Title: Modeling Bacterial Metabolism and Expression to Develop Biocontainment Strategies

Authors: Juan D. Tibocha-Bonilla[1]* (j1tiboch@eng.ucsd.edu), Rodrigo Santibanez[2]* (rsantibanezpalominos@health.ucsd.edu), Blake Estefan,[3] Brian Arianpour,[3] Kanksha Patel,[3] Cristal Zuniga,[2] **Karsten Zengler**[2,3,4], and **Michael Guarnieri**[5]

Institutions: [1]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA; [2]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0760, USA; [3]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA; [4]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0436, USA; [5]National Renewable Energy Laboratory, Golden, CO, USA.

Website URL: https://genomicscience.energy.gov/research/sfas/nrelimagine.shtml

Project Goals:

To develop computational tools for the predictive design of biocontainment strategies with enhanced stability and resilience in diverse microbial hosts while maintaining maximal fitness and bioproductivity of the engineered microbial strains. To do so, we are developing genome-scale metabolic and expression models (ME-models) for a selected group of industrially relevant bacteria and exhaustively reviewing developed metabolic models (M-models) to employ them. We will 1) predict the impact of biocontainment strategies in the fitness and productivity of industrial strains, 2) develop community metabolic models (CM-models) to determine possible microbe-microbe interactions and the subsequent capability of a native ecosystem to support the growth of an “escaped” industrial strain, and 3) develop novel biocontaiment strategies based on a conditional metabolism that could not be rescued on natural ecosystems.

Abstract Text:

Genome-scale metabolic models (GEMs) of industrial relevant bacteria are largely employed to determine genetic interventions for optimizing the production of a target metabolite. However, strategies for the biocontainment and eliminating the impact on native ecosystems if they escape from an industrial environment are lacking behind. Here, we present manually validated and curated GEMs of four industrially relevant bacteria and the first ME-model for *Pseudomonas putida* KT2440 (iJT1667-ME). To do so, we developed a bioinformatics pipeline to validate and curate those and other 25 published bacterial GEMs and obtain M-models having identical identifiers to be used in CM-models, and we established a methodology around the COBRAme software to develop bacterial ME-models.

The validation and curation pipeline obtains and combines the annotations derived from the KOFAM and the InterProScan software with the annotation of the best protein sequence

alignment against the UniProt, KEGG, Prokka, and TCDB databases. Our UniProt database contains all protein sequences from the SwissProt database and a filtered TrEMBL database, keeping only protein sequences with a "catalytic activity" annotation and removing sequences with a "fragment" annotation. Similarly, we keep only protein sequences larger than 11 amino acids and that are annotated with an ontology number from the KEGG database. The resulting UniProt and KEGG databases contain 24.1 and 18.2 million sequences, respectively, covering 10.7% and 52.1% of all proteins in the original databases. After bioinformatics analyses, we manually reviewed the annotations and assigned a "high-confidence" validation score when the annotation agreed and coincide with the reaction consulting BRENDA, KEGG, IntEnz, and RHEA databases. Finally, the GEM is curated and updated to accurately reflect information from databases and publications. In addition, genes that were not validated were discarded from the model.

We successfully designed a methodology to process data available in diverse databases, such as BioCyc, to COBRAme-ready files to aid ME-model reconstruction. This approach allowed for the rapid development of the first ME-model for *P. putida* KT2440. The *i*JT1667-ME model consists of 7,110 metabolites, 10,955 reactions, and 1667 genes, covering 30.8% of the 5,419 proteins of *P. putida* KT2440 proteome. *i*JT1667-ME includes a curated metabolic network of the *i*JN1462 model with an additional manual curation to incorporate ion and cofactor transport reactions, as well as resolution on the strain-specific gene expression machinery that resulted in the incorporation of 205 additional genes into the reconstruction. Simulations showed the correct minimization of metabolic loops, such as the ATP:polyphosphate phosphotransferase and L-lysine:pyruvate aminotransferase reactions, and significant differences in the prediction of metal ion transport, NAD synthesis, and pentose phosphate pathway fluxes compared to *i*JN1462. Next steps will involve the modeling of biocontainment strategies to determine *in silico* fitness and productivity.

Funding Statement: This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area (SFA) contract number DE-AC36-08GO28308.