Basic Info

Project Title: Profondeur
Project Repository: https://github.com/tcameronwaller/profondeur
Author Name: Thomas Cameron Waller
Author EMail: cameron.waller@biochem.utah.edu
Author Identifier: 00409400

Overview and Motivation

The biological study of metabolism has potential to promote the domains of biotechnology, pharmacology, and medicine. Thereby, this study has potential to benefit human health. Metabolism is a complex system. The experimental study of metabolism involves large and complex data sets that are difficult to interpret. Interpretation requires consideration for the complex metabolic system. Prospective users are investigators in the domains of biology, biotechnology, pharmacology, and medicine who study metabolism. Users have very diverse backgrounds and very variable familiarity and expertise in metabolism, and data analysis.

Users need to access information about biological entities and the relations between them. This information gives relevant context for experimental design and interpretation of experimental results. This information is especially important for interpretation of large-scale data sets that cover the metabolic system broadly.

Related Work

KEGG

MetaCyc

Escher

MetScape

GeneMania

Questions

How is it possible to represent the metabolic network accurately and clearly using custom subsets of the network and automatic layouts of node-link graphs?

Data

Systems biology is a mature sub-domain of biology that considers and studies biological systems from an holistic perspective. This sub-domain commonly develops intricate mathematical models and uses these in computational simulations. These models include information about metabolites (the small molecules of metabolism), chemical reactions between them, the rates and directionality of these reactions, the cellular compartments where these reactions happen, the protein enzymes that catalyze

these reactions, and the transcripts and genes that encode these proteins. The ultimate goal is for these models to represent the entirety of the biological system, and many models are comprehensive collections of current biological knowledge.

An active community in systems biology develops and curates these metabolic models as well as develops and maintains extensive resources for using them. The Systems-Biology-Markup-Language (SBML) (http://sbml.org) relates to the Extensible-Markup-Language (XML) and is an open standard and format for the representation of computational models for biological systems. Robust models in SBML are available for the metabolic systems of many organisms, including *Mus musculus* (mouse) and *Homo sapiens* (human). For this project, I chose to derive data from these metabolic models because they represent the most comprehensive and orderly (mass and charge-balanced reactions, compartments) collections of current biological knowledge on the metabolic system.

COBRApy ((https://opencobra.github.io/cobrapy/)) is an open-source tool from the community for using metabolic models in Python. I will use COBRApy along with libSBML (http://sbml.org/Software/libSBML) to convert data for metabolic models from SBML format to JavaScript Object Notation (JSON) format. I will use these data in JSON format in the web application, ultimately restructuring it for graph queries and representation in node-link graphs. To pilot the prototype of this web application, I used a small model of central metabolism in *Escherichia coli* (Gram-negative bacterium) that came as a model for demonstration in COBRApy. I modified this model to simplify it and derived the appropriate structure of nodes and links manually.

Exploratory Data Analysis

I am familiar with the metabolic network and visual representations of it from KEGG. I have also drawn many of my own visual representations of the metabolic network using InkScape. I gathered information for these representations from databases such as KEGG, MetaCyc, and UniProt.

Design Evolution

This project's repository (https://github.com/tcameronwaller/profondeur) has a directory with title "design". Within this directory are scans of sketches that I drew or recorded during the design process. Also in this directory is a text file with title note sketch. This file contains notes on several important or influential sketches and the relevant design concepts.

Implementation

Evaluation