

Probing metabolite essentiality through *in silico* genome scale analysis of *E. coli* production capabilities

Marcin Imieliński¹, Călin Belta², Ádám Halász³, Harvey Rubin⁴

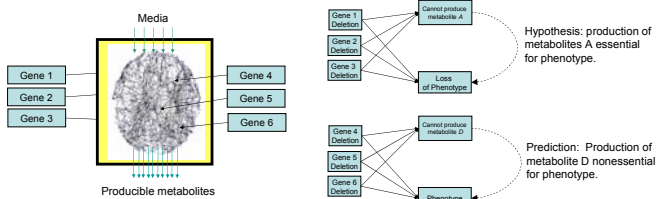
¹Genomics and Computational Biology Graduate Group, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104. ²Dept. of Mechanical Engineering, Drexel University, Philadelphia, PA, 19104. ³General Robotics, Automation, Sensing, and Perception Laboratory, University of Pennsylvania, Philadelphia, PA, 19104. ⁴Dept. of Infectious Disease, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104

Abstract

A phenotype mechanism is classically derived through the study of a set of mutants and comparison of their biochemical capabilities. However, manual assessment of the effects of a gene deletion for a large biochemical network is difficult, especially in the context of a complex media. For this purpose, we have developed a novel genome-scale computational approach that identifies the full set of biochemical species that are knocked out from the metabolome following a gene deletion. These results can be combined with data from *in vivo* mutant screens to examine the essentiality of metabolite production for a phenotype. This approach can also be a useful tool for metabolic network annotation validation and refinement in newly sequenced organisms. Combining an *in silico* genome-scale model of *E. coli* metabolism with *in vivo* survival data, we uncover possible essential roles for several cell membrane, cell wall, and quinone species. We also identify specific biomass components whose production appears to be non-essential for survival, contrary to the assumptions of previous models.

Introduction and Objectives

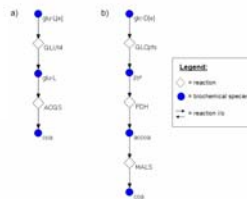
Using a novel computational approach, we sought to address the issue of metabolite essentiality in the context of a recent genome scale metabolic model of *E. coli*. An essential role for a metabolite in a phenotype is suggested if the knockout of the metabolite correlates consistently with the abolishment of the phenotype. Conversely, a metabolite is non-essential for a phenotype if the phenotype persists despite the knockout of that metabolite. The existence of a metabolite knockout can be predicted from the *in silico* analysis of a genome-scale metabolic model for a mutant. These predictions can then be used to suggest metabolites that are essential and non-essential for a phenotype. The latter approach can be an especially useful application of genome-scale metabolic modeling towards the interpretation of results from a large scale genotype to phenotype screen.



Results

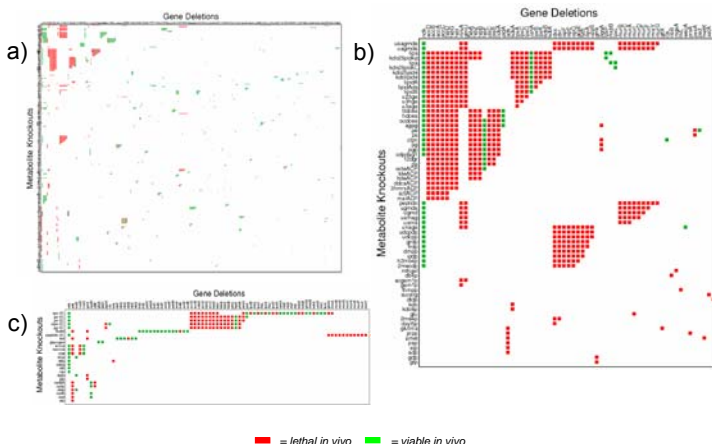
Connectedness ≠ Producibility

Graph theoretic notions of connectivity are not well suited to capture the biochemical property of producibility. Shown are paths in the *coaD* mutant metabolic network that link nutrients to the metabolic species *coa* (Coenzyme A). According to our algorithm, Coenzyme A is knocked out following the deletion of *coaD* in rich media. However, despite the lack of a biosynthetic pathway leading to the production of CoA in this mutant, there exist many paths that join nutrients to this species in a graph representation of the metabolic network. Graphs visualized using Pajek.



In silico gene to metabolite knockout maps for *E. coli* in rich media

Gene to metabolite knockout map for a) 895 *in silico* single gene deletion mutants in rich media. b) 67 metabolites whose knockout associates with lethality in ≥ 80% of cases and c) 81 gene deletions that knock out one or more biomass metabolites. The presence of a square in row *i* and column *j* represents the predicted knockout of metabolite *i* by the gene deletion *j* in rich media. Squares are colored green (red) depending on whether the *in vivo* mutant is viable (lethal) according to published experimental data. Only 40 of 81 mutants that knock out one or more biomass metabolites are lethal, and there exist multiple examples of viable mutants that knock out biomass species such as *lps* and *5mtf*, suggesting that these metabolites may not be essential. Our results suggest potential novel essential roles for metabolites such as *uaagmda* and 2-C-methyl-D-erythritol 4-phosphate, which are knocked out almost exclusively by lethal genes. Production of these species is not considered essential by previous FBA based models of survival. Metabolite notation is taken from Reed et al., 2003.



Methods

Mathematical Formulation

We represent the metabolic network in an $n \times m$ stoichiometry matrix S , whose rows correspond to intracellular metabolic species and columns represent transport and core metabolic reactions

In the cell, the rate of change of each metabolite concentration x_i is determined by two factors: production from the metabolic network s_i and consumption by other cellular processes c_i (where s_i corresponds to the i th row of S). Under the quasi-steady state assumption, the flux vector v will obey the following set of linear equalities and inequalities:

$$Sv - c = 0, 0 \leq v \leq u \quad (1)$$

Since certain metabolites may be produced in net by macromolecular and other cellular processes, we allow some components of c to be negative. For example, during a catabolic state, there may be net production of amino acids by proteolytic processes. However, given biologically reasonable assumptions regarding the physiology of a given cellular state we can identify a set of metabolite indices $P \subseteq \{1, \dots, n\}$ corresponding to species that do not contain any such sources, i.e. for which $c_i \geq 0$. We can then further restrict v to obey:

$$Sv = c, 0 \leq v \leq u, c_i \geq 0, p \in P \quad (2)$$

Given the constraints outlined above, we test the *producibility* of metabolite i by determining the existence of a feasible flux configuration that results in a positive production rate with respect to that metabolite. Formally, we test the existence of a v satisfying:

$$0 \leq v \leq u, s_i v > 0, s_i v \geq 0, p \in P, \quad (3)$$

where s_i represents the i th row of S .

Algorithm

- Calculate wild type producible metabolite set given a nutrient media
 - Configure constraints for transport fluxes.
 - Calculate producibility for each intracellular metabolite using (3)
- For each mutant
 - Set corresponding flux constraints to 0
 - For each wild type producible metabolite
 - If not producible, label as knocked out
 - Collect knocked out metabolites into *metabolite knockout profile* for that mutant
- Collect mutant metabolite knockout profiles into *gene to metabolite knockout map*

Implementation

- Genome scale metabolic model
 - iJR904 metabolic model (Reed et al. 2003) implemented as 618 x 1176 stoichiometry matrix in Matlab
 - iJR904 GPR associations (provided by B. Patsson) implemented as 1176 boolean rules
- Flux constraints
 - Active reactions given arbitrary positive upper bound
 - Active \Leftrightarrow all left hand side extracellular components present in the media AND Boolean rule based on GPR associations satisfied
 - Inactive reactions given 0 upper bound
 - Inactive \Leftrightarrow not Active
- Producibility algorithm
 - Matlab code using linear programming toolbox implementing flux constraints and producibility criteria
- Association rules mining
 - Agrawal et al (1993) implemented in Matlab
- In vivo* rich media survival data
 - PEC database (<http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp>)
 - genome-wide transposon mutagenesis study (Gerdes, Scholle et al. 2003).

Complex lethality associations

Though individual metabolites may have the property of being essential, more complex associations may exist between metabolite production and survival. Association rules data mining was used to discover Boolean combinations of metabolite knockouts that strongly associate with lethality. Shown below are graphs diagramming the gene deletion mutants supporting 2 complex lethality associations. Arrows link genes to metabolites knocked out as a result of their deletion in rich media. The knockout of any single gene in a rectangular box results in the knockout of all metabolites in the rounded boxes pointed to by the arrows. Genes colored red are essential while genes colored green are non-essential *in vivo*.

- a) knocking out (*ttfdeca* or *hdcea*) and (*12dgr* or *pa*) is lethal in 12 of 12 cases
- b) knocking out *hemeO* and (*mqr8* or *q8h2* or *2dmqr8* or *q8* or *2dmqr8* or *2omml* or *mql8* or *2ombz* or *2omp* or *2omhml*) is lethal in 9 of 10 cases



Conclusions

- We have developed a novel approach for *in silico* prediction of metabolite producibility and knockout.
- Our notion of producibility is more compatible with a biochemist's notion of a metabolite's connectedness to the media than a graph analytic criteria.
- Combining our results with *in vivo* data we can:
 - identify specific inconsistencies between a biomass-based model of survival and *in vivo* data.
 - suggest potential novel essential roles for specific metabolites in rich media survival.