

Model Credibility Plan

Our model credibility plan is focused on uncertainty quantification and sensitivity analysis. The largest uncertainties are due to the simulation parameters, which in this case are the chemical potentials of the metabolites. Chemical potentials are calculated from standard free energies of formation in solution using either component contribution methods or electronic structure calculations. Experimentally determined parameters are available for reactions of central metabolism, but these are *in vitro* assays, and *in vitro* conditions likely do not match *in vivo* conditions. Regardless, the data will serve as a useful benchmark.

The two largest sources of error are (1) the calculation of standard free energy of formation *in vacuo*, and (2) the calculation of the standard free energy of solvation. Of these two, the latter is likely the largest (especially for anions) [1]. The errors associated with the standard free energies of formation *in vacuo* are often due to electron correlation modeling and are likely small compared to solvation estimates. The errors in solvation estimates are due to the implicit solvation models used. These models use parameters similar to those used in Debye-Huckel theory and the largest uncertainty is in the parameter for the dielectric response of the solvent. Generally, the dielectric response of bulk water is used (~79), but estimates of the dielectric of the cell cytoplasm varies from 50 to 200 [2].

We will evaluate how the predictions of metabolite concentration and reaction flux vary as a function of the dielectric constant of the cytoplasm using ensemble modeling. Ensemble modeling is the use of multiple simulations using different plausible parameters to predict an outcome with an associated uncertainty. Ensemble modeling is widely used in weather forecasting but has also been applied to simulations of metabolism [3]. While varying the numerical value of the dielectric constant, the variance of the predicted fluxes amounts to a sensitivity analysis. Since we have experimental data on reaction fluxes, we will use the experimental flux values to constrain the range of metabolite concentrations, which will allow us to quantify the uncertainty of the predictions of metabolite concentrations.

For the ensemble modeling, we will generate free energies of solvation for the metabolites using different values of the dielectric constant using implicit solvent models in NWChem [4]. These parameters will be useful not just for this project, but will provide a benchmark data set for the community. Estimates of standard free energies of solvation depend on the level of theory used in the calculations. While we have targeted the major sources of uncertainty in these calculations, other researchers may want to evaluate different levels of theory, such as the use of explicit solvent models or different approaches to handling electron correlation.

We will provide a Matlab-based analysis package that analyzes the degree to which the simulation flux values agree with experiment. The agreement of the predicted flux values with the experimental observations relies on comparing the information contained in the two sets of flux values. We will use the Kullback–Leibler divergence of the predicted set of flux values q from the observed flux values p over each reaction i ,

$$D_{KL} = \sum_i p_i \log \frac{p_i}{q_i}$$

This is the expectation of the logarithmic difference between the sets of probabilities p and q , where the expectation is taken over the set of observed distributions. The exponent of the Kullback–Leibler divergence is a geometrically averaged likelihood ratio, $\exp(D_{KL}) = \prod_i (p_i/q_i)^{p_i}$. We will evaluate the agreement at the

individual reaction level as well as overall of metabolism. The significance of the likelihood ratio will be tested and corrections will be made for multiple hypothesis testing. We will determine the uncertainty in the predicted metabolite distributions free energy changes across reactions. We will determine how different chemical potentials impact the influence of metabolism on the clock proteins. We will package the code, parameters, initial conditions, experimental flux data, and other necessary metadata so that the simulations can be independently evaluated in collaboration with MSM members.

Timeline and Funding. Since it is difficult to estimate funding for an unknown party to evaluate and test our model credibility procedure, we have included 3 weeks of funding for credibility testing for a third party within the PNNL budget for years 4 or 5. We expect that the testing actually take less time. Setting up the funding in this way also provides flexibility for us to barter within the MSM consortium such that we could fund ourselves to evaluate the credibility of others models in exchange for others funding themselves to test ours.

- [1] J. J. Liu, C. P. Kelly, A. C. Goren, A. V. Marenich, C. J. Cramer, D. G. Truhlar, *et al.*, "Free Energies of Solvation with Surface, Volume, and Local Electrostatic Effects and Atomic Surface Tensions to Represent the First Solvation Shell," *Journal of Chemical Theory and Computation*, vol. 6, pp. 1109-1117, Apr 2010.
- [2] J. Gimsa, T. Muller, T. Schnelle, and G. Fuhr, "Dielectric spectroscopy of single human erythrocytes at physiological ionic strength: dispersion of the cytoplasm," *Biophys J*, vol. 71, pp. 495-506, Jul 1996.
- [3] L. M. Tran, M. L. Rizk, and J. C. Liao, "Ensemble modeling of metabolic networks," *Biophys J*, vol. 95, pp. 5606-17, Dec 15 2008.
- [4] H. J. J. van Dam, W. A. de Jong, E. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, *et al.*, "NWChem: scalable parallel computational chemistry," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, pp. n/a-n/a, 2011.