Biophysical Journal
Article

Biophysical Society

Acceleration Strategies to Enhance Metabolic Ensemble Modeling Performance

Jennifer L. Greene,[1] Andreas Wächter,[2] Keith E. J. Tyo,[1] and Linda J. Broadbelt[1,*]
[1]Department of Chemical and Biological Engineering and [2]Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois

ABSTRACT Developing reliable, predictive kinetic models of metabolism is a difficult, yet necessary, priority toward understanding and deliberately altering cellular behavior. Constraint-based modeling has enabled the fields of metabolic engineering and systems biology to make great strides in interrogating cellular metabolism but does not provide sufficient insight into regulation or kinetic limitations of metabolic pathways. Moreover, the growth-optimized assumptions that constraint-based models often rely on do not hold when studying stationary or persistor cell populations. However, developing kinetic models provides many unique challenges, as many of the kinetic parameters and rate laws governing individual enzymes are unknown. Ensemble modeling (EM) was developed to circumnavigate this challenge and effectively sample the large kinetic parameter solution space using consistent experimental datasets. Unfortunately, EM, in its base form, requires long solve times to complete and often leads to unstable kinetic model predictions. Furthermore, these limitations scale prohibitively with increasing model size. As larger metabolic models are developed with increasing genetic information and experimental validation, the demand to incorporate kinetic information increases. Therefore, in this work, we have begun to tackle the challenges of EM by introducing additional steps to the existing method framework specifically through reducing computation time and optimizing parameter sampling. We first reduce the structural complexity of the network by removing dependent species, and second, we sample locally stable parameter sets to reflect realistic biological states of cells. Lastly, we presort the screening data to eliminate the most incorrect predictions in the earliest screening stages, saving further calculations in later stages. Our complementary improvements to this EM framework are easily incorporated into concurrent EM efforts and broaden the application opportunities and accessibility of kinetic modeling across the field.

INTRODUCTION

Enabling kinetic and regulatory modeling of cellular metabolism is a major challenge in metabolic engineering and systems biology (1–6). Constraint-based stoichiometric modeling greatly aids in characterizing and improving strain designs, but without kinetic information, it is difficult to identify rate-limiting steps and interrogate regulatory behavior. Some studies using kinetic models for metabolic applications do exist, but they are limited. For example, individual kinetic models were developed for the red blood cells of 24 different patients to interrogate differences in metabolite levels and enzyme activities that are difficult to capture with constraint-based models alone (7). Another example is in strain design efforts, where the cellular objective of maximizing growth cannot be assumed (i.e., studying stationary, nongrowth phase metabolism), so the constraint-based stoichiometric methods are harder to utilize effectively. The ability to incorporate extensive regulatory behavior in kinetic models is also useful when studying systems where regulation heavily governs a cell's metabolism and even prevents cells from operating at maximum metabolic capacity (8–10). Constraint-based models cannot explicitly track metabolite concentrations, making regulation based on metabolites difficult. Ultimately, generating quality kinetic models of cellular metabolism will allow us to better resolve and interrogate cellular metabolism for strain design and biological discovery applications.

To build a kinetic model of metabolism, the rate laws and parameters are needed for each enzyme in the network. Some kinetic modeling methods combine rate laws and kinetic parameters from public databases or literature and combine them into a single metabolic model (11,12). Unfortunately, the in vitro derived kinetic parameters for enzymes most often reported in literature do not necessarily reflect true in vivo behavior and are often determined under varying experimental conditions without accounting for local

Submitted April 11, 2017, and accepted for publication July 11, 2017.
*Correspondence: broadbelt@northwestern.edu
Editor: Daniel Beard.
http://dx.doi.org/10.1016/j.bpj.2017.07.018

concentration effects (12). Moreover, for some enzymes these in vitro parameters and rate laws have not been determined, and exhaustive regulatory relationship studies have not been completed. Consequently, a single kinetic model combining these in vitro derived parameters, or several smaller models built on these parameters, are often unable to resolve experimentally observed in vivo data or describe metabolic states outside the immediate realm of the reference state (13–15).

The ensemble modeling (EM) framework was previously developed to address these hurdles by sampling kinetic parameters for the entire metabolic network simultaneously and screening them against an experimental dataset collected under consistent conditions (16–18). During the screening step, predictions from the sampled kinetic parameters are compared to experimental results, and kinetic parameters that predict the results poorly are rejected. Although estimates of individual kinetic parameters may not be strictly accurate, EM seeks to develop a network model that explains system behavior. Furthermore, the EM method constrains the large kinetic parameter sample space using readily available thermodynamic, stoichiometric, and steady-state flux data. The EM method has been successfully employed to model and resolve the kinetics of various metabolic pathway designs for bioproducts and to interrogate cancer metabolism (19–22). Additionally, improvements made to the screening methods used in EM led to the development of a kinetic core model of *Escherichia coli* metabolism and more recently a genome-scale model, k-ecoli457 (23,24). Lastly, the way reactions are defined in ensemble modeling allows for sampling different network structures to interrogate different potential governing or previously unknown regulation reactions as well as predict unresolvable flux distributions previously unconstrained in traditional constraint-based methods (21,25).

However, despite its numerous advantages, EM rapidly becomes computationally limiting with increasing network size and complexity. As our field increases the availability and uses of larger, genome-scale metabolic models for various organisms, it is imperative to improve our ability to generate larger kinetic models of these systems as well (3,4,26–28).

To enhance the computational efficiency of EM, it is critical to focus efforts on the step in which predictions created by the generated kinetic parameter sets are compared to the available experimental perturbation data, as it is the major rate-limiting step of the EM process. To compare to a single data set, each parameter set in the ensemble is perturbed by solving a system of ordinary differential equations (ODE) with different concentrations in one or more proteins (i.e., knockout or overexpression). Solving a single ODE calculation is time intensive, and screening parameter ensembles across the available perturbation data sets requires many iterations of these calculations. In total, these ODE-based screening steps account for $>99\%$ of the required time to complete the traditional EM method, and we observe that screening times scale nonlinearly with increasing metabolic network size.

Another prevalent challenge with the EM method is trying to resolve a large number of kinetic parameters using a limited number of experimental observations. The predictive power of the resulting kinetic model can be limited by the quality of the data used to train it. Generating more experimental data at different states is costly, time intensive, and, depending on the type of experiment, sometimes physically infeasible. Kinetic model systems are traditionally underdetermined as there are far more parameters to fit than training data sets available. Therefore, the problem is ill-posed, and several, unique kinetic parameter sets can be found with equal ability to fit any one flux distribution. At the same time, unavoidable experimental error or noise means it is likely there will be no one set that can describe all the flux distributions at the different cellular states. This challenge intensifies as the need to develop larger models with increasing numbers of kinetic parameters to identify grows, although the availability of experimental observations to screen against does not increase at the same rate. To alleviate this widening gap, additional constraints to narrow the kinetic parameter sample space and select more accurate kinetic parameters are warranted.

In this work, we addressed the computational limitations and parameter sampling difficulties inherent in the EM method. We have greatly reduced overall EM run times by utilizing additional parameter screening techniques and introducing previously developed methods to reduce structural model complexity during ODE integration (Fig. 1). Specifically, we have implemented a conservation analysis step to eliminate linear dependency present in the EM models used to date. We found eliminating linear dependency reduces the stiffness and screening time of the kinetic parameter sets. We have also elucidated a preferred method for selecting the order of screening data sets to reduce the number of incorrect parameter sets carried forward in each screening iteration. To improve kinetic parameter sampling, we have also further characterized and implemented a parameter screening step capitalizing on the known stability of wild-type cellular metabolism. This screening method was previously incorporated for robustness analysis of metabolic pathway designs and metabolic control analysis (29–32). It addresses observations of local instability in traditional EM solutions and reduces the kinetic parameter sampling space by removing parameter sets not locally stable at the initial wild-type (WT) steady-state condition. By incorporating this screen into the existing EM framework (Fig. 1), our final parameter set solutions more accurately reflect true biological behavior. In this work, we specifically demonstrate how local instability increases with network size and slows down parameter screening, making the case for incorporating this additional screen in all ensemble modeling efforts outside of its original use in ensemble modeling robustness analysis studies. Through our efforts to speed up and