Biochemical Network Identification: Considerations of Experimental Design, Data Requirements, Noise and Scalability for Linear and S-System Model Structures

Kenneth J. Kauffman[1], Ryan E. Altenbaugh[2], and Jeremy S. Edwards[2]

[1]University of California
Chemical & Environ Engr
A242 Bourns Hall
Riverside, CA 92521
kkauffma@ucr.edu

[2]University of Delaware
Chemical Engineering
150 Academy Street
Newark, DE 19711
edwards@che.udel.edu

The commonly used approach in biological experimentation is to make a step change in one variable and follow the progression of that change through a few data points. Biochemical engineers then try to use these data sets to develop mathematical models that describe the underlying biological phenomena. Often, these models are incapable of carrying out their desired function. Herein, we consider several issues that impact the quality of the resulting model.

Linear and s-function models are among the more commonly used classes of models for biochemical systems. In this work, we identified both linear and s-function models for several representative systems with similar, yet distinct nonlinear characteristics. We treat each system as an unknown and then attempt to identify the best model possible under various experimental conditions.

The key findings from these analyses on biochemical system identification are: (1) Even with high-throughput techniques that provide measurements for several species at once, standard, single variable step changes are unable to adequately characterize the dynamics of nonlinear biochemical systems. (2) Simultaneously changing multiple inputs at once in a pseudo-random manner increases the information content of the dataset and can lead to better models. (3) For the pseudo-random inputs, there exists an optimal sampling and perturbation frequency. However, for the cases analyzed here, the objective function is relatively flat around the optimum, thus allowing a wide range of experimental designs that are capable of providing additional insight. (4) For a given level of noise (2% to 32%), models based on random inputs were better able to describe the underlying dynamics than those based on step-changes. (5) Beyond a certain point, increasing the length of the experimental run provides diminishing returns with respect to improvement in the model behavior. (6) These findings were consistently observed for both reversible and irreversible branched systems consisting of either three or six metabolic states.