Of MODS and Models: Predicting and Validating Phenotypes from Pathway Tools Metabolic Models

Peter D. Karp

Bioinformatics Research Group
SRI International
pkarp@ai.sri.com
Overview

- Pathway Tools approach to metabolic modeling
- What’s coming up for Pathway Tools
Literate Modeling
I believe that the time is ripe for significantly better documentation of programs, and that we can best achieve this by considering programs to be works of literature. Hence, my title: “Literate Programming.”

Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to humans what we want the computer to do.
Literate Programming

- Combined a programming language with a document preparation language
- The resulting hyper-document integrated a program with well-styled documentation
Literate Modeling

• Collaboration around models will be impossible if models are as inscrutable as most software

• => Models and model results must be
  • Readily understandable
  • Web browsable
  • Connected to the genome
  • Connected to pathways
  • Connected to the regulatory network
  • Connected to multiple online databases
  • Queryable
  • Accessible through graphical visualizations
  • Beautiful
“The Database is the Model”

- Marriage of models and databases

- Generate steady state metabolic flux models directly from Pathway/Genome Databases such as EcoCyc
  - To update the model, update the database
  - To browse the model, browse the database
  - To view model results, use database-generated viewers
From MODS to Models

- The evolution of Model Organism Databases
- SGD, MGI, FlyBase, WormBase, etc.
- EcoCyc as MMOD
**Perspective 1:**

*EcoCyc as Textual Review Article*

- All gene products for which experimental literature exists are curated with a minireview summary
  - 3,730 gene products contain summaries
  - Summaries cover function, interactions, mutant phenotypes, crystal structures, regulation, and more

- Additional summaries found in pages for operons, pathways

- EcoCyc data derived from 24,000 publications
Perspective 2: EcoCyc as Computational Symbolic Theory

- Highly structured, high-fidelity knowledge representation provides computable information
- Each molecular species defined as a DB object
  - Genes, proteins, small molecules
- Each molecular interaction defined as a DB object
  - Metabolic and transport reactions
  - Regulation of enzyme activity, gene expression
- 220 database fields capture object properties and relationships
Perspective 3: EcoCyc as Predictive Metabolic Model

- A steady-state quantitative model of *E. coli* metabolism can be generated from EcoCyc.
- Predicts phenotypes of *E. coli* knock-outs, and growth/no-growth of *E. coli* on different nutrients.
- Model is updated on each EcoCyc release.
- Serves as a quality check on the EcoCyc data.
Two Paradigms of Flux-Balance Modeling

- FBA models as spreadsheets / SBML
- FBA models derived from MODs
Approach: FBA Model as a Database

- Store and update metabolic model within Pathway Tools PGDB
- Export to constraint solver for model execution
- 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
  - Metabolites, reactions, pathways, growth media
  - Visualize reaction flux and omics data using overviews
FBA Model as a Database

- Also store within a PGDB the growth observation data needed to validate and refine a PGDB
Curation is Critical to Systems Biology

- Common curation effort for MOD and systems-biology models

- Biological models undergo long periods of updating and refinement
  - New information from literature
  - To improve consistency of predictions with experimental data

- Methodologies from MODs can benefit systems-biology models
  - Evidence codes
  - Mini-review summaries
  - Literature citations
Pathway Tools Approach to Metabolic Modeling

- Power tools to accelerate modeling

- Debug/validate model using Pathway Tools
  - Multiple gap filling
  - Dead-end metabolite analysis
  - Reaction balance checking

- Modeling support – ptools-support@ai.sri.com
**Pathway Tools Software**

Annotated Genome + PathoLogic

MetaFlux

Pathway/Genome Database

Pathway/Genome Editors

Pathway/Genome Navigator

*Briefings in Bioinformatics 11:40-79 2010*
SRI Modeling Projects

- EcoCyc model for *E. coli*
- HumanCyc model for *H. sapiens*
- YeastCyc model for *S. cerevisiae*
Recent Pathway Tools Enhancements

- Version 16.5
- Save display state
  - File -> Save Display State to File
- Atom mappings
- Chemical radicals
- EC number changes
- Web Groups enhancements
Coming Soon

- Version 17.0 in late March

- Pathway prediction
  - Pathway abundance score for metagenomic pathway prediction
  - Improvements to enzyme name matcher

- Pathway search tool for metabolic engineering

- Web omics pop-ups

- Groups improvements
Coming Soon

• **Version 17.0**

• **Internals**
  • New faster Web image generation
  • Web image persistence for better caching
  • New installer
  • Relational DBMS performance improvements
Coming

- Better handling of compartments and cell types

- Modeling improvements
  - Automatically run model across many growth conditions and knock-outs
  - Hypothesize model changes to rectify prediction errors
  - Expanded gap filling
  - FBA for microbial communities
  - FBA for eukaryotes
Prediction of alternative growth media for an organism from its PGDB

- Method predicts 787 alternative anaerobic media for *E. coli*
  - 72.5% accuracy for 91 media
- Automatically partitions nutrients into equivalence classes
- Algorithm starts with all transportable compounds and exhaustively considers all combinations of nutrients
- Can take months to run for *E. coli*
Coming

- Redesign / modernize Navigator interface
- Add sequence operations
- Performance / scalability improvements
Minimal Information about a PM Expt

- markus.goeker@dsmz.de