Efficient simulation of macromolecular assembly

Dustin L. Maurer[1] and Eric J. Deeds[1,2]

***Short Abstract* — Molecular machines are essential for the survival and proliferation of cells. Despite their importance, the assembly dynamics of these machines are currently not well understood at a quantitative level. This is due in part to the lack of efficient computational tools for simulating their assembly. We have developed a computational framework for efficiently modeling the self-assembly of arbitrary macromolecular structures using both deterministic and stochastic techniques. We applied our method to simulate the assembly of the proteasome Core Particle (CP) on timescales of hours, demonstrating the generality and efficiency of our approach.**

***Keywords* — Macromolecular Machines, Self-assembly**

I. Introduction

MACROMOLECULAR machines such as the ribosome or the proteasome perform essential cellular tasks. These structures are constantly being lost from cells due either to degradation or cell division. In order to replace lost molecules, the cell must constantly synthesize the subunits that form these machines; the subunits subsequently assemble into active structures. In many cases, assembly occurs as a diffusive process and can be described by a series of reversible association reactions between assembly intermediates.

Most previous attempts at modeling assembly have focused on viral capsids and related systems [1,2]. These approaches construct physically or geometrically realistic structures from a set of monomer species, and rely exclusively on stochastic simulation approaches. As a result, in many cases these models are limited by the computational time needed to achieve biologically relevant timescales. They are also often not readily extensible, requiring significant modification to simulate different classes of structures.

In this work, we developed a novel simulation approach that can be applied to essentially any macromolecular structure of interest.

II. Results

We employ a binary representation of structures (either the fully assembled macromolecule or intermediates), as pioneered by Saiz and Vilar [3]. For structures that are not too large, one can use this representation to efficiently enumerate all possible intermediate species and the association reactions between them. The resulting Chemical Reaction Network (CRN) can be translated into a system of Ordinary Differential Equations (ODEs) using a few straightforward principles [4,5]. In cases where this is feasible, the efficiency of integrating ODEs provides incredible decreases in computational costs compared to stochastic approaches [1].

Some structures, however, generate far too many intermediates to be simulated using ODEs. The 20S CP of the proteasome, for instance, consists of a total of 28 subunits, 14 α and 14 β proteins arranged in an $\alpha_7\beta_7\beta_7\alpha_7$ stacked ring architecture [6]. We have shown that this structure can generate over 6 million unique intermediate structures, making an ODE-based approach impossible.

We thus developed a stochastic self-assembly simulator directly inspired by rule-based approaches to modeling cell signaling dynamics [7,8]. Since we encode structures using binary strings [3], most operations can be performed using bitwise operators, resulting in a very efficient codebase. We have applied this technique to the 20S CP structure from archaea: we can simulate 1000s of α and β subunits assembling for over 3 hours, which is the experimentally observed self-assembly timescale [6]. These simulations take minutes on a laptop computer.

III. Conclusion

We have developed a general simulation framework that can be applied to essentially any structure whose assembly is important to living systems. The framework allows users to generate CRNs/ODEs when possible, but provides efficient stochastic simulations when combinatorial complexity prevents the application of deterministic techniques.

References

[1] Sweeney, B, Zhang, T and Schwartz, R (2008) "Exploring the parameter space of complex self-assembly through virus capsid models." *Biophys J* **94**, 772-83.

[2] Nguyen, HD, Reddy, VS and Brooks, CL (2007) "Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids," *Nano Lett* **7**, 338-44.

[3] Saiz, L and Vilar, JMG (2006) "Stochastic dynamics of macromolecular-assembly networks," *Mol Syst Biol* **2**. 2006.0024.

[4] Deeds, EJ, Bachman, J, Fontana, W (2012) "Optimizing ring assembly reveals the strength of weak interactions," *PNAS* **109**, 2348-53.

[5] Hlavacek, WS *et al.* (2003) "The complexity of complexes in signal transduction," *Biotech Bioeng* **84**, 783-94.

[6] Marques, AJ *et al.* (2009) "Catalytic Mechanism and Assembly of the proteasome," *Chem Rev* **109**, 1509-36.

[7] Danos, V, Feret, J, Fontana, W and Krivine, J (2007) "Scalable simulation of cellular signaling networks," *Lect Notes Comput Sci* **4807**, 139-57.

[8] Sneddon, MW, Faeder, JR and Emonet, T (2011) "Efficient modeling, simulation and coarse-graining of biological complexity with NFsim," *Nat Meth* **8**, 177-83.

[1]Center for Bioinformatics, The University of Kansas, Lawrence, KS

[2]Department of Molecular Biosciences, The University of Kansas, Lawrence, KS. E-mail: deeds@ku.edu