

# Uncovering complex essential reaction sets in *E. coli* metabolism through pathway fragment analysis

Marcin Imielinski<sup>1,\*</sup>, Calin Belta<sup>2</sup>

1. MD/PhD Program, University of Pennsylvania School of Medicine, Philadelphia, USA
2. Bioinformatics Graduate Program, Boston University, Brookline, MA

\*E-mail: imielins@mail.med.upenn.edu

## Introduction

Genome scale metabolic network models enable *in silico* knockout experiment design for the purposes of metabolic engineering, drug discovery, and improving the systems-level understanding of metabolism. Currently, the primary method for genome-scale *in silico* knockout design is flux balance analysis (FBA), which uses linear programming (LP) to exhaustively test the metabolic capabilities of all single, double, triple, etc. knockout combinations [5]. In genome-scale models, this 'brute force' LP approach is computationally limited to the testing of small knockout combinations [2]. The minimal cut set (MCS) algorithm is an alternative "rational" approach for finding reactions sets (i.e. MCS) essential for an objective function; however, it is intractable for large (i.e. genome-scale) metabolic networks due to the computational complexity of elementary mode enumeration [4, 3].

In this paper we outline and apply a network-based approach for genome-scale metabolic knockout design that involves analysis of *pathway fragments* (PF). PF are analogues of extreme pathways, arising from partial satisfaction of network-level metabolite steady-state requirements. PF are computable at the genome-scale using a modification of *expa*, the standard algorithm for generating extreme pathways [1]. Since the cone generate by PF over-approximates the feasible flux cone, a simple traversal of PF can be used to generate cut sets for an objective reaction in the network (see Figure 1). We apply our PF-based knockout design method to uncover over 11,000 complex essential reaction sets for biomass production in an *in silico* genome-scale model of *E. coli* [6].

The 11,706 MCS discovered using our approach target 36 of 49 biomass components and employ 355 of the 1324 possible non-sink reactions in the *E. coli* network. Biomass components most often targeted by the MCS obtained using our approach are L-threonine (10655 MCS), dTTP (297 MCS), 5-methyl-THF (216 MCS), FAD (197 MCS), and dGTP (190 MCS). Though most (11218) of the MCS found with our approach have high cardinality (> 5 reactions), the vast majority (11343) carry out a precise "surgical strike" on biomass production, targeting the synthesis of only a single component. These MCS span 23 of 30 reaction subsystems in *E. coli* metabolic network, as defined in the original *E. coli* model annotation [6]. Superimposition of the MCS on the subsystem reaction classification yields 58 unique "subsystem signatures", which are shown in Figure 2. Only 289 of the MCS discovered target a single subsystem, while the vast majority (10838) target three or four subsystems. The most prevalent (7720 MCS) of these signatures target the combination of "Cell Envelope Biosynthesis", "Threonine and Lysine Metabolism", "Alternate Carbon Metabolism", and "Extracellular transport".

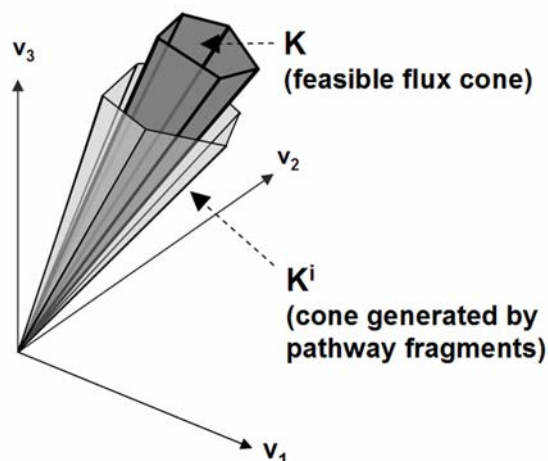


Figure 1: The feasible flux cone  $K$  of a metabolic network is contained inside the cone  $K^i$  generated by the pathway fragment collection  $E(K^i)$  obtained as the intermediate output from any iteration  $i$  of the standard extreme pathway algorithm. Because of this property, the knockout of a reaction set  $R$  that intersects all  $j$ -containing pathway fragments in  $E(K^i)$  will be guaranteed to "cut" the flux through an objective reaction  $j$ .

The minimal biomass knockouts discovered with our approach suggest combination drug targets and illuminate essential systems-level roles of reactions involved in robust *E. coli* subnetworks. Unlike the MCS algorithm, our approach yields results at the genome-scale. Compared to a brute force linear programming approach, we are able to generate many more minimal biomass knockouts that target a larger portion of the metabolic network.

## References

- [1] S. L. Bell and B. O. Palsson. Expa: a program for calculating extreme pathways in biochemical reaction networks. *Bioinformatics*, 21(8):1739–40, Apr 2005.
- [2] D. Deutscher, I. Meilijson, M. Kupiec, and E. Ruppin. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat Genet*, 38(9):993–998, Sep 2006.
- [3] S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2-3):233–247, 2006.
- [4] S. Klamt and E. D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–34, Jan 2004.
- [5] N. D. Price, J. L. Reed, and B. O. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2(11):886–97, Nov 2004.
- [6] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol*, 4(9):R54, 2003.

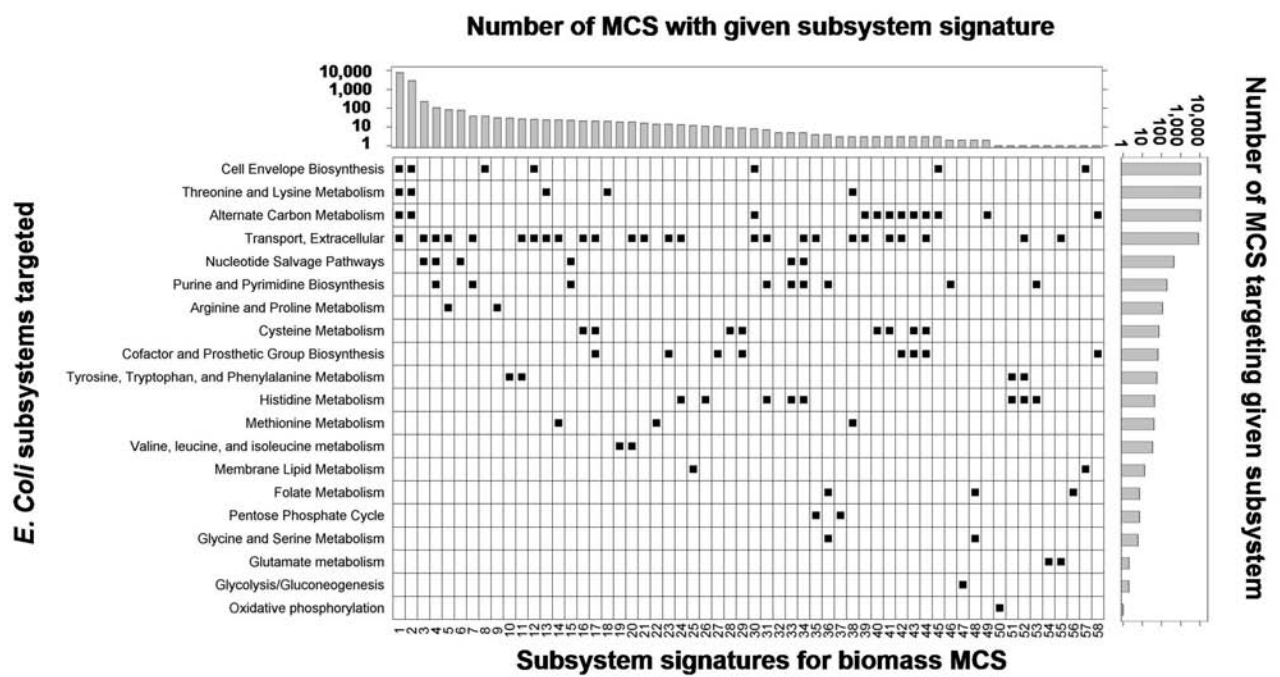


Figure 2: Subsystem signatures of 11,706 biomass MCS discovered using our approach. These MCS target 23 of 30 *E. coli* reaction subsystems. Subsystem annotations are taken from the original *E. coli* iJR904 model description [6].