

MetaFlux in Pathway Tools (Short Tutorial)

Mario Latendresse

SRI International

22 September 2015

Outline

- 1 Introduction to Flux Balance Analysis**
 - What is Flux Balance Analysis (FBA)?
 - Overview of MetaFlux
- 2 Introduction to Development Mode**
 - Single and Multiple Gap-Filling
 - User Input: Fixed and Try Sets, Weights
- 3 Genes and Reactions Knockout**

What is a Flux of a Reaction?

- The flux of a reaction is the rate of substrates consumed and produced, typically in steady state during growth
- Fluxes are typically expressed as mmol per gram dry weight per hour, denoted mmol/gDW/hr
- Solving an FBA model gives the fluxes of all reactions that are needed to create a non-zero flux for the biomass (set of metabolites necessary for growth)

What is Flux Balance Analysis (FBA)?

Metabolic Modeling Applications

Metabolic Modeling Applications

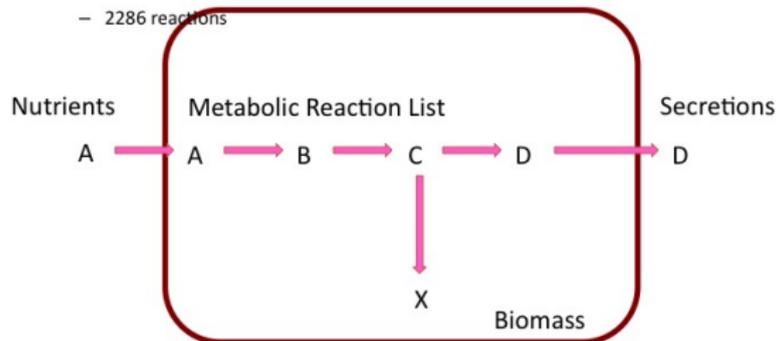
- Identify gaps in genome annotation
- Predict steady-state reaction fluxes for the metabolic network
- Predict growth rates, nutrient uptake rates

- Remove genes/reactions from model to predict knock-out phenotypes
 - EcoCyc model predicts knock-out phenotypes with 95.2% accuracy
- Supply alternative nutrient sets to predict growth phenotypes
 - EcoCyc model predicts growth under 431 growth media with 80.7% accuracy

- Metabolic engineering

Flux-Balance Analysis

- Steady state, constraint-based quantitative models of metabolism
- *E. coli* model derived from EcoCyc (*BMC Sys Biol* 2014 8:79):
 - 16 nutrients
 - 108 biomass metabolites
 - 2286 reactions



Flux Balance Analysis

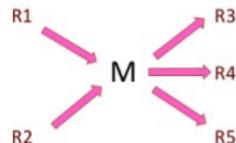
- Define system of linear equations encoding fluxes on each metabolite M

$$- R1 + R2 = R3 + R4 + R5$$

- Boundary reactions:

- Exchange fluxes for nutrients and secretions

- Biomass reaction $\text{L-arginine ... + GTP ... + ...} \rightarrow \text{biomass}$



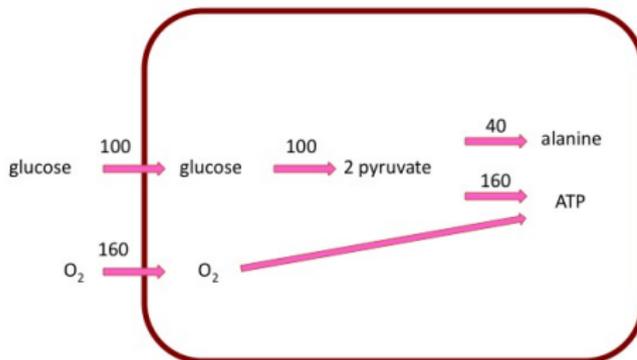
- Submit to linear optimization package
 - Optimize biomass production
 - Optimize ATP production
 - Optimize production of desired end product

What is Flux Balance Analysis (FBA)?

Simple Example of Fluxes

Example

Biomass: ATP:alanine, 4:1



The FBA Tool in Pathway Tools

- 1 The FBA Tool, MetaFlux, was introduced in version 15.0 of Pathway Tools (Feb 2011)
- 2 MetaFlux has three modes: solving, development, and gene knockout
- 3 Solving mode: compute the fluxes of reactions to produce the biomass
- 4 Development mode: trying different biomass, nutrients, secretions, and reactions to create a model
- 5 Gene knockout: deactivating gene(s) from the model and see the effect on growth (testing a model)
- 6 In version 19.5, dynamic FBA with community of organisms

Assumptions

MetaFlux Main Assumptions

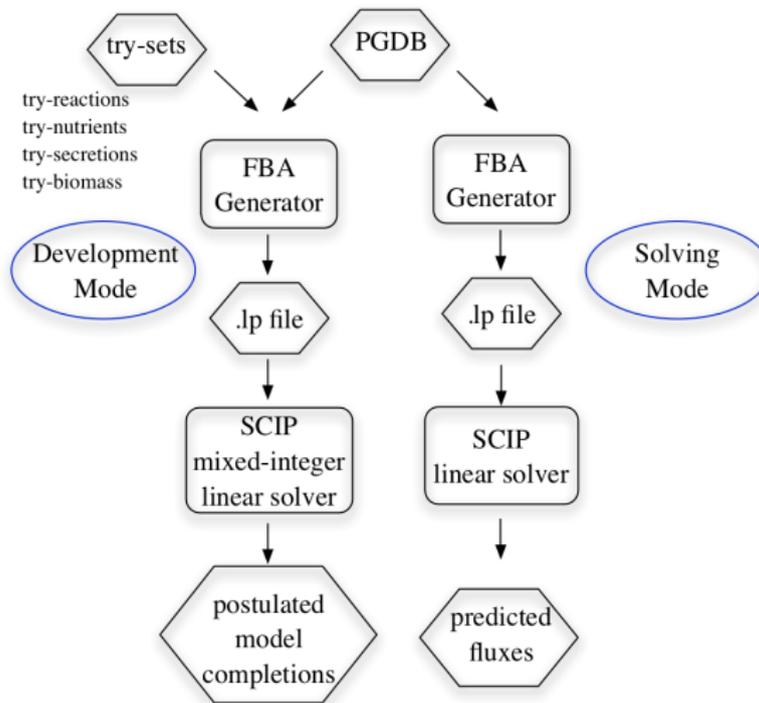
- Regulation is ignored
- Cofactors are ignored
- Compartments are not completely taken care of
- Some transport reactions must be explicitly specified (e.g., ATP synthase)

MetaFlux Modes of Operation

MetaFlux Modeling Tool: Modes of Operation

- **Solving mode**
 - Individual organisms, organism communities
 - Steady-state FBA, dynamic FBA
 - Single compartment, 2-D spatial grid with diffusion
 - Cellular-compartment aware
 - Removal of flux loops, inference of biomass reaction
- **Knock-out mode** (single/double gene/reaction knock-outs)
- **Model development mode**
 - Development mode (multiple gap filling)
 - Fast Development mode (reaction gap filling) [Latendresse 2014]
 - Identify dead-end metabolites and blocked reactions

Creating an FBA Model vs Solving an FBA Model



The Linear Programming (LP) Solver

- Solving a model is done by a Linear Programming (LP) solver
- There are many open source and commercial LP solvers: CPLEX, GLPK, SCIP, Gurobi, and more
- Pathway Tools uses SCIP
- Even with thousands of reactions, typical FBA/LP formulation can be solved in a few seconds
- Solving in development mode could be longer (e.g., 20 seconds) or much longer (e.g., one hour)
- MetaFlux is available on Mac OS X, Linux and Windows

Painting Fluxes on Metabolic Map

Painting *E. coli* Fluxes on Metabolic Map

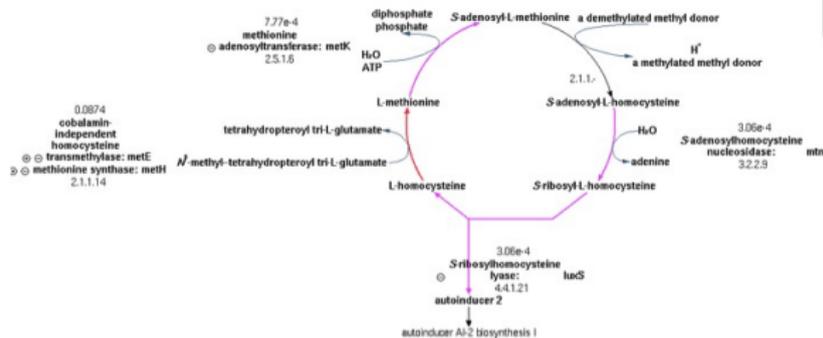


Fluxes on Pathway Diagram

E. coli Fluxes on Pathway Diagram

Escherichia coli K-12 substr. MG1655 Pathway: S-adenosyl-L-methionine cycle I 

[More Detail](#) [Less Detail](#) [Species Comparison](#)



Single and Multiple Gap-Filling

Typically "Gap-Filling" Means "Completing the Reaction Network"

- Gap-filling adds reactions from a reference database (e.g., MetaCyc) to the FBA model to produce missing biomass
- Model might still be infeasible due to a lack of reactions in MetaCyc, or lack of nutrients, or secretions

Solution: Gap-Filling Extended to Important Metabolites

Nutrients, secretions, and biomass metabolites can also be added or removed. For biomass metabolites, we try to include as many as possible while still getting a feasible solution.

Multiple Gap-Filling

Multiple Gap-Filling

Multiple gap-filling is done on reactions, nutrients, secretions, and biomass metabolites **at the same time**.

Objective

Try to add as many biomass metabolites as possible by adding a minimum number of nutrients, secretions, and reactions; and still get a feasible solution.

Usage

Speeds curation of a PGDB. It is a technique to complete a PGDB to do standard FBA analysis.

Fixed Sets for Multiple Gap-filling

The user provides fixed sets of reactions and metabolites “at no cost or gain”.

- Set of fixed reactions to use at no cost: typically all metabolic reactions of the PGDB are used
- Sets of nutrient and secreted metabolites that can be used at no cost
- But, it is recommended to start with an empty set of fixed biomass metabolites

Try-Sets and Weights for Multiple Gap-filling

The user provides four try-sets and weights to control the generation of the model.

- Set of reactions to try to add at a cost: typically all metabolic reactions of MetaCyc
- Sets of nutrients, secretions and biomass metabolites to try to add to the model
- Weights, as integers for gain and cost, for the reactions, nutrients, secretions and biomass metabolites
- Typically, adding a biomass metabolite is a gain, but adding a reaction or a nutrient is a cost. We have different weights for different type of reactions (e.g., spontaneous, in the taxonomic range, etc.)

User Input: Fixed and Try Sets, Weights

The Weights: Costs and Gains

Typical Weights

- Adding a biomass metabolite to the model is a **gain**.
- Adding any reaction, secretion, or nutrient has a **cost**.
- That corresponds to the usual goal: generating as many biomass metabolites as possible with the minimum number of nutrients, secretions, and added reactions

Variations

But other scenarios are useful: use as many nutrients and secretions as possible

Selecting the Right Weights for Reactions

There are many different weights for the reactions: taxonomic range, reversed, and more

User Input: Fixed and Try Sets, Weights

The Reaction Weights

- The basic weight for a reaction from MetaCyc
 - **outside the taxonomic range** of the PGDB is given by
`try-reactions-weight`
 - **in the taxonomic range** of the PGDB is given by
`try-reactions-in-taxa-weight`
 - **of unknown taxonomic range** is given by
`try-reactions-unknown-taxa-weight`

Testing a Model Using Genes Knockout

Knocking Out One Gene

- Knocking out a gene means to deactivate the reactions catalyzed by that gene
- Isozymes are taken into account

Multiple Knockouts

More than one gene might be knocked out simultaneously

Batch Knockouts

Typically, MetaFlux is used to run a batch of gene knockouts (e.g., all genes)

Examples of Knockout FBA Input Files

- Examples of gene knockout run on EcoCyc for
 - 1 A few genes: cysN, cysD, gltX
 - 2 All metabolic genes with summary solution file only (takes about one minute)
 - 3 All metabolic genes with all solution files generated (takes more than one minute)

Dynamic FBA (dFBA)

- Dynamic FBA (dFBA) goes beyond the steady state of FBA by solving a series of FBAs
- Each FBA solved changes the growth environment by keeping tracks of nutrients used and secretions produced
- dFBA replicates real growth with concentrations of metabolites in the growth environment
- Will be available in version 19.5 (November 2015)

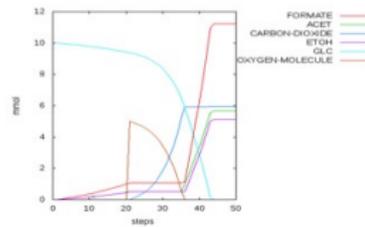
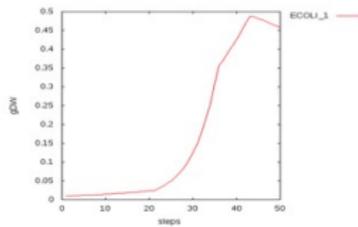
Dynamic FBA of community of Organisms

- A community of organisms can be represented by a set of FBAs exchanging metabolites in a compartment (e.g., extra-cellular)
- The organisms are in a real physical space represented as a grid
- The FBA of each organism is solved and the concentrations of secretions and nutrients are updated in each grid box
- When solving a FBA, the nutrient uptakes are bounded by the concentrations of metabolites

Dynamic FBA of *E. coli*

Dynamic FBA Modeling of *E. coli*

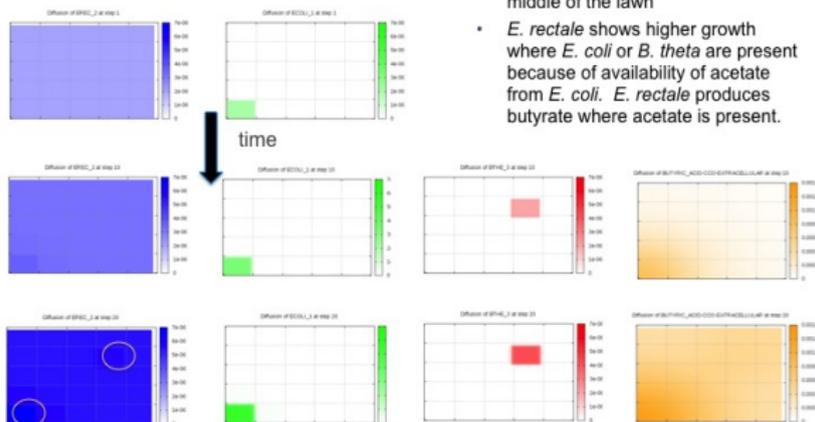
- Dynamic FBA modeling of *E. coli* growth under varying nutrient conditions
 - t=1-20: *E. coli* grows anaerobically on 10 mmol glucose
 - t=21-34: O₂ is added to the simulation; *E. coli* grows completely aerobically
 - t=34-35: O₂ availability becomes limiting; acetate forms
 - t=36-44: O₂ is exhausted; anaerobic growth resumes
 - t=45 onwards: glucose is exhausted, cells begin to die



Dynamic FBA Community of Organisms

Dynamic Grid Modeling of a Simple Microbial Community

Inspired by Segre's COMETS



Paper on MetaFlux

Mario Latendresse, Markus Krummenacker, Miles Trupp, and Peter Karp, *Construction and Completion of Flux-Balance Models from Pathway Databases*, Bioinformatics, doi: 10.1093/bioinformatics/btr681, January 2012. Open access