ABSTRACT

Network reconstruction has been a common denominator in systems biology. Taking advantage of the extensive characterization of E. coli, we reconstructed the metabolic network of E. coli K-12 MG1655 and examined its application for systems analysis. Specifically, we characterized the reconstruction, used the reconstruction for the computation of phenotypes and determined its use in interpreting high-throughput experimental data. This reconstruction is based on previous work [1,2] and additional information from published studies [4]. The reconstruction, iAF1255, is fully compartmentalized into three distinct cellular compartments and includes 1255 of the predicted 4578 ORFs in E. coli (27%). 2374 reactions, 1686 metabolites and thermodynamic estimates [6] for nearly all metabolites and reactions. The final reconstruction was generated through a collaborative effort with EcoCyc (SRI International, P.I., Peter Karp) and the Computational and Mathematical Biotechnology Laboratory (P.I., Vassili Hatziazmanikas). After converting the reconstruction to a computational model, we identified the important parameters needed for using the model as a predictive tool (using a constraint-based approach) and further performed a sensitivity analysis on these variables. Using the model, we analyzed the essential genes predicted under minimal media growth conditions [8] and identified failure modes that need further biochemical characterization and genetic analysis. Using flux variability analysis, we analyzed 174 different carbon source growth conditions and compared the resulting flux distributions to the estimated thermodynamic values. This work represents a significant enhancement from the previous work because of its expanded coverage of metabolism, inclusion of reaction and metabolite thermodynamic data and a thorough characterization of the variables that significantly affect modeling simulations.

THERMODYNAMIC ANALYSIS

Using the group contribution method and approximate intracellular metabolite activities, we estimated standard Gibbs free energy changes for each reaction in the reconstruction, iAF1255 [6].

IMPROVEMENTS IN IAF1255

(A) Coverage of ORFs from each of the clusters of orthologous groups (COGs) functional classes included in iAF1255 and five previous reconstructions. The total number of ORFs in each functional class from the E. coli genome [4] is shown.

(B) The number of reactions that are associated to ORFs from each COG functional class. Since ORFs can belong to multiple classes, the percentage of each class is listed.

CONVERSION TO PREDICTIVE MODEL

Method used to convert the reconstruction to a predictive model

(A) Represent the reconstruction in a computational network

(B) Constraining reactions that correspond to genes that are not transcribed under specific conditions

(C) Cap the maximum P/O ratio of high-energy phosphate bonds per mol of oxygen of the electron transport chain (ETC) by using observations and predictions from previous studies

(D) Estimated thermodynamic values used to predict growth rates for WT E. coli growth under different substrate conditions

HIGH-THROUGHPUT GENE ESSENTIALITY

We used iAF1255 as a framework to analyze the essential ORF predictions of E. coli K-12 [3,5] and compared these to computational predictions.

RESULTS

• Essential ORFs from amino acid (E) and nucleotide (F) metabolism are essential under minimal media conditions; cell wall (M) and lipid (L) synthesis ORFs are essential under rich media conditions

• Coenzymes (H) metabolism ORF essentiality mostly at the level of rich media conditions, some cofactors can not be transported into the cell (e.g., hemes)

• An overall agreement of 95% for both essential and non-essential ORFs contained in iAF1255

• Disagreement points to areas for further biochemical characterization, possible errors in experimental data / model and transcriptional regulatory effects (see [5]).

REFERENCES AND ACKNOWLEDGEMENTS

We would like to thank numerous colleagues & the NIH for funding during this project (Grant 5R01GM057089).

 Future Directions

The metabolic model will be integrated with additional cellular interactions in a two-step process:

1. With a reconstructed transcriptional and translational network to include additional major cellular processes such as protein and RNA synthesis. These additional processes will further constrain the possible physiological states of the cell and improve predicted phenotypes.

2. With a reconstructed regulatory network. These additional rules will be applied in the form of Boolean logic and will reduce the available states of the network given a defined growth media.

We have developed a method to design in silico mutants strains of E. coli which will be tested in the wet lab to determine their ability to overproduce desired metabolite and protein products.

The reconstructed model will be used to visually and interpret experimental data, such as gene expression, to better understand and characterize cellular processes (i.e. regulatory interactions).

REFERENCES AND ACKNOWLEDGEMENTS