Metabolic Network Analysis Demystified

Metabolic Network Analysis Demystified

Leonid Chindelevitch[1,2], Aviv Regev[1,2], and Bonnie Berger[1,2] ⋆

[1] Mathematics Department, Computer Science and Artificial Intelligence Laboratory; MIT, Cambridge, MA 02139
[2] Broad Institute, 7 Cambridge Center, Cambridge, MA 02142

Motivation. Metabolic networks are a representation of current knowledge about the metabolic reactions available to a given organism. These networks can be placed into various mathematical frameworks, of which the constraint-based framework [1] has received the most attention over the past 15 years. This results in a predictive model of metabolism. Metabolic models can yield predictions of two types: quantitative, such as the growth rate of an organism under given experimental conditions [1], and qualitative, such as the viability of a mutant [2]. Qualitative predictions, on which we focus, tend to be more robust and reliable than quantitative ones, yet experimentally testable and biologically relevant.

Here, we summarize new theoretical results related to metabolic models. These results are transformed into an algorithmic pipeline that reveals key structural properties of metabolic networks, such as blocked reactions, enzyme subsets, elementary modes, essential reactions and synthetic lethality. While the constraint-based approach to metabolic network analysis is over 15 years old, our work is, to our knowledge, the first time the theory of linear programming is used to reveal structural elements of metabolic models, rather than just predict a growth phenotype. We believe that a deeper understanding of these models will ultimately result in their wider applicability to biological questions.

Methods. We describe new theorems which reveal key structural properties of stoichiometric matrices. Theorems 1 and 2 state that cut sets and modes are closely related in networks with only reversible or only irreversible reactions. This relationship is based on the duality between the rowspace and the nullspace of a matrix, which is different from the Boolean duality described by Klamt and Gilles [3]. The same relationship holds between minimal cut sets and elementary modes. Theorems 1 and 2 provide a characterization of cut sets, and yield both an efficient method for identifying such sets as well as several important structural insights.

Theorem 3, based on a special case of Theorem 2, provides an efficient method for identifying all blocked reactions in a metabolic network. It is

⋆ To whom correspondence should be addressed: bab@mit.edu

helpful to identify these both for simplifying subsequent analysis (they can be deleted) and for pinpointing the areas in which our knowledge of metabolism may currently be incomplete. Theorem 4 is another application of Theorem 2 and provides an efficient method for identifying all enzyme subsets in a metabolic network. This method was used previously by Gagneur and Klamt [4], but the fact that it actually identifies all enzyme subsets had not been established to our knowledge.

Theorem 5 is a result about the reduction of a stoichiometric matrix to canonical form. We say that a matrix is in *canonical form* if it contains no blocked reactions, no effectively unidirectional reactions (reversible reactions which can only proceed in the forward or only in the reverse direction), no enzyme subsets, and no linearly dependent rows. We propose a 4-step reduction process that eliminates each of these undesirable structural elements in turn. The highly technical Theorem 5 states that this 4-step reduction process we propose is guaranteed to converge after a single iteration, unlike the one proposed by Gagneur and Klamt [4]. Finally, Theorem 6 is an auxiliary result about the numerical stability of blocked reactions. It says that if a reaction is blocked, then any reaction obtained from it by a small perturbation will be blocked as well.

Results. We have applied the algorithmic pipeline based on the methods above to each of the 52 genome-scale metabolic networks representing 37 different species, downloaded from the UCSD Systems Biology group website [5] and parsed to create the stoichiometric matrices. The most significant result is our finding that, of the 45 networks containing a well-defined biomass reaction, 20 are certifiably unable to exhibit growth. Another remarkable result for these networks is that their canonical form (obtained using Theorem 5) tends to be significantly smaller than their original stoichiometric matrix, providing an average 23-fold reduction in size. Additional details will be provided in the full version of the paper.

References

1. Varma, A. and Palsson, B. Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type Escherichia coli W3110. Applied and Environmental Microbiology (1994), 3724-3731.
2. Wunderlich, Z. and Mirny, L. Using the topology of metabolic networks to predict viability of mutant strains. Biophysics Journal (2006), Volume 91 (6), 2304-2311.
3. Klamt, S. and Gilles, E. Minimal cut sets in biochemical reaction networks. Bioinformatics (2004), Volume 20, 226-234.
4. Gagneur, J. and Klamt, S. Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics (2004), Volume 5 (175).
5. In Silico Organisms. *http://gcrg.ucsd.edu/In_Silico_Organisms/Other_Organisms*