Computational studies of ribosomal biogenesis: the small subunit

Blue Waters progress report 2013–2014

July 8, 2014

Abstract

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 a bacterial cell's dry mass. Biogenesis of the ribosome, together with all cellular activities involved in translation consume more than half of the cell's total energy. However, the cell's effort to enforce such balance between metabolism and macromolecular synthesis is yet to be recognized. Therefore, the study of ribosome biogenesis is crucial for our understanding of the cell growth and how it is regulated in response to environmental changes (Figure 1a,b). We investigate this process at both the molecular level by molecular dynamics simulations of individual steps in the assembly pathway, and through whole-cell models of the assembly network. The combination of these methods is crucial for our understanding of the cell growth and how it is regulated in response to environmental perturbations.

1 Molecular dynamics

Our first step towards a whole-cell description of ribosome biogenesis is to explore the highly coordinated assembly of the 30S ribosomes, which requires the precise association of 20 ribosomal proteins (r-proteins) with the 16S ribosomal RNA (rRNA). In collaborations with Taekjip Ha (UIUC, Physics & CPLC) and Sarah Woodson (JHU, Biophysics), we have reconstructed the folding landscape of a critical five-way junction in the 16S rRNA upon binding of the primary r-protein uS4 [1,2], and illustrated how uS4 enforces the hierarchy of protein addition by modulating a conformational switch in the 5′ domain rRNA [3]. In the following studies, we demonstrate the cooperation between primary binding proteins uS17 and bS20, and how the association of the secondary binding protein bS16 stabilizes the assembly intermediate by shifting the equilibrium between different conformations [4]. These studies not only guide our way through the early stages of the 30S ribosome assembly, but also greatly enrich our understanding of the protein:RNA recognition and interactions in general.

1.1 Simulations run

ID	Proteins included:	total # atoms	dimensions (Å)	Time (ns)
I	uS4, bS16, uS17, bS20	793,995	259 x 181 x 174	40
II	uS4, uS8, uS17, bS20	755,815	227 x 211 x 162	40
III	uS4, uS15, uS17, bS20	747,581	225 x 212 x 162	40
IV	uS4, uS8, bS16, uS17, bS20	772,630	223 x 212 x 168	40
V	uS4, uS15, bS16, uS17, bS20	742,847	229 x 212 x 158	40
VI	uS4, uS8, uS15, bS16, uS17, bS20	750,368	229 x 212 x 161	40

Table 1: Systems that were run on Blue Waters.

Atomic model of the 5′ domain was built using crystal structure of the *E. coli* ribosomal SSU (PDB code: 2I2P) [5]. In consistency with the previous footprinting experiments, nucleotides 21 to 562 (*E. coli* numbering) of the 16S rRNA, as well as various combinations of ribosomal proteins uS4, uS8, uS15, bS16, uS17, and bS20 (Table 1: *I* to *VI*) were included in the full 5′+ central domain model.

Proteins and nucleic acids were parameterized with the CHARMM36 force fields [6–8]. All systems, as summarized in Table 1, were prepared using the protocol established in Eargle *et al.* [9,10]. Systems were

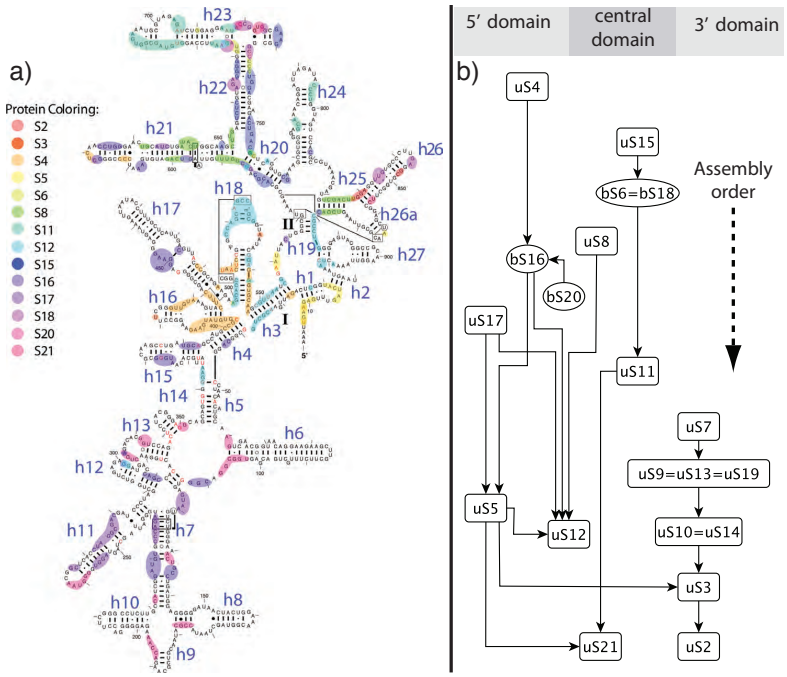

Figure 1: a) Secondary structure map of *Escherichia coli* small subunit with all r-proteins binding locations. b) Nomura map showing the order of r-protein binding in accordance to timing from pulse-chase quantitative mass spectrometry [12]. Order of r-protein binding listed from top to bottom (as shown with dashed arrows). R-proteins are sorted left to right by their primary binding site (from 5′ to the 3′ domain). Solid arrows shows which the thermodynamic binding dependencies.

solvated with at least 10Å of a 200 mM KCl solution. MD simulations were performed using the latest version of NAMD 2.9 [11]. Because the system were run under destabilizing conditions, a 1 fs timestep was used.

1.2 Results

As the r-proteins are removed from native 5′ and central domain, helices that were previously constrained will relax to a new conformation. Large structural changes at the interface fo the two domains might indicate large barriers in the assembly process; some of the biggest changes are found in the intrahelical angle of h21 (Figure 2 a). In system *I*, h21 flucutates between two states (Figure 2 b). With uS8, however, (see systems *II,IV*), h21 becomes more rigid. Similar results, not shown, are seen in systems *III,V,VI*, suggesting that the roles of uS8 and uS15 overlap.

Although uS8 (and uS15) bind to the 5′ domain., These proteins also have long range effects throughout the 5′and central domain. The increased rigidity in the h21, stabilizes the native contact between helices as shown in Figure 2 c. Similar results (not shown) are seen with uS15, suggesting that the primary binding r-proteins in the central domain serve to restrain h21. Additionally simulations are being run to see if the lost of h21 will lead to unfolding of other helices in the central domain and to its separation from the 5′ domain.

1.3 Software

NAMD 2.9 [11] is a highly parallelizable classical MD code designed for large scale biomolecular simulations; it runs on a multitude of different architectures and has been ported to all petascale machines—including Blue Waters. NAMD has near-linear scaling over hundred of thousands of cores and thousands of nodes [13].

To model a biomolecular system, NAMD moves the atoms in accordance to Newton's equation of motion given by:

$$m_\alpha \ddot{\vec{r}}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{total}(\vec{r}_1, \vec{r}_2, ...\vec{r}_N), \alpha = 1, 2, ..., N$$

where m_α is the mass of atom α, $\vec{r}_\alpha$ is its position in space, and U_{total} is its total potential energy as a function

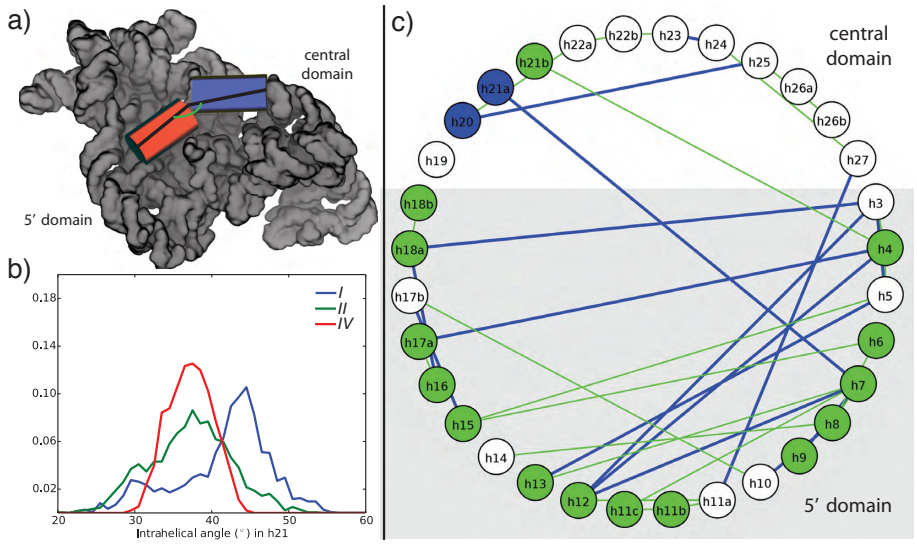

Figure 2: a) Intrahelical angle h21 in the 5′ domain and central domain of the small subunit. Red and blue cylinders shows the two halves of h21 used to calculate the intrahelical angle. Proteins are not shown for clarity. b) Distribution of h21 intrahelical angle. Without uS8, the intrahelical angle in h21 is a bimodal distribution, ranging from 20° to 55°. With uS8 and bS16, however, the helix angle shifts to a monomodal distribution centered at 36°. c) Native contacts between helices. Green lines show native contacts between helices in both simulations *I* and *IV*. Green circles show helices that have native contacts to r-proteins listed in system *I*. Blue lines shows native contacts unique to simulation *IV*. Blue circles show helices that have native contacts to r-proteins listed in system *IV*.

of spatial coordinates. The potential energy of an atom is determined from the CHARMM36 [6–8] force field. Integration of the partial differential equation is carried out by a multiple time stepping variant of the Velocity Verlet integrator.

2 Kinetic modeling

In bacterial cells, the precise synthesis and assembly of a ribosome (Figure 3 requires a number of critical steps [14]: (1) the transcription of pre-ribosomal RNA (rRNA) from the multiple copies of ribosomal DNA, which is the rate limiting step for the ribosome biogenesis in vivo [15]; (2) the synthesis of the ribosomal proteins (r-proteins), which is regulated on the translational level based on organization of the r-protein operons in the genome; (3) post-transcriptional processing and modification of both the rRNA and r-proteins; and (4) the highly coordinated assembly of r-proteins and rRNAs towards the mature ribosomal subunits. Ribosomal assembly requires the cooperation of many molecular components. It requires approximately 55 r-proteins, translated in different regions of the cell, to find and bind with the rRNA in the correct order of assembly. In addition, approximately 20 assembly cofactors are engaged to facilitate the process at various assembly stages.

The intricacy of ribosome assembly first attracted Nomura et al [16], who originally mapped out the hierarchical dependency of the r-proteins binding to the E. coli 16S rRNA using equilibrium reconstitution experiments (see figure 1b). Progress in biophysical approaches boosted our understanding of in vitro ribosomal self-assembly mainly for the protein assisted dynamics of RNA folding [3,17,18], and the kinetic cooperation of protein binding [19–21]. Both aspects agreed that assembly of the E. coli 30S subunit proceeds through multiple parallel pathways nucleated at different positions on the 16S rRNA, and a 5′ to 3′ directionality during 30S ribosomal assembly is always present. However, protein binding orders derived from thermodynamic and kinetic experiments do not always agree with each other, which hampers our investigation of the assembly under in vivo environment. Therefore, a comprehensive model that captures the topology of the protein-RNA interaction network is needed to decipher the underline rules governing the assembly of the ribosome.

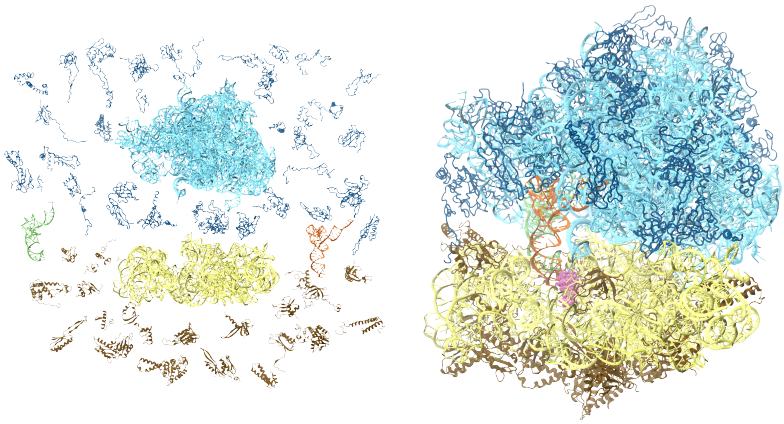

Figure 3: During the assembly of the small subunit of the bacterial ribosome, approximately 20 ribosomal proteins must bind with the 16S rRNA.

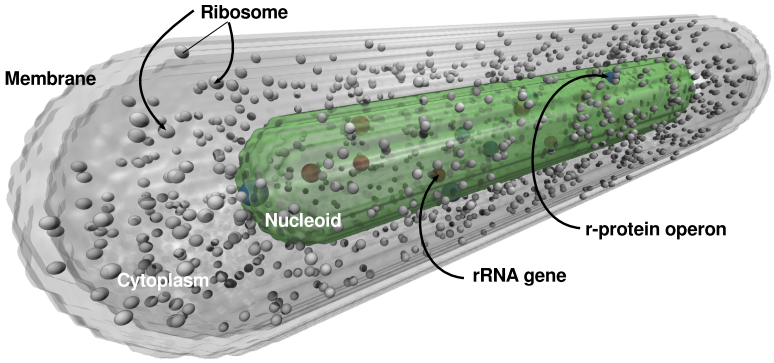

Figure 4: The small subunit assembly network is simulated within the environment of a living cell. Using our Lattice Microbes whole-cell simulation software [22,23], we are investigating the spacio-temporal effects of a complicated cellular environment on ribosomal biogenesis. The genes encoding the ribosomal proteins and rRNA are placed according to their location in the genome and allowed to diffuse throughout the nucleoid region. Ribosomes are placed within the cytoplasm according to their experimental distribution. In addition to the ribosomal protein binding network, transcription of r-proteins and rRNA, and translation of r-protein mRNA are simulated. With the effect of transcription and translation, a realistic simulation of in vivo ribosome biogenesis can be performed.

2.1 Accomplishments

To address the global complexity in in vivo ribosome biogenesis, we simulated a kinetic model of the 30S ribosome assembly using the Lattice Microbe software (LM) on Blue Waters. LM [22–24] is a package of stochastic solvers for simulating the kinetics of biological systems. There are two classes of simulations: those solved using the Chemical Master Equation (CME) that are assume that all molecules in the simulation are able to interact with each other at any time, and those solved with the Reaction Diffusion Master Equation (RDME) that incorporates spatial information and molecules can only interact others that are nearby. Diffusion allows molecules to move around its environment and may be subject to different kinetics depending on where it is. It allows for a more realistic description of biological systems in vivo than the CME can provide (figure 4). Molecular crowding and initial distributions of ribosomes within the cells are obtained from proteomics and cryo-electron tomography reconstructions. [25]

The construction of the small subunit requires the binding of 20 individual proteins to the 16S rRNA. If all possible binding reactions are treated fairly, a combinatorial explosion of network complexity results. This leads to on the order of 106 intermediates and 1018 reactions. We instead construct a model that explicitly uses the dependencies of the Nomera map [16] to decreases the size of the network to a manageable level of 1,633

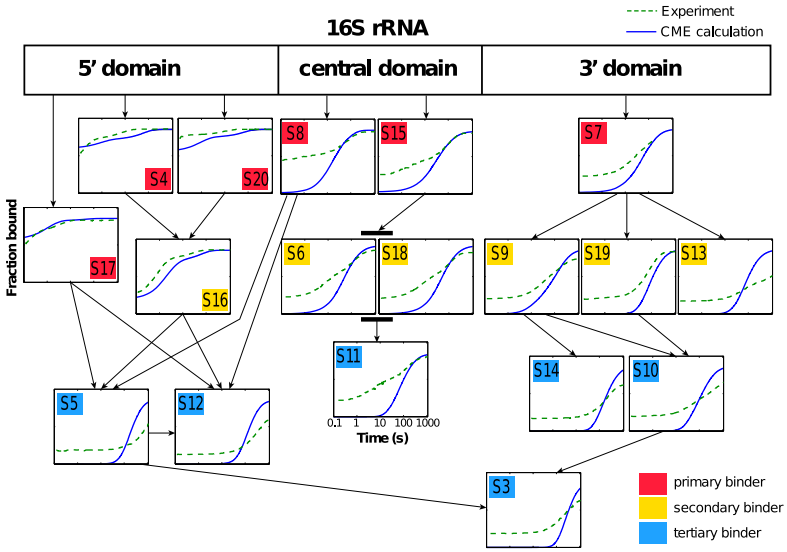

Figure 5: The binding rates of the ribosomal proteins are highly interdependent. Each curve is the fraction of a protein bound to the nascent ribosome as a function of time on a logarithmic scale. The experimental results from [21] are plotted against the predicted curves of our model showing a reasonable level of agreement. The arrows point from an r-protein to later proteins which are dependent on their binding.

species and 7,000 reactions, taking effective r-protein binding rates published by the Williamson lab [21] (figure 5). However this is still too complex to deal with on a spatially resolved level. To further reduce the network size, we use well-stirred stochastic simulations to identify the intermediates in which the majority of the reaction flux flows through. Intermediates which are underutilized by the reaction network are removed from the network along with their associated reactions. This analysis allowed us to reduce the assembly network to 62 species and 69 reactions.

To test the validity of this severely pruned network, we compared the well-stirred simulations using the full network to simulations of the reduced network. The experimentally observable parameters here are the protein binding curves, which show the fraction of r-protein bound to intermediates as a function of time. We saw no greater than 0.1 % root-mean-square error between the two sets of binding curves. After subsequent tuning with the competing folding conformations identified in our previous studies [1–4], the model has successfully reproduced the structural intermediates reported in the single particle electron microscopy experiments [12]. Furthermore, the model predicted new assembly intermediates that will guide further experimental discoveries (figure 6).

Stochastic simulations require multiple trajectories of the simulation be performed in order to adequately make statistical inferences on the results. Since we sample the RDME on a lattice, the problem is well suited for taking advantage of the high parallelism offered by GPU computing. Our software requires a CUDA-capable GPU for accelerated computation of every trajectory, thus requiring a potentially large number of GPUs to finish the simulation in a timely manner.

In order to fully understand how the system reacts to changing parameters a thorough parameter sensitivity analysis was performed. This required an array of simulations (at least 10) to be run for every parameter (69 parameters). We could not have run this without Blue Waters and the ample number of accelerated nodes outfitted with state-of-the-art GPUs. Our long range goal is to unite the kinetic model of translation with other cellular networks extending over several cycles of cell division so that whole cell simulations of bacteria responding to various stimuli and environmental factors can be achieved.

We have begun preliminary whole-cell simulations. Lattice Microbes is compiled and works well on Blue Waters. We are not utilizing the full potential of BW at this point since our whole-cell simulations are only using a single GPU per replicate. However, a version that spans multiple nodes over MPI is in development, and features such as Blue Water's high-speed interconnect, GPU-to-fabric DMA, and a highly parallel filesystem will be necessary for a performant RDME solver.

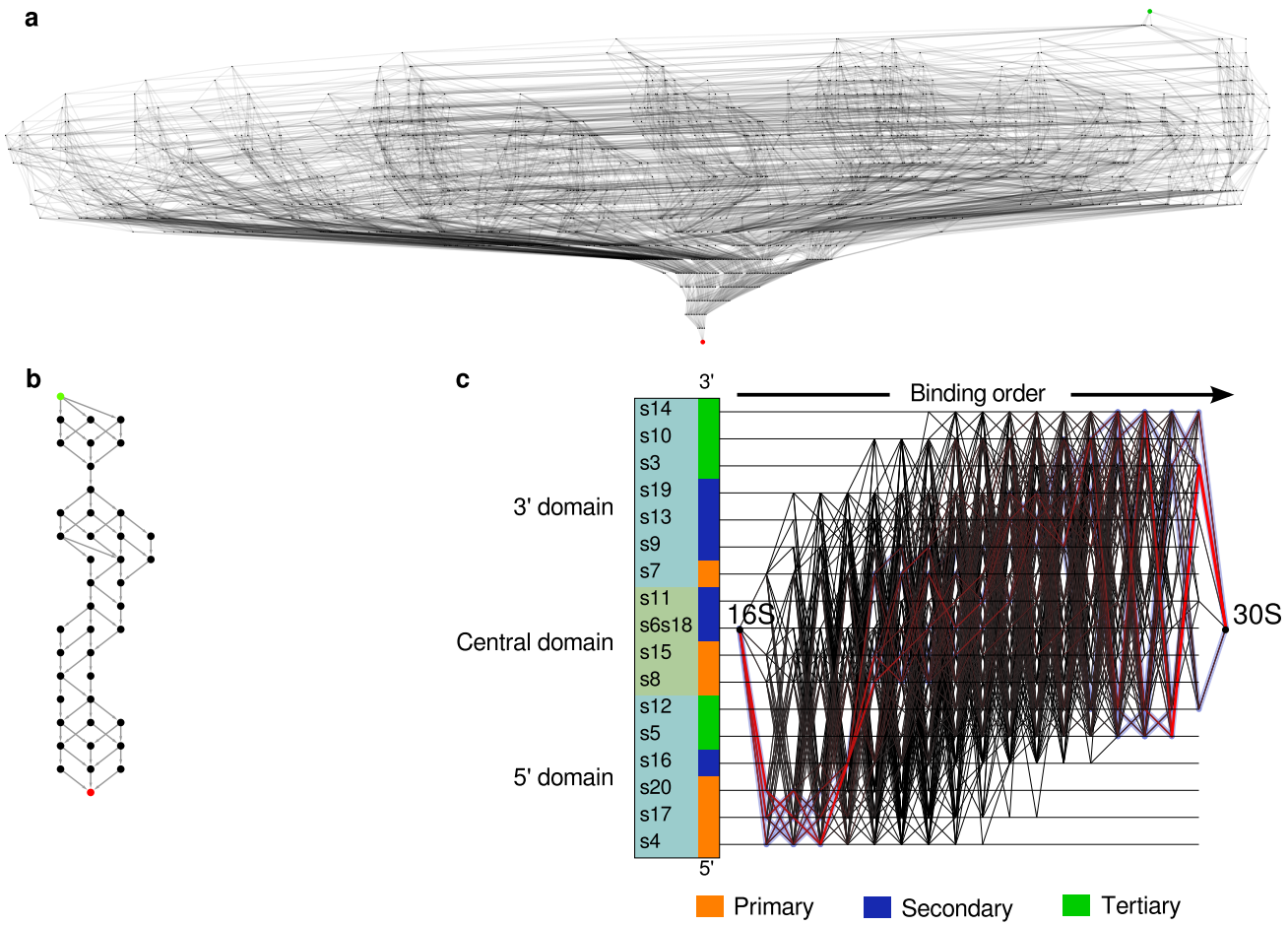

Figure 6: (a) The full reaction network of ribosomal intermediates during the binding of ribosomal proteins to the 16S rRNA (green node) to form the 30S subunit (red node). There are 1633 intermediates connected by 7000 reactions. (b) The full network can be reduced significantly by the analysis of flux through each intermediate. By eliminating intermediates whose flux is less than 0.7% of that of the most active species, we are able to prune the network down to 62 species and 69 reactions. (c) From kinetic simulations of the assembly of single small subunits we plot the order of r-protein binding. Each vertical line represents a ribosomal protein, ordered by binding domain (5′ to 3′) and by experimental binding order (primary [orange] to secondary [blue] to tertiary [green]). The binding paths are drawn according to reaction flux increasing from light gray to red. The highlight shows the paths from the reduced network. This assembly order agrees with the kinetic assembly maps determined in [4].

References

[1] Chen, K, Eargle, J, Sarkar, K, Gruebele, M, Luthey-Schulten, Z (2010) Functional role of ribosomal signatures. *Biophys. J.* 99:3930 – 3940.

[2] Chen, K, Eargle, J, Lai, J, Kim, H, Ha, T, Abeysirigunawardena, T, Mayerle, M, Woodson, S, Luthey-Schulten, Z (2012) Assembly of the fiveway junction in the ribosomal small subunit using hybrid MD/Gō simulation. *J. Phys. Chem. B* 116:6819–6831.

[3] Kim, H, Abeysirigunawarden, SC, Chen, K, Mayerle, M, Ragunathan, K, Luthey-Schulten, Z, Ha, T, Woodson, SA (2014) Protein-guided rna dynamics during early ribosome assembly. *Nature* 506:334–338.

[4] Lai, J, Chen, K, Luthey-Schulten, Z (2013) Structural intermediates and folding events in the early assembly of the ribosomal small subunit. *J. Phys. Chem. B* 42:13335–13345.

[5] Berk, V, Zhang, W, Pai, RD, Cate, JHD (2006) Structural Basis for mRNA and tRNA Positioning on the Ribosome. *Proc. Natl. Acad. Sci. USA* 103:15830–15834.

[6] Best, RB, Zhu, X, Shim, J, Lopes, PEM, Mittal, J, Feig, M, MacKerell, AD (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone , and Side-Chain 1 and 2 Dihedral Angles. *J. Comput. Theor. Chem.* 8:3257–3273.

[7] Denning, EJ, MacKerell, AD (2012) Intrinsic Contribution of the 2'-Hydroxyl to RNA Conformational Heterogeneity. *J. Am. Chem. Soc.* 134:2800–2806.

[8] Hart, K, Foloppe, N, Baker, CM, Denning, EJ, Nilsson, L, MacKerell, AD (2012) Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Comput. Theor. Chem.* 8:348–362.

[9] Eargle, J, Black, AA, Sethi, A, Trabuco, LG, Luthey-Schulten, Z (2008) Dynamics of recognition between tRNA and elongation factor Tu. *J. Mol. Biol.* 377:1382 – 1405.

[10] Eargle, J, Luthey-Schulten, ZA (2012) RNA 3D Structure Analysis and Prediction, eds Leontis, N, Westhof, E pp 213–238.

[11] Phillips, JC, Braun, R, Wang, W, Gumbart, J, Tajkhorshid, E, Villa, E, Chipot, C, Skeel, RD, Kale, L, Schulten, K (2005) Scalable Molecular Dynamics with NAMD. *J. Comp. Chem.* 26:1781–1802.

[12] Mulder, A, Yoshioka, C, Beck, A, Bunner, A, Milligan, R, Potter, C, Carragher, B, Williamson, J (2010) Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30S Subunit. *Science* 330:673–677.

[13] Mei, C, Sun, Y, Gengbin, Z, Bohm, EJ, Kale, LV, Phillips, JC, Harrison, C (2011) Enabling and Scaling Biomolecular Simulations of 100 Million Atoms on Petascale Machines with a Multicore-optimized Message-drive Runtime.

[14] Kaczanowska, M, Rydén-Aulin, M (2007) Ribosome biogenesis and the translation process in escherichia coli. *Microbiol. Mol. Biol. Rev.* 71:477–494.

[15] Paul, BJ, Ross, W, Gaal, T, Gourse, RL (2004) rrna transcription in escherichia coli. *Ann. Rev. Gen.* 38:749–770 PMID: 15568992.

[16] Held, WA, Ballou, B, Mizushima, S, Nomura, M (1974) Assembly mapping of 30 s ribosomal proteins from escherichia coli : Further studies. *J. Biol. Chem.* 249:3103–3111.

[17] Adilakshmi, T, Ramaswamy, P, Woodson, SA (2005) Protein-independent folding pathway of the 16s rrna 5′ domain. *J. Mol. Biol.* 351:508–519.

[18] Adilakshmi, T, Bellur, DL, Woodson, SA (2008) Concurrent nucleation of 16s folding and induced fit in 30s ribosome assembly. *Nature* 455:1268–1272.

[19] Talkington, MW, Siuzdak, G, Williamson, JR (2005) An assembly landscape for the 30s ribosomal subunit. *Nature* 438:628–632.

[20] Sykes, MT, Williamson, JR (2009) A complex assembly landscape for the 30s ribosomal subunit. *Ann. Rev. Biophys.* 38:197–215.

[21] Bunner, AE, Beck, AH, Williamson, JR (2010) Kinetic cooperativity in escherichia coli 30s ribosomal subunit reconstitution reveals additional complexity in the assembly landscape. *Proc. Natl. Acad. Sci. USA* 107:5417–5422.

[22] Hallock, MJ, Stone, JE, Roberts, E, Fry, C, Luthey-Schulten, Z (2014) Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-gpu workstations. *Parallel Computing* 40:86–99.

[23] Roberts, E, Stone, JE, Luthey-Schulten, Z (2013) Lattice microbes: high-performace stochastic simulation method for the reaction-diffusion master equation. *J. Comp. Chem.* 3:245–255.

[24] Peterson, JR, Hallock, MJ, Cole, JA, Luthey-Schulten, ZA (2013) A Problem Solving Environment for Stochastic Biological Simulations, *Supercomputing 2013*.

[25] Roberts, E, Magis, A, Ortiz, JO, Baumeister, W, Luthey-Schulten, Z (2011) Noise contributions in an inducible genetic switch: A whole-cell simulation study. *PLoS Comput. Biol.* 7:e1002010.