Uniform sampling of metabolic networks

Ben Cousins¹, Hulda S. Haraldsdóttir²* (hulda.haraldsdottir@uni.lu), Santosh Vempala¹ and Ronan M. T. Fleming²

¹School of Computer Science, Algorithms and Randomness Center, Georgia Institute of Technology, Atlanta, Georgia; ²Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

http://stanford.edu/group/SOL/multiscale/

Project Goals: Development of a framework for mass conserved elementary kinetic modelling of metabolic networks [1, 2, 3, 4, 5]. This collaborative project was motivated by a need for data to constrain and validate metabolic models. Sampling algorithms have demonstrated applications in measurement and estimation of kinetic parameters, steady state fluxes and metabolite concentrations for biochemical systems [6, 7, 8].

Constraint-based metabolic modelling provides a framework to explore feasible steady state fluxes in metabolic networks. Physicochemical constraints imposed, e.g., by network topology, mass conservation and substrate availability are formulated as linear equalities and inequalities that define a high-dimensional convex set. Uniform sampling of this set provides an unbiased characterisation of the metabolic capabilities of a cell or organism [9]. However, uniform sampling of steady state metabolic flux sets has proven algorithmically challenging due to their high dimensionality and inherent anisotropy. Here, we evaluate the performance of a recently published sampling algorithm [10] on metabolic networks of increasing size. The algorithm is based upon the provably efficient hit-and-run random walk [11] and crucially uses a rounding preprocessing step to place the set of feasible metabolic fluxes in near-isotropic position. This algorithm converges to a uniform sampling distribution up to 25 times faster than a popular artificial centering hit-and-run algorithm [12]. We demonstrate the effects of improved convergence on predictions of the metabolic capabilities of E. coli.

References


This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.