

Parameter Estimation Revisited (Again!): Low SSE and Speed are Not Enough

Eberhard O. Voit

Integrative BioSystems Institute and Wallace H. Coulter
Department of Biomedical Engineering, Georgia Institute of Technology
313 Ferst Drive, Atlanta, GA 30332-0535, USA;
eberhard.voit@bme.gatech.edu

Mathematical models are at the center of computational systems biology. They have become crucially important because the extent of quantitative data and of qualitative information characterizing biological systems has grown so much in recent years that the unaided human mind is no longer capable of keeping track of all pieces of information and integrating them in an efficacious fashion. The bottleneck in the construction of systems models is the identification of model parameters that make the model consistent with observed data. Parameter estimation may be approached from the bottom up, in the case of metabolic pathway analysis using information on metabolites, enzymes and modulators, or from the top down, using metabolic time series data, which have become more prevalent in recent years. The success criteria of these estimation attempts have in the past been the quality of fit, expressed as a residual error between model results and data, and the computational time needed to obtain the data fits. In this presentation, I will present some new findings suggesting that these criteria are certainly necessary but not sufficient. An important additional criterion is the ability of the model to model new data with a sufficient degree of reliability. This extrapolability is inversely related to error compensation within a fitted model, as well as the non-uniqueness (sloppiness) of models yielding acceptable fits. The presentation will review the concepts of our recently proposed method of Dynamic Flux Estimation (DFE) and discuss its two distinct phases, namely the initial model-free and assumption-free data analysis and the subsequent modelbased mathematical characterization of process representations. The model-free phase reveals inconsistencies within the data, and between data and the alleged system topology, while the model-based phase allows quantitative diagnostics of whether—or to what degree—the assumed mathematical formulations are appropriate or in need of improvement. The presentation will then describe some extensions of DFE, such as the simultaneous use of time series and kinetic data. As in earlier studies of this genre, I will use *in vivo* metabolite data characterizing the glycolytic pathway in the lactic acid bacterium *Lactococcus lactis*.