele was named a highly cited researcher by Clarivate, and received the NUI Galway President's award in research excellence. She is an author of over 100 international scientific papers and reviewer for multiple journals and funding agencies.

Title: Integrative Modeling of Enzymatic Coordination of Saccharification and Fermentation in Consolidated Bioprocessing of Cellulose

Author: Firnaaz Ahamed[1], Hyun-Seob Song[2], and Yong Kuen Ho[1]

Primary Affiliation: [1]Chemical Engineering Discipline, School of Engineering, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, Selangor, Malaysia; [2]Department of Biological Systems Engineering, University of Nebraska–Lincoln, Lincoln, NE, United States

Abstract

Consolidated bioprocessing (CBP) allows direct conversion of cellulosic biomass to valuable biochemicals in a single step using microbial cell factories. We currently lack a fundamental understanding of governing processes underlying complex dynamics of saccharification and fermentation in CBP. Due to the intrinsic complexity and high nonlinearity in CBP, no fully mechanistic models have been available to date, which in turn poses the development of economically competitive processes to remain challenging. Here, we propose the integration of two complementing frameworks, namely population balances [1–4] and metabolic network-based cybernetic model [5,6], to simulate coupled dynamics of cellulose saccharification and fermentation. The resulting framework, termed Unified Cybernetic-Population Balance Model (UC-PBM) [4,7], features a coordinated regulation of extracellular cellulases for saccharification and intracellular enzymes for fermentation at the network level through closed-loop interactions. UC-PBM uniquely accounts for the extensive heterogeneity in cellulose properties and robust regulation of cellular resources for competing biomass growth and cellulase secretion. Through a case study on *Clostridium thermocellum*, UC-PBM not only demonstrates decent fit and prediction of various experimental exometabolomic data but also provides a better representation of the nature of flux regulations and cellulolytic ability of the microorganism. As major findings, UC-PBM shows that: (1) the starting cellulose concentration controls the overall flux distribution to competing phenotypic functions of the microorganism and (2) growth–decoupled cellulase–secreting pathways are only activated during famine conditions to promote the production of growth substrates. Our findings also indicated that the previously unrecognized cellulosome–decoupled growth metabolic diversity may explain the growth–independent downregulation of cellulases under nutrient–rich conditions. Overall, UC-PBM predicts the effects of a closed-loop control mechanism across multiple scales of process features on CBP, which would vastly benefit rational reaction engineering and process optimization efforts.

References

[1] Y.K. Ho et al., "Interlinked population balance and cybernetic models for the simultaneous saccharification and fermentation of natural polymers", *Biotechnol. Bioeng*., 2015, 112:2084–2105.

[2] Y.K. Ho et al., "Why Are Two Enzymes Better than One for the Efficient Simultaneous Saccharification and Fermentation (SSF) of Natural Polymers? Hints from Inside and Outside a Yeast", *Ind. Eng. Chem. Res*., 2015, 54:10228–10244.

[3] F. Ahamed et al., "Modelling heterogeneity in cellulose properties predicts the slowdown phenomenon during enzymatic hydrolysis", *Chem. Eng. Sci.*, 2019, 206:118–133.

[4] F. Ahamed, "Modelling Consolidated Bioprocessing of Cellulose via Population Balances Coupled with Cybernetic Models", Monash University, 2020.

[5] D. Ramkrishna, H.-S. Song, "Dynamic models of metabolism: Review of the cybernetic approach", *AIChE J*., 2012, 58:986–997.

[6] D. Ramkrishna, H.-S. Song, "Cybernetic Modeling for Bioreaction Engineering", Cambridge University Press, 2019.

[7] F. Ahamed et al., "Modeling coordinated enzymatic control of saccharification and fermentation by Clostridium thermocellum during consolidated bioprocessing of cellulose", *Biotechnol. Bioeng*., 2021, 118:1898–1912.

Title: Predicting the Effect of Carbon Substrate on Bioplastic Production in *Rhodopseudomonas palustris*

Author: Adil Alsiyabi and Rajib Saha

Primary Affiliation: University of Nebraska-Lincoln, USA

Abstract

Rhodopseudomonas palustris is a metabolically versatile Purple Non-Sulfur Bacterium (PNSB). Depending on growth conditions, *R. palustris* can operate on either one of the four different forms of metabolism: photoautotrophic, photoheterotrophic, chemoautotrophic, and chemoheterotrophic. *R. palustris* is also a facultative anaerobe, meaning it can operate both aerobically and anaerobically. Furthermore, the organism is capable of fixing nitrogen and subsequently producing hydrogen and the bioplastic precursor polyhydroxybuyrate (PHB). Recent experimental findings revealed that PHB yields in *R. palustris* were highly dependent on the characteristics of the utilized carbon source. Interestingly, PHB production significantly increased when grown on the carbon- and electron-rich lignin breakdown product *p*-coumarate ($C_9H_8O_3$) compared to acetate when the same amount of carbon was supplied. However, the maximum yield did not improve further when grown on the higher molecular weight substrate coniferyl alcohol ($C_{10}H_{12}O_3$). Therefore, a model-driven investigation was performed to obtain a systems-level understanding of the factors driving PHB yield. A thermo-kinetic analysis of the PHB synthesis pathway identified how the relative concentration of various metabolites and cofactors in the pathway influenced overall productivity. These findings were incorporated into a recently constructed genome-scale metabolic model of the bacterium [1] to understand how characteristics of the utilized carbon substrate affected PHB productivity. This model-guided approach yielded several engineering design strategies for PHB over-production, including utilizing reduced, high molecular weight substrates that bypass the thiolase reaction [2]. Overall, these findings uncover key thermodynamic and enzyme saturation limitations controlling PHB production and lead to design strategies that can potentially be transferrable to other PHB producing bacteria.

References

[1] A. Alsiyabi, C. M. Immethun, and R. Saha, “Modeling the Interplay between Photosynthesis, CO2 Fixation, and the Quinone Pool in a Purple Non-Sulfur Bacterium,” *Sci. Rep.*, vol. 9, no. 1, p. 12638, 2019.

[2] A. Alsiyabi, B. Brown, C. Immethun, M. Wilkins, and R. Saha, “Synergistic Experimental and Computational Approach Identifies Novel Strategies for PHB Overproduction,” *Res. Sq.*, 2021

Title: Thermodynamics of Evolving Metabolic Reaction Networks

Author: John Barrett and Friedrich Srienc

Primary Affiliation: Department of Chemical Engineering and Materials Science, and BioTechnology Institute, University of Minnesota, Minneapolis/St.Paul, MN, USA

Abstract

Numerous methods have been proposed to predict the magnitude of fluxes through individual reactions in a metabolic network. These methods are typically based on linear programming which involves arbitrary assumptions for objective functions such as the maximization of the biomass yield or other subjectively defined constraints. We developed a rigorous theory for predicting the rate structure in a metabolic network by combining results from elementary mode analysis with the principles of statistical thermodynamics. We show that the system is characterized by two types of entropy production rates. The first is maximized when the system reaches a steady state and the Gibbs free energy is minimized corroborating the Maximum Entropy Production principle. The second is based on reaction entropies of individual elementary modes and is maximized when the frequency distribution of active elementary modes reaches a Boltzmann distribution resulting in a maximum of the system entropy. Evolution appears to favor this state which seems to be a general property as it can be observed not only in the metabolic network of *E. coli* but also in the metabolism of other bacteria such as *Thermoanaerobacterium saccharolyticum*.

Title: MODELING Lignin Biosynthesis Using Proteomics and Isotopic Labeling Data in *Brachypodium distachyon*

Author: Jaime Barros[1 2], Him K. Shrestha[3], Juan C. Serrani-Yarce[1], John I. Hendry[4], Paul E. Abraham[2 3], Maria Magallenes-Lundback[5], Guifen Li[6], Yuhong Tang[6], Costas D. Maranas[2 4], Dean DellaPenna[5], Robert L. Hettich[2 3], and Richard A. Dixon[1 2 7]

Primary Affiliation: [1]BioDiscovery Institute and Department of Biological Sciences, University of North Texas, Denton, TX, USA; [2]Center for Bioenergy Innovation (CBI), Oak Ridge National Laboratory, Oak Ridge, TN, USA; [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA; [4]Department of Chemical Engineering, The Pennsylvania State University, State College, PA, USA; [5]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, [28824], USA; [6]Noble Research Institute, Ardmore, OK, USA; [7]BioEnergy Science Center (BESC), Oak Ridge National Laboratory, Oak Ridge, TN, USA.

Abstract

Lignin is a phenolic polymer found in plant cell walls that plays a central role in plant physiology, nutrient transport and adaptation to biotic and abiotic stresses [1]. Despite the abundance and potential of this biopolymer as raw material for conversion into renewable chemicals [2,3], our current understanding of lignin formation in plants is still incomplete.

Lignin synthesis proceeds via a series of enzymatic reactions leading to the formation of three major monomers or monolignols: *p*-coumaryl, coniferyl, and sinapyl alcohols. The aromatic amino acid L-phenylalanine is the common monolignol precursor in all plants. However, commelinid monocots, including the grasses, possess bifunctional L-phenylalanine/tyrosine ammonia-lyase (PTAL) enzymes providing the unique ability to use L-tyrosine as an efficient precursor of lignin and other phenolic compounds [4,5]. Additionally, some grasses lack bona fide caffeoyl shikimate esterase (CSE) homologs, a central lignin enzyme in dicots [6,7]. These observations have prompted a need to reevaluate lignin formation in grasses, and recent studies [8] have shown that, in addition to PTAL for direct formation of *p*-coumaric acid, a soluble bifunctional ascorbate peroxidase/*p*-coumarate 3-hydroxylase (C3H/APX) provide a parallel route to caffeic acid that bypasses the main enzymes of the 'shikimate shunt' or 'esters pathway'. However, the biological significance of the existence of parallel pathways for lignin biosynthesis has not been fully investigated, and it is unclear whether the routes operate preferentially for the synthesis of specific monolignols.

To further investigate lignin metabolism in grasses, we have undertaken a combined genetic, proteomic, and 13C-isotopic labeling approach in the model grass *Brachypodium distachyon*. We identified the complete set of monolignol pathway proteins, revealing the key enzymes in the formation of p-coumaric, caffeic, ferulic and sinapic acids (PTAL, C3H/APX and caffeic acid O-methyltransferase (COMT)) to be among the top 1% most abundant proteins in actively lignifying stem tissues. Attempts to downregulate simultaneously pairs of early lignin pathway genes, but not single genes, resulted in broad changes in protein abundances, including large reductions in key players of the RNA interference pathway cyclophilin 40 (CYP40) and argonaute 1 (AGO1) and increases in tocopherol cyclase (VTE1) and overall vitamin E levels. 13C-labeling data supported a metabolic separation of flux from the aromatic amino acid precursors into different lignin subunits and flavonoid classes, and in situ hybridization experiments showed a differential tissue-specific expression of major ammonia-lyases, suggesting cellular compartmentalization of different phenylpropanoids synthesis via PAL and PTAL respectively. Metabolic flux analysis revealed the relative contribution of PAL and TAL activities for lignin biosynthesis.

Overall, we show that the soluble acids route via PTAL is present and it is a major pathway for lignin deposition in mature *Brachypodium* stems, providing evidence for a reconsideration of a model of lignin biosynthesis via free phenolic acids in the economically important Poaceae family of the monocotyledons. Our results provide a deeper understanding of the lignification process in grasses, the world's most important food, feed, and potential bioenergy crops.

References

[1] J. Barros et al., "The cell biology of lignification in higher plants", *Annals of Botany*, 2015; 115:1053-1074.

[2] A.J. Ragauskas et al., "Lignin valorization: improving lignin processing in the biorefinery", *Science,* 2014; 344: 6185.

[3] W. Schutyser et al., "Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading", *Chemical Society Reviews*, 2018; 47: 852-908.

[4] J. Barros et al., "Role of bifunctional ammonia-lyase in grass cell wall biosynthesis", *Nature plants*, 2016; 2: 1-9.

[5] J. Barros and R.A. Dixon, "Plant phenylalanine/tyrosine ammonia-lyases", *Trends in Plant Science*, 2020; 25:66-79.

[6] R. Vanholme et al., "Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*", *Science,* 2013; 341: 1103-1106.

[7] C.M. Ha et al., "An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*", *The Plant Journal*, 2016; 86: 363-75.

[8] J. Barros et al., "4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase", *Nature communications*, 2019; 10: 1-1.

Title: ASTHERISC: A MEthod to Design Microbial Communities with MaximAl Thermodynamic Driving Force for the Production of Chemicals

Author: Pavlos Bekiaris and Steffen Klamt

Primary Affiliation: Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Abstract

Microbial communities have become a major research focus due to their importance for biogeochemical cycles, biomedicine and biotechnological applications. While some biotechnological applications, such as anaerobic digestion, make use of naturally arising microbial communities, the rational design of microbial consortia for bio-based production processes has recently gained much interest. One class of synthetic microbial consortia is based on multiple strains of one species. A typical design principle for these consortia is based on division of labor, where the entire production pathway is divided between the different strains to reduce the metabolic burden caused by product synthesis.

Here, we present ASTHERISC (Algorithmic Search of THERmodynamic advantages in Single-species Communities), a new computational approach for designing multi-strain communities of a single species with the aim to divide a production pathway between different strains such that the thermodynamic driving force for product synthesis is maximized. ASTHERISC exploits the fact that compartmentalization of segments of a product pathway in different strains can circumvent thermodynamic bottlenecks arising when operation of one reaction requires a metabolite with high and operation of another reaction the same metabolite with low concentration. We implemented the ASTHERISC algorithm in a dedicated MATLAB package, which works in concert with a new Python module, CommModelPy, that automatically generates community models from single-species models and associated thermodynamic reaction data. We applied ASTHERISC on E. coli core and genome-scale models, using different settings, e.g., regarding number of strains or minimum product yield. These calculations showed that, for each scenario, many target metabolites (products) exist where a multi-strain community can provide a thermodynamic advantage compared to a single-strain solution. In some cases, a production with sufficiently high yield is thermodynamically only feasible with a community.

In summary, ASTHERISC provides a promising new principle for designing microbial communities for the bio-based production of chemicals.

Title: Metabolic Modeling of Cellular Senescence THROUGH MULTI-OMICS INTEGRATION

Author: Marian Breuer[1], Emiel Slaats[2], Yvar Hulshof[3], Eva Thielecke[4], Sonia Bălan[5], Dimna Barbosa[6], Michalis Papachristos[7], Michiel Adriaens[8]

Primary Affiliation: [1 3 4 5 6 7 8] Maastricht Centre for Systems Biology, Maastricht University, Netherlands; [2] Faculty of Sciences, Leiden University, Netherlands

Abstract

Cellular senescence is considered a hallmark of human ageing and plays a suspected role in numerous age-related diseases. Senescent cells stopped dividing in response to certain stresses; when they persist and accumulate in a tissue, they can exert a harmful effect on their surroundings through their senescence-associated secretory phenotype. The development of drugs to selectively destroy senescent cells or ameliorate their harmful effects in humans however is challenging due to the complexity and heterogeneity of the senescence phenotype, and the need for high specificity biomarkers. A promising direction to better understand the senescence phenotype and tackle the deleterious effects of senescent cells seem to be the significant ties of senescence to metabolism, which however are still not fully understood.

To address this knowledge gap, we here study the metabolism of human senescent cells through constraint-based modeling of genome-scale metabolic networks. In order to obtain sound models, we combine both publically available metabolomics and transcriptomics data to construct context-specific models of senescent and proliferating fibroblasts. Analysis and comparison of condition-specific fluxes in these models allows us to identify metabolic characteristics specific to the senescent state that would be challenging to study in experiment.

Title: EFMlrs: Massive Parallel and Scalable Enumeration of Elementary Flux Modes

Author: Bianca A. Buchner[1] and Jürgen Zanghellini[2]

Primary Affiliation: [1]University of Natural Resources and Life Sciences, Austria; [2]University of Vienna Department of Analytical Chemistry, Austria

Abstract

The mathematical characterization of cellular metabolism by minimal metabolic pathways is a key problem in systems biology. These pathways are also known as elementary flux modes (EFMs). EFMs are minimal sets of reactions that represent unique biological pathways in a metabolic network. In fact, all feasible steady-state phenotypes can be described as a non-negative superposition of elementary pathways. However, the enumeration of unique metabolic pathways in typical genome-scale metabolic models remains currently intractable, as the number of EFMs in a metabolic network explodes combinatorically with its size [1]. This in turn causes an explosion in the memory demand of the current standard algorithm, the double description method (DDM) [2].

Here we show that EFM enumeration in highly degenerated (metabolic) networks--contrary to common doctrine--is also possible with reverse search-based methods. In fact, we examined and compared different implementations of these algorithms (*efmtool* [3], polco [] and *mplrs* [4]) as well as different mathematical proplem formulations with respect to their computational performance and their suitability for systems biology applications. Although *efmtool* is much faster than *mplrs*, it scales rather poorly and requires a huge amount of random access memory, even for small models. In contrast, *mplrs*, is almost ideally parallelizable with neglible memory demands per thread. With several hundred threads *mplrs* outperforms *efmtool*. Thus, in a high-performance computing environment mplrs offers new possibilities for an unbiased analyze of substantial larger metabolic models, see Figure 1.

To harness these new possibilities, we developed a Python package, called *EFMlrs*. *EFMlrs* uses *COBRApy* and seamlessly integrates into existing workflows. *EFMlrs* provides functionalities to process metabolic model, to perform loss-free compressions of the stoichiometric matrix, and to generate suitable inputs for *mplrs* as well as *efmtool* for the enumeration of EFMs and elementary flux vectors. It is freely available on GitHub, comes together with a designated workflow and can be easily installed via pip.

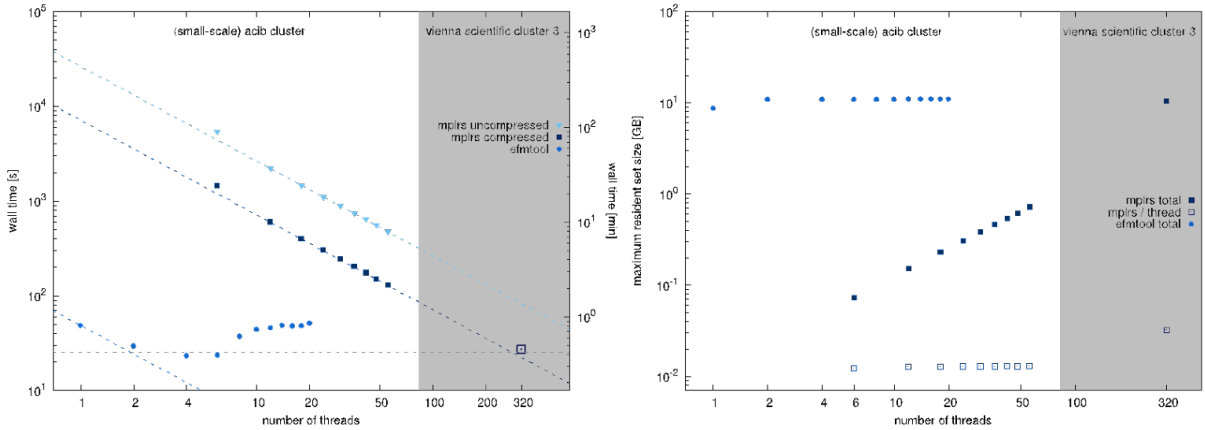


Figure 1 Runtime (left) and memory consumption (right) comparison of an E. coli core model [5] between efmtool and mplrs as function of the number of available threads

References

[1] S. Klamt, “Calculating as Many Fluxes as Possible in Underdetermined Metabolic Networks”, Mol. Biol. Rep., 2002, 29, 177-198

[2] K. Fukuda, “Double description method revisited,” M. Deza, Springer Berlin / Heidelberg, 1996, 1120, 91-111

[3] M. Terzer, “Large-scale computation of elementary flux modes with bit pattern trees”, Bioinformatics, 2008, 24,

2229--2235

[4] D. Avis, “mplrs: A scalable parallel vertex/facet enumeration code”, arXiv:1511.06487, 2017, 1511.06487

[5] C.T. Trinh, “Minimal Escherichia coli Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses”, AEM, 2008, 74, 3634 -3643

Title: The Role of RDCVFL in a Mathematical Model of Photoreceptor Interactions

Author: Erika T. Camacho[1], Emmanuelle Clérin[2], Kathryn Wifvat[3], Thierry Léveillard[4]

Primary Affiliation: [1] [3]Arizona State University, School of Mathematical & Natural Sciences, USA; [2] [4]Vision Institute of Paris, France

Abstract

The cone photoreceptors are the highest energy consuming cell in the body. They meet their high metabolic demands through aerobic glycolysis, like cancer cells but in a controlled manner. Through both aerobic glycolysis and the photo-oxidative damage due to light, damaging reactive oxygen species (ROS) are created. Antioxidants and redox (reduction-oxidation reaction) systems are in place to reduce ROS and thus minimize cone damage and loss that can result in blindness. Recent experiments suggest that failures within these processes may be to blame for retinitis pigmentosa in which cones begin to die after the vast majority of rod photoreceptors have died. While it is known that rod death is mediated by apoptosis, the manner of death in the cones is still unknown. Rods produce rod-derived cone viability factor (RdCVF) that stimulates aerobic glycolysis in the cones for the renewal of their outer segment, the structure that contains the light sensing molecule, opsin. RdCVF long form (RdCVFL), a thioredoxin expressed by rods and cones, is involved in cell protection against oxidative damage. These and other discoveries led to the concept of metabolic and redox signaling between both classes of photoreceptors because the activity of the thioredoxin RdCVFL. We created mathematical models (MMs) that are driven by and corroborate this experimental work. Experiments together with mathematical modeling offer the best hope of mitigating the progress of blindness. We investigate the role of RdCVFL in the maintenance of glycolic flux and in the repair of oxidative damages. We created the RdCVFL-/- mouse model to examine the potential role of RdCVFL in the secondary degeneration of cones. The RdCVFL-/- mouse is the most appropriate model to test experimentally the possible MMs on the role of RdCVFL in preventing loss of photoreceptor function and integrity mediated by oxidative damage. The MMs will be able to effectively simulate the process with their respective contributions to cone degeneration and test multiple hypotheses through simulation in a matter of minutes, without incurring extravagant cost.

Title: Dissecting the Metabolic Reprogramming of Maize Root Under Nitrogen Limiting Stress Condition

Author: Niaz Bahar Chowdhury[1], Wheaton L. Schroeder[1], Debolina Sarkar[2], Costas D. Maranas[2], and Rajib Saha[1]

Primary Affiliation: [1]Department of Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, USA; [2]Department of Chemical Engineering, The Pennsylvania State University, USA

Abstract

The growth and development of maize (*Zea mays* L.) largely depends on its nutrient uptake through root. Hence, studying its growth, response, and associated metabolic reprogramming to stress conditions is becoming a seminal research direction. Although there exist experimental studies on different aspects of maize-root metabolism under nutrient-starved conditions, a holistic approach for the systems-level study of maize-root metabolism under different stress conditions is missing. In this work, a genome scale metabolic model (GSM) for the maize-root of line B73 was reconstructed to study maize-root growth under nitrogen starvation conditions. The model was constructed based on the available information from public databases such as UniProt, KEGG, and MaizeCyc. This maize-root model included 6368 genes, 4002 reactions, and 4419 metabolites and a detailed gene-protein-reaction association. Transcriptomic data derived from the roots of hydroponically grown maize plants was used to incorporate regulatory constraints in the model and simulate nitrogen-non-limiting (N+) and nitrogen-deficient (N-) feeding conditions. Model-predicted result achieved 70% accuracy when compared to the experimentally observed metabolite levels. In addition to predicting important metabolic reprogramming in central carbon, fatty acid, amino acid, and other secondary metabolism, maize root GSM also predicted several metabolites (e.g., L-methionine, L-asparagine, L-lysine, cholesterol, and L-pipecolate) playing critical regulatory role in the root biomass growth. Furthermore, this study revealed eight phosphatidyl-choline and phosphatidyl-glycerol metabolites which were not coupled with biomass production and played a key role in the increased biomass production under N-. Overall, the omics-integrated GSM provides a promising tool to facilitate stress-condition analysis for maize root and hence to engineer a better stress-tolerant maize line.

Title: Integrated Knowledge Mining, Genome-Scale Modeling, and Machine Learning for Predicting *Yarrowia lipolytica* Bioproduction

Author: Jeffrey J. Czajka, Tolutola Oyetunde, and Yinjie J. Tang

Primary Affiliation: Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, USA

Abstract

Predicting bioproduction titers from microbial hosts has been challenging due to complex interactions between microbial regulatory networks, stress responses, and suboptimal cultivation conditions. This study integrated knowledge mining, feature extraction, genome scale models (GSMs), and machine learning (ML) to develop a model for predicting *Yarrowia lipolytica* chemical titers (i.e., organic acids, terpenoids, etc.). First, *Y. lipolytica* production data, including cultivation conditions, genetic engineering strategies, and product information, was collected from literature (~100 papers) and stored as either numerical (e.g., substrate concentrations and biosynthesis enzyme steps) or categorical (e.g., bioreactor modes and media type) variables. For each case recorded, central pathway fluxes were estimated using GSMs and flux balance analysis (FBA) to provide metabolic features. Second, a ML ensemble learner was trained to predict titers. Accurate predictions were obtained for instances with production titers >1 g/L (R^2=0.92). However, the model had reduced predictability for low performance strains (0.01–1 g/L, R^2=0.36), due to biosynthesis bottlenecks not captured in the features. Feature ranking indicated that the FBA fluxes, the number of enzyme steps, the substrate inputs, and thermodynamic barriers were the most influential factors. Third, the model was evaluated on other oleaginous yeasts and indicated there were conserved features for some hosts that can be exploited by transfer learning. The platform was also designed to assist computational strain design tools (such as OptKnock) to screen genetic targets for improved microbial production in light of complex experimental conditions.

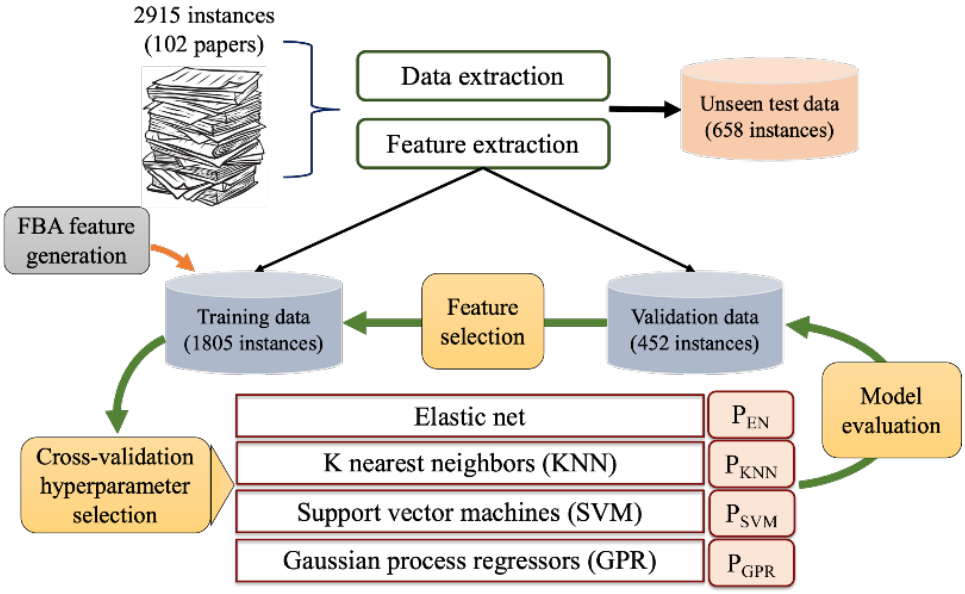


Figure. Schematic of data processing and ensemble model training procedure. *Data was split into Test-Validate-Train datasets. 10-fold cross validation was utilized for hyper-parameter selection and feature selection was performed on the validation data. Predictions of four base models (PEN, PKNN, PSVM, PGPR) were used in the ensemble model.*

References

1. Oyetunde T, Bao FS, Chen JW, Martin HG, Tang YJ. Leveraging knowledge engineering and machine learning for microbial bio-manufacturing. Biotechnology Advances. 2018. 36(4), 1308-1315

2. Oyetunde T, Liu D, Martin HG, Tang YJ. Machine learning framework for robust assessment of microbial factory performance. Plosone. 2019. 14(1):e02105584

Title: Flux Balance Network Expansion Predicts Stage-Specific Human Preimplantation Embryo Metabolism

Author: Andisheh Dadashi[1], Derek Martinez[2]

Primary Affiliation: [1]University of New Mexico, USA; [2]Kansas State University, USA

Abstract

Metabolism is an essential cellular process for the growth and maintenance of organisms. A better understanding of metabolism during embryogenesis may shed light on the developmental origins of human disease. Metabolic networks, however, are vastly complex with many redundant pathways and interconnected circuits. Thus, computational approaches serve as a practical solution for unraveling the genetic basis of embryo metabolism to help guide future experimental investigations. To our knowledge, this study presents the first model of human peri-implantation embryo metabolism via flux balance network expansion. The model accounts for 8 separate reaction compartments (extracellular, peroxisome, mitochondria, cytosol, lysosome, endoplasmic reticulum, Golgi apparatus, nucleus) including transport and exchange reactions between compartments. We found that modulating oxygen uptake promotes lactate diffusion across the outer mitochondrial layer, providing in silico support for a proposed lactate-malate-aspartate shuttle. While RNA-sequencing and other profiling technologies make it possible to elucidate metabolic genotype-phenotype relationships; yet, our understanding of metabolism during early mammalian development is currently limited. Very few studies have examined the temporal or spatial metabolomics of the human embryo, and prohibitively small sample sizes traditionally observed in human embryo research have presented logistical challenges for metabolic flux studies, hindering progress towards the reconstruction of the human embryonic metabolome. Thus, we developed a stage-specific model to serve as a proof-of-concept for the reconstruction of future metabolic developmental models. Our work shows that it is feasible to model human metabolism with respect to time-dependent changes characteristic of peri-implantation development.

Title: Metabolic Pathway Analysis in Presence of Biological Constraints

Author: Philippe Dague

Primary Affiliation: LMF, Université Paris-Saclay, CNRS, ENS Paris-Saclay, Gif-sur-Yvette, France

Abstract

Elementary flux modes (*EFMs*) or vectors (*EFVs*) enumeration, despite progress in the implementation of algorithms (computing billions of *EFMs*), is still intractable for genome-scale networks. Considering biological constraints *C* in addition to stoichiometric and irreversibility constraints both reduces the number of solutions and avoids unfeasible solutions. Using constraint satisfaction techniques, promising attempts have been made to directly compute the 'classical' *EFMs* that satisfy *C*. However, to interpret the results accurately, the following questions must be dealt with. How does the solution space Sol_C (fluxes satisfying *C*) look like? What is its mathematical structure, its geometric characterization (no reason to still be a flux cone *FC*)? Are support-minimality and conformally non-decomposability still identical in Sol_C? If not, is there a more relevant concept of non-decomposability and how do all those minimality criteria compare? Can we reconstruct Sol_C from such 'minimal' elements only and what is their relationship with what is computed? The aim of this work is to provide comprehensive answers to the above questions, by revisiting the classical concepts in this new framework, taking into account thermodynamic (*tCs*), kinetic (*kCs*) and regulatory (*rCs*) constraints, which will pave the way towards developing algorithms to compute the objects that biologists are interested in.

We show that the fundamental results depend entirely on general properties of compatibility of the constraints with the vector signs: sign-monotonicity (verified by *tCs* and *kCs*) or the weaker sign-invariance (verified by *rCs*). For sign-monotone *C*, Sol_C is a finite union of polyhedral cones, which are faces of the flux topes (defined by fixing the signs of reversible reactions in all possible ways) of the original *FC*. Support-minimal solutions $EFMs_C$ and conformally non-decomposable solutions $EFVs_C$ are still identical and coincide with $EFM_S{}^C$ and $EFV_S{}^C$, i.e., those original *EFMs* or *EFVs* that satisfy *C*. Furthermore, and importantly, they can be enumerated by the incremental Double Description algorithm, by filtering out at each step those newly built extreme rays that do not satisfy *C*. But reconstructing Sol_C as a union of cones requires identifying the largest constraint-consistent sets of *EFMs*, i.e., those maximal subsets of *EFMs* whose conformal conical hull is included in Sol_C. For sign-invariant *C*, Sol_C is a finite disjoint union of semi-open faces (cones without some of their faces of lesser dimension) of the flux topes of *FC*, and $EFVs_C$ coincide with $EFMs^C$ and $EFVs^C$. But now $EFMs^C$ is (in general strictly) included in $EFMs_C$, itself (in general strictly) included in conformally support-wise non-strictly-decomposable solutions $swNSDFVs_C$, i.e., those fluxes that cannot be conformally decomposed into two fluxes of smaller supports (this weaker concept of non-decomposability proving to be the relevant one). The precise geometric characterization and relationships of those three different types of 'minimal' solutions are given. More detailed results are provided for the specific cases of *tCs*, *kCs* and *rCs*. It is proved in particular that $Sol_{kC} = Sol_{tC}$ in absence of bounds on concentrations of enzymes. Self-contained paper [1] includes all results, proofs and references. These results (except some concerning *tCs* and *rCs*) are entirely novel.

References

[1] P. Dague, "Metabolic pathway analysis in presence of biological constraints", *BioRXiv*, ID#: BIORXIV/2020/175455, 2020.

Title: Growth Mechanics: The Economy, Control, and Optimality of Self-Replication

Author: Hugo Dourado[1], Wolfram Liebermeister[2], Oliver Ebenhöh[3], and Martin J. Lercher[4]

Primary Affiliation: [1] [3] [4]Heinrich Heine University Düsseldorf, Germany; [2]INRAE, France

Abstract

The growth rate of self-replicating cells is a major determinant of their fitness. In situations where evolutionary competition occurs mostly during phases of steady-state growth, evolution towards a cellular resource allocation pattern that maximizes the growth rate μ is expected. This expectation is the rationale behind many constraint-based methods such as flux balance analysis (FBA) [1] and resource balance analysis (RBA) [2]. These methods ignore the non-linear dependencies between reaction fluxes *v* and the concentrations *c* of cellular components, and hence require an a priori declaration of the biomass composition. In contrast, self-replicator models that account for non-linear enzyme kinetics and the production of catalytic proteins can directly relate biomass composition with the growth rate [3]. Different mathematical formulations for this self-replicator growth optimization problem have been proposed; in all of these, the number of free variables equals *m+n*, the number of biomass components (*m*) plus the number of reactions (*n*) [3,4].

Here, we propose a reformulation of the self-replicator problem, which reduces the number of free variables to *n* by stating the problem in terms of dimensionless, scaled reaction fluxes ω := v/μρ, where ρ is the cell's dry mass density and fluxes *v* are in units of mass per time and volume. This formulation allows a generalization of many results that before could only be proven for full-rank stoichiometric matrices, corresponding to a minimal reaction set (a single elementary flux mode) [6]. The growth rate μ(ω) can be expressed as an explicit function of the scaled fluxes ω, leading to explicit expressions for the corresponding partial derivatives, which can be understood as the marginal benefits [5,6] or growth control coefficients [6,7] of those scaled fluxes. The analytical necessary conditions for maximal growth rate show that active enzymes and the ribosome have zero marginal value at optimality. These results move us closer to a systems-level understanding of cellular growth; due to the much-reduced dimensionality, the reformulation may help to perform numerical optimizations of otherwise inaccessibly large cellular reaction systems. Finally, we introduce an equivalent formulation based on the singular value decomposition of the corresponding stoichiometric matrix, representing another step toward a systemic perspective on cellular metabolism and growth.

References

[1] N. E. Lewis et al., "Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods" *Nat Rev Microbiol*, 2012;10: 291–305.

[2] A. Goelzer et al., "Quantitative prediction of genome-wide resource allocation in bacteria" *Metab Eng*, 2015; 32: 232–243.

[3] D. Molenaar et al., "Shifts in growth strategies reflect tradeoffs in cellular economics" *Mol. Syst. Biol*, 2009; 5: 323

[4] D. H de Groot et al., "Elementary growth modes provide a molecular description of cellular self-fabrication" *Plos Comput Biol,* 2020; 16.

[5] W. Liebermeister, "The value structure of metabolic states" *bioRxiv,* 2018.

[6] H. Dourado et. al., "An analytical theory of balanced cellular growth" *Nat Comm*, 2020; 11: 1226.

[7] I. T. Kleijn et al., "Noise propagation in an integrated model of bacterial gene expression and growth" *Plos Compt Biol*, 2018; 14: 1–18.

Title: Comparing Glutamine and Glucose as Fuels of Cancer and Yeast Cells: A Linear Optimization Model

Author: Jan Ewald, Ziyang He, Wassili Dimitriew, and Stefan Schuster

Primary Affiliation: Dept. of Bioinformatics, Friedrich Schiller University of Jena, Ernst-Abbe-Platz 2, Jena, Germany

Abstract

Energy metabolism is crucial for all living cells and especially during fast growth or stress scenarios. A common metabolic response of eukaryotic cells in stressful conditions or fast growth, which require fast energy supply, is a switch from glucose respiration to fermentation even under sufficient oxygen supply, leading to lactate or ethanol and generating ATP. In human cancer cells this was first described by Otto Warburg [1]. The term Warburg effect is nowadays also used for similar observations in other human cell types like activated lymphocytes and microglia cells [2]. For yeast cells, the term Crabtree effect is used. In recent years, several mathematical models have been proposed to explain the Warburg effect on theoretical grounds [3,4]. Besides glucose, glutamine is a very important substrate for eukaryotic cells.

Here, we adapt a minimal model for explaining the Warburg effect so as to include the experimentally observed utilization of glutamine (WarburQ effect). It involves six combined reactions and takes into account the NADH and FADH2 balance. Our linear optimization model enables one to calculate the amount of ATP per substrate and per enzyme mass, reflecting the pathway costs, generated by glucose or glutamine respiration as well as fermentation pathways. A major advantage of linear optimization models is their scalability and straightforward mathematical analysis and biological interpretation. While our calculation supports glucose fermentation as the superior energy generating pathway in human cells with respect to ATP production rate, different enzyme characteristics in yeasts reduce this advantage or predict glucose respiration to be more effective. The latter is observed for the fungal pathogen *Candida albicans*, which is a Crabtree-negative yeast. Further, optimization results show that glutamine is a valuable energy resource and important substrate under glucose limitation, in addition to its role as carbon and nitrogen source of biomass in eukaryotic cells. In conclusion, our model explains well the observed parallel uptake of glucose and glutamine based on their role in energy metabolism.

References

[1] O. Warburg et al., "Versuche an überlebendem Carcinom-Gewebe", *Klin Wochenschr*, 1923; 2: 776-777.

[2] H. Stark et al., "Causes of upregulation of glycolysis in lymphocytes upon stimulation. a comparison with other cell types", *Biochimie*, 2015; 118: 185–194.

[3] T. Shlomi, T. Benyamini, E. Gottlieb, R. Sharan, E. Ruppin, "Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect", *PLoS Comput. Biol*., 2011; 7: e1002018.

[4] S. Schuster et al., "Mathematical models for explaining the Warburg effect: a review focussed on ATP and biomass production", *Biochem Soc Trans*, 2015; 43: 1187–1194.

Title: ModelSEED Release 2: High Throughput Genome-Scale Metabolic Model Reconstruction and Analysis

Author: José P. Faria, Filipe Liu, Janaka N. Edirisinghe, Samuel M.D. Seaver, Qizh Zhang, Nidhi Gupta and Christopher S. Henry

Primary Affiliation: Argonne National Laboratory, United States

Abstract

First released in 2010, the ModelSEED[1] genome-scale model reconstruction pipeline has now built over 200k draft metabolic reconstructions and supported hundreds of publications. Here we are describing the first major update to this model reconstruction tool with important new features including: (1) a dramatically improved representation of energy metabolism ensuring models produce accurate amounts of ATP per mol of nutrient consumed; (2) new templates for Archaea and Cyanobacteria; and (3) greatly improved curation of all metabolic pathways mapping to RAST[2] and other annotation pipelines.

We improved our model reconstruction procedure to start by constructing core models, test for proper ATP production from this core, then ensuring that ATP production does not incorrectly explode when expanding the core model to a genome-scale model. While other approaches aim to correct ATP overproduction in models, these new procedures in the ModelSEED pipeline aim to ensure that ATP overproduction does not happen in the first place. To handle the necessary expansion of templates, we developed machine learning (ML) classifiers to determine automatically which template most correctly applies to a new genome being modeled. This ML approach allows for the rapid introduction of additional modeling templates, enabling researchers working with unclassified species or metagenome-assembled genomes extracted to achieve more specific reconstructions.

To improve metabolic pathway annotation completeness and accuracy in ModelSEED models, we first updated our biochemistry database to include the latest reaction data from KEGG, MetaCyc, BIGG, and published models. We then manually curated the major pathways in our reconstruction templates to reconcile pathway representation across these multiple databases.Finally, we curated our mapping of RAST functional roles to this reconciled biochemistry based on data mined from KEGG and published metabolic models.Within the KBase[3] platform, we demonstrate our improved model reconstruction pipeline on a phylogenetically diverse set of approximately ~1600 genomes and constructed draft genome-scale metabolic models (GEMs). We show how the gene counts and modeling metrics (ATP production, biomass yields, reaction classification, pathway representation) are improved with this new release of the ModelSEED. We also selected genomes for which comprehensive Biolog data is available, and we compare model predictions of all data with experimental results, showing significant improvement compared to models generated by the original ModelSEED.

The listed improvements will be available as an update to the ModelSEED reconstruction pipeline. The new release will be made available across all web platforms currently supported; KBase, ModelSEED, and PATRIC resources.

References

[1] Henry, Christopher S., et al. “High-throughput generation, optimization and analysis of genome-scale metabolic models.” Nature biotechnology 28.9 (2010): 977-982.

[2] Aziz, Ramy K., et al. “The RAST Server: rapid annotations using subsystems technology.” BMC genomics 9.1 (2008): 1-15.

[3] Arkin, Adam P., et al. “KBase: the United States department of energy systems biology knowledgebase.” Nature biotechnology 36.7 (2018): 566-569.

Title: What is a STOICHIOMETRIC MATRIX?

Author: Ronan M.T. Fleming

Primary Affiliation: Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands

Abstract

Characterising biochemical reaction network structure in mathematical terms enables the inference of functional biochemical consequences from network structure with existing mathematical techniques and spurs the development of new mathematics that exploits the peculiarities of biochemical network structure. The structure of a biochemical network may be specified by reaction stoichiometry, that is, the relative quantities of each molecule produced and consumed in each reaction of the network. A biochemical network may also be specified at a higher level of resolution in terms of the internal structure of each molecule and how molecular structures are transformed by each reaction in a network. The stoichiometry for a set of reactions can be compiled into a stoichiometric matrix $N \in Z\ m \times n$, where each row corresponds to a molecule and each column corresponds to a reaction. We demonstrate that a stoichiometric matrix may be split into the sum of m-rank (N) moiety transition matrices, each of which corresponds to a subnetwork accessible to a structurally identifiable conserved moiety [1]. The existence of this moiety matrix splitting is a property that distinguishes a stoichiometric matrix from an arbitrary rectangular matrix. This result has fundamental importance from a theoretical perspective, because it provides a bridge between stoichiometric analysis and compartmental analysis, e.g., pharmacokinetic modelling. This result is also practically important, e.g, for the development of efficient algorithms for high dimensional flux balance analysis, or for the development of optimization algorithms for genome-scale kinetic modelling.

References

[1] Ghaderi, S., Haraldsdóttir, H. S., Ahookhosh, M., Arreckx, S. & Fleming, R. M. T. Structural conserved moiety splitting of a stoichiometric matrix. Journal of Theoretical Biology 499, 110276 (2020).

Title: Metabolic Complementarity Applied to the Screening of Microbiota and the Identification of Key Species

Author: Clémence Frioux[1], Arnaud Belcour[2], Méziane Aite[3], Anthony Bretaudeau[4], Falk Hildebrand[5] and Anne Siegel[6]

Primary Affiliation: [1]Inria Bordeaux Sud-Ouest, France; [2]Univ Rennes, Inria, CNRS, IRISA, France; [3]Univ Rennes, Inria, CNRS, IRISA, France; [4]INRAE IGEPP Rennes, Inria IRISA GenOuest core facility Rennes, France; [5]Quadram Institute, Earlham Institute Norwich, United Kingdom; [6]Univ Rennes, Inria, CNRS, IRISA, France

Abstract

Deciphering the complex interactions occurring within host-associated and environmental microbiomes is of utmost interest for health and ecological applications. As the culture of all these microbes is impractical, omics experiments and in particular shotgun metagenomics are the main sources of information to analyse the web of interactions in these large communities [1]. The number of available metagenomes is increasing, and in addition to the identification of species in the samples, the characterisation of the functions they carry is becoming as much valuable [2]. Indeed, these functions can be critical to understand the interactions between species, notably those related to cross-feeding [3]. In order to capture information on both the taxonomy and the functions, the reconstruction of metagenome-assembled genomes (MAGs) is becoming routine [4–6]. Combined to methods for metabolic network reconstruction, MAGs enable to dive into the metabolic landscape of a metagenome [4, 7]. However, the remaining imperfection of automatically-generated data - both for reconstructed MAGs and reconstructed metabolic networks [8, 9] - still impairs the applicability of metabolic modelling at the scale of a metagenome.

Our previous work in metabolic modelling and community selection [10] and recent developments led to the creation of a pipeline, Metage2Metabo (M2M), dedicated to the metabolic screening of large communities of microbes (Figure 1a) [11]. M2M automatically reconstructs and screens the metabolic potential of thousands of microbes, considered both individually and as a community, in order to evaluate the metabolic gain brought by mutualistic interactions. We use the network expansion algorithm to robustly predict producibility of metabolites within metabolic networks [10, 12, 13]. Metabolites predicted to be produced collectively only, or other sets of compounds can be used as an objective to identify minimal communities predicted to sustain their producibility. As up to millions of equivalent minimal communities can exist due to the combinatorics of the problem, we solve it using the logic paradigm Answer Set Programming in order to retrieve the key species (KS), that are all microbes occurring in at least one of such communities (Figure 1b). We can further distinguish KS by identifying those that occur in every minimal community, thereby targeting the metabolic key players within the original microbiome. We illustrated our methods and demonstrated their versatility using various genomic and metagenomic datasets. Applied to 1,520 high-quality draft reference genomes of the human gut microbiota, we established the screening potential of M2M and determined key species for several categories of metabolic end-products, that we further compared to literature knowledge. In addition, we established the robustness of M2M predictions on degraded MAGs demonstrating that M2M is applicable to metagenomics. Finally, we used M2M to screen the metabolism associated to the gut microbiota of individuals in the context of diabetes.

Functional description and complexity reduction of large communities are critical matters in the journey towards a better understanding of microbiota organisation. Our work provides a step in that direction by identifying functions and species of interest in microbial ecosystems.

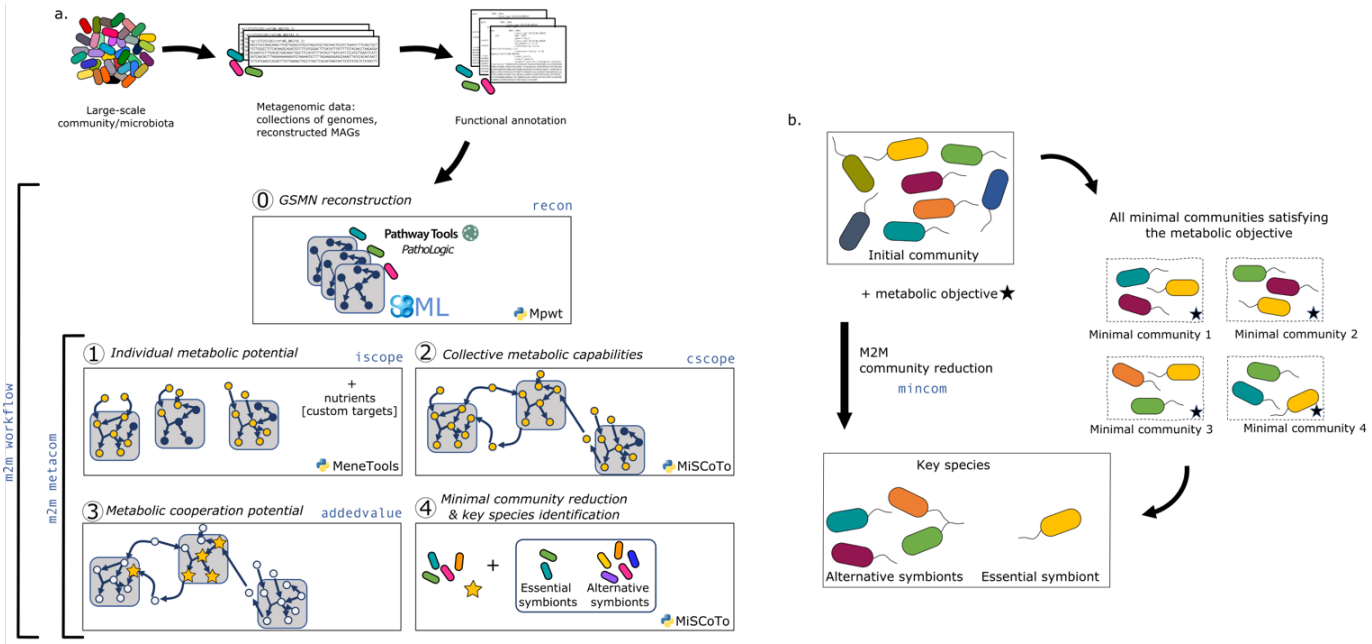


Figure 1 –Metabolic screening of a community as implemented in Metage2Metabo. a) Main steps of the pipeline, all can be performed independently. b) Illustration of the key species concept. Key species are those appearing in at least one minimal community predicted to sustain a given metabolic objective. GSMN = Genome-scale metabolic network. MAGs = Metagenome-assembled genomes. Figure adapted from [11].

References

[1] Nayfach, S. et al.: New insights from uncultivated genomes of the global human gut microbiome. *Nature*, **568** (7753), 2019, p. 505–510.

[2] Heintz-Buschart, A., Wilmes, P.: Human Gut Microbiome: Function Matters. *Trends Microbiol*, **26** (7), 2018, p. 563–574.

[3] Zelezniak, A. et al.: Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc National Acad Sci,* **112** (20), 2015, p. 6449–6454.

[4] Zorrilla, F. et al.: metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Biorxiv*, 2021, p. 2020.12.31.424982.

[5] Li, D. et al.: MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31** (10), 2015, p. 1674–1676.

[6] Kang, D.D. et al.: MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *Peerj*, **7**, 2019, p. e7359.

[7] Frioux, C. et al.: From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes. *Comput Struct Biotechnology J*, **18**, 2020, p. 1722–1734.

[8] Thiele, I., Palsson, B.Ø.: A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, **5** (1), 2010, p. 93–121.

[9] Karp, P.D. et al.: How accurate is automated gap filling of metabolic models? *Bmc Syst Biol*, **12** (1), 2018, p. 73.

[10] Frioux, C. et al.: Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics*, **34** (17), 2018, p. i934–i943.

[11] Belcour, A. et al.: Metage2Metabo, microbiota-scale metabolic complementarity for the identication of key species. *Elife*, **9**, 2020, p. e61968.

[12] Ebenhöh, O. et al.: Structural analysis of expanding metabolic networks. *Genome Informatics Int Conf Genome Informatics*, **15** (1), 2004, p. 35–45.

[13] Kruse, K., Ebenhöh, O.: Comparing flux balance analysis to network expansion: producibility, sustainability and the scope of compounds. *Genome Informatics Int Conf Genome Informatics*, **20**, 2008, p. 91–101.

Title: Metabolic Profiling of Cultured Erythroblast for the Large Scale Production of Red Blood Cells

Author: Joan S. Gallego M[1 2], Nurcan Yagci[2], Angelo D'Alessandro[3], Luuk van der Wielen[1], Emile van den Akker[2], Marieke von Lindern[2], Aljoscha Wahl[1]

Primary Affiliation: [1]Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands; [2]Sanquin Research, Department of Hematopoiesis, The Netherlands; [3]Department of Biochemistry and Molecular Genetics, University of Colorado Denver Anschutz Medical Campus, United States of America

Abstract

1. Introduction

Transfusion of donor-derived red blood cells (RBCs) is the most common form of cell therapy. Nevertheless, this fully donor-dependent system faces challenges such as emerging blood-borne diseases, and supply limitations. Production of cultured RBCs (cRBCs) has been proposed as a potential alternative. In this process, erythroid precursors can be cultured from hematopoietic progenitors and differentiated into transfusion-ready cRBCs.

However, the large number of RBCs required for a single transfusion unit, current limitations in maximum cell concentrations in culture, and the high cost of culture medium drive the production of cRBCs to market-incompatible high costs [1]. Understanding the metabolic changes of erythroblasts during their expansion and differentiation can allow to customize medium formulations and to develop feeding strategies to optimize these processes. In this work, we compared the metabolic profile of cultured erythroblasts under expansion conditions at low and high cell concentrations, and during differentiation into reticulocytes. We aim at identifying the origin of the previously described cell density limitations [2] and to propose medium improvements to increase the cell yield of the process.

2. Methods

Untargeted exo- and endometabolomics measurements were performed for erythroblast expansion and differentiation cultures. During expansion, cultures at varying starting cell densities were compared at 0, 12, 24 and 36h of culture to identify depleted metabolites and/or accumulating metabolic inhibitors. For differentiation cultures, samples were taken daily for 12 days in culture conditions without any medium refreshment. Metabolomics data was processed using MetaboAnalystR [3] to identify metabolic pathways potentially related to metabolite limitations or accumulation of toxic growth inhibitors. Following this, medium modifications were proposed and tested on culture, and the effect on the metabolome was evaluated by further metabolomics measurements.

3. Results and discussion

Media reconstitution experiments indicated that components smaller than 3kDa are responsible for growth inhibition of erythroblasts under expansion conditions. After 36h of culture, we observed degradation of nucleosides, a strong depletion of essential lipids and amino acids, and a decrease in intermediates of the glutathione-ascorbate, γ-glutamyl and cysteine-methionine cycles. The latter pathways are involved in glutathione metabolism, a key intracellular antioxidant. Transient accumulation of hypoxanthine and the accumulation of allantoate in high cell concentration cultures suggested a high flux through the purine degradation pathways and a potential accumulation of oxidative stress. Removal of nucleosides from the medium led to an increase in the growth rate of cultured erythroblasts. During erythroblast differentiation, most of the changes in exo- and endometabolome took place in the first 2 days of culture, resembling the patterns observed in erythroblast expansion, followed by a relatively unchanged metabolic profile during terminal differentiation.

4. Conclusions

Accumulation of and/or depletion of component(s) smaller than 3 kDa seems to inhibit and/or limit growth in high cell density erythroblast cultures. Metabolites involved in purine degradation and oxidative stress response pathways were identified from untargeted semiquantitative metabolomics measurements. To further identify critical pathways and potential limiting metabolites we are currently developing an erythroblast-specific genome-scale metabolic model (GSMM) using the Human1 model, together with proteomics and transcriptomics data from cultured erythroblasts.

References

[1] A. Zeuner et al., "Concise Review: Stem Cell Derived Erythrocytes as Upcoming Players in Blood Transfusion", *Translational and Clinical Research*, 2012; 30: 1587-1596.

[2] K. Glen et al., "A mechanistic model of erythroblast growth inhibition providing a framework for optimisation of cell therapy manufacturing", *Biochemical Engineering Journal*, 2018; 133: 28-38.

[3] Pang et al., "MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics", *Metabolites*, 2020; 10: 186.

Title: Identification of Metabolic Pairs Allow a Reliable and Quantitative Analysis of the FINGER SWeat Metabolome

Author: Mathias Gotsmy, Julia Brunmair, Andrea Bileck, Astrid Slany, Samuel Matthias Meier-Menches, Christopher Gerner, and Jürgen Zanghellini

Primary Affiliation: Department of Analytical Chemistry, University of Vienna, Austria

Abstract

Typically, clinical metabolome analysis is performed on blood samples. However, drawing blood is not only a cumbersome procedure for patients but requires qualified personnel which impairs measurement during real life settings. A promising alternative is the analysis of the metabolome from finger sweat where sampling is as simple as holding filter paper between fingertips. High-resolution orbitrap MS/MS hyphenated with UHPLC then enabled metabolomic phenotyping from minute amounts of the collected sweat. This method drastically simplifies sampling at short intervals which is valuable for time course studies.

However, a major obstacle to finger sweat analysis is the inability to control or measure the amount of sweat produced by the sweat glands at any given time. Even conservative estimates put the variability of the sweat flux on finger tips between 0.05 and 0.62 mg cm^{-2} min^{-1}, depending on multiple endo- and exogenous factors. Not addressing this problem prevents a reliable quantification of metabolites in finger sweat. Here we present a computational method based on the identification of metabolic pairs in the sweat metabolome that allows us to quantify sweat volumes and enables an individualized, accurate quantitative finger sweat analysis for clinical applications.

In a proof-of-principal application we use short interval sampling of sweat from fingertips to monitor the dynamic response of 43 indifiduals after caffeine consumption. We not only identified corresponding xenobiotics, but extracted individualized kinetic parameters of caffeine metabolites from sweat and show the long-time stability of these parameters. Moreover, based on the computationally recovered sweat volumes we identified marker metabolites that are correlated to the sweat volume, which in turn allows us to predict sweat volumes of future metabolome measurements.

In conclusion, this work highlights the feasibility of individualized and reliable biomonitoring using sweat samples from fingertips which may have far reaching implications for personalised medical diagnostics and biomarker discovery.

Title: Proteome Efficiency of Metabolic Pathways in *Escherichia coli*

Author: Xiao-Pan Hu and Martin J. Lercher

Primary Affiliation: Heinrich Heine University Düsseldorf, Germany

Abstract

What determines the allocation of proteome fractions to different cellular processes? Answering this question is crucial to understand the organizing principles of bacterial physiology. Phenomenological "bacterial growth laws" can describe the partitioning of the proteome into 3-7 sectors that scale linearly with changes in growth rate; corresponding coarse-grained models describe a wide range of growth-related phenomena quantitatively [1,2]. The three most important sectors have been identified with protein translation, anabolism, and catabolism; however, while more than half of the proteome is involved in metabolism [3], there is no simple mapping between these sectors and metabolic pathways.

Here, we explore if the proteome allocation to different metabolic pathways can also be expressed as simple linear functions of the growth rate, and if this allocation can be explained by optimal proteome efficiency. We partitioned metabolic activities into biosynthesis pathways for amino acids, nucleotides, cofactors, lipid, peptidoglycan, and lipopolysaccharides; pathways for the generation of precursors and energy, including glycolysis, pentose phosphate pathway, TCA cycle, glyoxylate shunt, and oxidative phosphorylation; transporters; and other pathways. Analyzing quantitative proteomics data [3], we found that for most pathways, the allocated proteome fraction is a linear function of the growth rate. To estimate the growth-rate dependent optimal protein demand for each pathway, we used a modified version of flux balance analysis (FBA) with molecular crowding [4]. We incorporated a growth rate-dependent biomass composition, based on the experimentally observed scaling of the RNA/protein mass ratio [1] and the cell volume [5] in nutrient-limited conditions. Parameterizing the model with recently published in vivo effective enzyme turnover numbers [6], we predicted the minimal proteome demand in a given nutrient condition at the observed growth rate.

Comparing the predicted optimal proteome demand of individual pathways with the experimental data [3], we found that proteome efficiency can qualitatively explain the growth rate-dependent expression of biosynthesis pathways, glycolysis, and the pentose phosphate pathway, but is unable to explain the expression of other pathways. Unexpectedly, by mass, more than half of the metabolic pathways show a growth rate-dependence opposite of that expected from optimal demand; this includes transporters, the TCA cycle, and the glyoxylate shunt. Overall, growth rate-dependent proteome efficiency increases along the carbon flow through the metabolic network (Fig. 1): transporters scale contrary to optimal demand; different pathways of central metabolism scale contrary to optimal demand, are independent of optimal demand, or show a scaling that agrees qualitatively with optimality; the growth rate-dependence of most biosynthesis pathways scales qualitatively as predicted from optimal demand. While our work thus provides a bird's-eye view of proteome efficiency at the pathway level, future work will have to elucidate why proteome allocation evolved this way, and how it gives rise to the widely used bacterial growth laws.

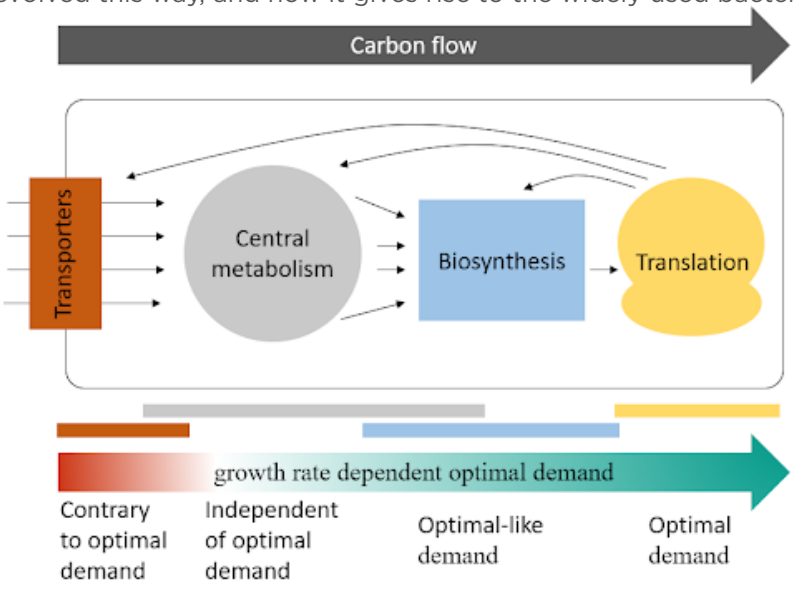


Figure 1. The growth rate dependent proteome efficiency increases along the carbon flow through metabolic network.

References

[1] Scott M et al., Interdependence of cell growth and gene expression: origins and consequences, *Science*, 2010; 330: 1099–102.

[2] Hui S et al., Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria, *Mol Syst Biol,* 2015; 11: 784.

[3] Schmidt A et al., The quantitative and condition-dependent Escherichia coli proteome, *Nat Biotechnol*, 2016; 34: 104–10.

[4] Adadi R et al., Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *Plos Compt Biol*, 2012; 8: e1002575.

[5] Si F et al., Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. *Curr Biol*, 2017; 27: 1278–87.

[6] Heckmann D et al., Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proc Natl Acad Sci* U S A, 2020;117:23182–90.

Title: Metabolic NETWORK Modeling of Degrader-Cheater Interactions in A Chitin-DECOMPOSING Model Soil CommunitY

Author: Aimee K. Kessell[1], Ryan S. McClure[2], Yuliya Farris[2], William C. Nelson[2], Christopher S. Henry[3], Janaka N. Edirisinghe[3], Kirsten S. Hofmockel[2], Janet K. Jansson[2] and Hyun-Seob Song[1 4]

Primary Affiliation: [1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, United States; [2]Pacific Northwest National Laboratory, United States; [3]Argonne National Laboratory, United States; [4]Department of Food Science and Technology, Nebraska Food for Health Center, University of Nebraska–Lincoln, Lincoln, NE, United States

Abstract

Chitin – one of the most abundant biopolymers in nature – is a major structural component in fungi, as well as insects and crustaceans. In natural environments, including soil systems, chitin is decomposed into N-acetyl-D-glucosamine (NAG) monomers through the chitinolytic cascade. Chitin decomposition is a complex and energetically expensive process that involves the induction of several dozens of proteins [1]. Since not all microorganisms are able to degrade chitin, there are non-degraders (or cheaters) that take advantage of hydrolysis products (NAGs) without contributing to chitin decomposition. It is therefore critical to understand how degraders and cheaters interact with each other and how their interactions affect community function and dynamics across different contexts. Here we present genome-scale metabolic network models of tractable microbial communities to predict interspecies interactions during chitin-degradation. For this purpose, we used a model soil consortium (MSC) developed by the Pacific Northwest National Laboratory (PNNL)'s Soil Microbiome Scientific Focus Area (SFA) [2], which is composed of eight genome sequenced isolates that includes both degraders and cheaters. Using the DOE's KnowledgeBase (KBase) modeling pipeline [3], we first built metabolic networks of the eight isolates that show consistency with phenotypic growth data from 11 different carbon sources in minimal media. These individual metabolic networks were subsequently merged in KBase into compartmentalized-community models. Models of sub-communities composed of a few species of degraders and cheaters predicted the provision of NAG from a chitin-degrader, e.g., *Variovorax paradoxus*, to non-degraders such as *Streptomyces venezuelae* or *Neorhizobium galegae*. Sub-community models also predicted the membership-dependent shifts in pathway activation and interspecies interactions. We are continuing to increase the complexity of the model by incorporating the full 8-member community to understand context-dependent microbial interactions in response to variations in substrate inputs. The resulting mechanistic knowledge of the relationships between the membership, interspecies interactions, and community phenotypies, 'metaphenomes', acquired through the MSC modeling provides a fundamental basis for deciphering governing processes underlying the complex dynamics of soil microbiomes.

References

[1] S. Beier and S. Bertilsson, "Bacterial chitin degradation – mechanisms and ecophysiological strategies", Frontiers in Microbiology, 2013; 4: 1-12.

[2] R.S. McClure et al., "Development and analysis of a stable, reduced complexity model soil microbiome", Frontiers in Microbiology, 2020; 11: 1-15.

[3] A.P. Arkin et al., "KBase: The United States Department of Energy Systems Biology Knowledgebase", Nature Biotechnology, 2018; 36: 566 – 569.

Title: GC-MS-based 13C-Metabolic Flux Analysis in Pseudomonads and Related Species

Author: Michael Kohlstedt[1], Stephen Kevin Dolan[2] and Christoph Wittmann[1]

Primary Affiliation: [1]Institute of Systems Biotechnology, Saarland University, Germany; [2] Cambridge Infectious Diseases, Interdisciplinary Research Centre, University of Cambridge, United Kingdom

Abstract

Pseudomonas species are of high industrial relevance owing to their broad product spectrum ranging from polyhydroxyalkanoates, over high value natural products to bulk chemicals, as well as their catabolic potential to make use of various cheap and renewable feedstocks [1,2]. Other members of the clade are human, insect and plant pathogens intensively studied to develop novel drugs and therapies. Pseudomonads and many related bacteria are known to assimilate glucose via three convergent peripheral pathways. After uptake into the periplasm, glucose can be either directly phosphorylated into glucose 6-phosphate or it can be oxidized into gluconate, which is then taken up in the cytoplasm. Additionally, gluconate can be further converted into 2- ketogluconate, subsequently phosphorylated and finally reduced via the 2-ketogluconate loop. All three catabolic steps converge at the level of 6-phosphogluconate, which is next cleaved into glyceraldehyde 3-phosphate and pyruvate via the Entner-Doudoroff pathway.

Here we present a simple and solely GC-MS-based approach to resolve the complex metabolic architecture of Pseudomonads as demonstrated for the two model species *P. putida* KT2440 and *P. aeruginosa* PAO1. We combined parallel labeling experiments using different 13C tracer mixtures with GC-MS measurements of labeling data obtained from proteinogenic amino acids, cell glycogen and peptidoglycan, as well as fragments thereof. The additional information for glucose 6-phosphate and fructose 6-phosphate from cell glycogen and glucosamine resulted in an extensive data set (534 mass isotopomer distributions) with sufficient power to resolve all relevant fluxome features in the investigated bacteria. Excellent fits between experimental and simulated data (R^2 = 0.9995) were obtained.

As a proof of concept, different *P. putida* producer strains were stressed by the degradation of lignin-derived aromatic compounds to produce *cis,cis* muconic acid [4]. Distinct flux adjustments in carbon, energy and redox metabolism to cope with these toxic compounds were observed In addition to that, when applied to *P. aeruginosa* growing on various carbon sources, highly resolved metabolic flux maps for this human pathogen could be generated for the first time [5].

References

[1] Weimer et al. “Industrial biotechnology of *P. putida*: advances and prospects”, *Appl Microbiol Biotechnol*, 2020, 104, 7745–7766

[2] Poblete-Castro et al. “Host Organism: *P. putida*”, In: Industrial Biotechnology: Microorganisms, Eds.: Liao & Wittmann, Wiley-VCH, Weinheim

[3] Kohlstedt et al. „GC-MS-based 13C metabolic flux analysis resolves the parallel and cyclic glucose metabolism of *P. putida* KT2440 and *P. aeruginosa* PAO1” *Metab Eng*, 2019, 54:35–53.

[4] Kohlstedt et al. “From lignin to nylon: Cascaded chemical and biochemical conversion using metabolically engineered *P. putida*” *Metab Eng*, 2018, 47:279–293.

[5] Dolan et al. „Contextual flexibility in *P. aeruginosa* central carbon metabolism during growth in single carbon sources”, *mBio*, 2020 11:e02684-19.

Title: Efficient Dry Mass Utilization May Explain Bacterial Growth Laws that Relate Catalyst Concentrations with Growth Rate

Author: Martin J. Lercher[1], Hugo Dourado[2], Matteo Mori[3], Xiao-Pan Hu[4], Terence Hwa[5]

Primary Affiliation: [1]Heinrich Heine University Düsseldorf, Germany; [2]Heinrich Heine University Düsseldorf, Germany; [3]University of California at San Diego, U.S.A.; [4]Heinrich Heine University Düsseldorf, Germany; [3]University of California at San Diego, U.S.A.

Abstract

Much recent progress has been made to understand the impact of proteome allocation on bacterial growth. The observed linear relationships between the growth rate and the cellular investment into different protein sectors has led to the formulation of phenomenological bacterial growth laws [1], [2]. These relationships include substantial fractions of catalytic proteins at zero growth, which are not expected to carry any notable reaction fluxes. While these protein "offsets" may in part result from the maintenance of protein reserves that accelerate the transition into states of faster growth [3], [4], it is not clear if such pre-emptive investment is indeed their predominant underlying cause. In previous work, we have shown that the relationship between the concentrations of enzymes and their substrates in *E. coli* may be the consequence of optimal resource allocation, subject to an overall constraint on the biomass density [5]. Here, we explore to what extent the same organizing principle may explain the observed growth laws.

We show that the optimal protein concentration has a linear-plus-square root relationship to the reaction flux, with the enzyme becoming progressively less saturated as flux is reduced. This predicted optimal relation accounts quantitatively for the decreasing abundances of the two most expensive catalytic molecules in *E. coli* grown in progressively poorer minimal growth conditions [6]: the ribosome and the enzyme metE, which catalyzes the last step in methionine biosynthesis. Under simplifying assumptions, the non-linear relationship between proteome allocation and growth rate can be approximated through a linear function, where the concentration of a metabolite in a reference condition relates to the expected combined offset of all enzymes capable of consuming the metabolite. We find that this expectation is consistent with observations when combining the offsets of all enzymes [6] consuming one of the 43 metabolites assayed in Ref. [7], but not when considering individual metabolites and their consuming enzymes.

We speculate that regulatory evolution in *E. coli* has implemented a near-optimal scaling of protein abundance with reaction fluxes only for the most expensive catalysts (metE, ribosome); yet even in these two cases, observations do not support the expected downward curvature as the growth rate approaches zero, suggesting that cellular regulation approximates the optimal non-linear scaling with a – likely more robust – linear dependence. For other enzymes, the regulatory implementation of the optimal scaling may be so approximate that it is clearly visible only when combining many enzymes into large protein sectors.

References

[1] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, "Interdependence of cell growth and gene expression: Origins and consequences," *Science*, 2010; 330:1099–1102

[2] S. Hui *et al*., "Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria," *Mol. Syst. Biol*., 2015; 11: e784

[3] M. Mori, S. Schink, D. W. Erickson, U. Gerland, and T. Hwa, "Quantifying the benefit of a proteome reserve in fluctuating environments," *Nat. Commun*., 2017; 8: 1225

[4] D. W. Erickson, S. J. Schink, V. Patsalo, J. R. Williamson, U. Gerland, and T. Hwa, "A global resource allocation strategy governs growth transition kinetics of Escherichia coli," *Nature*, 2017; 551:119–123

[5] H. Dourado, V. G. Maurino, and M. J. Lercher, "Enzymes and Substrates Are Balanced at Minimal Combined Mass Concentration in vivo," *bioRxiv*, 2017; doi: https://doi.org/10.1101/128009

[6] A. Schmidt *et al*., "The quantitative and condition-dependent Escherichia coli proteome," *Nat. Biotechnol*., 2016; 34:104–110

[7] L. Gerosa *et al*., "Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data," *Cell Syst*., 2015; 1: 270–282

Title: Advances in Metabolic Modeling Tools in KBase and ModelSEED Enable Integration of Multiomics Data for Pathway Discovery

Author: Filipe Liu, Jose Faria, Janaka Edirisinghe, Claudia Ortiz, Sam Seaver, Valerie De Crecey, Andrew Hanson, Oliver Feihn, Christopher Henry

Abstract

Increasingly multi-omics data is becoming more accessible for the study of a wide range of complex biological systems. Today, large-scale metagenomes can be readily obtained from soil microbiome systems, while the instruments and protocols surrounding the collection of metabolomic and proteomic data are constantly improving. Yet analysis methods still struggle to annotate these individual datasets, let alone combine them to discover new biological principles. For example, one of the great challenges associated with the use and interpretation of metabolomics data is the large portion of observed peaks that cannot be readily associated with known biochemical compounds. With the lack of clear identities for peaks, and with many identified peaks lacking known pathways, analysis is often limited to correlations alone. In this talk, we will discuss recent advances in tools and workflows in KBase and ModelSEED that are expanding the possibilities and opportunities for the use of metabolic models to integrate multi-omics data for the discovery of novel biochemical pathways.

Specifically, we have made significant improvements to our pipeline for the rapid reconstruction of metabolic models from sequence data, including isolate genomes and metagenomes. Now models have hundreds of additional genes and reactions, produce energy in biologically relevant ways, and include tailored templates for archaea, bacteria, plants, fungi, and cyanobacteria. We also offer a fully integrated pipeline for the prediction of novel biochemical compounds and reactions using cheminformatics approaches, including prediction of novel promiscuous enzymatic reactions and spontaneous chemical reactions. Finally, we have flux balance analysis workflows for combining genomic-based and novel chemical networks together to predict pathways to explain metabolomics data.

Scientifically, we will explore how these improved tools permit us to study pathway variation across the microbial tree of life, learn insights about microbial diversity and variation from microbiome data, and study evolutionary implications over how potential spontaneous reactions occur across the known metabolic pathways. We'll demonstrate our multi-omics integration tools to discover new pathways in the JCVI minimal genome and to mechanistically map metabolites to microbes within the human microbiome. Our exploration of microbiome data demonstrates organizing principles for the assembly and function of microbiome systems.

Finally, we will discuss an exciting new way of accessing ModelSEED reconstruction capabilities and analyses through a new ModelSEEDpy module, designed to be completely compatible with COBRApy.

While still in development, we will have a prototype release of this module at MPA.

Title: ModelSEEDpy: A Python Library for Automated Model Reconstruction and Analysis

Author: Filipe Liu and Christopher Henry

Primary Affiliation: Computing, Environment, and Life Sciences Division, Argonne National Laboratory, Lemont, IL, USA

Abstract

Genome-scale metabolic models (GEMs) have come a long way, from single specie models to community models of thousand of species. As result of many years of research, tools and platforms were developed to build and manipulate GEMs with the goal to better utilize this technology for biological research.

The SEED microbial models have been widely used for a variety of studies from modelling single microbes to large scale microbial communities [1]. A key feature of the SEED models is the ability to fast scaffold draft models from any prokaryote genome, allowing to quickly generate the initial data to bootstrap these studies.

As of today, there are a few web services that allows users to create SEED models, however a programmable interface is still nonexistent, making it difficult to include in other methods and tools.

Here we present the ModelSEEDpy, a Python framework of a set of libraries that provide programable functions to build, manipulate and evaluate SEED models. The core modelling functionalities are built on top of the COBRApy [2] python library, thus making ModelSEEDpy models fully compatible with other tools that are also built with COBRApy.

The initial release of ModelSEEDpy includes prokaryote GEM reconstruction, gapfilling, community model assembly, visualization, and ontology translation between BiGG and SEED and a KBase Object API.

The model reconstruction pipeline is based on RAST [3] annotations, ModelSEEDpy provides functionalities to annotate genomes with RAST and subsequently generate draft SEED models from these genomes.

To map proteins to reactions ModelSEEDpy uses a curated template library that maps RAST annotations to protein complexes that subsequently maps to the SEED biochemistry [4]. We currently maintain a core template that contains only essential reactions for energy metabolism, a gram positive and negative template for genome-scale metabolism and later adding the cyanobacteria and archaea specialized templates. With the ModelSEEDpy the users can also create, edit, and manipulate our existing templates or generate new ones to redesign the reconstruction knowledge base to fit their needs.

We develop our flux analysis and gapfilling methods for reconstruction and model analysis purposes, the library includes several methods to evaluate the draft GEMs such as biolog, gene essentiality and pathway presence, etc.

For visualization we use the Escher widgets [5] to include built in pathway visualization within our model object allowing to visualize metabolic pathways within the jupyter notebooks.

To enable compatibility with the BiGG models [6] we implemented an ontology conversion between the BiGG and SEED representations, this allows the use of BiGG models or other tools using BiGG representation with the SEED modelling framework.

Finally, a KBase [7] Object API module is included, enables the interface with data from the KBase Narratives, which allows ModelSEEDpy to read data generated from KBase or to upload data in order to use with other available tools in KBase.

References

[1] Almut *et al*. AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities, 2020; bioRxiv 2020.11.09.375451

[2] Ebrahim, A., Lerman, J. A., Palsson, B. Ø., & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Systems Biology, 2013; 7(1), 74.

[3] Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., ... Zagnitko, O. The RAST Server: Rapid

Annotations using Subsystems Technology. BMC Genomics, 2008; 9(1), 75. https://doi.org/10.1186/1471-2164-9-75

[4] Seaver *et al*.. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Research, 2021; 49, D575-D588

[5] King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. Ø. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLOS Computational Biology, 2015; 11(8), e1004321.

[6] King *et al*. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Research, 2016; 44(D1), D515–D522.

[7] Arkin *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnology, 2018; 36(7), 566–569.

Title: Applications of Answer Set Programming to the Analysis of Constraint-Based Models

Author: Maxime Mahout[1], Laurent Simon[2] and Sabine Peres[1]

Primary Affiliation: [1]Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, Orsay, France; [2]Bordeaux-INP, Université Bordeaux, LaBRI, Talence Cedex, France

Abstract

Answer Set Programming (ASP) is a declarative logic programming formalism for NP-Hard problems with a highly active community. It has been used to solve a variety of biological problems including in the metabolic pathways analysis field, such as in Gebser et al. [1] and Frioux et al [2]. The language has been extended to handle other types of constraints, including linear constraints over integer and real numbers.

Elementary Flux Modes (EFMs) provide a rigorous basis to systematically characterize steady state fluxes, cellular phenotypes, as well as metabolic network robustness and fragility on constraint-based models.

Minimal Cut Sets (MCSs) are a dual framework to EFMs providing the sets of reactions that should be removed from the network to disable a certain target reaction such as the biomass reaction.

The number of EFMs typically grows exponentially with the size of the metabolic network, leading to prohibitive computational demands and, unfortunately, a large fraction of these EFMs is not biologically feasible. This combinatorial explosion generally prevents the complete analysis of genome-scale metabolic models (GSMs), but partial analyses are possible by restricting the solution space with additional constraints. Traditionally, EFMs are computed by the double description method, an efficient algorithm based on matrix calculation; however only a few constraints can be integrated into this computation. Thus, simpler methods for enumerating relevant subsets of solutions on the fly are needed.

We present *aspefm* [3], a hybrid computational tool based on ASP with linear constraints that permits the computation of EFMs considering additional logical and linear constraints. We apply our methodology to the *Escherichia coli* core model, which contains 226 × 106 EFMs [4]. In considering transcriptional and environmental regulation, thermodynamic constraints, and resource usage considerations, the solution space is reduced to 1118 EFMs that can be computed directly with *aspefm*. The solution set, for *E. coli* growth on O2 gradients spanning fully aerobic to anaerobic, can be further reduced to four optimal EFMs using post-processing and Pareto front analysis.

In addition, we show that our tool *aspefm* can be simply adapted to compute MCSs using the dual network formalism presented by Ballerstein et al [5].

Finally, we present several new methods using ASP aiming to retrieve sets of genes that enable or disable a set of reactions from the gene product reaction associations retrieved in recent GSMs. Such methods can be used to compute the minimal cut sets of genes from minimal cut sets of reactions, hence obtaining the sets of genes to be knocked out, which are of great biological interest.

References

[1] M. Gebser, T. Schaub, S. Thiele, B. Usadel, and P. Veber, "Detecting inconsistencies in large biological networks with answer set programming," *International Conference on Logic Programming*, pp. 130–144, 2008.

[2] C. Frioux, T. Schaub, S. Schellhorn, A. Siegel, and P. Wanko, "Hybrid Metabolic Network Completion," *CoRR*, vol. abs/1808.04149, 2018.

[3] M. Mahout, R. P. Carlson, and S. Peres, "Answer Set Programming for Computing Constraints-Based Elementary Flux Modes: Application to Escherichia coli Core Metabolism," *Processes*, vol. 8, no. 12, p. 1649, Dec. 2020.

[4] C. Jungreuthmayer, D. E. Ruckerbauer, and J. Zanghellini, "regEfmtool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic," *Biosystems*, vol. 113, no. 1, pp. 37–39, 2013.

[5] K. Ballerstein, A. von Kamp, S. Klamt, and U.-U. Haus, "Minimal cut sets in a metabolic network are elementary modes in a dual network," *Bioinformatics*, vol. 28, pp. 381–387, 2012.

Title: *Pseudomonas aeruginosa* Reverse Diauxie is a Multidimensional, Optimized, Resource Utilization Strategy

Author: S. Lee McGill[1], Yeni Yung[2], Kristopher A. Hunt[1], Michael A. Henson[3], Luke Hanley[2], Ross P. Carlson[1]

Primary Affiliation: [1]Department of Chemical and Biological Engineering, Center for Biofilm Engineering, Montana State University, Bozeman MT; [2]Department of Chemistry, University of Illinois at Chicago, Chicago, IL, U.S.A.; [3]Department of Chemical Engineering and the Institute for Applied Life Sciences, University of Massachusetts, Amherst, MA, U.S.A.

Abstract

Pseudomonas aeruginosa is a globally-distributed bacterium often found in medical infections. The opportunistic pathogen uses a different, carbon catabolite repression (CCR) strategy than many, model microorganisms. It does not utilize a classic diauxie phenotype, nor does it follow common systems biology assumptions including preferential consumption of glucose with an 'overflow' metabolism. Despite these contradictions, *P. aeruginosa* is competitive in many, disparate environments underscoring knowledge gaps in microbial ecology and systems biology. Physiological, omics, and *in silico* analyses were used to quantify the *P. aeruginosa* CCR strategy known as 'reverse diauxie'. An ecological basis of reverse diauxie was identified using a genome-scale, metabolic model interrogated with *in vitro* omics data. Reverse diauxie preference for lower energy, nonfermentable carbon sources, such as acetate or succinate over glucose, was predicted using a multidimensional strategy which minimized resource investment into central metabolism while completely oxidizing substrates. Application of a common, *in silico* optimization criterion, which maximizes growth rate, did not predict the reverse diauxie phenotypes. This study quantifies *P. aeruginosa* metabolic strategies foundational to its wide distribution and virulence including its potentially, mutualistic interactions with microorganisms found commonly in the environment and in medical infections.

Title: Functional Analysis of Metabolism Quantifies Protein Costs for the Synthesis of Biomass Components in *E. coli*

Author: Matteo Mori[1], Chuankai Cheng[2] and Terence Hwa[3]

Primary Affiliation: [1] [3]Department of Physics, University of California San Diego , La Jolla, California, USA; [2]Department of Biological Sciences, University of Southern California, Los Angeles, California, USA

Abstract

The cost of protein synthesis is known to be an important determinant of metabolic strategies. However, currently there is no method that can systematically assign the abundance of each enzyme to distinct metabolic functions. Here we describe a novel computational framework, termed Functional Analysis of Metabolism (FAM), that allows disentangling fluxes and proteome components associated with the production of energy and of individual biosynthetic precursors. After integrating genome-scale models of metabolism for *Escherichia coli* with experimental data on fluxes and cellular biomass composition, we can assign experimental protein abundances to function-specific protein shares, thereby quantifying the protein burden associated with each metabolic activity. This approach provides the foundation for a function-based coarse graining of the cellular proteome, leading to a systematic way to formulate predictive models of protein allocation.

The fundamentally interconnected nature of metabolic networks makes it very difficult if not impossible to define functional roles and properties for individual proteins. For instance, a metabolic reaction might be used for both energy and biomass biosynthesis, but quantifying its contributions to these two biological functions is nontrivial. This hinders the assignment of reactions and enzymes to specific biological functions, which is commonly used in the analysis of –omics data and evaluation of gene expression patterns, e.g., via Gene Ontology (GO)-terms [1]. This problem also affects the determination of costs and yields for specific pathways, which are performed on simplified networks to cut through the complexity [2].

This challenge is resolved by our framework. Based on the canonical Flux Balance Analysis, FAM relies on general linear properties of the optimal solutions to quantify the contribution of each reaction to different metabolic functions (Fig. 1A). We applied this approach to carbon-limited and translationally inhibited *Escherichia coli* cells, by integrating genome-scale models of metabolism with experimental data on exchange fluxes and cellular biomass composition. Our approach allowed us to quantitatively determine the allocation of metabolic enzymes towards the production of energy and every metabolic precursor. Notably, this is done without introducing any *ad hoc* parameters, a common limitation in other data integration approaches [3]. In this way, we determined energetic and carbon costs/yields for the biosynthesis of every biomass component, providing an accurate picture of the energy budget of the cell across conditions. FAM allowed us to assign experimental protein abundances according to metabolic function, thus quantifying the protein burden associated with each metabolic activity across conditions. Together with a GO-term based classification for the non-metabolic component of the proteome, we show how a global function-based partitioning of the proteome can be defined systematically (Fig. 1B). We used this global classification to define a function-based coarse graining of the proteome, reconciling two different protein allocation models based on orthogonal approaches [4,5] into a single consistent model.

Owing to the generality of the approach, the minimal number of assumptions, and the lack of ad hoc parameters, FAM is a powerful tool for the quantitative analysis of metabolism and proteome allocation in diverse systems.

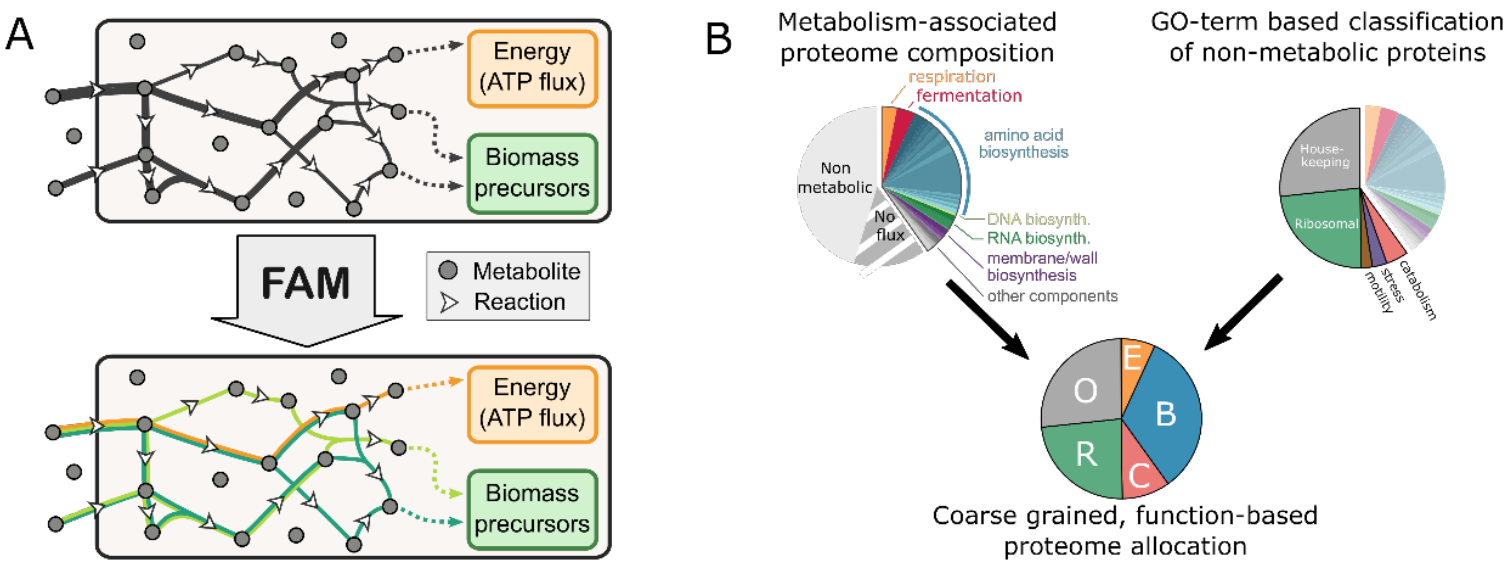


Figure 1. (A) Functional Analysis of Metabolism (FAM) allows to decompose metabolic fluxes into distinct flux components.

(B) Functional decomposition of the proteome for E. coli in glucose minimal medium.

References

[1] The Gene Ontology Consortium, "The Gene Ontology Resource: 20 years and still GOing strong", *Nuc Acid Res*, 2018

[2] C. Kaleta et al., "Metabolic cost of amino acid and protein production in Escherichia coli", *Biotech J*, 2013

[3] Y. Chen et al., "Mathematical modelling of proteome constraints within metabolism", *Curr Op Sys Bio*, 2021

[4] S. Hui et al., "Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria", *Mol Syst Biol*, 2015

[5] M. Basan et al., "Overflow metabolism in Escherichia coli results from efficient proteome allocation", *Nature*, 2015

Title: Stoichiometric Modeling of String Chemistries

Author: Devlin C. Moyer[1], Alan R. Pacheco[1], David B. Bernstein[3], and Daniel Segrè[1]

Primary Affiliation: [1] Bioinformatics Program, Boston University, Boston, MA; [3] Biological Design Center, Boston University, Boston, MA

Abstract

Uncovering the general principles that govern the structure of metabolic networks is key to understanding the emergence and evolution of living systems. Artificial chemistries can help illuminate this problem by enabling the exploration of chemical reaction universes that are constrained by general mathematical rules that can be as complex or as simple as desired. Here, we focus on string chemistries: artificial chemistries in which strings of characters represent simplified molecules, and string concatenation and splitting represent possible chemical reactions. Using a novel Python package, ARtificial CHemistry NEtwork Toolbox (ARCHNET), we have explored the topological characteristics of different string chemistry networks. We have also developed a network-pruning algorithm that can generate minimal metabolic networks capable of producing a specified set of biomass precursors from a given assortment of environmental nutrients using flux balance analysis. We found that the compositions of these minimal metabolic networks were influenced more strongly by the metabolites in the biomass reaction than the identities of the environmental nutrients. This finding has important implications for the reconstruction of organismal metabolic networks, and could help us better understand the rise and evolution of biochemical organization. More generally, our work provides a bridge between artificial chemistries and stoichiometric modeling, which can help address a broad range of open questions, from the spontaneous emergence of an organized metabolism to the structure of microbial communities.

Title: PICKaxe, JN1224min, & the MINE: Unearthing Enzyme Promiscuity for New Applications

Author: Joseph Ni*, Kevin Shebek*, Jon Strutz*, Linda Broadbelt and Keith Tyo

Primary Affiliation: Dept. of Chemical & Biological Engineering, Northwestern U., USA

Abstract

We are just now beginning to understand the prominence of enzyme substrate promiscuity in metabolism. It is estimated that 30% of *E. coli* enzymes carry out promiscuous reactions under physiological conditions. This promiscuity is of critical importance as we (a) seek to model metabolic networks, (b) introduce enzymes from other organisms to a non-native host, and (c) push the boundaries of chemical products that can be produced by biomanufacturing. To better understand the consequences of enzyme promiscuity, a number of biocheminformatics programs have been developed that predict enzyme substrate promiscuity. Here we will describe the biocheminformatics program Pickaxe, the chemical reaction rule set JN1224min, and the MINE 2.0. We believe these tools will be useful to a broad swath of the community and aid in incorporating enzyme promiscuity effects into many studies.

Pickaxe is an open-source, python-based enzyme promiscuity prediction software built on the well-supported cheminformatics package RD-kit. Pickaxe combines operational simplicity with powerful computational algorithms and flexible configurations. Like other biocheminformatics packages, this software relies on operators or reaction rules that specify chemical substructures required for a promiscuous reaction. We have advanced these capabilities by developing modules for calculating the $\Delta G^{\circ'}_f$ for arbitrary compounds on the fly using Equilibrator and probabilistically filtering searches based on molecular weight, formula, or chemical similarity to user-specified targets. Through software engineering, Pickaxe can routinely predict and store >10^6 molecules/reactions in a Mongo database. We are not aware of software that can create larger compound libraries. To facilitate use of Pickaxe by the community, we have prepared a conda package for simple installation and a well-documented github site for collaboration. To compliment Pickaxe, we have developed a well curated set of chemical reaction rules to use with Pickaxe. Rule set JN1224min is a set of 1,224 reaction rules that are designed to predict the greatest number of known enzymatic reactions in the least number of chemical rules. These rules were curated from MetaCyc and only specify the atoms involved in the reaction. Finally, leveraging both Pickaxe and JN1224min, we have released an update to the Metabolic In-silico Network Expansions (MINE). MINEs are databases of products of promiscuous reactions with a website frontend. The MINE 2.0 is an update on the 2015 effort and contains 10x more promiscuous products, resulting in better coverage of Mass bank mass spectrometry data. To improve its utility in untargeted metabolomics, we have upgraded the LC-MS fragmentation predictions (for MS2) from CFM-ID 2.0 to CFM-ID 4.0. Together, Pickaxe, JN1224min, and MINE2.0 are powerful, yet simple tools for predicting enzyme promiscuity for use by the community.

* equal contribution

Title: Is the Cytosolic Density of a Prokaryotic Cell Optimized for Metabolic Efficiency?

Author: Tin Yau Pang and Martin Lercher

Primary Affiliation: Heinrich Heine University Düsseldorf, Germany

Abstract

The fluxes of the physicochemical reactions within a prokaryotic cell determine the rate at which the cell converts nutrients into biomass, a major determinant of its growth rate. These reactions are catalyzed by enzymes, the activities of which are sensitive to other molecules of the volume excluding co-solutes in the cytosol. The cytosolic dry mass, comprising these enzymes, their substrates and products, and other macromolecules, occupies a substantial fraction of the cytosolic volume. While increased dry mass concentration in crowded cells may increase cellular efficiency by increasing biochemical fluxes, crowding may also affect cellular efficiency by slowing down diffusion and by changing Gibbs free energies.

Previous modelling studies addressing this trade-off hypothesized that optimal fitness is achieved by maximizing the reaction fluxes. However, these modelling frameworks cannot account for the differential effects on reactions with catalysts of a wide range of sizes; the latter is necessary when comparing cells in distinct physiological states with different molecule size distributions, e.g., fast growth in rich media, where the cytosol is dominated by large ribosomal molecules and tRNAs, and slow growth in minimal media, where the cytosol has a much larger share of small globular proteins.

Here we revisit this question and study how cytosolic density constrains the growth rate of a cell. Our mathematical model accounts systematically for (1) the volume exclusion effect that changes reaction Gibbs free energies and (2) the slowdown of diffusion caused by molecular crowding. Our model predicts a downshift of the optimal cytosolic density when the cell switches from slow to fast growth. Our model further predicts that the decrease in growth rate is very moderate even when the deviation from the optimal cytosolic density is large. The predicted optimal density, as well as the shift in density associated with switching between poor and rich media is consistent with empirical observations.

Title: Dynamic Flux Balance Analysis Reveals the Distribution of Biochemical Subtypes in Cam Photosynthesis

Author: Antonio Rigueiro-Mesejo, Lea Kloss, and Martin J. Lercher

Primary Affiliation: Heinrich Heine University Düsseldorf, Germany

Abstract

Crassulacean acid metabolism (CAM) is one of the three major pathways in photosynthesis, present in 6% of higher plants. This pathway has evolved convergently in many different plant families distributed along different continents. Its time-dependent metabolism presents advantages in water and nitrogen use efficiency that make it very advantageous in scenarios where water or nutrients are very scarce, such as in deserts. While considered for a long time a botanical curiosity useful only for marginal lands, experiments in the 1980s revealed that CAM crops have a large potential for high productivity when cultivated on prime soil [1]; their combination of high productivity and water use efficiency have spurred considerable interest in understanding this pathway and engineering it in species of economic interest.

CAM photosynthesis follows the unique strategy of inverting the diel stomatal opening cycle of other plants, opening the leaf stomata during the night instead of during the day. As this change decouples initial carbon fixation from light-dependent photosynthesis, it requires a large reorganization of central carbon metabolism, which generally occurs without the acquisition of new genes. A large organic acid pool serves as transitory storage for the captured carbon, predominantly in the form of malic acid, which is stored in enlarged cell vacuoles. The corresponding CAM cycle consumes a substantial part of cellular resources, and its optimization in terms of energy and nitrogen is a crucial part of CAM evolution. Similar to C4 photosynthesis, there are three possible main variants of the cycle, depending on the enzyme used to decarboxylate the stored acid: NAD-ME CAM, NADP-ME CAM, and PEPCK CAM. NADP-ME and PEPCK CAM appear to be the main pathways observed in vivo. It is currently unclear whether this biased distribution represents an evolutionary relic or reflects metabolic advantages.

Here, we offer a possible solution to this question based on modeling results. A genome-scale constraint-based model from Zea mays was modified to invert stomatal behavior coupled with malate accumulation, generating the observed diel profiles of CAM metabolism. Dynamic FBA simulations with this network reproduce the experimental observations, with PEPCK CAM and NADP-ME CAM emerging as alternative main pathways, while NAD-ME is unfavored. We validate the computational analysis with experimental data [2]. We analyze the role of factors such as light, nitrogen and assimilation rates. Our results show that PEPCK is favored at higher fluxes through the CAM cycle, hinting at a tradeoff between energy and nitrogen and suggesting that the decarboxylase subsystem that generates the CAM subtypes is not a remnant of its evolutionary history, but instead plays an important role in the energy-nitrogen-water tradeoff of the plant.

References

[1] Nobel, P. S. (1991). Achievable productivities of certain CAM plants: basis for high values compared with C3 and C4 plants. *New Phytologist*, 119(2), 183-205

[2] Reddy, A.R, Das, V.R. (1978). The decarboxylating systems in fourteen taxa exhibiting CAM pathway. *Zeitschrift für Pflanzenphysiologie,* 86(2), 141-146

Title: Modelling of Glycogen Metabolism and Glycogen-Related Disorders

Author: Yvan Rousset and Oliver Ebenhöh

Primary Affiliation: Quantitative and Theoretical Biology, Heinrich Heine University, Germany

Abstract

Metabolic diseases that affect both energy storage and consumption are a burden on the European population and health care system. Among them, glycogen storage diseases (GSDs) are rare metabolic disorders caused by enzyme defects that disturb both synthesis and degradation of glycogen as well as export of glucose to the blood.

A major challenge in the investigation of glycogen metabolism is that the structural properties of the molecule directly affect the dynamics of the biochemical pathways involved in its synthesis and degradation. Indeed, different structures are achieved during glycogen synthesis depending on enzymes' activities or type. Likewise, during glycogen breakdown the availability of glucose chains and their branching pattern affect the release of glucose.

We developed a model that reproduces the synthesis and the degradation of glycogen using stochastic simulations of the involved enzymatic reactions, coupled with a 3D structural model for glycogen.

We present the model and demonstrate its capability to reproduce experimentally observed properties of glycogen *in vivo*, despite the small number of enzymes involved in our model. Then we show that different structures can be obtained depending on enzyme concentrations and properties. Increasing branching activity leads to highly packed structures that are easily degraded, while increasing elongation of glucose chains leads to extended, hard to degrade structures.

Different scenarios will be discussed in the context of glycogen diseases.

We extend later the framework to other reactions using intermediary glucose metabolites such as glucose-6-phosphate, glucose-1-phosphate or UDP-glucose. This allows increasing the spectrum of diseases which can be studied and interpreted with our model, focusing on conversion of glycogen to blood glucose.

Title: INTRODUCING New APPROACHES to GapFilling and DYNAMIC fLUX BALANCE ANALYSIS for Genome-Scale Models

Author: Wheaton L. Schroeder and Rajib Saha

Primary Affiliation: Department of Chemical and Biomolecular Engineering University of Nebraska - Lincoln, United States of America

Abstract

Genome-scale Models (GSMs) of metabolism have become important tools for the *silico* study and design of metabolism *in silico*. The model reconstruction process typically involves collecting information from databases such as NCBI, UniProt, KEGG, ModelSeed, and KBase; however, incomplete systems knowledge leaves gaps in any genome-scale reconstruction. Current tools for addressing gaps, use databases of biochemical functionalities to address gaps on a per-metabolite basis and can provide multiple solutions. However, their major limitation is that they cannot avoid Thermodynamically Infeasible Cycles (TICs), invariably requiring lengthy manual curation. This is in part due to their per-metabolite approach. To address these limitations, an optimization-based multi-step method named OptFill is developed, which performs TIC-avoiding, whole-model (holistic) gapfilling. OptFill, as with other methods, uses a database of biochemical functionalities to address metabolic gaps, in contrast, it uses a three-step approach to maximize metabolites connected, minimize the number of reactions, and maximize the number of reversible reactions in each solution. Additionally, OptFill can be readily adapted to automate inherent TICs identification, aiding manual curation. OptFill was first applied to three fictional prokaryotic "toy" models of increasing sizes and to a published GSM of *Escherichia coli*, *i*JR904. This application resulted in holistic and infeasible cycle free gapfilling solutions. Following this, OptFill was applied to the reconstruction of a genome-scale model of *Exophiala dermatitidis*, *i*Ede2091, which was used to study the cost of polyextremotolerance and the similarities of human and *E. dermatitidis* melanin synthesis. Overall, OptFill can address critical issues in automated development of high-quality GSMs.

As GSMs are under-defined systems of equations, optimization-based tools are required for their analysis, most frequently Flux Balance Analysis (FBA). An adaptation of FBA, dynamic FBA (dFBA), is a tool which allows for the study of a modeled system across time. Introduced here is a generalized **O**ptimization- and explicit **R**unge-**K**utta-based **A**pproach (**ORKA**) to perform dFBA, which is more accurate and computationally tractable than existing approaches, namely the Static and Dynamic Optimization Approachs (SOA and DOA, respectively). A four-tissue (leaf, root, seed, and stem) model of *Arabidopsis thaliana*, p-ath773, is analyzed using ORKA. P-ath773 uniquely captures the core-metabolism of several stages of growth from seedling to senescence while strongly emphasizing plant-scale behavioral agreement between *in silico* results and *in vivo* data. Using ORKA, p-ath773 takes metabolic "snapshots" at hourly intervals throughout the lifecycle of an individual plant. This analysis shows the transition of metabolism and whole-plant growth, such as the evolution of sulfur metabolism and the diurnal flow of water throughout the plant. Specifically, p-ath773 shows how transpiration drives water flow through the plant and how water produced by leaf tissue metabolism may contribute significantly to transpired water. Investigation of sulfur metabolism shows frequent cross-compartment exchange of a standing pool of amino acids which is used to regulate proton flow. Additionally, p-ath773 has shown broad agreement with published plant-scale properties such as mass, maintenance, and senescence. Overall, p-ath773 serves as a scaffold for lifecycle models of other plants to further increase the range of hypotheses which can be investigated *in silico*.

Title: Systematizing the Different Notions of Growth-Coupled Strain Design and a Single Framework for Their Computation

Author: Philipp Schneider[1], Radhakrishnan Mahadevan[2], Steffen Klamt[1]

Primary Affiliation: [1]Max Planck Institute for Dynamics of Complex Technical Systems, Germany; [2]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Canada

Abstract

A commonly used design principle for the metabolic engineering of microorganisms, aims to introduce interventions that enforce growth-coupled product synthesis (GCP) such that the product of interest becomes a (mandatory) by-product of growth. Although applied in numerous strain design approaches, different variants and partially contradicting notions of this principle exist. We therefore propose a standardized ontology to categorize the existing notions. This ontology distinguishes between four classes of GCP with gradually increasing coupling degrees:

1. potentially growth-coupled production (pGCP), demanding production potential at maximal growth,
2. weakly growth-coupled production (wGCP), demanding product synthesis at maximal growth,
3. directionally growth-coupled production (dGCP), demanding product synthesis whenever the cell grows,
4. substrate-uptake coupled production (SUCP), demanding product synthesis whenever substrate is taken up by the cell.

This unified nomenclature not only classifies previous strain optimization approaches, but also supports the formalization of strain design demands for GCP. We then show that the framework of Minimal Cut Sets (MCS) can be employed to compute strain designs for all four coupling degrees. While the MCS computation of dGCP and SUCP strain designs is straightforward and has been routinely used before, we extended the MCS framework to embed optimality constraints in the design specifications, a feature that is required to compute pGCP and wGCP strain designs. This extension closes the gap between MCS-based and bilevel-based strain design approaches and allows for the computation of all classes of GCP within a single framework. This also simplifies a direct comparison of the different strategies (e.g., regarding computation time etc.).

We used this framework to compute exemplary strain designs for the growth-coupled production of a range of relevant products and illustrate the strict hierarchical relationships of strain design solutions with increasing coupling degree. Furthermore, while stronger coupling degrees usually require more interventions, the time needed to compute them may be lower. In particular, calculating strain designs for SUCP was often faster than for wGCP and dGCP.

Finally, we show that the principle of coupling (at different degrees) can be generalized to couple product synthesis with other biological functions than growth, e.g., with net ATP formation (ATP-coupled production; ACP). We discuss the relationships between GCP and ACP strain designs and demonstrate how strain designs for ACP can be computed via MCS in an analogous manner.

Title: Metabolism of *Campylobacter jejuni:* an Integrated in silico and in vitro Study

Author: Dipali Singh[1], Noemi Tejera[1], Emily Stoakes[2], Andrew Grant[2] and John Wain[1]

Primary Affiliation: [1]Microbes in the Food Chain, Quadram Institute Biosciences, Norwich Research Park, UK; [2]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

Abstract

Campylobacter, Gram-negative, microaerophilic, spiral-shaped bacteria, are the leading cause of acute bacterial gastroenteritis worldwide [1]. The most common species associated with campylobacteriosis is *Campylobacter jejuni,* with *Campylobacter coli* reported in 10–25% of cases. Symptoms of campylobacteriosis such as diarrhoea, fever, nausea often last between 2 to 5 days, however, the sequelae include colitis, reactive arthritis, and Miller-Fisher and Guillain-Barré syndromes [2]. The disease is economically important, with an estimated cost to the UK economy at around £1 billion per annum. Thus, making them a serious and growing public health threat and a priority to identify novel control strategies.

C. jejuni and *C.coli* are fastidious organism when grown in the laboratory. They require microaerophilic environment and temperature above 30o C for growth. They are susceptible to oxidative stress and yet can widely survive in environment and can be readily isolated from food, water, poultry, and milk which on ingestion can lead to human infection. In this study, we have combined an *in silico*, metabolic modelling technique, and *in vitro*, growth experiments and BIOLOG assays, to establish the link between genotype and phenotype and explore the metabolic requirements for growth and survival of *C. jejuni* M1 [3].

We have constructed a fully curated, genome-scale metabolic model (GSM) for the reference organism *C. jejuni* M1 (our variant is M1cam [4]) and validated it theoretically and through laboratory experiments. The model was analysed, to identify cellular functions or metabolic routes that are crucial for growth and/or survival, using Flux Balance Analysis. We identified potential auxotrophies (L-methionine, niacinamide, and pantothenate) and substrates that may improve growth. Results were confirmed experimentally. These results resulted to design of simple defined media for our strain of interest [5]. Substrate auxotrophy retained by *C. jejuni* M1cam is different to that of other well studied *C. jejuni* strains such as NCTC11168, and 81-176 thus, suggesting the strain specific growth requirement and metabolic diversity in *C.jejuni.*

Further, we have used GSM and BIOLOG assays to investigate the flexibility in energy metabolism of *C. jejuni* M1cam, to use wide varieties of substrates and gases as electron donors/acceptors. Here, we report that *C. jejuni* M1cam can to produce energy solely using hydrogen as an electron donor and oxygen as an electron acceptor even in the absence of any organic carbon sources. In addition, we have identified a set of organic substrates from BIOLOG plates that can be utilised by this strain. Oxygen requirement for the catabolism of these substrates has a wide range (0.5-5.5 mol of oxygen per mol of substrate) suggesting that the ability of this organism to use various substrate is not only useful for energy production but also to adapt at different oxygen level and cope with oxidative stress and thus, survive in wide range of environments.

References

[1] A.J. Lastovica et al., *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*, E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt and F. Thompson, Springer Berlin Heidelberg: 2014, 307-335

[2] M.J. Blaser et.al., *Campylobacter, 3rd edn*, I. Nachamkin, C. Szymanski, and M. Blaser, John Wiley & Sons, Ltd: 2008, 99–121

[3] C. Friis et al. Genomic characterization of *Campylobacter jejuni* strain M1. *PloS one*. 2010; 5: e12253.

[4] S.P.W. de Vries, et al. Motility defects in *Campylobacter jejuni* defined gene deletion mutants caused by second-site mutations. *Microbiology*. 2015; 161: 2316-27.

[5] N. Tejera et al., Genome-Scale Metabolic Model Driven Design of a Defined Medium for *Campylobacter jejuni* M1cam. *Front. Microbiol*. 2020; 11:1072

Title: An Improved Kinetic Model to Explain the Physiological Response of *Escherichia coli* Under High ATP Demand

Author: Giulia Slaviero, Simon Boecker, and Steffen Klamt

Primary Affiliation: Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

Abstract

Escherichia coli is one of the most studied model organisms and pioneer of genetic modification and biotechnological applications. Despite the extensive knowledge of its metabolism, availability of multiple omics data and genome-scale models, a comprehensive understanding of how *E. coli* responds and adapts to perturbations is still lacking in many cases. Kinetic models can be of great help to understand cell metabolism and to predict metabolic phenotypes. However, construction of larger and robust kinetic models is complicated and still only a limited number of kinetic models of the central metabolism of *E. coli* can be found in the literature, each of them having some advantages but lacking also some key features (e.g., considering aerobic and anaerobic conditions, correct stoichiometry of biomass synthesis, SBML compatibility). Therefore, taking the models presented in [1,2] as a starting point, we developed a comprehensive kinetic model that (1) is capable to simulate *E. coli* metabolism under aerobic as well as under fermentative conditions, (2) uses a growth rate kinetics that correctly considers stoichiometric requirements of the precursors, (3) includes all energy cofactors such as ATP, ADP, NADP(H) as explicit metabolite, (4) includes major metabolic and genetic regulations, (5) describes proton translocation and ATP synthesis in the electron transport chain and (6) maintains a model structure that is easy to use and can be stored in SBML format.

The developed model was employed to understand what happens in the cell when its energy levels are perturbed. It has been demonstrated that enforcing a high turnover of ATP (by overexpressing the ATP-hydrolyzing F_1-portion of the ATPase [3]) considerably increases the specific glucose uptake rate and the production of ferementation products [3,4]. On the other hand, we found that the increasing trend is not monotone: the glycolytic flux reaches a maximum for a medium ATPase level and drops markedly when this level is exceeded [5]. With the support of the model, we could propose mechanisms responsible for the observed fluxes and metabolite concentration profiles. Particularly, the phosphofructokinase was identified as responsible for the biphasic flux behavior, due to its dual dependency on ATP as substrate and ADP as activator. To elucidate what limits the metabolism by high ATPase level, we focused on anaerobic conditions. The unexpected measured accumulation of glycolytic metabolites such as pyruvate, phosphoenolpyruvate and fructose-1,6-bisphosphate and the considerable production of lactate, led to new hypotheses on the regulation of key metabolic steps such as pyruvate formate-lyase (PFL) and pyruvate kinase (PYK). To verify our hypotheses and model predictions, we overexpressed the genes of the enzymes catalyzing the supposed rate-limiting reactions (*pfkA* and *pflB*) in the strain with high ATPase levels and a significant increase in glucose uptake could indeed be confirmed. Our work provides new insights on the functioning of glycolysis under extreme (ATP-demanding) conditions and is of great value for rational engineering of cell factories with high productivity.

References

[1] P. Millard et al., "Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*.", *PLOS Computational Biology*, 2017; 13(2): e1005396.

[2] Y. Matsuoka et al., "Modeling and simulation of the redox regulation of the metabolism in Escherichia coli at different oxygen concentrations", *Biotechnol Biofuels*, 2017; 10:183.

[3] B. J. Koebmann et al., "The Glycolytic Flux in *Escherichia coli* Is Controlled by the Demand for ATP.", *J. Bacteriol.*, 2002; 184, 3909-3916.

[4] S. Boecker et al., "Broadening the Scope of Enforced ATP Wasting as a Tool for Metabolic Engineering in *Escherichia coli*." *Biotechnol. J.*, 2019; 14: 1800438.

[5] S. Boecker et al., "An Integrated Approach to Decipher the Physiological Response of Escherichia coli Under High ATP Demand". Manuscript submitted for publication.

Title: Integration of Metabolic, Ecological, and Kinetic Models to Predict the Dynamics in Context-Dependent Microbial Interactions

Author: Hyun-Seob Song[1], Aimee K. Kessell[1], Na-Rae Lee[2], Zhou Kang[3], Dong-Yup Lee[4]

Primary Affiliation: [1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE; [2]Department of Bioscience and Biotechnology, Konkuk University, Seoul, Republic of Korea; [3]Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore, Singapore; [4]School of Chemical Engineering, Sungkyunkwan University, Suwon-si, Gyeonggi-do, Republic of Korea

Abstract

Microbial communities are complex adaptive systems that dynamically rewire interspecies interaction networks in response to environmental variations. Dynamic reorganization in interaction networks is a key mechanism underlying robust metaphenomes in microbial communities subject to various biotic and abiotic perturbations. Despite criticality in predicting complex dynamics and emergent properties in microbial communities, we currently have poor understanding of context dependency in microbial interactions and consequently poor representation in microbial community models [1,2]. To address this issue, we present a novel modeling framework that enables accurately estimating dynamic shifts in microbial interactions in varying environments. For this purpose, we combined two complementary modeling approaches: kinetic and a generalized-Lotka-Volterra (gLV)-type of ecological models. Kinetics models can simulate environment-dependent dynamics in microbial growth and community behaviors but are unable to determine interspecies interactions. By contrast, gLV models allow us to quantify interspecies interactions (e.g., through model-data fitting), but its utility is limited to the case where interaction coefficients are constant (therefore not suitable for modeling context-dependent interactions). By synergistic integration of these two models, we derived an analytical form of interspecies interaction coefficients as a function of environmental variables and kinetic parameters, which is termed here Kinetics-based Inference of Dynamic variation in microbial Interactions (KIDI). We tested the effectiveness of KIDI through the case study of a co-culture of two *Escherichia coli* mutants that are auxotrophic for amino acids. These two organisms have mixed relationships, i.e., cooperate for cross-feeding amino acids (i.e., tryptophan and tyrosine), but compete for a shared nutrient (i.e., glucose); depending on environmental conditions, their relationships turn into various different types of interactions: cooperation, competition, commensalism, and amensalism. Through coordinated design of experiments where controlling substrates switch between amino acids and glucose, we demonstrated that KIDI can predict dramatic variations in interspecies interactions as nutrient concentrations in media change over time. For deeper insights on how these two amino acid auxotrophs metabolically interact with each other, we also performed flux balance analysis (FBA) using genome-scale metabolic networks of the two strains. By dynamic adjustment of constraints on exchange fluxes in FBA as informed from KIDI, we were able to simulate how metabolic shifts occur at pathway levels in individual species. The integrative approach such as KIDI with metabolic network modeling is expected to serve as a useful *in silico* tool for engineering microbial interactions through environmental manipulation.

References

[1] A. K. Kessell et al. “Predictive interactome modeling for precision microbiome engineering,” *Current Opinion in Chemical Engineering*, 2020; 30: 77 – 85.

[2] H.-S. Song et al., “Mathematical modeling of microbial community dynamics: a methodological review,” *Processes,* 2014; 2(4): 711 – 752.

Title: Improving Quantitative Genome-scale Metabolic Modeling of CHO

Author: Diana Széliová[1 2], David E. Ruckerbauer[1 2], Jerneja Štor[1 3], Dmytro Iurashev[1 2], Harald Schoeny[2], Michael Hanscho[1], Michael Melcher[1], Stephan Hann[1 4], Gunda Koellensperger[3], Isabel Rocha[5 6], Nicole Borth[1 3] and Jürgen Zanghellini[1 2]

Primary Affiliation: [1]acib - Austrian Centre of Industrial Biotechnology, Vienna, Austria, EU; [2]Department of Analytical Chemistry, University of Vienna, Vienna, Austria, EU; [3]Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, EU; [4]Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria, EU; [5]Centro de Engenharia Biológica, Universidade do Minho, Braga, Portugal, EU; [6]ITQB-NOVA—Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal, EU

Abstract

Chinese Hamster Ovary (CHO) cells are the leading mammalian production hosts for complex protein biopharmaceuticals. Currently, the cell line and bioprocess development still rely mostly on trial-and-error approaches and high-throughput screening, which is time consuming and costly. Constraint-based metabolic modeling can help to overcome these bottlenecks and enable more targeted development approaches. The reconstruction of the genome-scale metabolic model (GSMM) of CHO, iCHO1766 [1] (and recent updates), was an important and necessary step towards applying these methods to complex mammalian production systems. However, key to the success of COBRA methods is not only a reliable reconstruction but also highly accurate experimental data that feed the model. Here we improve the predictive qualities of iCHO by integrating measured (i) cell line- and condition-specific biomass compositions (ii) uptake and secretion spectra and (iii) non-growth associated maintenance requirements (mATP). We measured biomass composition of thirteen cell lines/conditions (dry mass, amino acid and lipid composition, total protein, lipid, RNA, DNA and carbohydrate content). This data improved FBA predictions of growth rates but only if highly accurate exchange rates were used as input [2]. To achieve these, accurate concentration measurements (of essential amino acids with low uptake rates) and dense sampling frequency (every 6 h) are needed [3]. Next, we compared intracellular flux predictions by parsimonious FBA with experimental 13C flux data from six publications. Most fluxes of the central carbon metabolism were grossly underestimated if mATP was not accounted for (R2 = 0.45). We estimated mATP computationally as no experimental data was available for CHO. Constraining the flux of mATP demand reaction to this value significantly improved the flux predictions (R2 = 0.93). Furthermore, we confirmed mATP experimentally for CHO-K1 cell line [4]. In summary, this work improved predictions of CHO GSMM which will increase its usefulness for cell line and bioprocess development.

References

[1] H. Hefzi et al., “A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism”, *Cell Systems*, 2016; 3: 434–443.e8

[2] D. Széliová et al., “What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines”, *Metabolic Engineering*, 2020; 61: 288-300

[3] D. Széliová et al., “Error propagation in constraint-based modeling of Chinese hamster ovary cells”, *Biotechnology Journal*, 2020; 16

[4] D. Széliová et al., “Inclusion of maintenance energy improves the intracellular flux predictions of CHO”, *bioRxiv*, 2020;

Title: Modular Design Principles in Biological Systems: Theory, Computation, and Experimental Validation

Author: Cong T. Trinh

Primary Affiliation: Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, USA

Abstract

Modular design is a governing principle in natural and engineered systems, enabling rapid, efficient, and reproducible construction and maintenance. While the theory and application of modular design are developed in non-biological engineering disciplines, the field is still at a nascent stage to understand and harness modular design across biological scales. In this talk, we present our recent research progress on the development of theory, mathematical formulation and computation, and experimental validation of modular biosystems design. Microorganisms can be engineered to function as microbial cell factories to produce a large space of fuels, chemicals, and materials in a sustainable and renewable manner. The current technology for microbial biocatalyst development, however, remains laborious and costly to implement, precluding its widespread adoption. To overcome this roadblock, we propose to adapt the principles of modular design that drive innovation, efficiency, and predictability across modern engineering disciplines to the fields of synthetic biology and metabolic engineering for microbial biocatalyst development. These microbial biocatalysts can be rapidly and systematically built from a modular (chassis) cell strongly coupled with exchangeable pathway modules that enable programmed functions for overproduction of desirable chemicals with minimal requirement of iterative strain optimization cycles. We formulate a mathematical and computational framework that enables systematic design of modular cells based on the metabolic pathway analysis (e.g., elementary mode analysis, flux balance analysis) and Pareto optimization theory for tenths to hundreds of modules in various organisms. Using *Escherichia coli* as a testbed, we demonstrate it is feasible to design a modular cell(s) capable of synthesizing in a large, biochemically diverse library of molecules at high yields and rates with minimal tradeoff of modularity, efficiency, and robustness. By identifying reaction usage patterns for different modules in the modular cell, we elucidate modular organization of the designed cells, defining the interfaces between a modular cell and its production modules. Our analysis reveals the broad pathway compatibility of the designed modular cell is enabled by the natural modularity and flexible flux capacity of endogenous core metabolism. To validate the modular cell design experimentally, we design and create an E. coli modular cell for switchable efficient biosynthesis of designer alcohols and esters from fermentable sugars that have broad utility as fragrances, flavors, solvents, and fuels. We demonstrate metabolic coupling between the modular cell and production modules can be modulated to enhance target product production and have important implications for enzyme pathway selection and evolution.

Title: Bringing to Bear Retrosynthesis Tools and Enzyme Engineering for Pathway Synthesis

Author: Vikas Upadhyay, Veda Sheersh Boorla, Deepro Banerjee, and Costas D. Maranas

Primary Affiliation: Department of Chemical Engineering, The Pennsylvania State University

Abstract

Metabolic pathway design is a cross-cutting task present in a wide range of endeavors, from biofuel and bio-renewables design to the biodegradation or funneling of complex aromatics, and chemicals of industrial interest. It is also a multi-disciplinary task which requires knowledge and encoding of the relevant biochemistry of pathways, understanding of the thermodynamic feasibility of individual enzymatic steps, biophysical understanding of activity and specificity of the enzymes involved, and subsequent host selection and metabolic engineering. In this talk, we will discuss recent progress from our group on a variety of fronts starting with pathway retrosynthesis allowing for novel steps (i.e., novoStoic), prediction of Gibbs free energy of change for novel and uncharacterized reaction steps (i.e., DGPredictor) for ensuring thermodynamic feasibility of pathways, rank-ordering of enzyme candidates in terms of their potential for adaptability for the targeted novel conversion, and computational enzyme redesign (i.e., IPRO +/-) for tuning substrate specificity. Both algorithmic developments and software implementations will be discussed along with applications and future perspectives.

Title: Linear Energy Converter Inspired Metabolic Analysis Couples Rate and Yield in Metabolic Models

Author: St. Elmo Wilken[1], Victor V. Frazão[2], Nima P. Saadat[3], Oliver Ebenhöh[4 5]

Primary Affiliation: [1 2 3 4]Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, Düsseldorf, Germany; [5]Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, Düsseldorf, Germany

Abstract

Before the widespread adoption of inexpensive genomic sequencing, so-called black box models were developed to understand microbial metabolism and to predict biotechnologically important parameters, like biomass yield. These techniques made use of precise measurements of substrate, product, biomass, and heat exchange with the environment. Importantly, by incorporating thermodynamic constraints into black box models, accurate yields could be predicted using just these measurements [1,2]. Recently, much more detailed genome-scale metabolic models have been developed, seemingly obviating the utility of these black box approaches. However, a well-known drawback of analysis techniques, like flux balance analysis, used to interrogate these detailed metabolic models, is that they do not consider the tradeoff between rate and yield. As such, it is challenging to formulate a robust measure of system efficiency, exacerbating difficulties found in designing or quantifying the optimality of metabolic pathways [3].

Thermodynamic analysis is a primary tool used to investigate efficiency limitations of natural and engineered systems. Indeed, its development was driven partly by the need to optimize steam engines in the 19th century. This suggests that the application of thermodynamic principles can be used to complement metabolic model analyses and to inform efficiency considerations. Indeed, thermodynamic and loopless flux balance analysis were developed specifically to ensure that metabolic models obey the laws of thermodynamics. Unfortunately, these techniques are computationally expensive, preventing their widespread use, and, importantly, they also do not address the rate/yield tradeoff [4]. Since economic competitiveness is seen as a major hurdle for bioprocesses [5], it is important to develop a robust efficiency measure of metabolic pathways when using modern techniques. Given the well-developed nature of black box approaches, and their tight and fruitful integration with thermodynamic principles, we investigate whether it is possible to combine some of these techniques with the analysis of metabolic models at the genome-scale.

Here, we extend the linear energy converter model, and associated non-equilibrium thermodynamic framework of coupled processes [6], to investigate the rate/yield tradeoff of metabolic pathways. Our investigation ranges across three levels of detail: single pathways, whole cell metabolism as captured by genome-scale metabolic models, and microbial community interactions. At the pathway level, the energy coupling highlighted by this framework directly leads to a quantitative description of the rate/yield tradeoff. Extending this pathway analysis framework to both single cell, and community metabolic models leads to similar quantitative rate/yield tradeoff arguments. This suggests that these techniques could be integrated with constraint-based analyses. In all three cases the importance of thermodynamic constraints is highlighted, suggesting a common framework could be used to define a robust measure of pathway efficiency for use in bioengineering and systems optimization.

References

[1] Heijnen, J. J., et. al., “A black box mathematical model to calculate auto- and heterotrophic biomass yields based on Gibbs energy dissipation.”, Biotech. & Bioeng. 1992.

[2] McCarty, P. L., “Thermodynamic electron equivalents model for bacterial yield prediction: modifications and comparative evaluations.”, Biotech. & Bioeng, 2007.

[3] Flamholz, A., et al., “Glycolytic strategy as a tradeoff between energy yield and protein cost.”, PNAS, 2013.

[4] Wilken, S. E., et. al., “The view of microbes as energy converters illustrates the trade-off between growth rate and yield”, bioRxiv, 2021.

[5] Issa, I., et. al., “Bioeconomy from experts’ perspectives–Results of a global expert survey.”, PloS one, 2019.

[6] Stucki, J. W., “The optimal efficiency and the economic degrees of coupling of oxidative phosphorylation.”, Eur. J. Biochem., 1980.

Title: Determination of Metformin Transport Parameters Between Plasma and Red Blood Cells of Humans in Different Scale Models

Author: Darta M. Zake[1], Janis Kurlovics[2], Linda Zaharenko[1], Krsitaps Berzins[2], Vitalijs Komasilovs[2], Janis Klovins[1] and Egils Stalidzans[1]

Primary Affiliation: [1]Latvian Biomedical Research and Study Centre, Riga, Latvia; [2]Institute of Microbiology and Biotechnology, University of Latvia, Riga, Latvia

Abstract

1. Introduction

Metformin has been prescribed to treat Type 2 Diabetes mellitus (T2D) since the 1960-s [1]. It is currently the most commonly prescribed drug for T2D, as it is recommended as the first-line medication in nearly all newly diagnosed T2D patients by international treatment guidelines [2], [3]. It is assumed that metformin is used by over 120 million patients worldwide.

Given the lipophobic nature of metformin, it has been assumed that it cannot cross the phospholipid bilayer via diffusion [4]. At the same time, metformin enters the red blood cells (RBC) where relevant transporters are not expressed, suggesting a marginal transport via diffusion [5]–[8]. The maximal metformin concentration in RBC is reached later than in plasma and is considerably lower [7], confirming that the transport to RBC is not intensive.

Considering that there are no active transporters expressed in RBC and plasma membrane permeability should be equal in both directions, we hypothesize that concentration difference-driven passive transport can explain the distribution of metformin between plasma and RBC. We test the hypothesis using large-scale and small-scale models.

2. Methods

Two ordinary differential equation (ODE) based models have been developed for metformin pharmacokinetics. The large scale model is a physiologically based pharmacokinetic whole body model [9] with 21 tissues and body fluids compartments and can simulate metformin concentration in the stomach, small intestine, liver, kidney, heart, skeletal muscle adipose, and brain depending on the body weight, dose, and administration regimen. The second, the small-scale model, consists just of two compartments (Fig.1). The proposed model is based on Fick's Law of diffusion that describes the time course of the transfer of a solute between two compartments that are separated by a membrane [10].

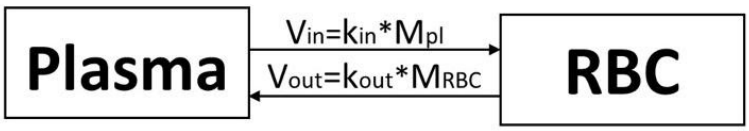


Fig. 1: Two-compartment model of metformin transport between plasma and RBC with absorption (vin) and release (vout) fluxes, where kin – transport rate coefficient to RBC from plasma (h-1); kout – transport rate coefficient from RBC to plasma (h-1), Mpl – metformin concentration in plasma (ng/mL); MRBC -metformin concentration in RBC (ng/mL).

In the small model, the metformin concentration in plasma (Mpl) is approximated by a biexponential decline [11], while in the large model, it is calculated by the ODE system. Two small models for each of the 35 patients have been developed and parametrized during parameter estimation: 1) independent transport coefficients (kin and kout) for uptake and release and 2) single coefficient k used for transport in both directions (uptake and release coefficients are equal).

3. Results and discussion

The large model including a number of tissues was a compromise between measurements in different tissues and the fit of model simulations to measurements of plasma was relatively poor. Consequently, the passive transport coefficients to and from plasma become influenced. The small, plasma/RBC transport centered model enabled a better fit of experimental plasma results with model simulation by biexponential decline and, as a consequence, led to better fitness of RBC measurements to model simulations.

Using the small model, a single passive transport rate coefficient is sufficient for the description of metformin exchange between plasma and RBC and independent transport coefficients do not give a better fit between experimental and model simulation data (Fig 2). That supports our hypothesis of passive transport (diffusion) being the main process ensuring metformin transportation between RBC and plasma.

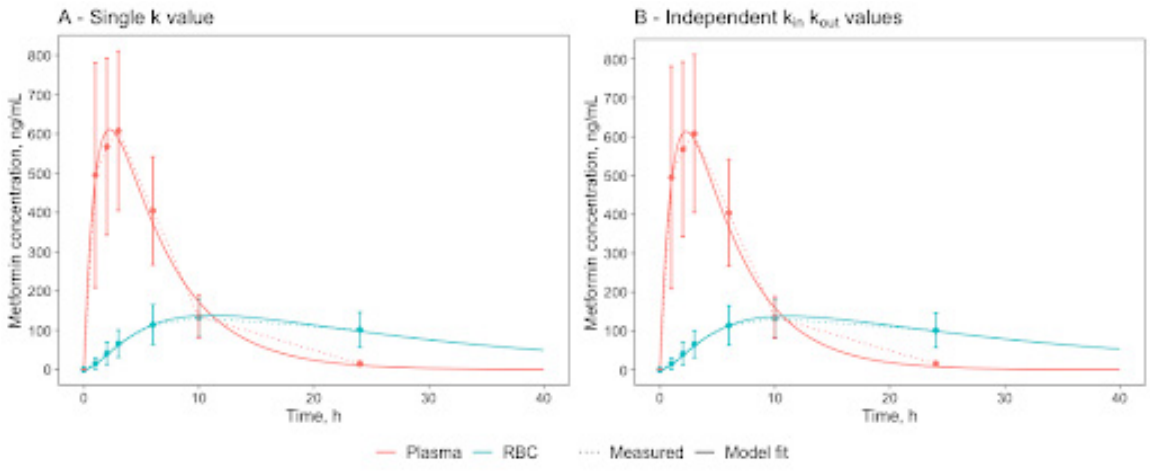


Fig. 2: Parameter estimation results of the average curve. Dots show experimental values while lines represent model simulation. A: single k value (k=0.044 h-1, mean square error = 286) and B: independent kin and kout values (kin=0.044 and kout=0.039 h-1, mean square error = 273).

References

[1] C. Wilcock and C. J. Bailey, "Accumulation of metformin by tissues of the normal and diabetic mouse," *Xenobiotica*, vol. 24, no. 1, pp. 49–57, 1994.

[2] D. M. Nathan *et al*., "Management of hyperglycemia in type 2 diabetes: A consensus algorithm for the initiation and adjustment of therapy. A consensus statement from the American diabetes association and the European association for the study of diabetes," *Diabetes Care*, vol. 29, no. 8, pp. 1963–1972, 2006.

[3] S. M. Marshall, "60 years of metformin use: a glance at the past and a look to the future," *Diabetologia*, vol. 60, no. 9, pp. 1561–1565, Sep. 2017.

[4] X. Liang and K. M. Giacomini, "Transporters Involved in Metformin Pharmacokinetics and Treatment Response," *J. Pharm. Sci.*, vol. 106, no. 9, pp. 2245–2250, 2017.

[5] D. Detaille, B. Guigas, X. Leverve, N. Wiernsperger, and P. Devos, "Obligatory role of membrane events in the regulatory effect of metformin on the respiratory chain function," *Biochem. Pharmacol*., vol. 63, no. 7, pp. 1259–1272, 2002.

[6] G. Tucker, C. Casey, P. Phillips, H. Connor, J. Ward, and H. Woods, "Metformin kinetics in healthy subjects and in patients with diabetes mellitus.," *Br. J. Clin. Pharmacol*., vol. 12, no. 2, pp. 235–246, Aug. 1981.

[7] F. Robert, S. Fendri, L. Hary, C. Lacroix, M. Andréjak, and J. D. Lalau, "Kinetics of plasma and erythrocyte metformin after acute administration in healthy subjects," *Diabetes Metab*., vol. 29, no. 3, pp. 279–283, 2003.

[8] J. D. Lalau and C. Lacroix, "Measurement of metformin concentration in erythrocytes: Clinical implications," *Diabetes, Obes. Metab*., vol. 5, no. 2, pp. 93–98, 2003.

[9] D. M. Zake, J. Kurlovics, L. Zaharenko, V. Komasilovs, J. Klovins, and E. Stalidzans, "Physiologically based metformin pharmacokinetics model of mice and scale-up to humans for the estimation of concentrations in various tissues," *PLoS One*, vol. 16, no. 4, p. e0249594, Apr. 2021.

[10] J. D. Enderle, "Compartmental Modeling," in *Introduction to Biomedical Engineering,* Elsevier Inc., 2011, pp. 359–445.

[11] R. W. Seabury and C. M. Stork, "Pharmacokinetic and Toxicokinetic Modeling," in *Encyclopedia of Toxicology*, Elsevier, 2014, pp. 856–861.

Title: A Cybernetic Trait-Based Model Identifies Microbial Regulation as a Key Driver of Priming Effects

Author: Firnaaz Ahamed[1], James C. Stegen[2], Emily B. Graham[2], Timothy D. Sheibe[2] and Hyun-Seob Song[1]

Primary Affiliation: [1]Department of Biological Systems Engineering, University of Nebraska–Lincoln, Lincoln, NE, United States; [2]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, United States.

Abstract

Natural ecosystems often develop highly nonlinear dynamics. Even a minimal treatment to soils or river sediments, for example, can induce significant changes in the decomposition rate of organic matter (OM). The treatment that often constitutes influx of fresh substrates can accelerate (positive priming) or retard (negative priming) the rate of OM decomposition. Priming effects significantly influence the OM cycles in natural ecosystems, however, we poorly understand the mechanisms and the factors that govern the priming magnitude and the direction. To address these limitations, we propose a trait-based model of two interacting microbial groups that enables predicting various forms of priming effects in diverse contexts. As a key feature, our model accounts for dynamic regulation of microbial growth and interspecies interactions based on the cybernetic approach [1–3]. The cybernetic model views microorganisms as dynamic systems that optimally control metabolic behaviors to fulfill their survival. Motivated by field observations from the hyporheic corridor of a riverine ecosystem [4], we formulated our model by treating the OM as recalcitrant polymeric compounds that must be degraded to labile compounds for microbial assimilation so that priming effects are induced by introducing exogenous labile compounds to the system. Model simulations show that (1) the direction and magnitude of priming effects are highly reliant on the OM richness and the extent of treatments (level of exogenous labile compounds) and (2) interspecies interactions between the microbial groups with distinct metabolic traits also majorly dictate the priming effects. This study not only demonstrates the capability of our model that simulates both positive and negative priming effects using the same model structure, but also identifies key metabolic parameters (including population turnover rate, sensitivity to labile compounds, and efficiency in degrading OM) that potentially drive the priming effects. Therefore, our model offers itself as a unified framework for studying priming effects in diverse biological and environmental contexts.

References

[1] J.D. Young et al., “On the matching and proportional laws of cybernetic models”, *Biotechnol. Prog.*, 2007, 23:83–99.

[2] D. Ramkrishna and H.-S. Song, “Dynamic models of metabolism: Review of the cybernetic approach”, *AIChE J.*, 2012 58:986–997.

[3] D. Ramkrishna and H.-S. Song, “Cybernetic Modeling for Bioreaction Engineering”, Cambridge University Press, 2019.

[4] J.C. Stegen et al., “Influences of organic carbon speciation on hyporheic corridor biogeochemistry and microbial ecology”, *Nat. Commun.*, 2018, 9:1–11.

Title: Methanotroph Acclimation to Cultivation Stresses and the Role of Byproduct Synthesis

Author: Adrienne D. Arnold[1] and Ross P. Carlson[2]

Primary Affiliation: [1]Department of Microbiology and Cell Biology, Montana State University, USA; [2]Center for Biofilm Engineering, Montana State University, USA

Abstract

Methanotrophs are organisms that can use methane as their sole carbon and energy source. They play a crucial role in carbon cycling in the environment and are also of economic interest, as they have the potential to upgrade methane to chemical feedstocks or other value-added byproducts like polyhydroxybutyrate [1]. There are two main classifications of methanotrophs: gammaproteobacterial (type I) and alphaproteobacterial (type II). Type I methanotrophs use the ribulose monophosphate pathway (RuMP) for methane assimilation, allowing for an easy link into glycolysis and thus production of the storage compound glycogen [2]. Type II methanotrophs use the serine cycle for methane assimilation and accumulate polyhydroxybutyrate using the ethymalonyl-CoA pathway.

In this study, we constructed two core metabolic models, one for type I methanotrophs and one for type II methanotrophs. We used elementary flux mode analysis to detail both the accumulation and the utilization of the storage compounds glycogen and polyhydroxybutyrate in the type I and type II models, respectively. Elementary flux modes were analyzed according to the availability of carbon substrate and electron acceptors as well as a proxies for enzyme investment. These analyses allowed us to make predictions of potential excreted byproducts under culture conditions like nitrogen and oxygen limitation, with implications for product synthesis and for microbial interactions in the environment. We also included the pathway for aerobic denitrification in the predictions, which provides additional insight into methanotroph metabolism at oxic/anoxic interfaces [3]. We show that oxygen limitation is linked to increased production of nitrous oxide, a more potent greenhouse gas than methane.

References

[1] Pieja, A. J.; Rostkowski, K. H.; Criddle, C. S. Distribution and Selection of Poly-3-Hydroxybutyrate Production Capacity in Methanotrophic Proteobacteria. Microb. Ecol. 2011, 62 (3), 564–573. https://doi.org/10.1007/s00248-011-9873-0.

[2] But, S. Y.; Dedysh, S. N.; Popov, V. O.; Pimenov, N. V.; Khmelenina, V. N. Construction of a Type-I Metanotroph with Reduced Capacity for Glycogen and Sucrose Accumulation. Appl. Biochem. Microbiol. 2020, 56 (5), 538–543. https://doi.org/10.1134/S0003683820050063.

[3] Stein, L. Y.; Klotz, M. G. Nitrifying and Denitrifying Pathways of Methanotrophic Bacteria. Biochem. Soc. Trans. 2011, 39 (6), 1826–1831. https://doi.org/10.1042/BST20110712.

Title: Prediction Of Phenolic Compound Metabolism In The Human Gut Microbiota Using Machine Learning Methods

Author: Francesco Balzerani[1], Xabier Cendoya[2], Telmo Blasco[3], Iñigo Apaolaza[4], Francisco J. Planes[5]

Primary Affiliation: [12345]Biomedical Engineering Center, University of Navarra, Campus Universitario 31009 Pamplona, Navarra, Spain

Abstract

In recent years, the studies about phenolic compounds in human diets have increased, particularly due to their dietary importance and their role as natural antioxidants and chemopreventive agents. Phenolic compounds are mainly metabolized in the human body by the gut microbiota. Nevertheless, their metabolism is not well represented in public databases and existing reconstructions. AGORA[1], the most accurate and extensive reconstruction of gut microbiota metabolism, includes little information about phenolic compounds. In a parallel work, we integrated different sources of knowledge to address this issue. However, it was not sufficient to represent the diverse metabolic space of this family of molecules. In particular, for 371 compounds in Phenol-Explorer[2], the first comprehensive database on polyphenol content in food, we found reactions related to 107 of them in the metabolic network. In this context, the aim of this work is to use computational methods to predict new degradation pathways involving phenolic compounds, included in the Phenol-Explorer[2] database, which are currently under-represented. To that end, we used RetroPath RL[3], a retrosynthesis algorithm that applies the Monte Carlo Tree Search reinforcement learning method. We identified tentative pathways for 303 metabolites in the Phenol-Explorer2 database to connect them to the metabolic network. Relevant examples and pathways are presented.

References

[1] Magnúsdóttir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat. Biotechnol. 35, 81–89 (2017)

[2] Rothwell, J. A. et al. Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. Database 2013, (2013)

[3] Koch, M., Duigou, T. & Faulon, J. L. Reinforcement learning for bioretrosynthesis. ACS Synth. Biol. 9, 157–168 (2020)

Title: Integration of Regulatory and Signaling Networks with Genome-Scale Metabolic Networks for the Prediction of Drug Targets

Author: Naroa Barrena, Iñigo Apaolaza, Luis V. Valcárcel and Francisco J. Planes

Primary Affiliation: University of Navarra, Tecnun School of Engineering, San Sebastian, Spain

Abstract

Cancer is one of the major causes of death nowadays and, therefore, huge effort is being directed towards the identification of therapeutic targets in this disease. In this context, Constraint-Based Reconstruction and Analysis (COBRA) has emerged as a promising approach providing the scientific community with a plethora of algorithms which aim at predicting novel targets. Apaolaza and colleagues [1] introduced the concept of genetic Minimal Cut Sets (gMCSs) for the identification of novel metabolic drug targets in cancer. gMCSs define minimal genetic interventions which block cellular proliferation and, when integrated with -omics, can be used for elucidated with metabolic vulnerabilities. However, this approach is limited to the prediction of metabolic targets and their application to regulatory and signaling networks is not straightforward due to the complex interactions underlying Boolean networks.

Here, we present a work that integrates regulatory and signaling networks with genome-scale metabolic networks to calculate essential genes for cellular metabolism. This novel algorithm overcomes the issues introduced by networks of Boolean nature and permits the combinatorial computation of the genetic interventions whose deletion render a given biological target impossible. Results are contrasted with those obtained by state-of-the-art iterative approaches, particularly SPIDDOR [2] and BoolNet [3], which depend on the initial conditions on the problem. The advantages of this novel approach and future lines are presented.

References

[1] Apaolaza, I. et al. “An in-silico approach to predict and exploit synthetic lethality in cancer metabolism.” Nat. Commun. 8, 1–9 (2017).

[2] Irurzun-Arana, I. et al. “Advanced Boolean modeling of biological networks applied to systems pharmacology.” Bioinformatics 33, 1040–1048 (2017).

[3] Müssel, C. et al. “BoolNet—an R package for generation, reconstruction and analysis of Boolean networks.” Bioinformatics 26, 1378–1380 (2010).

Title: A Programming Interface to Resource Balance Analysis, Enabling Development of Workflows on Cellular Resource Allocation

Author: Oliver Bodeit[1], Wolfram Liebermeister[2] and Anne Goelzer[2]

Primary Affiliation: [1]Heinrich-Heine-Universitaet Düsseldorf, Germany, [2]INRAE - Unité MaIAGE, France

Abstract

Optimised allocation of resources underlies an organism's fitness in various environments and facilitates success in competition. The advent of techniques to model the allocation of cellular resources, to various molecular processes, facilitated investigations on the underlying economic principles of microbial growth. *Resource Balance Analysis (RBA)*, as a computational framework, enables the analysis of an organism's growth-optimal configurations in various environments, at genome-scale [1]. Building upon available software [2], *rbatools* has been developed as a programming interface to *RBA*, facilitating the flexible implementation of analyses on microbial resource allocation, transcending the representation of growth-optimal cellular configurations. The tool utilises the flexible formulation of *RBA* as constraint-based linear problem, to facilitate the design of custom analysis-workflows with various implemented computational methods to programmatically modify and simulate the *RBA* model (gene knock-outs, variability analysis, screening substrate concentrations, imposing gene-expression level etc.) and allows export of simulation results to different output formats, such as *SBtab, JSON* and visualisation input for metabolic fluxes and protein abundance. Functionality has been exemplified by implementation of various workflows on a genome-scale *RBA*-model of *Escherichia coli* [2], during the development process of the tool. Including the simulation of (double) gene knock-outs and their effect on fitness. The evaluation of the variability of cellular configurations at (sub-) optimal growth-rates, such as the maximal and minimal feasible machinery concentrations or reaction fluxes. The effect of over- and under expression of proteins on fitness and an estimation of the relative impact of different protein cost-factors on fitness.

The *rbatools* software utilises the flexible formulation and mathematical representation of *Resource Balance Analysis* to provide an accessible and user friendly interface to the modelling and investigation of cellular resource allocation in *Python*.

References

[1] : A. Goelzer, V. Fromion, and G. Scorletti, "Cell design in bacteria as a convex optimization problem" Automatica, vol. 47, 2011.

[2] : A. Bulović, S. Fischer, F. Golib, W. Liebermeister, C. Poirier, L. Tournier, E. Klipp, V. Fromion, A. Goelzer. Automated generation of bacterial resource allocation models. Metabolic Engineering, vol. 55, 2019.

Title: Tuatara: an Automated Tool for Constructing and Handling Multiple Genome-Scale Metabolic Models

Author: Cailean Carter, Gemma Langridge, and Dipali Singh

Primary Affiliation: Quadram Institute, United Kingdom

Abstract

1. Introduction

We are developing a tool for reducing the level of curation required for newly built genome-scale metabolic models (GSMs) of bacteria by using pre-existing high-quality GSMs as a template. We are developing this into a tool called tuatara, which is an add-on to the publicly available metabolic modelling tool ScrumPy [1]. Tuatara is imported into ScrumPy as a Python package and interacted with using the same Python IDLE interface through the Python programming language. It can be run on a laptop or PC and performs its core functions within seconds on a quad-core Linux desktop computer. The pre-requisite tools Prokka and Roary can be run on the same Linux desktop computer in less than an hour [2-3]. The source code and documentation is publicly available on GitHub under an AGPL-3.0 License.

2. Aims

Our aims are to:

1. reduce the degree of curation needed for newly built high-quality GSMs and
2. implement intra-species diversity in GSMs, replacing the standard model organism approach.

Our system means microbial communities can be modelled and simplified to build.

3. Methods

Using a pre-existing high-quality GSM, we can use annotated sequences of strains from the same species to build 'pseudo-models'. The method requires gene presence/absence results from Prokka annotated genome assemblies [2] of bacteria using the pan genome software Roary [3]. Tuatara translates these results into reaction presence/absence data using several publicly available organism databases from BioCyc [4]. Only the unique and absent reactions in each sample are stored relative to the existing model. Reactions in the existing GSM can then be seamlessly swapped to the 'pseudo-model' while retaining the original GSM. These models can be analysed individually for strain specific metabolic properties or together and compared without having to switch between models.

4. Application

Tuatara is being applied to investigate diversity in uropathogenic Escherichia coli (UPEC) strains, the main cause of urinary tract infections (UTIs), in relation to antimicrobial resistance. The pseudo-models produced are being used to interpret the effects of bactericidal (killing) vs bacteriostatic (replication-preventing) antibiotics on ATP, a biomarker for antimicrobial resistance.

References

[1] M. G. POOLMAN et al., "ScrumPy: metabolic modelling with Python", *IEE Proceedings - Systems Biology*, 2006; 153.

[2] T. SEEMANN, "Prokka: rapid prokaryotic genome annotation", *Bioinformatics*, 2014; 30: 2068-2069.

[3] A. J. PAGE et al., "Roary: rapid large-scale prokaryote pan genome analysis", *Bioinformatics*, 2015; 31: 3691-3693.

[4] P. D. KARP et al., "The BioCyc collection of microbial genomes and metabolic pathways", *Briefings in Bioinformatics*, 2017; 20: 1085-1093.

Title: Applying Genome-Scale Metabolic Modelling to Understand the Metabolism of Staphylococcus Epidermidis RP62A

Author: Teresa Díaz Calvo[1], Noemi Tejera[2], Iain McNamara[3], Gemma C. Langridge[4], John Wain[5], Mark Poolman[6] and Dipali Singh[7]

Primary Affiliation: [1]Quadram Institute Bioscience, Norwich Research Park, UK, [2 4 5 7]Microbes in the Food Chain, Quadram Institute Bioscience, Norwich Research Park, UK, [3]Norfolk and Norwich University Hospital NHS foundation trust, Norwich, UK, [6]Cell System Modelling Group, Oxford Brookes University, Oxford, UK

Abstract

Staphylococcus epidermidis is an ubiquitous commensal of the mucous membranes and the skin with the potential to cause harm under specific circumstances. This opportunistic pathogen is currently a major cause of nosocomial infections such us foreign body-related infections, endocarditis and neonatal infections. Its pathogenic potential and resistance to antibacterial drugs is on the rise. Traditionally considered less important than the well-known pathogen *Staphylococcus aureus*, we have identified a gap in knowledge related to the metabolic mechanisms that allow *S. epidermidis* to adapt to different niches. With an aim to improve our understanding on the nutritional requirements of these organisms, we have constructed, curated and analyzed a genome-scale metabolic model for the *S. epidermidis* strain RP62A. Model analysis focused on identifying possible amino acid auxotrophies and investigating nutrient utilization and preferences for growth and energy production. The results obtained *in-silico* were compared with the outcome of experimental work that assessed the ability of this organism to grow in the absence of specific amino acids and the effect of nutrient deprivation on bacterial growth. We identified those amino acids preferentially utilized by RP62A for energy and biomass production and compared these results with published data on the matter. We have also shown an interesting link between amino acid and anergy metabolism. Further investigations are now needed in clinically relevant conditions to shed light on the factors that allow *S. epidermidis* to colonize specific body sites and ultimately cause disease.

Title: Modeling and Deciphering the Regulation of Microbial Central Metabolism by Dynamic Optimization

Author: Wassili Dimitriew, Stefan Schuster and Jan Ewald

Primary Affiliation: Department of Bioinformatics, Friedrich Schiller University Jena, Germany

Abstract

Proper regulation of metabolism is necessary for microorganisms to prevail in fluctuating environments and during interactions with other cells. Hence, evolutionary processes have shaped optimal strategies to regulate the flux through metabolic pathways. A wide variety of biological sciences and industrial applications benefit from a better understanding of the dynamic regulatory strategies in metabolism. Recently, dynamic optimization of metabolic regulation has provided insights into activation and regulation of metabolic pathways [1-3], allowing one to study the aforementioned strategies more thoroughly. Further, it was shown that these strategies are crucial in preventing accumulation of toxic intermediates [2]. However, the dynamic regulation of larger and more complex metabolic networks is still elusive [3].

Taking into account the aforesaid, we constructed a model of *Escherichia coli's* central metabolism, using the existing kinetic model described in [4]. Enzyme abundances were introduced to serve as controls, and some reactions were lumped for the sake of the ability to distinguish between enzymes and metabolites. For the situations where enzymes still could be considered both as metabolites and controls, enzyme synthesis rates were used as the latter. Unlike in constraint based methods or Metabolic Control Analysis, we did not use the steady state assumption. All the aforementioned allows us to simulate a wide variety of the bacterium's dynamic regulatory responses during different environmental changes like carbon source fluctuations. Using enzymes as control variables and minimization of both initial enzyme concentrations and deviations from these concentrations as an objective functional, our dynamic optimization approach is able to determine highly regulated enzymes and hence key points of regulation in central metabolism. As in previous studies, we expect to observe similarities and differences between highly regulated enzymes and enzymes with high control in the view of Metabolic Control Analysis [3]. Based on the regulation determinants deciphered by dynamic optimization, our approach will provide a more holistic picture of the relation between metabolic control and regulation.

References:

[1] F. Wessely et al., "Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs.", *Mol Syst Biol*, 2011; 7:1-13.

[2] J. Ewald et al., "Optimality principles reveal a complex interplay of intermediate toxicity and kinetic efficiency in the regulation of prokaryotic metabolism.", *PLoS Comput Biol*, 2017; 13:1-19.

[3] J. Ewald et al., "Deciphering the regulation of metabolism with dynamic optimization: an overview of recent advances.", *Biochem Soc Trans*, 2017; 45:1035-1043.

[4] P. Millard et al., "Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*.", *PLoS Comput Biol*, 2017; 13:1-24.

Title: Automated Genome-Scale Fungal Model Construction in Kbase: Towards Eliminating Redundancies in Fungal Biochemistry Template

Author: Janaka N. Edirisinghe, Claudia Lerma, Filipe Liu, José P. Faria, Sam Seaver and Christopher S. Henry

Primary Affiliation: Argonne National Laboratory, United States

Abstract

Fungal genome-scale metabolic models are an efficient way of predicting phenotypes across various environmental conditions. These metabolic models are a key utility in understanding overall metabolism, fungal-bacterial, and plant-fungal community behavior. However, (i) automating the construction of high-quality fungal draft models and (ii) the consolidation of redundant fungal biochemistry that is derived from different published models has been a challenge.

In KBase, users can easily construct draft fungal metabolic models via the "Build Fungal Model" app based on any fungal genome with structural annotations. This methodology uses the computation of orthologous protein families to propagate a curated set of biochemistry that is derived from 14 published fungal metabolic models.

In addressing the challenge of significantly lowering the redundancy that occurs in fungal biochemistry template; we have been working towards a manually curated, non-redundant, compartmentalized template based on ModelSEED biochemistry, that the compartmentalized biochemical pathways are fully hand-drawn representing the entire fungal biochemistry utilizing the Escher mapping environment.

To demonstrate this methodology and the recent progress, we show the representation of the yeast biochemistry template in this presentation. As a community effort, we are currently expanding this template capturing the unique fungal biochemistry beyond yeast. With this curation approach, we able to see the genuine pathway variations, quickly spot gaps in biochemistry by different compartments and also the redundancies that we can eliminate from the template which results in the construction of high-quality draft models. Fungal model construction based on this recent yeast template is available via the 'Build Fungal Model' app in KBase as the default option.

Title: Dynamic Modeling of the Shikimate Pathway and Central Carbon Metabolism of *E. coli*

Author: André Fonseca, Isabel Rocha

Primary Affiliation: ITQB (Universidade Nova de Lisboa), Portugal

Abstract

The shikimate pathway is the central carbon metabolism's (CCM) gateway to aromatic amino acids and other high value compounds important in industry and pharmaceuticals[1][2]. The biological synthesis of these metabolites is highly sought after due to its sustainability[2], however, the commercial viability of large- scale fermentation processes depend enormously on the microorganism's efficiency of producing the desired products[3].

While the constant development of metabolic engineering (ME) has been continuously increasing the widespread use of such processes, the sheer complexity of metabolic regulation makes it difficult to predict the ideal targets for ME[4]. To tackle this issue, mathematical models that can describe the biological systems behavior have become an essential tool[5]. These models can quantitively predict intracellular concentrations of metabolites, as well as fluxes, from ordinary differential equations based on kinetic equations that depict enzymatic reactions[5][6]. Because of the many biochemical details of metabolic networks, along with a lack of kinetic information on the dynamics of reactions, there are no dynamic models that encompass a cell's entire metabolism, focusing on specific pathways or cellular mechanisms[7][8][9].

In this study, we present a dynamic model for *Escherichia coli* that simulates the CCM: glycolysis, pentose phosphate pathway, citric acid cycle (TCA), anaplerotic reactions and the glyoxylate pathway, along with the shikimate pathway, aromatic amino acids synthesis pathway and the electron transport chain, including ATP synthesis. Our model is composed of 156 reactions, 73 metabolites and over 500 kinetic parameters. Fluxomic and metabolomic data was used to validate the model's behavior, with simulations of wild-type and genetically modified strains (single knockouts). The model acts in accordance with experimental data, qualitatively speaking, however, quantitatively there is still the need for fine tuning. Nevertheless, the results we obtained indicate that the present model can simulate the metabolic fluxes of E. coli and could therefore be a great tool for studying the control of such fluxes.

References

[1] Bongaerts, J. et al., "Metabolic engineering for microbial production of aromatic amino acids and derived compounds", Metabolic Engineering, 2001, 3(4), 289–300.

[2] Liu, Q. et al., "Rewiring carbon metabolism in yeast for high level production of aromatic chemicals", Nature Communications, 2019, 10(1), 1–13.

[3] Crater, J. S. et al., "Scale-up of industrial microbial processes", FEMS Microbiology Letters, 2018, 365(13), 1–5.

[4] Papagianni, M., "Recent advances in engineering the central carbon metabolism of industrially important bacteria", Microbial Cell Factories, 2012, 11, 1–13.

[5] Kim, O. D. et al., "A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering", Frontiers in Microbiology, 2018, 9(July), 1–22.

[6] Almquist, J. et al., "Kinetic models in industrial biotechnology - Improving cell factory performance", Metabolic Engineering, 2014, 24, 38–60.

[7] Peskov, K. et al., "Kinetic modelling of central carbon metabolism in Escherichia coli", FEBS Journal, 2012, 279(18), 3374–3385.

[8] Chassagnole, C. et al., "Dynamic modeling of the central carbon metabolism of Escherichia coli", Biotechnology and Bioengineering, 2002, 79(1), 53–73

[9] Bruggeman, F. J. et al., "The multifarious short-term regulation of ammonium assimilation of Escherichia coli: Dissection using an *in silico* replica", FEBS Journal, 2005, 272(8), 1965–1985.

Title: Computational Design and Biological Insights into Modular Cell Design for Large Libraries of Exchangeable Product Synthesis Modules in *Escherichia Coli*

Author: Sergio Garcia and Cong T. Trinh

Primary Affiliation: Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, USA

Abstract

Modular cell (ModCell) design enables rapid generation of production strains by assembling a modular (chassis) cell with exchangeable production modules to achieve overproduction of target molecules. To enable modular cell design for large libraries with hundreds of products, we developed a new computational method, named ModCell-HPC, with a highly-parallel and multi-objective evolutionary algorithm that allows us to analyze modular design properties. We demonstrated ModCell-HPC to design E. coli modular cells towards a library of 161 endogenous production modules. From these simulations, we identified E. coli modular cells with few genetic manipulations that can produce dozens of molecules in a growth-coupled manner with different types of fermentable sugars. These designs revealed key genetic manipulations at the chassis and module levels to accomplish versatile modular cells, involving not only in the removal of major by-products but also modification of branch points in the central metabolism. We further found that the effect of various sugar degradation on redox metabolism results in lower compatibility between a modular cell and production modules for growth on pentoses than hexoses. To better characterize the degree of compatibility, we developed a method to calculate the minimal set cover, identifying that only three modular cells are all needed to couple with all of 161 production modules. By determining the unknown compatibility contribution metric, we further elucidated the design features that allow an existing modular cell to be re-purposed towards production of new molecules. Overall, ModCell-HPC is a useful tool for understanding modularity of biological systems and guiding more efficient and generalizable design of modular cells that help reduce research and development cost in biocatalysis.

Title: Multiscale Analysis of Autotroph-Heterotroph Interactions in a High-Temperature Microbial Community

Author: KA Hunt, RdeM Jennings, WP Inskeep, RP Carlson

Primary Affiliation: Montana State University

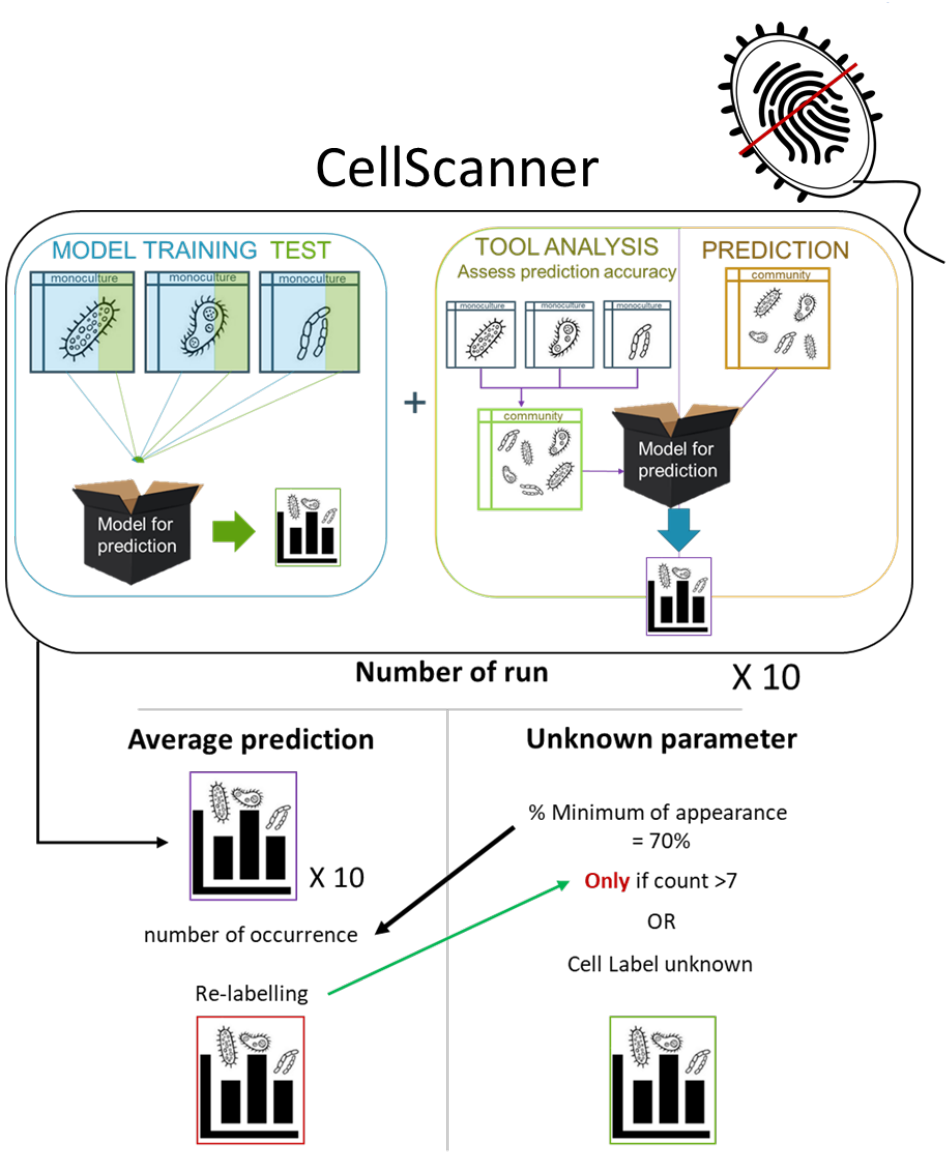
Abstract

Interactions among microbial community members can lead to emergent properties, such as enhanced productivity, stability, and robustness. Iron-oxide mats in acidic (pH 2 – 4), high-temperature (> 65 ºC) springs of Yellowstone National Park (YNP) contain relatively simple microbial communities and are well-characterized geochemically. Consequently, these communities are excellent model systems for studying the metabolic activity of individual populations and key microbial interactions. The primary goals of the current study were to integrate data collected in situ with in silico calculations across process-scales encompassing enzymatic activity, cellular metabolism, community interactions, and ecosystem biogeochemistry, and to predict and quantify the functional limits of autotroph-heterotroph interactions. Metagenomic and transcriptomic data were used to reconstruct carbon and energy metabolisms of an important autotroph (Metallosphaera yellowstonensis) and heterotroph (Geoarchaeum sp. OSPB) from Fe(III)-oxide mat communities. Standard and hybrid elementary flux mode and flux balance analyses of metabolic models predicted cellular- and community-level metabolic acclimations to simulated environmental stresses. In situ geochemical analyses, including oxygen depth-profiles, Fe(III)-oxide deposition rates, stable carbon isotopes and mat biomass concentrations, were combined with cellular models to explore autotroph-heterotroph interactions important to community structure-function. Integration of metabolic modeling with in situ measurements, including the relative population abundance of autotrophs to heterotrophs, demonstrated that Fe(III)-oxide mat communities maximize total community growth rate, as opposed to the net community growth rate, as predicted from the maximum power principle. Integration of multiscale data with practical ecological theory provides a basis for predicting autotroph-heterotroph interactions and community-level cellular organization.

Title: CellScanner, a User-Friendly Tool to Identify and Enumerate Cells in Flow Cytometry Data

Author: C. Joseph and K. Faust

Primary Affiliation: KU Leuven, Belgium



CellScanner: a tool to enumerate and identify cells from flow cytometry data with supervised classification
The tool has two principal functions: *Tool analysis* (for in-silico communities) and *Prediction* (for in-vitro communities) both based on a model trained with monoculture file annotated by the user. The tool produces statistical data such as Accuracy, F1 score, confusion matrix, 3D graphs, or prediction values, represented with the graph icons. The parameters *Number of run* and *Average prediction* allows the user to increase the accuracy of a prediction, the *Unknown parameter*, increase the specificity and precision of the prediction.

Abstract

CellScanner is a python app that aims to identify different microbial species in flow cytometry data. Users can carry out classification or clustering to predict the microbial composition of their samples or to assess tool accuracy on samples with known composition. The tool will learn species characteristics with a machine learning technique (neural network, random forest or logistic regression) and then find cells with the same characteristics in an unknown mix. CellScanner also includes a new method to differentiate debris from cells using supervied classification, offering to the user an automated gating option. With a user-friendly interface, users can design their experiments and analyse results efficiently.

Title: Metabolic Network Modeling to Determine Microbially-Driven Biogeochemical Functions in Globally-Distributed River Corridors

Author: Aimee K. Kessell[1], Fangchao Song[2], Janaka N. Edirisinghe[3], Christopher S. Henry[3], Kelly Wrighton[4], James C. Stegen[5], Timothy D. Scheibe[5], and Hyun-Seob Song[1]

Primary Affiliation: [1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, United States, [2]Lawrence Berkeley National Laboratory, United States, [3]Argonne National Laboratory, United States, [4]Colorado State University, Fort Collins, CO, United States, [5]Pacific Northwest National Laboratory, United States

Abstract

Microbial processes in river corridors drive the material and energy cycles in the natural environment. Model-data integration is critical for determining microbial activities and biogeochemical hotspots and hot moments. Collection of omics profiles from globally distributed river corridors through the Worldwide Hydrobiogeochemical Observation Network for Dynamic River Systems (WHONDRS) consortium provides valuable resources in this regard [1]. The 48 Hour Diel Cycling Study performed by WHONDRS in 2018 used a time-series sampling approach across seven river systems with different drivers or sub-daily river stage variation [2]. Through a week-long summer school hosted by Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL) in July 2020, we analyzed the omics data (including metagnomes, metagenome-assembled genomes, and metabolomics) from four of these river systems. Full integration of omics data into conventional lumped models is however ineffective due to their oversimplified representation of reaction pathways. In this work, we demonstrate the utility of genome-scale metabolic models in improving our understanding of biogeochemical reactions. Using the U.S. DOE's KBase (www.kbase.us) modeling pipeline, we constructed metabolic networks of metagenomes and their high-quality extracted bins [3]. Through comparative analysis of community- and individual taxon-level metabolism and biogeochemical potentials, we identified anaerobic pathways and auxotrophs in pore and surface water samples from multiple river systems and determined both conserved and distinct metabolic pathways within and across river systems. Significant differences in metabolic functions across river systems were predicted by incorporating high-resolution metabolomics obtained from Fourier transform ion cyclotron resonance mass spectrometry. For integrating this high-resolution metabolomic data, we used KBase chemoinformatics tools to expand the known chemistry of reference compounds through iterative addition of reactions while pruning to avoid a combinatorial explosion of potential compounds and reactions. The resulting models provide a comprehensive description of potential biogeochemical reactions with unprecedented detail.

References

[1] Stegen, J. C., and Goldman, A. E. "WHONDRS: a community resource for studying dynamic river corridors." *mSystems*; 2018, 3:e0151-18.

[2] R.E. Danczak et al., "WHONDRS 48 hour diel cycling study at the Altamaha River in Georgia, USA", *Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems (WHONDRS)*, 2019.

[3] A.P. Arkin et al., "KBase: The United States Department of Energy Systems Biology Knowledgebase", *Nature Biotechnology,* 2018; 36: 566 – 569.

Title: Elementary-Mode Analysis of the Methylcitrate Cycle

Author: Lukas Korn[1], Matthias Brock[2] and Stefan Schuster[1]

Primary Affiliation: [1]Dept. of Bioinformatics, University of Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany, [2]School of Life Sciences, University of Nottingham, United Kingdom

Abstract

The aim of this work is to analyse the methylcitrate cycle of several fungi, such as *Aspergillus fumigatus* [1], and bacteria such as *Escherichia coli* [2] via Elementary-mode analysis [3]. The methylcitrate cycle is a metabolic pathway that catalyses an alpha-oxidation of propionyl-CoA and oxaloacetate to pyruvate [4]. A reaction model has been developed, using information from databases and the literature. This reaction network was then used to calculate the elementary modes, with the help of the simulation software Metatool 4.3 [5].

These elementary modes were analysed systematically, for example, with respect to their ATP yield. Additionally the robustness of the network has been computed [6] and some predictions were made for the impact of enzyme deletion on the network [7].

As result of this work, some previously known pathways have been confirmed theoretically and additionally some new ones have been found, for example, a different way of oxaloacetate regeneration. Promising targets for pharmaceuticals against pathogenic fungi in this network are identified.

References

[1] O. Granet et al., "Methylcitrate synthase from *Aspergillus fumigatus* is essential for manifestation of invasive aspergillosis", *Cell Microbiol* 2008; 10: 477 – 492.

[2] M. Brock et al., "Oxidation of propionate to pyruvate in *Escherichia Coli:* involvement of methylcitrate dehytratase and aconitase", *Eur. J Biochem*, 2003; 269: 8184 - 94.

[3] S. Schuster et al., "Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering.", *Trends Biotechnol*, 1999; 17: 53 – 60.

[4] M. Brock et al., "Methylcitrate synthase from *Aspergillus nidulans:* Implications for propionate as an antifungal agent", *Mol Microbiol*, 2000; 35: 961 - 73.

[5] T. Pfeiffer et al., "Metatool: for studying metabolic networks.", *Bioinformatics,* 1999; 15: 251 – 257.

[6] T. Wilhelm et al., "Analysis of structural robustness of metabolic networks.", *Systems Biol*, 2004; 1: 114 – 20.

[7] P. Unrean et al., "Rational design and construction of an efficient *Escherichia coli* for production of diapolycopendioic acid.", *Metab Eng*, 2009; 12: 112 - 22.

Title: Metabolic Modeling Deciphers Interactions in a Cheese Bacterial Community

Author: Maxime Lecomte[1], David Sherman[2], Hélène Falentin[1], Clémence Frioux[2] and Simon Labarthe[2]

Primary Affiliation: [1]INRAE UMR STLO, France, [2]INRIA BSO, France

Abstract

Bacterial communities occur in every environment, from the deep ocean to mammal hosts, and are also important in industrial processes, such as fermented food production. Yet, understanding the interactions within complex bacterial ecosystems remains a major issue in the modeling world. In this work, we focus on a bacterial community composed of two lactic acid bacteria (LAB), Lactoccocus lactis and lactiplantibacillus plantarum, and one propionic bacterium Propionibacterium freudenreichii, in an industrial cheese environment. These bacteria, by releasing organoleptic compounds, participate to the flavour of the cheese. As within any microbial community, the bacteria from the cheese ecosystem interact one with each other, constantly altering the environment they share and thereby possibly impacting the growth of the other species. Such interactions [1] occur mainly by metabolite exchange in the intercellular space, which makes the study of the bacterial metabolism adequate. Computational models enable us to identify them through two main steps: building a genome-scale metabolic network (GEM) for every member of the community to predict substrate consumption and metabolite production, then, revealing interactions between members. A metabolic network represents a set of chemical reactions inferred from the genome of an organism.

In this study, GEMs of three strains of the cheese ecosystem were built using a top-down approach with the CarveMe tool [2]. GEMs were used as a resource for the identification of key metabolites and pathways involved in organoleptic compounds synthesis, and regulation elements were detected with mathematical formalisms [3,4]. In addition, we constrained GEMs with multi-omics data such as metatranscriptomics, providing more in-depth predictions on the organization of the community by characterizing pathways of interest for the production of organoleptic metabolites. In order to identify putative interactions over time, dynamical systems [5] were used to model the behavior of the GEMs at the community scale. The resulting model was in accordance with the literature and showed trophic interactions within the bacterial community. Lactic acid and acetaldehyde were identified as mediators of such trophic interactions and the competition for lactose was characterized among the lactic bacteria. This works illustrates how omics data integration combined with genome-scale metabolic modeling can help decipher the interactions within a bacterial community. Challenges reside in scaling up to larger communities while preserving the accuracy of predictions.

References

[1] Faust et al, Microbial interactions: from networks to models. Nat Rev Microbiol 10, 538–550 (2012).

[2] D. Machado et al, “Fast automated reconstruction of genome-scale metabolic models for microbial species and communities”, Nucleic Acids Research, gky537, 2018.

[3] Orth JD et al. What is flux balance analysis?. Nat Biotechnol. 2010;28(3):245-248.

[4] Ebenhöh O et al. Structural analysis of expanding metabolic networks. Genome Inform. 2004;15(1):35-45.

[5] Radhakrishnan Mahadevan et al, Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli,Biophysical Journal,Volume 83, Issue 3,2002

Title: Omics and Metabolic Network Integration to Predict the Impacts of Copper Dysregulation on the Liver Functions and Energy Metabolism

Author: Naeun Lee[1], Joseph Brake[2], Heejeong Kim[2], Jaekwon Lee[2] and Hyun-Seob Song[1]

Primary Affiliation: [1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, United States, [2]Department of Biochemistry, University of Nebraska-Lincoln, United States

Abstract

Copper (Cu) is an essential mineral for our body. Diverse biological processes in mammals, including humans energy generation, conversion of superoxide anions to hydrogen peroxide, iron transport, are dependent on Cu-containing enzymes. Deficiency, excess, and uncontrolled chemical reactions of Cu cause or contribute to the onset and progression of human diseases, such as Menkes disease, Wilson Disease, cancer, neurodegenerative diseases, and metabolic disorder. By excreting and storing excess Cu, the liver maintains the whole-body Cu homeostasis. Moreover, the functions of liver cells are dependent on Cu-requiring enzymes. This research aims to discover the impacts of Cu dysregulation in the liver in conjunction with metabolic dysfunction and fatty liver on hepatic functions and systemic energy metabolism. To achieve the goal, we employ a co-designed modeling and experimental study. A mouse strain in which the CTR1 gene encoding a high-affinity Cu importer is knocked out in hepatocytes induced Cu limitation in the liver specifically. To identify metabolic changes caused by Cu deficiency, we incorporated gene expression data (collected from the liver tissues in control and knockout mice) into a human cell genome-scale metabolic network (Human-GEM) [1]. We added Cu-dependent cellular pathways to the metabolic network. Integration of transcriptomic data with the metabolic network was implemented using E-Fmin, which estimates flux distributions as indicated by expression levels based on the flux minimization principle [2]. Our model presented that Cu deficiency leads to (1) compromised detoxification of reactive oxygen species (ROS) in the cytosol through the decreased activity of superoxide dismutase 1 (SOD1) and (2) an increase in mitochondria net activity as indicated by amplified fluxes through oxidative phosphorylation and TCA cycle, as well as by an increase of the respiratory chain. These predictions were consistent with experimental data. We will extend our model-data integration approach to provide deeper mechanistic insights, eventually developing better therapeutic strategies for liver diseases and metabolic dysfunction.

References

[1] Robinson, J. L. et al., "An atlas of human metabolism", *Science Signaling*, 2020; 13: 1-12.

[2] Song, H.-S. et al., "Prediction of Metabolic Flux Distribution from Gene Expression Data Based on the Flux Minimization Principle", *PLoS ONE,* 2014; 9: 1-11.

Title: A Protein-Enabled Genome-Scale Metabolic Model of Maize In PlantSEED

Author: Claudia Lerma-Ortiz[1], Samuel MD Seaver[2], Fatima Foflonker[2], Filipe Liu[2], Donald R McCarty[1], Andrew D Hanson[1] and Christopher S Henry[2]

Primary Affiliation: [1] University of Florida, USA; [2] Argonne National Laboratory, USA

Abstract

Recent advances in proteomics methods and technologies are greatly improving the quality of proteomics data while simultaneously reducing costs. However, limited approaches exist today for using this data to improve our understanding of metabolism and metabolic variation across conditions. Here we present a new flux balance analysis (FBA) approach, called proteomics-driven FBA, where we integrate data about known kinetic constants, measured or curated metabolic fluxes, and quantitative proteomics measurements to improve overall metabolic model quality and gain insights into metabolic flux profiles, metabolic variation, and kinetic bottlenecks. We apply this approach to study comprehensive quantitative protein abundance measurements from sixteen distinct tissues in the C4 grass, Zea mays [1]. We began our studies using an enhanced PlantSEED model of Zea mays [2], which we subsequently refined to better fit our experimental data. Refinements included improvements to the biomass objective function, improvements to gene-protein-reactions associations, adjustments to metabolic pathways and subcellular compartmentation, and addition of a pseudo-platidial compartment to better represent C4 metabolism within a single model. We used proteomics-driven FBA with this improved model to simultaneously solve for optimal flux and apparent kinetic constants across all sixteen tissues and cell types with available proteomics data, exposing in detail how flux and kinetic bottlenecks vary from tissue to tissue in Maize. We find this this simultaneous solution process improves consistency of results across all tissues, while also reveal key areas of variation (e.g. photosynthesis, nitrogen acquisition, and byproduct excretion pathways).

The code for this new approach is implemented in COBRApy and available in a Jupyter notebook. Models and datasets are available in KBase.

References

[1] Walley, Justin W., et al. “Reconstruction of protein networks from an atlas of maize seed proteotypes.” Proceedings of the National Academy of Sciences 110.49 (2013): E4808-E4817.

[2] Seaver, Samuel MD, et al. “Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm.” Crop Breeding. Apple Academic Press, 2016. 277-314.

Title: A Computational and Experimental Analysis: Information About PAM Instrument Parameters may Affect our Research

Author: Tim Nies[1], Yuxi Niu[2] , Oliver Ebenhöh[1], Shizue Matsubara[2], Anna Matuszyńska[1]

Primary Affiliation: [1]Institute of Quantitative and Theoretical Biology, Heinrich Heine University Düsseldorf; [2]IBG-2: Pflanzenwissenschaften, Forschungszentrum Jülich, Germany

Abstract

Pulse Amplitude Modulation (PAM) fluoremetry is a widely used experimental technique in the field of plant science, using Chlorophyll *a* fluorescence. Frequently combined with computational and modeling approaches, PAM measurements can give valuable information about the photosynthetic energy conversion in plants and photosynthetic microorganisms. Unfortunately, many scientific articles using the PAM technique do not provide detailed parameter settings. However, small variations in these technicalities can influence measurements. Thus, neglecting technical details may lead to misinterpretation and hamper the reproduction of scientific results, especially *in silico*. By employing a combined modeling and experimental approach, we show the effects of parameter settings on PAM measurements. We used a validated model of non-photochemical quenching (NPQ) to investigate how NPQ induction and relaxation are influenced by changes in often non-disclosed parameters. By doing so, we hope to raise the awareness of the issue of non-disclosed parameters and urge to establish minimum standards for the documentation of PAM measurement protocols. This will help theoretical and experimental biologists to advance a system-wide understanding of photosynthesis and help to launch standardized databases for the plant and data science communities.

Title: Emergence of Metabolic Cooperation Between Cells in Schematic Balanced Growth Models

Author: Tobias Pfennig, Martin J. Lercher, and Peter Schubert

Primary Affiliation: Heinrich Heine University Düsseldorf, Germany

Abstract

Metabolic cooperation emerges frequently in microbial communities. In many cases, such cooperation may be beneficial to one and selectively neutral for the other cell; in others, two cells with complementary deficiencies may complement each other. It is unclear, however, if there are situations where cooperation is fundamentally superior, i.e., where two self-sufficient and equally efficient cells can raise both of their growth rates by exchanging metabolites with each other. Models that ignore metabolite concentrations, such as Flux Balance Analysis (FBA) and Resource Balance Analysis (RBA) are fundamentally unable to predict such behaviour. Here, we thus approach this question through growth balance analysis (GBA) [1-4] of minimal, interacting cell models, explicitly considering metabolite concentrations and non-linear enzyme kinetics. Each autosynthetic cell can import two nutrients and convert them to precursors for protein synthesis via independent enzymatic pathways. Alternatively, the cells can also exchange one or both protein precursors via dedicated transporters. We then maximized the simultaneous growth rates of the cells. We find that when both cells possess equally efficient metabolic systems, mutualistic symbiosis emerges only when the costs of metabolite exchange are sufficiently small compared to the costs of the circumvented enzymatic pathways. In such cases, the benefit of cooperation is related to the saturation of the remaining enzymes. Cooperation emerges more easily if the two metabolic systems complement each other, such that each cell is more efficient at producing one of the two protein precursors.

References

[1] D. Molenaar et al., “Shifts in growth strategies reflect tradeoffs in cellular economics”, *Mol Syst Biol*, 2009; 5: 1–10.

[2] H. Dourado et al., “Enzymes and Substrates Are Balanced at Minimal Combined Mass Concentration *in vivo*”, *BioRxiv*, 2017

[3] H. Dourado et al., “An analytical theory of balanced cellular growth”, *Nat Commun*, 2020; 11: 1226.

[4] D. Sezer et al., “Structure of autosynthetic models of balanced cell growth and numerical optimization of their growth rate”, *BioRxiv,* 2020

Title: Analysis of Global Changes of the Lipid Metabolism on Lipid Metabolic Networks with the Lipid Network Explorer

Author: Tim Daniel Rose, Nikolai Köhler, and Josch Konstantin Pauling

Primary Affiliation: LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Germany

Abstract

Advances in mass spectrometry allow researchers to identify and quantify an increasing number of lipids from biological samples. Lipids play an important role in biological systems such as membrane compositions, signaling cascades, or energy storage and are highly relevant for metabolic diseases. However, approaches to analyze changes in lipid metabolism functionally, e.g. by incorporating biochemical reactions or lipid metabolic networks on the species level are still rare [1]. A common strategy to analyze associations between lipids are correlation networks, which do not contain knowledge about functional associations about lipids [2–4].

A challenge in lipid pathway analyses is the multispecificity of many enzymes catalyzing lipid-reactions. Lipids are grouped in categories, classes, and species [5]. Glycerophospholipids commonly consist of a glycerol backbone, a headgroup, and fatty acids. Enzymes can convert lipids e.g. by modifying headgroups or removing fatty acids independently of the length or saturation of fatty acids attached to the lipid. However, many databases such as Reactome [6] contain lipid metabolic reactions only on the class level and can therefore not directly be applied to lipidomics data, where lipids are usually measured on the sum or molecular species level.

We developed the Lipid Network Explorer (LINEX), a webapp for the analysis of lipid metabolic networks. LINEX makes it possible to dynamically generate metabolic networks of lipids based on quantified lipidomics data. Data specific lipid networks are then combined with statistical properties that can be visualized on the networks, to allow researchers to analyze the network structure and changes of the lipidome between conditions. With this global changes of the lipidome can be connected to local enzymatic reactions. We applied LINEX to publicly available lipidomics data and could gain new insights in the regulation of lipid metabolism and homeostasis. We observed differences in the availability and associations of lipid species of different classes indicating a tight regulation of lipid enzymatic reactions.

LINEX gives a different view on the lipidome and provides the basis for the analysis of the lipid species metabolism. We are working on extending LINEX to integrate lipid reaction databases to give more insights about lipid reactions and their regulation, and develop algorithms to extract dysregulated lipid and enzyme subnetworks.

References

[1] C Gaud C et al. “BioPAN: a web-based tool to explore mammalian lipidome metabolic pathways on LIPID MAPS” *F1000Res*, 2021;10: 4.

[2] MS Köberlin et al. A Conserved Circular Network of Coregulated Lipids Modulates Innate Immune Responses. Cell. 2015;162: 170–183.

[3] L Yetukuri et al. “Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis” *BMC Syst Biol*, 2007;1: 12.

[4] G Wong et al. “LICRE : unsupervised feature correlation reduction for lipidomics” *Bioinformatics*, 2014;30: 2832–2833.

[5] G Liebisch et al. “Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures” *J Lipid Res*, 2020;61: 1539–1555.

[6] B Jassal et al. “The reactome pathway knowledgebase” *Nucleic Acids Res*, 2020;48: D498–D503.

Title: Information- and Communication-Centric Approach in Cell Metabolism

Author: Zahmeeth Sakkaff[1], Nidhi Gupta[2 4], Massimiliano Pierobon[3], and Christopher S. Henry[1]

Primary Affiliation: [1]Data Science and Learning Division, [2]Biosciences Division, Argonne National Laboratory, Lemont, IL; [3]Department of Computer Science and Engineering, University of Nebraska, Lincoln, NE, USA; [4]University of Chicago, Chicago, IL, USA

Abstract

In natural environments, microorganisms may exist in their single/isolate-celled form or as mixed colonies of cells that may interact readily with one another and perform complex interactions and communication mediated by the exchange of chemical substrates. Understanding the mechanism underlying natural cell-to-cell and intracellular communication is one of the challenging research problems addressed by scientific communities. In this study, we introduce information- and communication-centric computational approaches to estimate the information flow in biological cells underlying the acquisition of extracellular information and its impact on single-species and multiple-species behavior. The focus of this poster is to present the limits in the fine-tuning and controllability of the behavior of natural biological cell communication with the environment that has many diverse applications. We adopt a molecular communication abstraction of cell metabolism and fundamentals from Shannon information theory to understand variations in the amount of information that propagates (information flow) through the genome to the metabolic network of individual species, as well as the information exchanged among species. We study the models growing separately, growing together in merged ("mixed-bag") form as if both species were combined into a single species, or growing together in a "compartmentalized" manner as if both species were growing in close proximity. We utilize the gold standard models of *Escherichia coli* (*E. coli*) and *Bacteroides thetaiotaomicron* (*B. theta*) to study the bacteria occupancy at different niches in the gut and to evaluate their impact on a range of applications. We introduce an open-source computational tool, named RFMIA, that estimates the amount of information flow that occurs through a single-cell or multi-cell metabolic network as nutrients in the environment are consumed and transformed. Our study shows that, overall, information flows are more efficient through community than with single models. All the tools and data related to this study are publicly available at the DOE Systems Biology Knowledgebase.

Title: A Python Tool Making SBML Kinetic Modelling More Accessible

Author: Peter Schubert

Primary Affiliation: Computational Cell Biology, Heinrich-Heine-University Düsseldorf, Germany

Abstract

SBML (Systems Biology Modelling Language), a community standard for metabolic modelling, facilitates model reuse and exchange in a standardized format based on XML. libSBML is an open-source software library that provides an application programming interface (API) for the SBML format.

sbmlxdf is a Python package developed inhouse using libSBML API to access SBML encoded models, but shielding the complexities of directly interfacing with libSBML. sbmlxdf should make it easier for the kinetic modeler to create SBML kinetic models with maximal control over the resulting XML content. The tool also facilitates access to SBML model data for downstream processes, like optimizers.

Basically sbmlxdf converts between the SBML XML format and a set of Pandas dataframes, one for each SBML component. This allows creation and modification of metabolic models using Excel spreadsheets as an intermediate format. Filtering features and 'helper' functions in Excel can speed up model creation and reduce errors. Simple kinetic models can be created/modified from templates without detailed knowledge of SBML syntax and semantics. Compliance with respect to the SBML specification can be validated and errors corrected before writing out the corresponding SBML XML file. Downstream systems can access the set of Pandas dataframes, e.g. automatically create optimization problems.

We used the tool to create smaller kinetic 'toy' models in SBML. Subsequenlty we optimized these SBML models using balanced growth optimization methods based on [1] and [2].

sbmlxdf supports, with few exceptions, all functionality of SBML L3V2 core package and extension packages flux balance constraints, groups, distributions. sbmlxdf is free software released under the terms of the GNU General Public License as published by the Free Software Foundation; version 3 of the License.

References

[1] H. Dourado et al., "An analytical theory of balanced cellular growth", Nature Comm, 2020; 11:1129.

[2] D. Sezer et al., "Structure of autosynthetic models of balanced cell growth and numerical optimization of their growth rate", bioRxiv, 2020

Title: Engineering Modularity of Ester Biosynthesis Across Biological Scales

Author: Hyeongmin Seo and Cong T. Trinh

Primary Affiliation: Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, USA

Abstract

Metabolic engineering and synthetic biology enable controlled manipulation of whole-cell biocatalysts to produce valuable chemicals from renewable feedstocks in a rapid and efficient manner, helping reduce our reliance on the conventional petroleum-based chemical synthesis. However, strain engineering process is costly and time-consuming that developing economically competitive bioprocess at industrial scale is still challenging. To accelerate the strain engineering process, modular cell engineering has been proposed as an innovative approach that harnesses modularity of metabolism for designing microbial cell factories. It is important to understand biological modularity and to develop design principles for effective implementation of modular cell engineering. In this study, the modularity of ester biosynthesis was engineered from the molecular to the microbial community levels. Specifically, three important features of modularity (i.e., robustness, efficiency, and compatibility) were engineered and quantitatively analyzed across different scales. At the molecular (enzymatic) level, thermostability and promiscuity of alcohol acyltransferases were engineered to develop a robust designer ester biosynthesis. At the metabolic network (cellular) level, metabolism of *Escherichia coli* was rewired to overproduce isoamyl acetate through metabolic engineering and synthetic biology strategies. Also, by harnessing the engineered robust alcohol acyltransferase, a non-model thermophilic bacterium *Clostridium thermocellum* was engineered to produce medium chain esters directly from recalcitrant lignocellulosic biomass at elevated temperatures. Finally, at the microbial community level, a syntrophic *E. coli* co-culture was engineered for isobutyl butyrate production from a mixture of glucose and xylose. The successful engineering of the modularity of ester biosynthesis not only sheds light into the modular design principles of biological systems, but also seeks to develop industrially relevant ester production platforms.

Title: Exploring Valorisation of Fermentation By-Products Through Metabolic Modelling

Author: Anna Stikane[1], Elina Dace[2], and Egils Stalidzans[3]

Primary Affiliation: Biosystems Group, University of Latvia, Latvia

Abstract

Global challenges such as climate change, loss of biodiversity and growing global human population are driving the shift from fossil-based economy to bioeconomy. A key pillar of bioeconomy is industrial fermentation (IF), which produces metabolites (e.g., ethanol, amino acids, vitamins), protein (e.g., enzymes, biologics) and polymers (e.g., PLA, PHB, bacterial cellulose), as well as fermentation by-products (FerBP) such as spent microbial biomass (SMB) and wastewaters. Increasing volume and diversity of IF products lead to growing amount of FerBP, which could itself be used as feedstocks or feed-supplements for subsequent fermentation processes.

We aim to use metabolic modelling tools (e.g., flux balance analysis, FBA) to explore strategies for valorisation of FerBP. FBA approach will be applied on constraint based stoichiometric models (CBSM) using the modelling tool COBRA v3.0[1]. The algorithms in COBRA toolbox will enable metabolic engineering of strains by: 1) optimizing the substrate composition for production of particular metabolite, 2) improving the set of byproducts, and 3) optimizing the strain by deletions, insertions or modulating regulation of enzymes.

Thus, the project focuses on sustainable growth of bioeconomy, paving way for predictive models to aid the design of more integrated bioconversion processes and biorefineries.

References

[1] L. Heirendt et al., “Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.”, Nature Protocols, 2019; 14: 639–702

Title: dGPredictor: Automated Fragmentation Method for Gibbs Energy Change Prediction of Metabolic Reaction and *de novo* Pathway Design

Author: Vikas Upadhyay, Lin Wang, and Costas D. Maranas

Primary Affiliation: Department of Chemical Engineering, The Pennsylvania State University, University Park, PA

Abstract

Thermodynamic analysis of metabolic pathways is crucial to designing novel pathways for biochemical synthesis. To this end, Group Contribution (GC) methods are used to estimate the standard Gibbs energy change of (rG'o) of enzymatic reactions using limited experimental measurements. However, there are drawbacks to such methods due to their dependence on manually curated functional groups and their inability to include stereochemical information, resulting in a low reaction coverage. Therefore, we propose a moiety-based automated fragmentation method using molecular fingerprints to design a thermodynamic analysis tool called dGPredictor. It enables the inclusion of stereochemistry within chemical structures and thus covers more biochemical reactions compared to state-of-the-art GC methods. dGPredictor shows higher prediction accuracy compared to the current GC methods and can capture Gibbs energy change for reactions that only undergo stereochemical changes such as isomerase and transferase reactions, which shows no overall group change. We show the ability of dGPredictor to predict the Gibbs energy change for reactions involving novel structures and integration with de novo metabolic pathway design tools such as novoStoic to eliminate the reactions steps for which the directionalities are infeasible thermodynamically. We also developed a graphical user interface to facilitate easy access to dGPredictor for predicting Gibbs energy change of reactions at different pH and ionic strengths. The user interface allows customized user input of molecules as KEGG IDs and novel metabolites as InChI strings (Github). Thus, dGPredictor can cover more reactions and integrate seamlessly with the pathway design tool, which involves reactions with novel molecules.

Title: Linear Model for the Integration Of Labeling Data With Genome-Scale Metabolic Networks

Author: Luis V. Valcárcel[1], Jon Pey,[2], Francisco J. Planes[1]

Primary Affiliation: [1]Tecnun School of Engineering, University of Navarra, San Sebastian, Spain; [2]Intelligent Biodata, San Sebastian, Spain

Abstract

The growth of constraint-based models has been mainly driven by high-throughput -omics data, particularly genomic and transcriptomic profiling experiments. Unfortunately, the elucidation of reaction fluxes in metabolic networks remains a challenging task, because they cannot be directly predicted using this type of data. Instead, isotopic labeling experiments, which measures the amount of isotopes in the different metabolites along time or in steady-state, provides a more accurate approach. However, the integration of labeling data with reactions networks is complex and requires non-linear relationships, which have typically restricted its use for networks of small size.

We present a novel constraint-based model that integrates 13C labeling data and reactions fluxes in a linear fashion. Our approach relies on linear programming, which presents several advantages: easy integration with other constraint-based models; substantial reduction of computation time; efficient scaling to larger networks. Our approach was applied to a medium-size network [1], comprising 350 reactions and 184 metabolites, finding similar accuracy to other non-linear methods. We also investigate additional methods to integrate mass-spectrometry data in the predictions, which could importantly expand the application of the methodology presented. In conclusion, we believe the results here presented constitutes an important advance in the field of Systems Biology, facilitating the use of labeling data and, consequently, providing a more reliable estimation of reaction fluxes.

References

[1] Suthers, Patrick F., et al. “Metabolic flux elucidation for large-scale models using 13C labeled isotopes.” Metabolic engineering 9.5-6 (2007): 387-405.

Title: Computational Models for the Identification of Metabolic Vulnerabilities in Multiple Myeloma

Author: Luis V. Valcárcel[1], Raquel Ordoñez[2], Iñigo Apaolaza[3], Ana Valcárcel[4], Leire Garate[5], Cem Meydan[6], Ari Melnick[7], Jesús San Miguel[8], Xabier Agirre[9], Felipe Prósper[10], Francisco J. Planes[11]

Primary Affiliation: [1 3 11]Tecnun School of Engineering, University of Navarra, San Sebastian, Spain; [2 4 5 9 10] Hemato-Oncology Program, Centro de Investigación Médica Aplicada, IDISNA, University of Navarra, Pamplona, Spain; [6 7] Tecnun Division of Hematology/Oncology, Dept. of Medicine, Weill Cornell Medical College, New York, NY; [8] Hematology Department, Clínica Universidad de Navarra, University of Navarra, Pamplona, Spain

Abstract

The concept of genetic Minimal Cut Sets (gMCSs) [1] has made possible the search of possible combinations of two or more genes which, when simultaneously knocked-out, render cell proliferation impossible, defining synthetic lethal genes. Thanks to the gMCS and its integration with gene expression profiling experiments, it is also possible to uncover essential genes without previous metabolic network contextualization.

In this study, first, we used the latest reconstruction of the human metabolism, Human1[1], to enumerate up to 160,000 gMCSs linked to 57 essential metabolic tasks, including biomass production. Secondly, we obtained three RNAseq data sources for multiple myeloma (MM): 1) the MMRF CoMMpass cohort, containing 615 MM patients, 2) our cohort of 35 samples from different subpopulations of B cells and plasma cells of 37 MM patient samples and 3) CCLE data from 7 MM cell lines. Third, in order to identify possible metabolic vulnerabilities and essential genes, we projected the RNAseq into the gMCSs. We filtered the resulting essential genes by restricting the number of samples of healthy tissue where they are predicted to be essential and maximizing the number of MM samples and cell lines. Our results predict 5 genes to be essential, which are involved in 6 different gMCSs that explain synthetic lethality in MM patients.

References

[1] Apaolaza, Iñigo, et al. “An in-silico approach to predict and exploit synthetic lethality in cancer metabolism.” *Nature communications* 8.1 (2017): 459.

Title: Modelling of Glucose Recycling Through Storage Metabolism in Saccharomyces cerevisiae Under Dynamic Substrate Conditions

Author: Koen J.A. Verhagen[1], Ana Helena Valdeira Caetano[1], David Lao Martil[2], Siem A. Eerden[1], Ilse H. Pardijs[1], Baukje J. Sikkema[1] and S. Aljoscha Wahl[1]

Primary Affiliation: [1]Delft University of Technology, the Netherlands; [2]Eindhoven University of Technology, the Netherlands

Introduction

Natural and industrial, large -scale cultivations dynamic: Substrate availability, temperature and pH can rapidly change in time. Such extracellular changes have an immediate impact on the cellular metabolism. Previous studies [1][2] showed that a fast increase in substrate availability trigger different metabolic response mechanisms in *Saccharomyces cerevisiae*. Key control mechanisms were related to the trehalose cycle. In fact, the simultaneous production and degradation of this carbohydrate is an important glycolytic "safety valve" for regulating the glycolytic flux [3] during the transition. During a substrate pulse, up to 28% of the substrate entered the trehalose cycle. To study whether such high fluxes are also observed in cells experiencing repetitive perturbations an experimental setup using a fast, repetitive, block-wise feeding regime was used (feast/famine regime).

Results

The feast/famine regime leads to repetitive dynamics conditions, allowing for accurate measurement [2]. Compared to a chemostat-grown culture, a culture grown in a 400 second feast/famine regime shows a different metabolic response to a substrate pulse, with 17% of the consumed glucose being recycled through the trehalose cycle compared to 28% during a pulse[3].

In this study, the adaptation to these repetitive substrate concentration changes was elucidated using a dynamic kinetic model of yeast central carbon metabolism based on the glycolysis model by Teusink [4] by evaluating changing in parameters from steady-state, pulse and feast/famine substrate conditions. Targets for parameter estimation were chosen using a minimal proteome-dependent kinetic model. This model was used to predict expected proteome composition at both steady-state and feast/famine conditions. Significant predicted changes in enzyme concentration were subsequently used as targets for parameters estimation in the larger dynamic kinetic model. The model was fitted against both steady-state, pulse and feast/famine metabolomics and proteomics datasets, as well as 13C labelling data.

A framework of optimization methods was utilized, which includes parameter estimation with combinations of parameters and combinations of weights assigned to the variables used for data-fitting. With these, the parameters of glycolysis and the trehalose cycle were estimated and a final parameter set was obtained. Changes in parameters of HXK/GLK and TDH were shown to be instrumental in fitting the feast/famine conditions compared to steady state/pulse conditions, which were also changes observed from proteome measurements. Putative transport and compartmentation of trehalose in the vacuole was shown to be essential for fitting the observed fluxes through the trehalose cycle with the measured intracellular trehalose concentration. For the current state of this model, its output presents strong correlation coefficients between the enzymatic fluxes of trehalose cycle simulated and observed experimentally.

Outlook

Compartmentation of trehalose was shown to be essential in fitting both fluxes and concentrations in the trehalose cycle under dynamic substrate conditions. However, fitting of the trehalose degradation fluxes shows that additional kinetic interactions, such as post-translational modifications, need to be implemented in this model to fully understand the dynamic metabolic response of *S.cerevisiae.*

References

[1] M.R. Mashego et al., "In vivo kinetics with rapid perturbation experiments in *Saccharomyces cerevisiae* using a second generation BioScope", *Metab Eng*, 2006; 8: 370-383.

[2] C.A. Suarez-Mendez et al., "Metabolic adjustment upon repetitive substrate perturbations using dynamic 13C-tracing in yeast", *Microb Cell Fact*, 2017; 16: 161.

[3] J.H. van Heerden et al., "Lost in Transition: Start-Up of Glycolysis Yields Subpopulations of Nongrowing Cells", *Science*, 2014; 343: 987.

[4] B. Teusink et al., "Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry", *Eur J Biochem*, 2000; 267: 5313-5329.

Title: Stress Induced Cross-Feeding of Internal Metabolites Provides a Dynamic Mechanism of Microbial Cooperation

Author: Kapil Amarnath, Avaneesh V. Narla, Sammy Pontrelli, Jiajia Dong, Tolga Caglar, Brian R. Taylor, Julia Schwartzman, Uwe Sauer, Otto X. Cordero, Terence Hwa

Primary Affiliation: Department of Physics, U.C. San Diego, La Jolla, CA

Abstract

Despite the ubiquity of microbial diversity observed across environments, mechanisms of cooperativity that enable species coexistence beyond the classical limit of one-species-per-niche have been elusive. Here we report the observation of a transient but substantial cross-feeding of internal metabolites between two marine bacterial species under acid stress, and further establish through quantitative physiological characterization of the individual strains that this cross-feeding is central to the survival and coexistence of these species in growth-dilution cycles. The coculture self-organizes into a limit cycle in which acid-stressed producers excrete various internal metabolites upon entering growth arrest, enabling the cross-feeders to grow, restore medium pH, and protect the producers from death. These results establish a rare, mechanistic example of inter-species cooperation under stress, as anticipated long ago by the stress gradient hypothesis. As the accumulation of acetate and other fermentation products occurs ubiquitously in habitats ranging from the gut to wastewater, and the excretion of internal metabolites is an obligatory physiological response by bacteria under weak acid stress, stress-induced cross-feeding provides a general physiological basis for the extensive sharing of metabolic resources to promote the coexistence of diverse species in microbial communities.

Title: Large-scale metabolic reconstruction of the human microbiome accounting for strain-specific drug metabolism

Author: Almut Heinken, Geeta Acharya, Dmitry A. Ravcheev, Johannes Hertel, Malgorzata Nyga, Onyedika Emmanuel Okpala, Marcus Hogan, Stefania Magnusdottir, Filippo Martinelli, German Andres Preciat Gonzalez, Ronan M.T. Fleming, and Ines Thiele

Primary Affiliation: School of Medicine, National University of Galway, Galway, Ireland

Introduction

The human microbiome influences the efficacy and safety of a wide variety of commonly prescribed drugs, yet comprehensive systems-level approaches to interrogate drug-microbiome interactions are lacking. Previously, we have published AGORA, a resource of semi-automatically curated genome-scale reconstructions of human gut microbes [1]. AGORA has been applied to modelling metabolic host-microbiome interactions in a variety of studies [2].

Methods

We reconstructed AGORA2, an expansion of AGORA in both size and scope, accounting for 7,206 strains, 1,645 species, and 24 phyla [3]. Like its predecessor, AGORA2 was extensively curated in a data-driven manner based on comparative genomics and literature searches.

Results

AGORA2 accounts for a microbial drug metabolism module that was manually formulated based on comparative genomics for over 5,000 strains and drug metabolite structures from literature searches and databases. We take advantage of this by mechanistically modelling the individual-specific drug conversion potential in a cohort of 616 Japanese colorectal cancer patients and controls. This analysis reveals that both qualitative and quantitative drug conversion potential is individual-specific, for instance, the capacity to perform azoreduction of prodrugs as well as digoxin reduction are present in only a subset of microbiomes. Moreover, the drug conversion potential of specific drugs correlates with clinical parameters, such as age and BMI.

Discussion

We present AGORA2, a genome-scale reconstruction resource with unpreceded taxonomic coverage of the human microbiome that enables the creation and interrogation of personalised host-microbiome models. Hence, AGORA2 paves the way towards personalised, predictive analysis of host-drug-microbiome interactions.

References

[1] Magnusdottir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 35, 81-89, doi:10.1038/nbt.3703 (2017).

[2] Heinken, A., Basile, A. & Thiele, I. Advances in constraint-based modelling of microbial communities. Current Opinion in Systems Biology, doi:https://doi.org/10.1016/j.coisb.2021.05.007 (2021).

[3] Heinken, A. et al. AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities. bioRxiv, 2020.2011.2009.375451, doi:10.1101/2020.11.09.375451 (2020).

Title: Micromaps: Microbiome Metabolic Network Visualization Using Systematic Team Empowerment

Author: Cyrille C. Thinnes, Eoghan Courtney, Eoghan Culligan, Katie Fahy, Ruby Ferrazza, Ciara Ferris, Cristina Gonzales, Angeline Lagali, Rebecca Lane, Colm Maye, Olivia Murphy, David Noone, Jonatanas Pesaitis, Saoirse Ryan, Renee Waschkowitz, Almut Heinken and Ines Thiele

Primary Affiliation: School of Medicine & Ryan Institute, National University of Ireland Galway, Ireland

Abstract

MicroMaps (www.thielelab.eu/micromaps) was a 10-week summer undergraduate research program inviting 30 undergraduate STEM researchers across the Republic of Ireland to collaborate virtually on tackling systems biomedicine challenges during the 2020 COVID-19 pandemic. MicroMaps systematically involved small, diverse, and interdisciplinary teams to cumulatively craft a valuable learning experience for each other, including, project design, execution, and management, science communications, research symposia, and career development, to facilitate community access to computational biology and digital health research.

The Visualization Taskforce focused on representing microbiome metabolism in a Google map-like representation enabling, e.g., the visualization of flux-balance analysis results using the COBRA Toolbox (opencobra.github.io). Inspecting the AGORA2 resource of human microbial genome-scale reconstructions containing 7,206 strains [1], the team designed a visualization methodology based on the pillars of content, design, and quality control. The content of ~8,000 biochemical reactions was organized into ~150 reaction classes, further subdivided along molecular characteristics, such as, reactions primarily related to proteins, carbohydrates, lipids, nucleic acids, or xenobiotics. The design outcome amounted to adopting a decentralized visualization approach inspired by an urban cartography analogy including several interconnected hubs. Quality control led to the establishment of standard operating procedures to facilitate team integration while ensuring prudent and faithful visual translation of the AGORA2 source content. Over the MicroMaps duration, the taskforce arrived at visualizing ~50% of AGORA2-contained reactions using the CellDesigner software (www.celldesigner.org), enabling graphical notation through Systems Biology Markup Language (SBML). Future steps will include comprehensive AGORA2 reaction implementation and open-access release to the scientific community on the Virtual Metabolic Human website (www.vmh.life), also enabling interfacing with the COBRA Toolbox.

In conclusion, MicroMaps provided a successful use case for systematic team empowerment to drive systems biomedicine research in a fully virtual remote-working environment, amounting to the creation of a biochemical network representation of microbiome metabolism.

Reference

[1] A. Heinken et al., "AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities", bioRxiv 2020.11.09.375451.

Title: Optimal resource allocation and the limits of phototrophic productivity

Author: Ralf Steuer

Primary Affiliation: Humboldt-University of Berlin, Berlin, Germany

Introduction

Phototrophic microorganisms allow to couple light-driven assimilation of atmospheric CO2 directly to the synthesis of carbon-based products, and are therefore attractive platforms for the synthesis of renewable raw materials, from animal feed to biofuels. The economic viability of phototrophic cultivation, however, also crucially depends on photobioreactor design and culture parameters. Our aim is to investigate the limits of phototrophic productivity using models of cellular resource allocation.

To this end, we integrate a coarse-grained model of cyanobacterial growth into a light-limited chemostat and its heterogeneous light gradient induced by self-shading of cells. Different from previous models based on phenomenological growth equations, our model provides a mechanistic link between intracellular protein allocation, population growth and the resulting culture productivity. Our results suggest novel strategies to maximize culture productivity with light as the limiting nutrient. Some of the resuts are highly non-intuitive and illustrate the utility of current metabolic resource allocation models to support green biotechnology.

References

[1] A Faizi M and Steuer R (2019) Optimal proteome allocation strategies for phototrophic growth in a light-limited chemostat. Microb Cell Fact 18, 165.

[2] Sharma S and Steuer R (2019) Modeling microbial communities using biochemical resource allocation analysis. J. R. Soc. Interface 16: 20190474.

[3] Tomáš Zavřel, Marjan Faizi, Cristina Loureiro, Gereon Poschmann, Kai Stühler, Maria Sinetova, Anna Zorina. Ralf Steuer, Jan Červený (2019) Quantitative insights into the cyanobacterial cell economy. eLife 2019;8:e42508.