Automated Generation of Metabolic Flux Models from PGDBs

Mario Latendresse, Markus Krummenacker, Peter Karp

Bioinformatics Research Group

SRI International

pkarp@ai.sri.com

Approach: FBA Model as a Database

- **Store and update metabolic model within Pathway Tools**
- **Export to constraint solver for model execution**

- **Fast generation of metabolic model from annotated genome**
- **Close coupling to genome and regulatory information**
- **Extensive PTools schema**
  - Associate a wealth of information with each model
  - Unique identifiers for each component of the model
- **Extensive query and visualization tools**
  - Visualize reaction flux and omics data using overviews
- **Debug/validate model using Pathway Tools**
  - Reachability analysis
  - Dead-end metabolite analysis
  - Visual inspection on cellular overview

Issues: Reaction Balance and Protonation

- **Reaction balancing**
  - Balance checked in reaction editor
  - Bulk balancing tool
  - Majority of MetaCyc reactions are balanced
  - Use MetaCyc update tool to propagate MetaCyc updates to your PGDB
- **Protonation**
  - Formerly our compound structures were protonated to inconsistent states
  - In MetaCyc 13.0 and forward, all structures are computationally protonated to cellular pH 7.3
    - Using Marvin software from ChemAxon
  - Reaction balances adjusted computationally by adding protons

Issues: Generic Reactions

- **Examples:**
- **1.1.1.182: NAD(P)+ + shikimate → NAD(P)H + 3-dehydroshikimate + H+**
- **1.3.99.3: a 2,3,4-saturated fatty acyl CoA + FAD → FADH2 + a 2,3-dehydroacyl-CoA**

- **Introduced many years ago to simplify descriptions of reactions and pathways**
- **Problem: Without special smarts, reactions involving instances and their classes are not connected within models**
  - a 2,3,4-saturated fatty acyl CoA
  - decanoyl-CoA

Generic Reactions: Solution

- **Generate instantiated reactions from generic reactions**
  - A + b → C
  - a1 + b → c1
  - a2 + b → c2
  - a3 + b → c3
- **Generate all reaction instantiations**
- **Prune those that are unbalanced**

Generation of FBA Models from PGDBs

- **Export PGDB to SBML**
  - (Thanks to Jeremy Zucker)
  - Coming soon: reaction instantiation
- **Export of PGDB to GLPK / CPLEX**

Export of PGDB to GLPK

- **A single Lisp function will:**
  - Generate a GLPK .lp file containing FBA constraints from
    - PGDB reactions
    - Supplied biomass components
    - Specified nutrients
    - Allowed waste products
    - Additional set of reactions to include or reject
  - Run GLPK on this file
  - Parse the GLPK output file
    - Determine if it found a solution
    - Generate another file mapping fluxes to PGDB reactions
  - Display the resulting fluxes on the Cellular Overview

Results: E. coli K-12

- ***E. coli*** **model generated from EcoCyc is solvable by GLPK**

- **Lipids are missing**
- **We have not yet verified magnitudes of fluxes**
- **Many reactions where fluxes appear are reasonable**
- **Flux is zero in unexpected places**
- **High fluxes are present in unexpected places**

Results: BioCyc Buchnera aphidicola

- **No solution found**
- **Search for largest subset of biomass components for which a non-zero flux can be found:**
  - 3 compounds found

Model Gap Filling

- Initially try gap filling on a partial E. coli model
- Full *E. coli* model F contains 1471 reactions

- Define base set B of 1,000 randomly chosen reactions from F
- Define extension set E of 471 remaining reactions of F
- Define optimization problem to GLPK to find minimal extension of B from E that yields non-zero solution
- GLPK found a set of 60 such reactions from E