Structure Identification of Biochemical Systems with Genetic Algorithms

Jaime Combadão[1], Jonas Almeida[2], and Eberhard O. Voit[2]
[1] Biomathematics Group, Instituto Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República (EAN), P.O.Box 127, 2781-901 Oeiras, Portugal
[2] Department of Biometry and Epidemiology, Medical University of South Carolina, Charleston, SC 29425 (*contact address*)

Modern methods of genomics and proteomics are beginning to yield data of a quantity and quality unimaginable just a decade ago. Already, sophisticated nuclear magnetic resonance and mass spectrometry techniques are able to quantify hundreds and sometimes thousands of simultaneously measured metabolites. These metabolic profiles are usually obtained as snapshots, but could be generated as dense time series that reveal the dynamic trends in many or all metabolites of an organism, following some stimulus. Metabolic profiles contain enormous information about the flux distribution, and also the regulation, of metabolic pathways *in vivo*. This information is not immediately explicit though, but requires adequate analytical and computational methods of retrieval and interpretation. We propose to address this information retrieval with a novel approach that combines an established modeling framework for metabolic systems, *Biochemical Systems Theory* (*BST*), with computational methods of parameter estimation that are based on a *Genetic Algorithm (GA)*. The key feature of this merging of techniques is that parameter values in biochemical systems models are directly and uniquely related to both the flow structure and the regulation of the modeled network, which allows interpretation of estimated parameter values in terms of possible and probable network structures.

BST, and particularly its implementation in the form of *S-system* differential equations, provides powerful tools for the modeling and analysis of complex metabolic networks. Of crucial benefit in the present context is that every S-system model has the same homogeneous mathematical structure, namely,

$$\dot{X}_i = a_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - b_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \qquad i = 1, \ldots, n.$$

Thus, the change in each variable X_i is given by two processes, one collecting all influences leading to the production or increase in X_i, and one collecting all influences contributing to the decrease or degradation of X_i. These processes are linearly approximated in logarithmic space, which corresponds mathematically to products of power-law functions in Cartesian coordinates. It is customary to denote the dependent variables, which are typically metabolites, as $X_1, \ldots, X_n$, and the independent variables, which are usually enzymes or constant controls, as $X_{n+1}, \ldots, X_{n+m}$. The non-negative parameters a_i and b_i are called *rate constants*, and the real-valued exponents g_{ij} and h_{ij} are *kinetic orders*. Each rate constant reflects the magnitude of a given process, and each kinetic order g_{ij} or h_{ij} describes uniquely and quantitatively the effect of X_j on the production or degradation of X_i, respectively. For instance, a kinetic order with negative value signifies an inhibitory effect.

We show preliminary results from artificial examples in which a GA is used to estimate S-system parameters from dynamic metabolic profiles. The novelty of this approach is that S-systems are globally and directly estimated (by GA) from time series of metabolite concentrations, as opposed to assumed distributions of fluxes and flux rates. Therefore, knowledge about the underlying stoichiometric structure is not *a priori* required, but is revealed by the algorithm. We analyze several scenarios that differ in the accuracy and resolution of the time series data. We also explore to what degree available stoichiometric or regulatory information provides added advantages for structure predictions about the metabolic network.