Multiple Gap-Filling of Flux-Balance Models

Mario Latendresse
and
Peter D. Karp

Bioinformatics Research Group
SRI International, Menlo Park, USA

COBRA 2011, Reykjavik, Iceland, June 25

Outline

Background

Many Pathway/Genome Databases but Few FBA Models

The number of pathway/genome databases is way above the number of FBA models

Time to Create a FBA Model

Developing one FBA model is time consuming (1-2 years, Thiele 2010)

Observation

"Automatic generation" of FBA models is not (yet) a complete answer. For example, gap-filling of reactions might suggest to add inappropriate reactions

Approach

- Accelerate curator development of FBA models
- Multiple gap-filling suggests model modifications
- Integration of FBA with Pathway Tools
  - Reaction balancing
  - Dead-end metabolite finder
  - Query and visualization of metabolic network
  - Visualization of fluxes on a metabolic map diagram

A Common Problem: infeasible FBA models

- An FBA/LP formulation requires that all **biomass** metabolites be produced
- Too many metabolites in the biomass reaction creates an infeasible model
- This is due to missing reactions, nutrients, secretions, or a combination of these
- It is better to offer the best **feasible** solution given a set of biomass metabolites, nutrients, and secretions to **try**
- **Typically, multiple Gap-filling proposes model modifications of minimum cost on reactions (and nutrients/secretions) and maximum gain on biomass**

Input: Fixed Sets and Try-Sets

Fixed Sets

Four fixed sets: reactions from organism, biomass metabolites, nutrients, and secretions. This is a FBA model to complete. Several fixed sets could be empty!

Try-Sets

Four try-sets: reactions from MetaCyc (optional reversed reactions from organism), biomass metabolites, nutrients, and secretions

Mixing Fixed Sets and Try-Sets

In general, when generating a model, a part of it is known (fixed), another part needs to be completed from try-sets

Multiple Gap-Filling

Multiple Gap-Filling

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

Typical 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

It is a technique to complete a PGDB to do standard FBA analysis

Our multiple gap-filling extends the reaction gap-filling idea developed by Costas Maranas

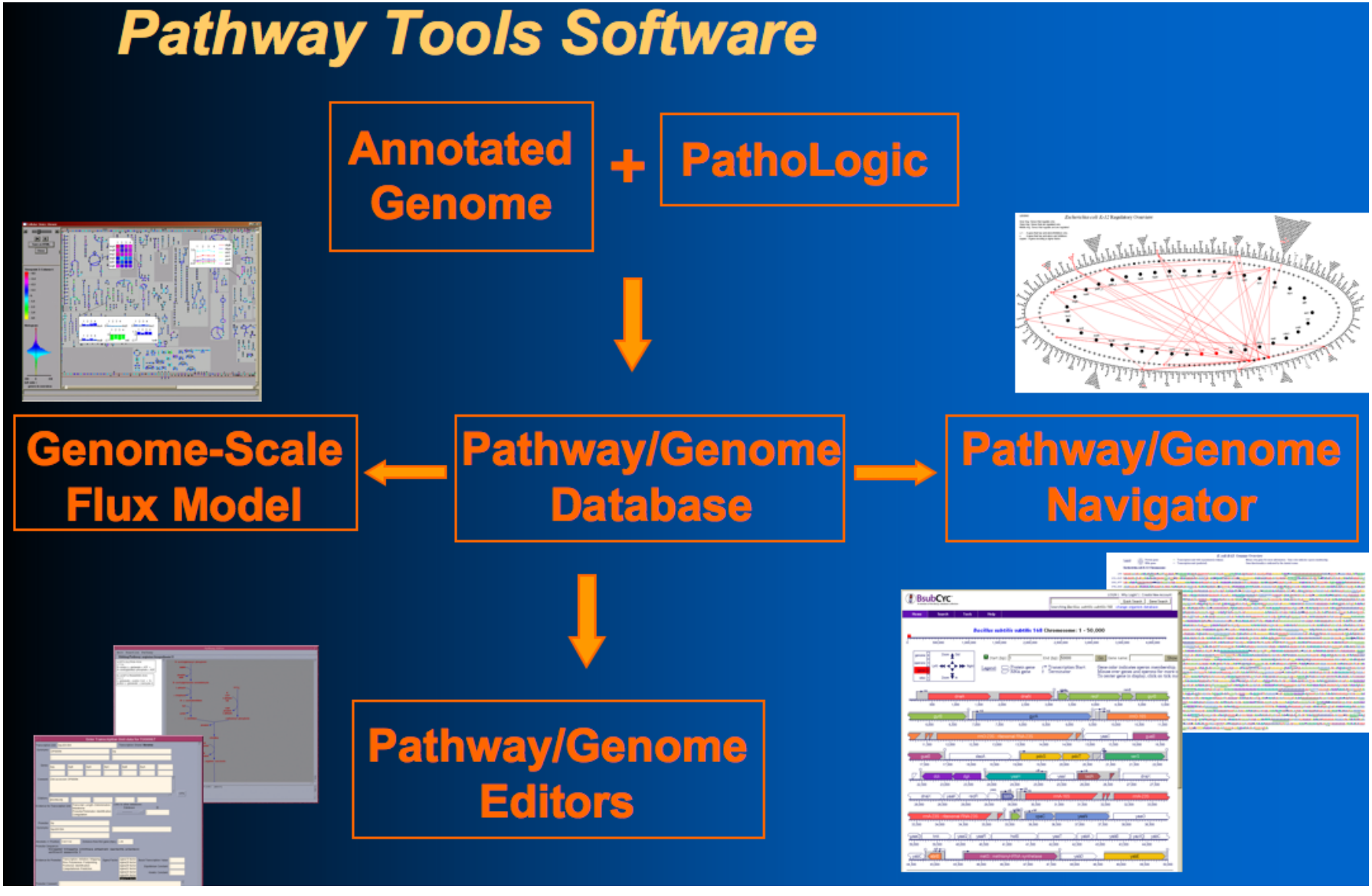
Mixed-integer Linear Programming (MILP) Formulation

- The nutrient, secretion, and biomass try-sets are converted to virtual try-reactions
- For try-biomass metabolites: **no major constraint as in the LP formulation**
- All fixed sets are fromulated as in the FBA/LP formulation
- Every (virtual or not) try-reaction flux r_i is controlled by a binary variable s_i:

$$0 \leq r_i \leq \mathbf{s}_i 1000$$

- The maximized objective function uses weights (cost and gain) on the s_i
- Weights take care of taxonomic range, reversing reactions, and more.

Pathway Tools



MetaFlux in Pathway Tools

- MetaFlux is a new module in version 15.0 of Pathway Tools (Feb 2011)
- MetaFlux has two modes: generating a model or solving a model
- First version: focused on providing a general tool to help create FBA models from pathway/genome databases (PGDBs)
- Future version: improve tool to generate model and help validate FBA models
- We are currently adding new functionalities for 15.5, in particular, gene knockout capability
- Pathway Tools (and MetaFlux) is available for download at `biocyc.org/download.shtml`

Four Files Generated

MetaFlux generates four files when generating/solving a model:

- A `.lp` file: the input to the solver (SCIP)
- A `.log` file: a trace of the output of the generation phase
- A `.sol` file: a summary of the solution and gap-filler actions
  - Fixed+try biomass metabolites produced
  - Fixed+try nutrients used
  - Fixed+try secretions produced
  - Reactions inserted and/or reversed
  - Fluxes for active reactions
- A `.dat` file: enables painting flux data on Pathway Tools Cellular Omics Viewer

Cellular Omics Viewer Example

First Step when Using MetaFlux

- Empty fixed-sets for biomass, nutrients, and secretions
- Specify as many try-biomass metabolites as possible
- No reactions to try (no gap-filling reaction at this step)
- Set the right weights to control the objective function to maximize the **number** of biomass metabolites
- MetaFlux will find the maximum set of biomass metabolites to produce suggesting no reactions to add
- Easier as a first step to figure out what the current reaction network can do

Iterative Steps when Using MetaFlux

- All biomass metabolites that can be produced in the first step are moved into the fixed set of biomass metabolites
- All metabolic reactions from MetaCyc are in the reaction try-set
- MetaFlux will find the maximum set of biomass metabolites to produce suggestinga minimum set of reactions to add
- This step can also be repeated by suggesting to add reversed reactions from the pathway/genome database and MetaCyc

MetaFlux applied to *Homo Sapiens* and *E. coli*

- MetaFlux was applied to HumanCyc and EcoCyc, two pathway/genome databases at biocyc.org
- HumanCyc: 53 biomass metabolites, 1720 reactions, 247 reactions carrying flux (about a month of effort)
- EcoCyc: 52 biomass metabolites, 1330 reactions, 371 reactions carrying flux (about a month of effort)
- For HumanCyc, the tool suggested many new reactions to add (from MetaCyc)
- EcoCyc is a well-curated database: new secretions were discovered, instantiation of reactions was improved

Acknowledgments

This is joint work with

Markus Krummenacker
(*E. coli*)

Miles Trupp
(*Homo Sapiens*)

This work was supported by NIH Grant GM080746

Questions

Thank You