^{13}C-assisted flux elucidation using genome-scale carbon mapping models

Saratram Gopalakrishnan[1*] and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA

Project Goals: This project aims to expand existing carbon mapping models from core models to genome-scale models for flux elucidation using 13C-MFA using tools, algorithms and reaction information contained within the MetRxn database. The constructed mapping models will be deployed for genome-scale flux elucidation to obtain insights into the impact of model scale-up and loss of information, sensitivity of flux distributions to biomass composition, and novel carbon backbone scrambling patterns and pathway usage in cyanobacteria.

Metabolic models used in ^{13}C metabolic flux analysis generally include a limited number of reactions primarily from central metabolism. They typically omit degradation pathways, complete cofactor balances, and atom transition contributions for reactions outside central metabolism. This study addresses the impact on prediction fidelity of scaling-up bacterial and cyanobacterial mapping models to a genome-scale. The core mapping model for *E. coli* employed in this study accounts for 75 reactions and 65 metabolites primarily from central metabolism. The genome-scale metabolic mapping model (GSMM) (*i*mEco726, 668 reaction and 566 metabolites) is constructed using as a basis the *i*AF1260 model upon eliminating reactions guaranteed not to carry flux based on growth and fermentation data for a minimal glucose growth medium. This GSMM model identifies all peripheral metabolic pathways contributing to small metabolite recycling, alternate routes to lower glycolysis and the TCA cycle, multiple transhydrogenase mechanisms. The cyanobacterial mapping models *i*mSyn711 and *i*mSyf608 detail reaction atom mapping for all carbon-balanced reactions in cyanobacteria *Synechocystis* PCC 6803 and *Synechococcus* PCC 7942. 221 unique cyanobacterial reactions contribute to 67 novel carbon paths identified using an EMU-based depth-first search algorithm spanning Calvin cycle, photorespiration, an expanded glyoxylate metabolism, and corrinoid biosynthetic pathways. Differences in metabolism around the 3-phosphoglycerate (3PG) metabolic node and glyceraldehyde-3-phosphate (G3P) are flagged as the primary source of carbon scrambling patterns unique to the cyanobacteria with acetate recycling from porphyrin biosynthesis as a secondary source. The interplay between enzymes of the non-oxidative pentose phosphate pathway and carbon fixation via the Calvin cycle, results in rearrangements of C1 and C2 carbons of the pentose phosphates allowing for the transfer of fixed ^{13}C-CO_2 to all atoms of G3P and 3PG metabolites. In contrast, in *E. coli* carbon scrambling occurs only in upper glycolysis, pentose phosphate pathway, and the Entner-Doudoroff pathways. Furthermore, cyanobacterial mapping models reveal the presence of an additional ^{13}C incorporation path via glyoxylate metabolism and photorespiration contributing to three unique carbon arrangement patterns on triose phosphates, absent in *E. coli*. Labeling data for 17 amino acid fragments obtained from cells fed with glucose labeled at the second carbon was used to obtain fluxes and ranges for E. coli, whereas labeling distributions of 15 central metabolites obtained via ^{13}C-labeled bicarbonate are used for flux and range elucidation in *Synechocystis*. Metabolic fluxes and confidence intervals are estimated, for both core and genome-scale mapping models, by minimizing the sum of square of differences between predicted and experimentally measured isotopic steady-state labeling distributions for *E. coli*, and isotope labeling dynamics for *Synechocystis*.

Overall, we find that both the topology and estimated values of the metabolic fluxes remain largely consistent between the core and GSMM models for *E. coli*. Stepping up to a genome-scale mapping model leads to wider flux inference ranges for 20 key reactions present in the core model. The glycolysis flux range doubles due to the possibility of active gluconeogenesis, the TCA flux range expanded by 80% due to the availability of a bypass through arginine consistent with labeling data, and the transhydrogenase reaction flux was essentially unresolved due to the presence of as many as five routes for the inter-conversion of NADPH to NADH afforded by the genome-scale model. By globally accounting for ATP demands in the GSMM model the unused ATP decreased drastically with the lower bound matching the maintenance ATP requirement. A non-zero flux for the arginine degradation pathway was identified to meet biomass precursor demands as detailed in the iAF1260 model. Inferred ranges for 81% of the reactions in the genome-scale metabolic (GSM) model varied less than one-tenth of the basis glucose uptake rate (95% confidence test). This is because as many as 411 reactions in the GSM are growth coupled meaning that the single measurement of biomass formation rate locks the reaction flux values. This implies that accurate biomass formation rate and composition are critical for resolving metabolic fluxes away from central metabolism and suggests the importance of biomass composition (re)assessment under different genetic and environmental backgrounds. In addition to better recapitulation of experimentally observed labeling distributions of all measured central metabolites, flux elucidation using the cyanobacterial mapping model predicts the existence of a serine biosynthesis route from 3PG and trace flux through the GABA shunt. The loss of information associated with mapping fluxes from MFA on a core model to a GSM model is quantified and its implications on inferences drawn on the metabolic capabilities of E. coli and cyanobacteria are discussed.

Supported by funding from the U.S. Department of Energy by grants DE-FG02-05ER25684 and DE-SC0012722.