Parameter Inference for Virus Capsid Assembly via Simulation-Based Data Fitting

Lu Xie[1,2], Gregory R. Smith[3], and Russell Schwartz[2, 3]

***Abstract* — Self-assembly is ubiquitous in biology but challenging to model because it involves enormous networks of possible reactions that we cannot experimentally monitor. Computer simulations offer a way to explore possible pathways, but require physical parameters one cannot directly measure. We developed a strategy to learn assembly models of specific viruses by fitting rule-based stochastic simulations to indirect measures of bulk assembly. Here, we examine the problem of optimizing these parameter fits with specific application to light scattering measurements of virus capsid assembly. We find that derivative-free optimization (DFO) methods offer important advantages in handling the computational challenges of these systems.**

***Keywords* — Self-Assembly, Stochastic Simulation, Data Fitting, Derivative Free Optimization, Virus Capsid.**

I. Motivation

VIRUS capsid assembly has attracted considerable interest across various modeling fields as a challenging model of macromolecular assembly [1]. There is, unfortunately, no experimental method to monitor fine-scale assembly pathways of large systems with high symmetries such as capsids. Simulation methods have therefore proven valuable for exploring possible pathways and understanding how they contribute to robust, efficient assembly. Such generic theoretical models, however, could only explore spaces of possible pathways, not characterize pathways of specific viruses, because they depend on detailed binding rate parameters that we cannot directly measure.

In prior work, we sought to extend these methods to models of specific real virus systems by fitting rule-based stochastic simulations of assembly [2] to experimental static light scattering measurements of bulk capsid assembly *in vitro* [3,4]. Applying this approach to real capsid assembly data provided for the first time a way to explore possible fine-scale assembly pathways of specific viruses.

This approach to simulation-based data fitting nonetheless faces substantial challenges due to computationally costly simulations, stochastic variation between simulation trajectories, and noisy, ambiguous experimental data. In the present work, we explore improved approaches to data-fitting in an effort to more efficiently and accurately characterize kinetic rate parameters and implied pathways.

Acknowledgements: This work was supported by NIH grant 1R01AI076318.

[1]Joint Carnegie Mellon – University of Pittsburgh Ph.D Program in Computational Biology, [2]Ray and Stephanie Lane Center for Computational Biology, and [3]Department of Biological Sciences, Carnegie Mellon University

II. Modeling and Methods

We develop rule models for viral capsid assembly as in our prior work and simulate them by a variant of the stochastic simulation algorithm. We then fit rate parameters to minimize root mean square deviation (RMSD) between simulated and real *in vitro* light scattering data. To improve data fitting, we apply derivative-free optimization (DFO) methods, a class of numerical optimization algorithm well suited to systems with computationally expensive, noisy function evaluations for which gradient-based optimizations tend to do poorly. We compare two DFO methods – multi-coordinate search (MCS) [5] and stable noisy branch and fit (SNOBFIT) [6] – to a heuristic gradient-based method custom developed for this application in our prior work [4]. We evaluate the methods by their ability to minimize RMSD on real data and to identify true parameters on simulated data. We further explore the consequences of inaccuracies in data fitting on pathway selection.

Conclusion

Application of the available methods to real and semi-simulated data suggests that DFO methods improve accuracy relative to the prior work. SNOBFIT and MCS yield lower RMSD and more precise fitting of simulated rates than our prior gradient-based method. Preliminary results, however, suggest that imprecision in the parameter fits leads to only minor changes in implied assembly pathways, consistent with prior observations that pathway selection for real viruses tends to be robust to relatively large changes in parameters or presumed assembly conditions.

References

[1] Whitesides GM, Grzybowski B (2002) Self-assembly at all scales. *Science* **295**, 2418-2421.

[2] Zhang T, Rohlfs R, Schwartz R (2005) Implementation of a discrete event simulator for biological self-assembly systems. *Proc. 37th Winter Simulation Conf.* pp. 2223-2231

[3] Kumar MS, Schwartz R (2010) A parameter estimation technique for stochastic self-assembly systems and its application to human papillomavirus self-assembly. Phys. Biol. 7, 45005-45016.

[4] Xie, L, Smith, G.R., Feng, X., Schwartz, R. (2012) Surveying capsid assembly pathways through simulation-based data fitting. *Biophys. J.* **103**, 1545-1554.

[5] Neumaier, A. MCS: Global Optimization by Multilevel Coordinate Search. http://www.mat.univie.ac.at/~neum/software/mcs/.

[6] Huyer, W., and A. Neumaier. 2008. SNOBFIT – Stable noisy optimization by branch and fit. *ACM Transactions on Mathematical Software* 35:1–25.