SIMULATING RIBOSOME BIOGENESIS IN WHOLE CELLS

Allocation: Illinois/0.60 Mnh
PI: Zaida Luthey-Schulten[1,2,3,4]
Co-PIs: Tyler Earnest[1,2], Jonathan Lai[1], Ke Chen[1], Michael Hallock[1]
Collaborator: James R. Williamson[5]

[1]University of Illinois at Urbana–Champaign
[2]NSF Center for the Physics of Living Cells
[3]NIH Center for Macromolecular Modeling and Bioinformatics
[4]Beckman Institute
[5]Scripps Research Institute

EXECUTIVE SUMMARY:

Central to all life is the assembly of the ribosome: a coordinated process involving the hierarchical association of approximately 50 proteins to the RNAs forming the small and larger ribosomal subunits. Through the synthesis of data from disparate sources such as *in vitro* kinetics, cryo-electron tomography, single-particle diffusion assays, mRNA expression, and genomics data, we developed a spatially-resolved stochastic model of the biogenesis of the ribosomal small subunit in *E. coli*. Using our GPU-accelerated Lattice Microbes software [1–3] on Blue Waters, we observed biogenesis in modeled cells at timescales up to one hour. Our model reproduced the correct assembly times and predicted the spatial distribution of assembly intermediates. We will report on the results we have achieved to date and outline how further work using Blue Waters will provide us a window into cellular processes from single cells to colonies.

INTRODUCTION

Translation is the universal process that synthesizes proteins in all living cells. Central to translation is the ribosome, which itself constitutes approximately one-fourth of the bacterial dry mass. Biogenesis of the ribosome together with all cellular activities involved in translation consume a significant budget of the cell's total energy. However, the mechanism enforcing the balance between metabolism and macromolecular synthesis is yet to be fully resolved. An overarching goal of our research program is the development of large-scale whole-cell models including metabolism, gene expression, signaling, DNA replication, and cell division. Modeling the biogenesis of the ribosome is a major milestone towards the achievement of this goal. These models are crucial for our understanding of cell growth and how it is regulated in response to environmental perturbations.

The intricacy of ribosome assembly first attracted Nomura et al. [4], who originally mapped out the hierarchical dependency of the r-proteins binding to the *E. coli* 16S rRNA to form the small subunit (30S). Progress in biophysical approaches boosted our understanding of *in vitro* ribosomal self-assembly mainly for the protein-assisted dynamics of RNA folding [5–7], and the kinetic cooperativity of protein binding [8–10]. Both aspects agreed that assembly of *E. coli* proceeds through multiple parallel pathways nucleated at different positions on the rRNA, and a 5' to 3' directionality during assembly is always present. However, the order in which the r-proteins bind to assembly intermediates as inferred from thermodynamic and kinetic experiments can be contradictory, hampering our investigation of the assembly under an *in vivo* environment. Therefore, a comprehensive model that captures the topology of the protein RNA interaction network is needed to decipher the underlying rules governing the assembly of the ribosome.

METHODS & RESULTS

The stochastic *in vivo* model is constructed upon a deterministic *in vitro* assembly model. We built a chemical reaction network describing the binding of proteins to assembly intermediates using kinetic data from mass spectrometry experiments [10,11]. Since the assembly involves the binding of 20 r-proteins, naively, this would result in 2^{20} intermediate species and 20! (2.43 quadrillion) reactions. To mitigate combinatoric explosion, we use Nomura's assembly map to only include species and reactions that are allowed by the binding hierarchy. This reduces the model to 1,612 intermediates and allows for a thorough search of parameter space to fit the deterministic model to the kinetic data.

The 1,612-species model can be simplified further by removing species from the network that contribute the least to the overall assembly flux. By iterating this procedure, we were able to generate an assembly network containing

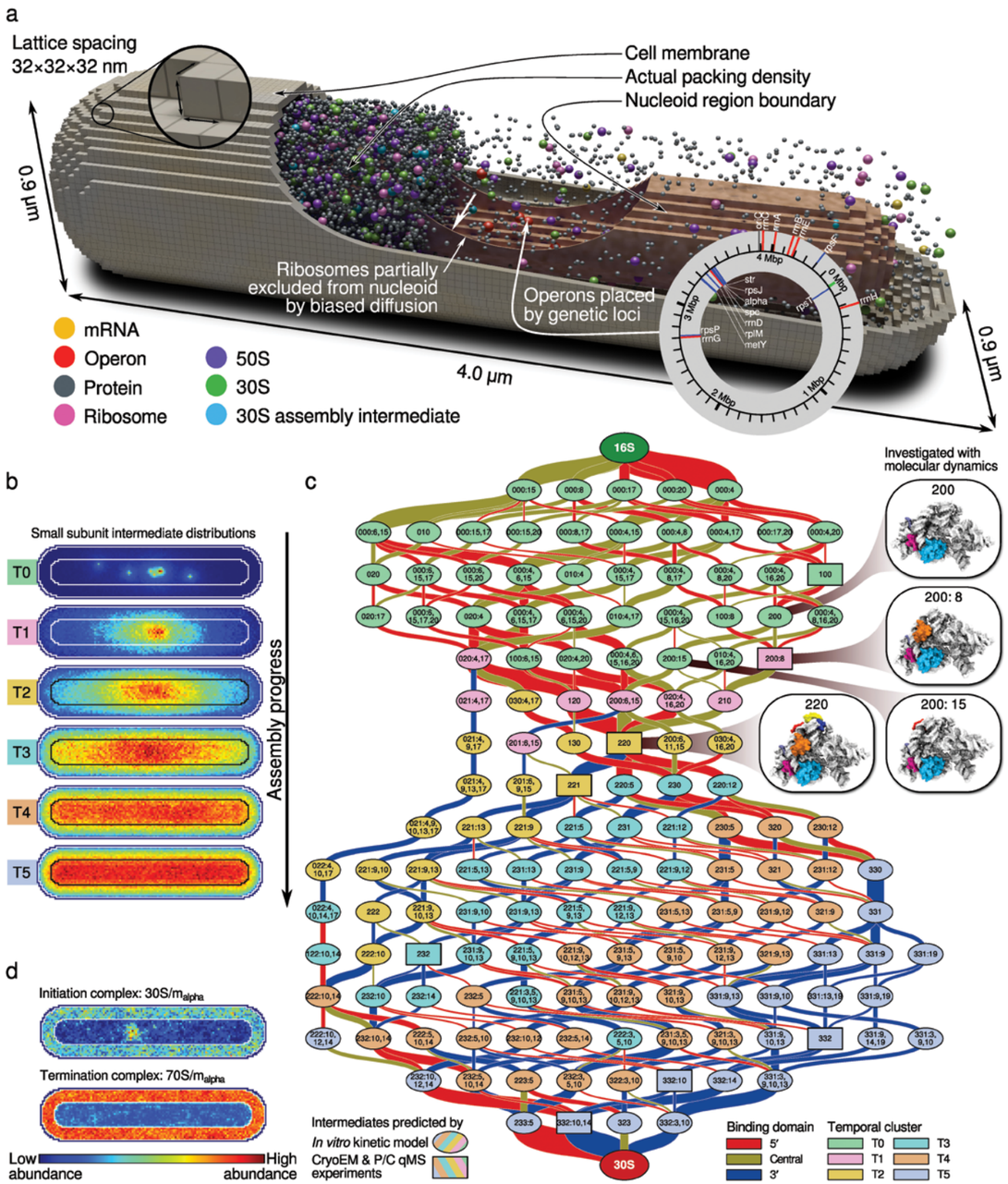

FIGURE 1: (a) Schematic of whole-cell stochastic model. (b) Heterogeneous distribution of 30S intermediates, clustered based on maximum abundance times. (c) Assembly network showing 30S intermediates as nodes and protein binding reactions as edges. Nodes are colored based on their clustering. (d) Distribution of translation initiation and termination events.

only 134 intermediates while reproducing the protein binding curves of the 1,612 parameter network up to an error of 1%. This network included the majority of the intermediates identified experimentally [11] and identified the observed 5' to 3' assembly order. We used NAMD to perform molecular dynamics simulations on these intermediates to investigate how they control the assembly process.

The cell architecture was based on slow-growing *E. coli* of dimensions 4.0 x 0.9 x 0.9 μm^3 with an inner DNA-containing nucleoid region. The cell was discretized into 32 nm^3 lattice sites. The operon sites for r-protein and rRNA were placed within the nucleoid region according to their position on the genome. Experimentally, ribosomes were observed to be excluded from the nucleoid region [12,13], so we biased the transition rates between the cytosol and nucleoid for ribosomes in order to recover this effect when DNA is omitted (fig. 1a). The diffusion constants for the mRNA, ribosomal subunits, full ribosomes, and proteins were taken from experiments [12,13].

We treated the transcription of ribosomal and messenger RNA explicitly and included the active decay of mRNA with rates derived from microarray experiments [14]. Translation of the r-proteins was modeled using explicit initiation and termination steps. The large subunit (50S) was included in the model, but participated only in assembly of the translating ribosome. Though cell division is a planned addition to this model, in the meantime we are treating the loss of species due to dilution as a simple first-order decay reaction. Finally, the assembly reactions were taken from the reduced model (fig 1c).

The resulting model contained 251 species and 1,200 reactions, which made it the most complicated Lattice Microbes model to date. We were able to optimize the code to deal with the larger species and reaction counts through the development of just-in-time code generation for the reaction kernel. Our optimizations and the preliminary just-in-time code increased the performance of Lattice Microbes, such that a 60-minute cell cycle is completed in approximately seven days.

The stochastic, spatially resolved model of ribosome biogenesis allowed us to measure the distribution of assembly times for the 30S, which matched well with experiments. The model predicted spatial heterogeneity in the distribution of assembly intermediates, with the majority of the 5' and central domain proteins bound before the intermediate diffused far from the ribosomal operon (fig. 1b). Binding of the mRNA to the 30S tends to only occur near the originating operon or in the cytoplasm following the dissolution of the translating ribosome (fig. 1d).

WHY BLUE WATERS?

Blue Waters was crucial to the investigation of this problem due to the size of system. Lattice Microbes requires NVIDIA GPUs, and was easily modified to take full advantage of the available K20X accelerators. These simulations are I/O intensive, generating trajectory files approximately 10 GB in size per trajectory, of which many are needed for a proper statistical analysis of the results. Blue Waters provided all the technical resources necessary for us to study this system.

Further work on this project will focus on the integration of other subsystems to build a highly realistic bacterium *in silico*. Using chromosome capture [15,16] and Brownian dynamics [17,18] we will build DNA dynamics into the cell with accurate loci positioning. This will eliminate the need for biased transition and allow for the development of replication. We will build metabolism and general gene expression into our model [18] to couple ribosome biogenesis with DNA dynamics. With metabolism included, exchange reactions between the cell and the environment will allow for the investigation of the system by perturbing the external nutriment conditions.

SIMULATING INFLUENZA HEMAGGLUTININ MEMBRANE ASSEMBLIES

Allocation: GLCPC/0.28 Mnh
PI: Peter Kasson[1]

[1]University of Virginia

EXECUTIVE SUMMARY:

Viruses such as influenza enter cells via a process of membrane fusion between the viral envelope and cellular membranes. Fusion is accomplished by the concerted action of specialized viral coat proteins, but how these proteins work together to achieve fusion is not known. We are using molecular dynamics simulations to understand how influenza fusion proteins interact in lipid membranes, predict structural features of this interaction, and test our understanding via biophysical experiments. Achieving this understanding requires the extensive compute power and fast networking that Blue Waters provides. Success in this endeavor will yield testable hypotheses of how fusion proteins interact and contribute to viral entry mechanisms.

INTRODUCTION

An influenza virus enters cells via a process of membrane fusion, mediated by the viral hemagglutinin protein (which is also the target of most influenza vaccines). Mutagenesis experiments have established that portions of the hemagglutinin protein that insert into host membranes are surprisingly sensitive to mutations, suggesting that protein-membrane interactions are key to viral infectivity. On the other hand, the membrane "anchor" that holds hemagglutinin in the viral envelope is critical to organizing and positioning the protein for fusion but is surprisingly robust to point mutations. The dynamics of protein-membrane interactions and loose protein assemblies in membranes are slow from an atomistic point of view but difficult to probe via experiments. Here, we have used large-scale molecular dynamics simulations to understand the formation of membrane assemblies by the hemagglutinin transmembrane "anchor" and pilot a methodology for exploring such assemblies on a larger scale.

METHODS & RESULTS

We first performed lower-resolution simulations of protein-protein assembly in the membrane. We then used these to calculate a diverse sample of assembly geometries and conducted higher-resolution, more computationally intense simulations. Each of these, a large simulation in itself, was run in parallel (50 simulations per condition) on Blue Waters. Finally, to test the determinants of protein assembly, we predicted a set of mutations to the influenza proteins that abrogated the most stable contacts in the hemagglutinin assemblies. We simulated each of these mutants both under equilibrium conditions and applying external forces to measure the strength of each of the mutated assemblies.

Our results suggest that hemagglutinin transmembrane anchors can form very stable assemblies in the viral envelope. Most interestingly, these assemblies are extremely robust to mutation: If we mutate away the most stable "core" of the assembly, it remains more or less intact. These results yield a structural model for the experimental observation that hemagglutinin function is robust to mutations of this nature; the next question we plan to address in future work is precisely how these assemblies are related to hemagglutinin protein function and how we might be able to perturb this to disable viral infection.

WHY BLUE WATERS

This work relies on having both long simulations and many simulations to achieve rigorous sampling of a complex and slow physical process. Blue Waters combines these features, as it provides many GPU-enabled nodes, each of which can be run efficiently in a few-node parallel configuration. Future Tier-1 systems that may offer more capability for running many moderate-sized simulations that could in turn exchange data efficiently will greatly accelerate both our research and many similar problems in biomolecular simulation.