THE EFFECT OF TERM ORDER ON IDENTIFYING MODELS OF GENE REGULATORY NETWORKS

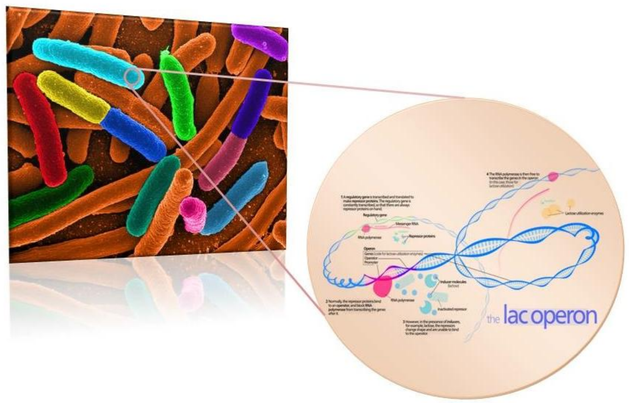
Brittani Boukather • Vy Nguyen • Pinky Patel (UTA) • Dr. Brandilyn Stigler

Introduction

Gene regulatory networks (GRNs) control many life processes. An important problem in systems biology is to identify a model of the network for a given set of laboratory data. A recently developed method builds a space of polynomial models from a given data set; however, it requires that an order on the variables (genes) and an order on the monomial terms (gene interactions) to be established. No rigorous analyses have been performed to determine the impact of these orders on the construction of the model space.

We chose a well-studied GRN in *E. coli* to study the effect of these orders on the identification of an existing model of the network. We also determined how much data, as well as which data explicitly are required for model identification.

Case Study: Lactose Metabolism in *E. coli*



The *lac* operon is a network of genes that controls the metabolism, or breakdown, of lactose in *E. coli.* We used a simplified version of a published Boolean model of the *lac* operon [1], in which glucose concentrations are fixed.

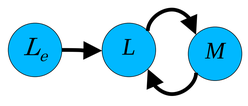
Models of GRNs

A polynomial model for a GRN is a collection of polynomials, in which the behavior of each gene is described by a polynomial function.

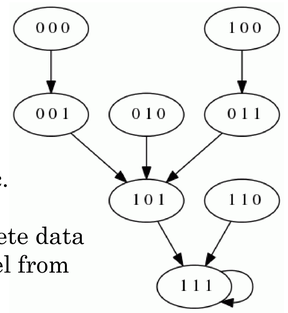
Model for the *lac* operon $F = (f_M, f_L, f_G)$ where

$$f_M = L + Le + L*L_e$$
$$f_L = M$$
$$f_G = 1$$

We simulate the GRN (L) by evaluating F on all possible inputs (R).



M = *lac* gene
L = lactose
G = glucose
L_e = extracell. lac.



Probing a GRN typically produces incomplete data sets. It is a problem to identify the model from these limited data.

Methods

The algebraic method in [3] computes all models that fit given data.

$$\text{Model space} = F_p + F_h$$

where F_p is a particular function that fits the data, and F_h is any function that is zero on the data. If I = {all possible F_h}, then

F_p mod I : = remainder of F_p upon division by all F_h

is a "minimal" model that fits the data. However, polynomial division is well defined only if I is written as a Groebner basis.

Definitions [2]

A **Groebner basis** is a multivariate nonlinear version of Gauss-Jordan Elimination and requires a canonical way to write polynomials.

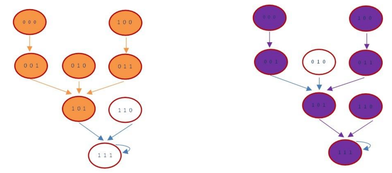
A **variable order (VO)** is a sorting of variables, and a **monomial order (MO)** is a rule for comparing monomials. A monomial order requires a variable order.

Ex If $x > y$, then $xy > y^3$ in lexicographic MO, *i.e.* dictionary order, while $y^3 > xy$ in graded lex., in which total degree is counted first. If $y > x$, then $y^3 > xy$ in both orders.

Results

We considered 3 MOs and all possible VOs on 3 variables (3! = 6). As I is computed from data, we created all subsets of the complete set of $2^3 = 8$ data points (see Models of GRNs). We computed minimal models using each of the $2^8 = 256$ subsets and reported which data sets, MOs, and VOs result in the original model F.

1. Any subset of 7 data points is **sufficient** to return F.
2. A minimum of 5 data points is **necessary** to return F.
   a. Only 32 of the 56 possible subsets of 5 data points return F (two such sets shown below).
   b. In these 32 subsets, each data point appears with the same frequency (20).



3. Each MO produced identical Groebner bases, resulting in identical minimal models, though not always F.
4. Only 2 VOs returned F.

Discussion

1. The results suggest that the choice of variable order has a greater impact on identifying models than the monomial order.
2. Solving this open problem will make significant contributions towards refined experimental design and more predictive models of GRNs.
3. To test the validity of the new hypothesis, further studies of monomial orders and models of higher degree will be conducted.

References

1. A. Veliz-Cuba, B. Stigler. *Boolean models can explain bistability in the lac operon.* 2011.
2. D. Cox, J. Little, D. O'Shea. "Ideals, Varieties, and Algorithms." 2006.
3. A. Jarrah, R. Laubenbacher, B. Stigler, M. Stillman. *Reverse engineering of polynomial dynamical systems.* 2007.

The *lac* operon and *E. coli* pictures were taken from WikiCommons.