

Article

Toward a Whole-Cell Model of Ribosome Biogenesis: Kinetic Modeling of SSU Assembly

Tyler M. Earnest,^{1,2} Jonathan Lai,³ Ke Chen,^{3,4} Michael J. Hallock,⁵ James R. Williamson,^{6,7,8} and Zaida Luthey-Schulten^{1,2,3,*}

¹Center for the Physics of Living Cells, ²Department of Physics, and ³Department of Chemistry, University of Illinois, Urbana, Illinois;

⁴Department of Bioengineering, University of California, San Diego, La Jolla, California; and ⁵School of Chemical Sciences, University of Illinois, Urbana, Illinois; ⁶Department of Integrative Structural and Computational Biology, ⁷Department of Chemistry, and ⁸Skaggs Institute for Chemical Biology, Scripps Research Institute, La Jolla, California

ABSTRACT Central to all life is the assembly of the ribosome: a coordinated process involving the hierarchical association of ribosomal proteins to the RNAs forming the small and large ribosomal subunits. The process is further complicated by effects arising from the intracellular heterogeneous environment and the location of ribosomal operons within the cell. We provide a simplified model of ribosome biogenesis in slow-growing *Escherichia coli*. Kinetic models of in vitro small-subunit reconstitution at the level of individual protein/ribosomal RNA interactions are developed for two temperature regimes. The model at low temperatures predicts the existence of a novel 5' → 3' → central assembly pathway, which we investigate further using molecular dynamics. The high-temperature assembly network is incorporated into a model of in vivo ribosome biogenesis in slow-growing *E. coli*. The model, described in terms of reaction-diffusion master equations, contains 1336 reactions and 251 species that dynamically couple transcription and translation to ribosome assembly. We use the Lattice Microbes software package to simulate the stochastic production of mRNA, proteins, and ribosome intermediates over a full cell cycle of 120 min. The whole-cell model captures the correct growth rate of ribosomes, predicts the localization of early assembly intermediates to the nucleoid region, and reproduces the known assembly timescales for the small subunit with no modifications made to the embedded in vitro assembly network.

INTRODUCTION

Translation is the universal process that synthesizes proteins in all living cells. Sequence (and structural) signatures in the ribosomal RNA were used to classify all living organisms into the three domains of life (1,2). Ribosomal proteins (r-proteins) can themselves be signatures of ribosomal evolution and, in the case of bacteria, roughly one-third of them are unique, with the remaining ones common to all three domains of life (2,3). Ribosomes constitute approximately one-fourth of a bacterial cell's dry mass, and biogenesis of the ribosome, together with the other cellular processes involved in translation, consume a significant fraction of the cell's energy budget. A whole-cell model of ribosome biogenesis is crucial for our understanding of cell growth, yet a comprehensive dynamical description of the biogenesis process is still missing.

In bacteria, the precise synthesis and assembly of a ribosome (4) involves at least four critical steps: transcription of ribosomal RNA from multiple ribosomal operons; synthesis of the r-proteins, which is regulated on the translational level based on organization of the r-protein operons in the genome; posttranscriptional processing and modification of both the ribosomal RNA (rRNA) and r-proteins; and

highly coordinated assembly of r-proteins and rRNA toward the mature ribosomal subunits. All these events occur constantly and in parallel throughout the cell cycle.

Ribosomal assembly involves the cooperation of many molecular components. The 30S small subunit (SSU), tasked with the initial binding of messenger RNA (mRNA) and its decoding, is composed of the 16S rRNA and 21 r-proteins. The 50S large subunit (LSU), tasked with channeling growth of the nascent polypeptide chain through peptide bond formation, is composed of the 5S and 23S rRNA and 33 r-proteins. These 54 proteins must diffuse through the cell to find their rRNA and bind in a well-defined assembly order. These proteins are classified by their order of binding to the rRNA. Primary proteins bind to the bare rRNA, secondary proteins require the presence of certain primary proteins to bind, and tertiary proteins require the presence of a secondary protein to bind. The r-proteins can compose 9–22% of the total protein counts in the cell (5,6). In addition, ~20 assembly cofactors are engaged to facilitate the process at various assembly stages.

The rich complexity of the 30S assembly process attracted Nomura et al. (7), who first observed how the binding stability of r-proteins can depend on the prior binding of other r-proteins. Using equilibrium reconstitution experiments at temperatures optimal for the growth of *Escherichia coli* (37°C), Nomura constructed a hierarchical dependency

Submitted March 20, 2015, and accepted for publication July 13, 2015.

*Correspondence: zan@illinois.edu

Editor: Ozlem Keskin.

© 2015 by the Biophysical Society

0006-3495/15/09/1117/19



<http://dx.doi.org/10.1016/j.bpj.2015.07.030>

map of the assembly process (Fig. 1). Progress in biophysical approaches has increased our understanding of in vitro ribosomal self-assembly through the protein-assisted dynamics of RNA folding (8–10) and the kinetic cooperativity of protein binding (11–15). All of the studies suggest that assembly of the *E. coli* 30S subunit proceeds through multiple parallel pathways, first binding the proteins associated with the 5' domain of the 16S rRNA, then the central-domain proteins, and finally the 3'-domain proteins.

Using the Nomura map of thermodynamic binding dependencies and kinetic data of protein incorporation, we have constructed comprehensive in vitro kinetic models that capture the topology of the r-protein/rRNA interaction network and reproduce the protein-binding kinetics of assembly, starting from the bare 16S rRNA or from preprepared assembly intermediates, at low and high temperatures (13,14). Both models are consistent with an assembly mechanism inferred from cryo-electron microscopy (cryoEM) of 30S assembly intermediates. Molecular dynamics (MD) simulations of the early intermediates in the in vitro assembly model suggest a molecular basis for the two distinct assembly pathways predicted by the low-temperature kinetic model. The low-temperature model reproduces all of the control and prebinding experimental kinetics (14,15). Furthermore, both models predict intermediates central to the assembly process that would be good candidates for further experimental and computational studies.

The in vivo biogenesis of the ribosome is further complicated by spatial segregation of the ribosomes from the nucleoid region (16–20). Cryo-electron tomograms and single-molecule experiments have indicated that the full 70S ribosomes (16,21) are partitioned such that 80% are found outside of the nucleoid region; however, the 30S and 50S subunits are found uniformly throughout the cell (20). In

slow-growing *E. coli* (grown in minimal media), roughly 3000 ribosomes accumulate at the cell poles and are almost entirely excluded from the nucleoid (16,17). In living *E. coli* cells, there can be as little as one copy of the gene coding for an r-protein. Due to the relatively small number of 30S particles in the process of assembly and the large range of possible intermediates, the counts of specific 16S/r-protein configurations can be of the order of one per cell. To describe the effects and fluctuations arising from the spatial segregation of ribosomes and the low copy number of genes and assembly intermediates, a spatially resolved representation accounting for the discreteness of chemical species is essential for a more realistic treatment of the problem (22).

We present a detailed reaction-diffusion master-equation (RDME) representation of the in vivo biogenesis of the SSU, incorporating the spatially inhomogeneous environment of the cell and the stochastic nature of chemical reactions. We have adapted our high temperature in vitro assembly model—developed from kinetic studies utilizing pulse/chase quantitative mass spectrometry (P/C qMS)—to an in vivo model of ribosome biogenesis including transcription of mRNA and rRNA from DNA localized at their genetic loci, translation of r-protein, and loss of species due to active degradation of mRNA and dilution arising from cell division. The cell is compartmentalized into cytoplasm and nucleoid regions, which can have different diffusion and intercompartmental transition rates for each chemical species. Our models of in vivo 30S biogenesis based on slow-growing *E. coli* (16,21) roughly reproduce the timescale for assembly seen in live cells and predict spatial inhomogeneity in the assembly process.

MATERIALS AND METHODS

Generation of assembly networks

The network of r-protein association reactions is constructed programmatically by iteratively adding species and reactions according to a rule list. The reaction rule list is a representation of the Nomura map of thermodynamic binding dependencies, in which the binding of a protein to an intermediate is thermodynamically stable only if all of that protein's upstream dependencies are bound. Starting with a stack containing only bare rRNA, an intermediate is removed from the top of the stack and stored in a list of visited species. All possible binding reactions from this species are computed using the reaction rules and their products are only added to the top of the stack if they have not been previously visited. This process is iterated until the stack is empty.

Another rule set is used to assign rate constants to the generated reactions (see Table 1). A sequence of rate rules is defined for each r-protein. These rules consist of additional requirements on the composition of the intermediate independent of the thermodynamic dependencies. To choose the rate parameter for that reaction, each rule is tested in order and the first to succeed is applied to the reaction. These rates are derived from kinetic experiments using preprepared intermediates with various proteins bound to the rRNA. For the low-temperature model, a rich variety of prebinding experiments are available from which to derive these rules. For the high-temperature model, no prebinding data are available, so only one parameter is used for the binding of a protein to any intermediate. Parameter values are given in Table S2 in the Supporting Material.

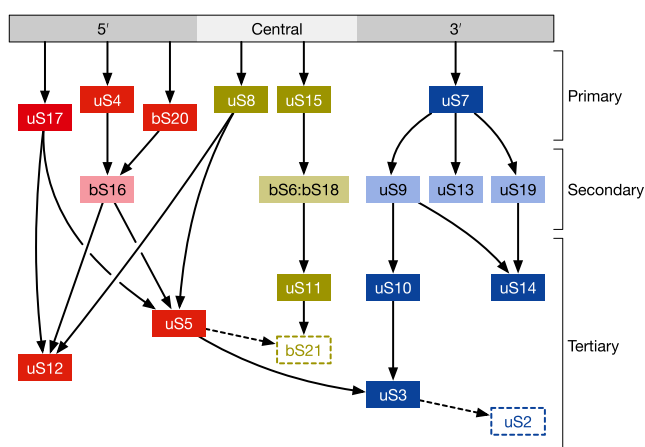


FIGURE 1 Graph of thermodynamic protein binding dependencies to the 16S rRNA (7). Only the major dependencies used in the in vitro model are depicted here. Arrows point from a protein to the protein that is dependent on it. uS2 and bS21, shown in open rectangles, are not included in these models, due to difficulties in acquiring their kinetic data (13). To see this figure in color, go online.

TABLE 1 Assembly rate constants for the in vitro ribosome biogenesis kinetic model at 15°C

Protein	Symbol	No. of Reactions	Experiment	Rate ($\mu\text{M}^{-1} \text{ s}^{-1}$)		Rules	
				Initial	Optimized	Present	Absent
5' Domain							
uS4	$k_{4,\text{o7}}$	32	uS7	1.713×10^{-1}	2.918×10^{-1}	uS7	uS9, uS13, or uS19
	$k_{4,\text{def}}$	512	control	8.383×10^{-2}	2.173×10^{-1}		
uS17	$k_{17,13\text{o}19}$	120	uS7 and uS19	5.285×10^{-2}	1.152×10^{-1}	uS13 or uS19	uS9
	$k_{17,\text{def}}$	560	control	1.421×10^{-1}	1.614×10^{-1}		
bS20	$k_{20,7}$	32	uS7	4.483×10^{-1}	9.325×10^{-1}	uS7	uS9, uS13, or uS19
	$k_{20,\text{def}}$	512	control	2.005×10^{-1}	4.968×10^{-1}		
bS16	$k_{16,\text{def}}$	272	1°	5.103×10^{-2}	7.655×10^{-2}		
uS5	$k_{5,\text{def}}$	136	1° and 2°	7.29×10^{-4}	1.701×10^{-4}		
uS12	$k_{12,\text{def}}$	160	1° and 2°	1.895×10^{-3}	1.806×10^{-4}		
Central Domain							
uS8	$k_{8,7\text{r}9}$	120	uS7 and uS9	2.223×10^{-2}	3.419×10^{-2}	uS7 or uS9	uS13 or uS19
	$k_{8,13}$	320	uS7 and uS13	6.488×10^{-3}	3.429×10^{-3}	uS13	
	$k_{8,\text{def}}$	240	control	1.531×10^{-3}	4.52×10^{-4}		
uS15	$k_{15,13\text{o}19}$	92	uS7 and uS13	5.176×10^{-4}	MIN	uS7, uS13, or uS19	uS9
	$k_{15,\text{def}}$	311	control	1.276×10^{-3}	1.265×10^{-3}		
bS6:bS18	$k_{6,\text{def}}$	403	1°	1.257×10^{-1}	2.89×10^{-1}		
uS11	$k_{11,\text{def}}$	403	1°	1.166×10^{-2}	2.441×10^{-2}		
3' Domain							
uS7	$k_{7,5\text{c}}$	1	5' and cent.	2.333×10^{-3}	5.146×10^{-3}	5' and cent.	
	$k_{7,\text{def}}$	91	control	7.654×10^{-4}	1.665×10^{-3}		
uS9	$k_{9,19}$	184	uS7 and uS19	1.786×10^{-1}	4.456×10^{-1}	uS19	
	$k_{9,13}$	92	uS7 and uS13	2.989×10^{-3}	3.007×10^{-3}	uS13	
	$k_{9,5\text{c}}$	8	5', cent., and uS7	4.374×10^{-3}	1.027×10^{-3}	1° and 2° of 5' and central	
	$k_{9,\text{pri}}$	7	1°	8.019×10^{-4}	MIN	1°	
	$k_{9,\text{def}}$	77	uS7	1.713×10^{-2}	2.572×10^{-2}		
uS13	$k_{13,19}$	476	uS7 and uS19	1.13×10^{-1}	1.134×10^{-1}	uS19	
	$k_{13,\text{pri}}$	51	1°	2.187×10^{-3}	MIN	1°	
	$k_{13,\text{def}}$	233	uS7	4.009×10^{-4}	MIN		
uS19	$k_{19,\text{pri}}$	102	1°	1.713×10^{-3}	1.838×10^{-3}	1°	
	$k_{19,\text{def}}$	466	uS7	1.093×10^{-3}	5.718×10^{-4}		
uS3	$k_{3,\text{def}}$	48	1° and 2°	1.13×10^{-3}	3.703×10^{-2}		
uS10	$k_{10,19}$	368	uS7 and uS19	4.592×10^{-2}	6.584×10^{-2}	uS19	
	$k_{10,\text{def}}$	184	uS7 and uS9	4.738×10^{-4}	4.008×10^{-4}		
uS14	$k_{14,\text{def}}$	384	1° and 2°	1.749×10^{-3}	1.173×10^{-3}		

The 32 parameters in the ribosome assembly kinetic model shown are separated by domains and listed in decreasing rule precedence. The initial reaction rate constants are estimated from (13), and the final reaction rate from global optimization are shown for each parameter. The parameters are sorted by decreasing rule precedence. If an intermediate does not satisfy the rules for a parameter (presence or absence of certain r-proteins), the next parameter in the list is tested. MIN indicates that the local optimizer has driven this parameter to the lower limit of $4 \times 10^{-6} \mu\text{M}^{-1}\text{s}^{-1}$.

Deterministic modeling and optimization of rate constants at low temperature

The in vitro binding process at 15°C is simulated using the same initial conditions used in the P/C qMS study, which had a 50% excess of r-protein over the 16S rRNA (0.458 μM r-protein versus 0.305 μM 16S rRNA) (13). The system of ordinary differential equations is solved numerically using the CVODES package (23) (solver equations derived in Section S1 of the [Supporting Material](#)). Goodness of fit to the experimental protein binding curves is measured using the objective function

$$\Phi(\{k_i\}) = \frac{1}{\mathcal{N}_{\text{expt}} \mathcal{N}_{\text{prot}} (T_1 - T_0)} \sum_{e \in \{\text{expts}\}} \int_{T_0}^{T_1} \frac{dt}{t} \times \sum_{s \in \{\text{r-prot}\}} \left[\chi(y_{e,s}(t)) - \chi_{e,s}^{\text{expt}}(t) \right]^2, \quad (1)$$

which computes the MSE between the experimental and simulated assembly progress curves for the parameters $\{k_i\}$. Here, $y_{e,s}(t)$ is the protein concentration, s , at time t starting from the initial prebinding intermediate, e , $\chi_{e,s}^{\text{expt}}(t)$ is a single exponential fit to the actual P/C qMS experiment, and

$$\chi(y) = \frac{p_0}{p_0^* + p_0} + \frac{p_0^*(p_0^* - r_0 + p_0)}{r_0(p_0^* + p_0)} \left(\frac{p_0 - y}{p_0^* + y} \right), \quad (2)$$

converts protein concentrations to an idealized pulse/chase fraction where p_0 is the concentration of labeled protein due to the pulse, p_0^* is the concentration of unlabeled protein due to the chase, and r_0 is the initial rRNA concentration. This assumes that binding is irreversible and all rRNA is converted to intermediates (derived in Section S2 of the [Supporting Material](#)). The integration is performed over the same time interval as the experiment, with a weighting of $1/t$ to treat each decade in time equally. Using the adjoint sensitivity analysis capabilities of the CVODES package, we are able to compute the gradient of Eq. 1 with respect to the reaction rates to

enable rapid minimization of the objective function using a gradient-based optimization algorithm.

The rate constants are derived from single-exponential fits to the kinetic data. This exponential rate is converted into a second-order rate constant by assuming that the protein concentration remains constant over the assembly process. At 50% excess, this is a poor approximation and the converted second-order rate constant will not be measuring the binding rate directly for the secondary and tertiary proteins, but instead will measure a composite rate that includes the time for the dependent proteins to bind. We will use local optimization from these initial values using the L-BFGS method (24) informed with true gradient information from CVODES to find proper second-order rate constants for these reactions.

Reduction of the kinetic model

To increase the speed of our whole-cell simulations, the assembly network must be pruned of species which do not contribute significantly to the assembly process. This is accomplished by iteratively removing the species, s , that contributes the least to the total amount of 30S assembled. This contribution is quantified as the total reaction flux consuming that species, \mathcal{F}_s , which is computed from the integral

$$\mathcal{F}_s = \sum_{r \in \mathcal{R}_s} \int_{T_0}^{T_1} dt k_r [\mathbf{P}_r][\mathbf{I}_s], \quad (3)$$

where the summation is over all reactions consuming species s . The quality of the reduced low-temperature model is monitored by computing the root MSE (RMSE) of the protein binding curves between the initial and modified networks. The modified network with the minimal number of intermediates not exceeding the error tolerance of 2×10^{-2} is accepted. Due to the limited data available for the high-temperature model, we instead monitor the difference in free protein half-lives between the reduced and unmodified models and accept the smallest network that does not exceed an average of 6% \log_{10} difference in half-lives.

Construction of the ribosomal biogenesis network in vivo

The in vivo biogenesis model consists of the assembly network determined from the in vitro data at 40°C, as well as transcription, translation, mRNA degradation, and dilution reactions, along with the cellular geometry and diffusion constants for all species. Transcription is modeled as a first-order birth process, where RNA production is localized at points in the cell representing their originating operon in the genome. The rates of the mRNA and rRNA birth processes are tuned to an intended expression level, with no gene regulation included in the model. SSU components are produced from nine r-protein and seven rRNA operons placed throughout the cell according to their genomic position. Assembly of the LSU is not included in this model. Instead, the LSU is introduced into the system as a zeroth-order birth process that creates LSU species uniformly throughout the cell at a rate matching 16S rRNA expression to ensure that the 30S and 50S copy numbers remain balanced.

The rates for translation depend on the operon structure taken from the *E. coli* K-12 MG1655 genome (accession number U00096 (25); genomic data processed using Biopython (26)). Translation elongation is modeled by a series of reactions. Each reaction represents the combination of the formation of an r-protein associated and the advancement of the ribosome along the transcript to the next r-protein gene. The transition rate between positions along the mRNA is simply the translation rate per nucleotide divided by the number of bases between the start of the protein created during this step and the beginning of the next protein to be produced (or the end of the transcript). The lengths of intervening genes that code for proteins not included in the model are included in the genomic distance used to compute the transition rate. Rates of transcription from the operons considered in our

model are chosen such that the proteins reach a realistic steady-state concentration. The values of parameters used in the in vivo model are summarized in Table 2. All parameter values are reported in Table S3.

Spatially resolved simulations of the in vivo biogenesis network

Spatially resolved chemical reaction trajectories are sampled from the solution to the RDME describing the in vivo network and cell geometry discretized onto a lattice. The RDME is

$$\begin{aligned} \frac{dP(\mathbf{x}, t)}{dt} = & \sum_v \sum_r \left[-a_r(\mathbf{x}_v) P(\mathbf{x}_v, t) \right. \\ & \left. + a_r(\mathbf{x}_v - \mathbf{S}_r) P(\mathbf{x}_v - \mathbf{S}_r, t) \right] \\ & + \sum_v \sum_{\xi} \sum_{\alpha} \left[-d_v^{\alpha} x_v^{\alpha} P(\mathbf{x}, t) \right. \\ & \left. + d_{v+\xi}^{\alpha} (x_{v+\xi}^{\alpha} + 1) P(\mathbf{x} + \mathbf{1}_{v+\xi}^{\alpha} - \mathbf{1}_v^{\alpha}, t) \right], \quad (4) \end{aligned}$$

where $P(\mathbf{x}, t)$ is the probability distribution to find a configuration \mathbf{x} at time t . The configuration vector \mathbf{x} contains the number of species present at each individual lattice site. The first term in Eq. 4 describes the flow of probability between different copy-number states at every lattice site. The reaction propensities $a_r(\mathbf{x}_v)$ give the transition probabilities for reaction r at site v . The r row of the stoichiometry matrix \mathbf{S} is the change in species counts when reaction r occurs. The second term describes the flow of probability due to diffusion between neighboring lattice sites, indexed by ξ . Here, d_v^{α} is the diffusive propensity for species α in volume v to leave its lattice site. Lattice Microbes (27), a software package designed to simulate stochastic reaction-diffusion systems using the multiparticle-diffusion RDME (MPD-RDME) algorithm (28–30), is used to sample trajectories from the solution to Eq. 4. This software is highly optimized to take advantage of GPGPU computing on NVIDIA hardware, allowing for simulation times reaching cell-cycle timescales.

Since this is the most complex RDME model simulated by Lattice Microbes to date, modifications to the code base were necessary to increase the performance of models with many chemical species and reactions. The reaction kernel, responsible for selecting the reaction and performing the update of species counts at each time step, was replaced with a programmatically generated code with all loops unrolled and all constant factors to the propensity calculations replaced with immediate values. This leads to a speed-up allowing for an hour of simulation time to complete within ~3 days.

Lattice Microbes (LM 2.2.1) simulations were executed on the XK7 nodes of NCSA Blue Waters (AMD 6276 Interlagos/NVIDIA Tesla K20X GPU accelerators using CUDA 6.5) for short trajectories (<10 min) over 64 simultaneous replicates. Replicates covering an entire cell cycle were performed on a local machine (2× Intel Xeon CPU E5-2640/4× NVIDIA GeForce GTX 980 GPUs using CUDA 6.5) allowing for four simultaneous replicates.

MD simulations of early intermediates

Atomic models of the assembly intermediates are built using the crystal structure of the *E. coli* ribosomal SSU (PDB 212P) (31). Proteins and nucleic acids are parameterized with the CHARMM36 (32,33) force fields. All systems are prepared using the protocol described in Section S3 of the Supporting Material. Systems are neutralized with sodium ions. A total of 840 ns of MD simulation on the 16S intermediates are reported.

Production runs are conducted using NAMD 2.10 (34) under the NPT ensemble at 1 atm and 300 K. Periodic boundary conditions are applied, and a 1-fs-2-fs-4-fs multiple-time-stepping approach was used. Long-range

TABLE 2 Summary of reactions and rate constants for the *in vivo* ribosome biogenesis model

Type	Reaction	Parameter Values	Units	Compartments
Assembly	$I_i + P_j \rightarrow I_{i+1}$ (1° prot.)	0.041–1.69	$\mu\text{M}^{-1} \text{s}^{-1}$	cytoplasm, nucleoid
	$I_i + P_j \rightarrow I_{i+1}$ (2° prot.)	0.24–31.	$\mu\text{M}^{-1} \text{s}^{-1}$	cytoplasm, nucleoid
	$I_i + P_j \rightarrow I_{i+1}$ (3° prot.)	0.025–1.75	$\mu\text{M}^{-1} \text{s}^{-1}$	cytoplasm, nucleoid
Degradation	$\text{mRNA}_i \rightarrow \emptyset$	1.0×10^{-3} – 1.4×10^{-3}	s^{-1}	cytoplasm, nucleoid
Dilution	$x \rightarrow \emptyset$	9.6×10^{-5}	s^{-1}	cytoplasm, nucleoid
Transcription	$\text{DNA}_{\text{mX}} \rightarrow \text{DNA}_{\text{mX}} + 16\text{S}$	0.062	s^{-1}	nucleoid
	$\text{DNA}_x \rightarrow \text{DNA}_x + \text{mRNA}_x$	4.9×10^{-3} –0.012	s^{-1}	nucleoid
Translation	$\text{mRNA}_x + 30\text{S} \rightarrow \text{Rib}_{\text{init}}^x$	1.0×10^2	$\mu\text{M}^{-1} \text{s}^{-1}$	cytoplasm, nucleoid
	$\text{Rib}_{\text{init}}^x + 50\text{S} \rightarrow \text{Rib}_0^x$	3.0	$\mu\text{M}^{-1} \text{s}^{-1}$	cytoplasm, nucleoid
	$\text{Rib}_i^x \rightarrow \text{Rib}_{i+1}^x + \text{P}_{x_i}$	0.019–0.27	s^{-1}	cytoplasm, nucleoid
	$\text{Rib}_{\text{term}}^x \rightarrow 30\text{S} + 50\text{S} + \text{mRNA}_x$	0.015	s^{-1}	cytoplasm, nucleoid
LSU birth	$\emptyset \rightarrow 50\text{S}$	3.1×10^{-4}	$\mu\text{M} \text{s}^{-1}$	cytoplasm, nucleoid
Dimerization	$\text{bS6} + \text{bS18} \rightarrow \text{bS6:bS18}$	1.0	$\mu\text{M}^{-1} \text{s}^{-1}$	cytoplasm, nucleoid
	$\text{bS6:bS18} \rightarrow \text{bS6} + \text{bS18}$	8.7×10^{-3}	s^{-1}	cytoplasm, nucleoid

interactions are calculated using particle-mesh Ewald with 10 Å switching/12 Å cutoffs. Each run uses ~40,000 node hours on NCSA Blue Waters XE6 nodes (2× AMD 6276 Interlagos).

RESULTS

Modeling the *in vitro* SSU assembly

Construction of the in vitro low-temperature kinetic model of SSU assembly

The assembly process of the *E. coli* SSU can be described by a network of binding reactions of the 21 r-proteins to the 16S rRNA and subsequent assembly products. We are omitting bS1 in this model because it is not an integral part of the mature 30S particle, and uS2 and bS21 due to the lack of kinetic data owing to their transient binding nature. We have adopted nomenclature for the r-proteins that emphasizes their homology or lack thereof between the three domains of life (3). Because bS6 and bS18 form a stable heterodimer in solution (35), they are treated singly as the dimer bS6:bS18 in all the binding reactions, and this dimer is assumed to have already formed. The naïve assumption is that these proteins can bind in any order. If this is the case, then the network will include 2^{17} (10^5) species and $17!$ (10^{14}) reactions. To reduce this complexity, the Nomura map of thermodynamic dependencies among r-proteins (7) is used to determine under which circumstances a protein can bind to an intermediate. Imposing this requirement leads to 1612 SSU assembly intermediates and 6997 reactions.

Initially, the rate constants are taken from a P/C qMS study of the reconstitution of the SSU *in vitro* (13). Curves tracking the progress of r-protein binding to assembly intermediates were measured starting with no proteins bound initially (control experiment) and proceeding to various r-protein/16S intermediate configurations, i.e., prebinding experiments (Fig. 2 a). From single-exponential curves fit to these data, an initial rate constant is approximated by assuming that the exponential rate is a pseudo-first-order rate constant and converting it to a proper second-order rate constant using

the initial protein concentration. The rates are chosen from the prebinding experiments where the protein binds directly without requiring the presence of any dependent proteins. This study revealed that the rates for several protein binding reactions are significantly increased for initial intermediates configured with proteins on which the binding protein is not thermodynamically dependent. These situations are referred to as kinetic cooperativity to differentiate the phenomenon from the thermodynamic cooperativity observed by Nomura (7). For binding reactions exhibiting kinetic cooperativity, an ancillary rate constant is used to take this behavior into account. New rates are only introduced if there is a twofold or greater difference compared to the slowest rate observed for binding of that protein. This criterion ensures that the general character of kinetic cooperativity is represented in the model while minimizing the set of unnecessary parameters. A summary of the fold increases due to this phenomenon is provided in Table S1 for all P/C qMS experiments used in this model.

The proteins uS3, uS5, bS6:bS18, uS11, uS12, uS14, and bS16 show no significant kinetic cooperativity. In this model, each of these proteins binds to allowed intermediates at a rate independent of the intermediate composition. All other proteins bind using some manner of kinetic cooperativity. The rate rules for assigning parameters to reactions are derived by considering the kinetic data for each protein individually. When all rules fail to apply to a reaction, a default rate is used. This rate is chosen from the prebinding experiment in which the initial intermediate satisfies all of the dependencies with the least total number of proteins bound.

The most significant examples of kinetic cooperativity were observed in binding to the 3' domain. For uS9, its binding rate is increased by over 200-fold if the intermediate it binds to includes uS19 (and uS7 from Nomura dependencies). The minimum rate was observed for binding to the intermediate with all primary proteins prebound. If uS7 is present alone, the rate is 20 times the minimum, but if uS7 and uS13 are both present, the rate drops to four times the minimum. Finally, if all 5'- and central-domain proteins and uS7 are prebound, the rate is five times the minimum

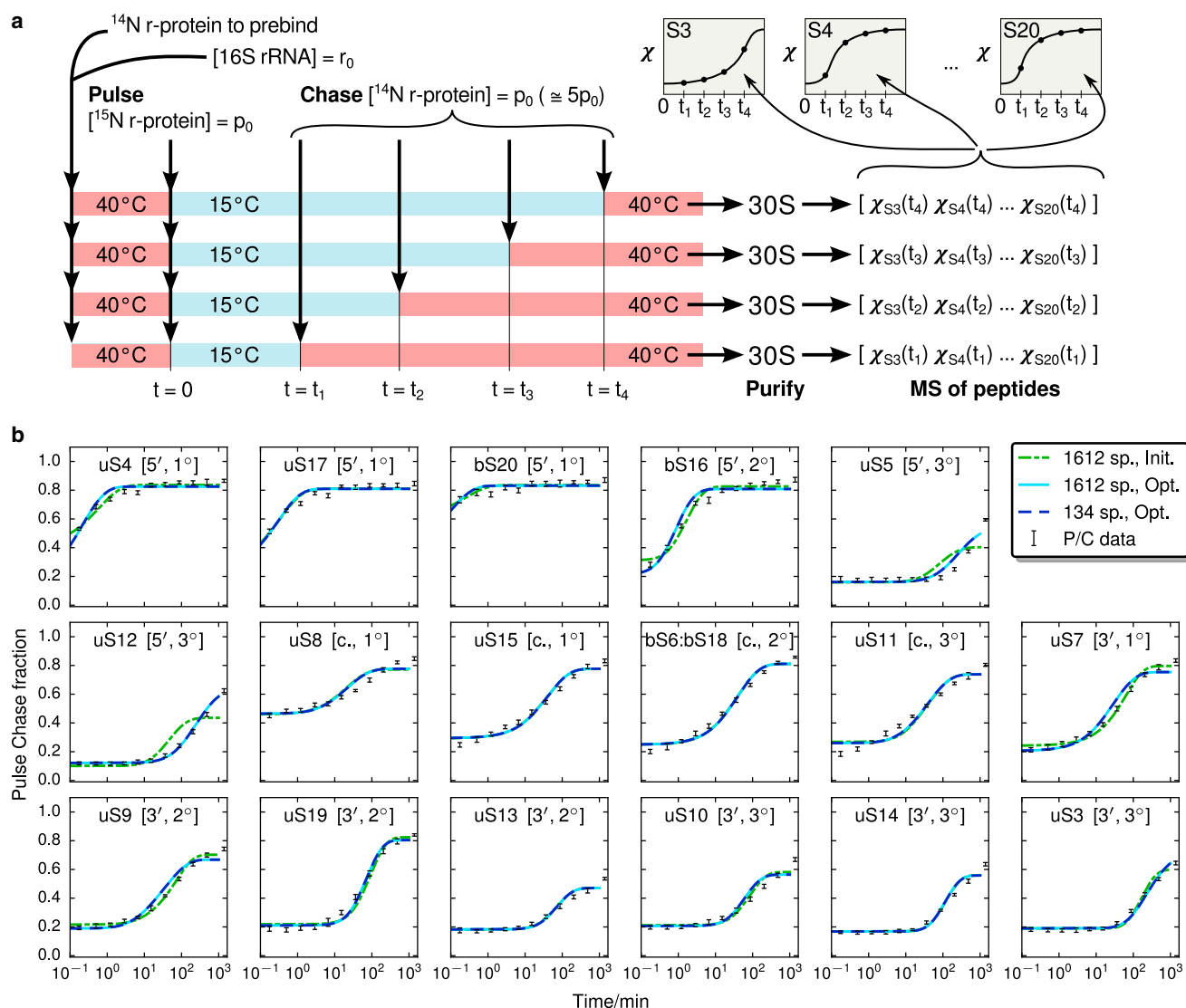


FIGURE 2 (a) Schematic of pulse/chase experiments. The prebinding intermediate is constructed initially from rRNA and the initial set of unlabeled r-proteins by incubation at 40°C . The labeled proteins are added and incubated at 15°C until the chase of fivefold molar excess of unlabeled proteins is added. This is incubated at 40°C again to allow all binding to complete. The 30S particles are purified, and mass spectrometry is used to analyze the fraction of labeled proteins, χ , for all r-proteins simultaneously. This process is performed many times to build up the pulse/chase curves. (b) Comparison of experimental pulse/chase measurements of ribosome assembly starting with bare 16S rRNA (error bars) to the 15°C model (curves). Raw concentration data from the model is transformed into an idealized pulse/chase curve assuming the same ratios of labeled to unlabeled species used in the experiments (13). Using the rates estimated by fitting to the experimental curves yields the dash-dot green curve. Improvement on this curve is made by optimizing the model parameters over pulse/chase experiments starting with nine different initial intermediates (solid cyan curve; for fitting to all initial intermediates, see Fig. S1). By reducing the intermediate count from 1612 to 134 by removing the least important intermediates, a simplified model (dashed blue curve) is generated that quantitatively matches the full model. To see this figure in color, go online.

rate, implying that some or all of the secondary and tertiary proteins binding to the 5' and central domains increase the binding efficiency. Assuming that the effect of uS19 is dominant, the rate rule list for uS9 is developed by first testing for the presence of uS19, ignoring any species nondependent on uS9, such as the 5'- and central-domain proteins. Each rule defines a new rate parameter for the model. The value of this parameter is taken from the prebinding study that the rate originates from. Second, the presence of uS13 is tested, since this appears to decrease the binding efficiency

compared to the case of uS7 bound alone. Third, the presence of all primary and secondary 5'- and central-domain proteins is tested for, ignoring the tertiary proteins. Fourth, the presence of all primary binding proteins is tested, and finally, the default rate is chosen to be from the uS7 prebinding experiment, since this prebinding intermediate minimally satisfies the thermodynamic dependencies for uS9. The parameter assignment rules are developed similarly for all other proteins. A summary of the 32 parameters and their rules is provided in Table 1. This method gives rise to an

enormous reduction of the parameter space dimensionality, leading to 15 parameters describing kinetic cooperativity, and 17 default rates. Since we are fitting to 107 curves that are all parameterized by a single rate constant, overfitting of the model is not a concern.

The 32 parameters in the ribosome-assembly kinetic model shown in Table 1 are separated by domains and listed in decreasing rule precedence. The initial reaction-rate constants are estimated from Bunner et al. (13), and the final reaction rates from global optimization are shown for each parameter. If an intermediate does not satisfy the rules for a parameter (presence or absence of certain r-proteins), the next parameter in the list is tested. MIN indicates that the local optimizer has driven this parameter to the lower limit of $4 \times 10^{-6} \mu\text{M}^{-1} \text{s}^{-1}$.

The initial conditions are chosen to match the experimental conditions used in the pulse/chase experiments: $0.305 \mu\text{M}$ of 16S rRNA and $0.458 \mu\text{M}$ of each r-protein. The model is integrated from 6 s to 2000 min. Fig. 2 b (red curve) compares the protein binding curves from the model to the control experiment. The experimental pulse/chase curves do not compare directly to the simulated ideal pulse/chase curves, since experimentally the reactions are not 100% efficient. To correct for this, a linear transformation is applied to the simulated data to match the starting and ending fractions of the experimentally measured curves. To compute the initial second-order rate constants, a single exponential is fit to the experimental assembly progress curves for the proteins and experiments referenced in Table 1. The exponential rate from this fit is then used to compute a second-order rate constant assuming pseudo-first-order conditions with constant protein concentration. This is not necessarily a good approximation in this situation, but it is sufficient to compute an initial parameter set to perform a local optimization.

Optimization of assembly parameters and kinetic rules

Since there is some variability between rates taken from different experiments and our initial rates were derived using a pseudo-first-order approximation, it is justified to perform optimization on our network to tune the parameters toward a better fit. Biologically reasonable limits on the parameter space were used: $4 \times 10^{-6} \mu\text{M}^{-1} \text{s}^{-1}$ for the lower limit, which corresponds to a reaction timescale an order of magnitude larger than the duration of the P/C qMS experiments, and $3.5 \times 10^3 \mu\text{M}^{-1} \text{s}^{-1}$ for the upper limit, corresponding to the fastest diffusion-limited association of r-protein to the 16S rRNA. By minimizing Eq. 1, we reduced the mean-square error (MSE) between the pulse-chase experiments and our model to 6.5% of the error computed from the initial rates (Fig. 2 b, blue curve). The majority of parameters change within an order of magnitude or less, but significant deviations in the parameters for uS3 and uS5 were observed between the estimated and optimized rates.

Analysis of the low-temperature binding-reaction network

To gain a better understanding of the core of the binding-reaction network, we simplified the full kinetic model by eliminating species with the smallest contribution to the overall integrated flux (Eq. 3) through the assembly network. The network was reduced from 1612 species to 134 species. Using a simple MSE metric, the protein binding curves of the reduced network match that of the full network with an average error of 1.8×10^{-2} (Fig. 2 b, green curve). With the network thinned out, one can readily visualize the distribution of reaction fluxes by drawing a network diagram (Fig. 3) where the thickness of each edge from intermediate A to intermediate B represents the integrated fluxes or, equivalently, the total amount of species A converted to B over the entire assembly time (summand of Eq. 3).

To discuss individual assembly intermediates, we must first develop a concise nomenclature to uniquely specify its protein/rRNA configuration. The states are labeled by the symbol $xyz : s_1, s_2, \dots, s_k$, which consists of two parts. The first part indicates the level of completion of the 5' domain (x), the central domain (y), and the 3' domain (z). The letters here are placeholders for integers that indicate that not all primary proteins are bound to that domain (0), all primary proteins are bound (1), all primary and secondary proteins are bound (2), or all proteins for that domain are bound (3). The second term indicates the specific proteins bound in the intermediate that were not included in the first domain label. For example, state 000:4 describes the 16S rRNA with only the primary 5'-domain protein uS4 bound, and state 100 describes the state with all primary 5'-domain proteins—uS4, uS17, and bS20—present.

A dominant pathway emerges from the reduced network diagram (Fig. 3) where the 30S is assembled in the order $5' \rightarrow \text{central} \rightarrow 3'$. This result confirms the observed $5'$ - to $3'$ binding order seen in experiments (11,36–38). This main pathway contains intermediates seen in cryoEM maps of in vitro SSU assembly at higher temperatures: states 100; 232; 232:5,10,14; 233:5; and 332:10,14 (14). With the exception of state 100, these intermediates are all found late in the assembly process. An ensemble of binding-order sequences can be constructed through random walks over the network using the amount of intermediate converted to weight the transition probabilities. These sequences cluster well into two classes. The first cluster is associated with the dominant $5' \rightarrow \text{central} \rightarrow 3'$ ordering and contributes 70% of the total reaction flux. The other appears to assemble in a general $5' \rightarrow 3' \rightarrow \text{central}$ binding sequence and contributes the remaining 30%.

Both binding-order clusters start out by binding all of the primary and secondary r-proteins in the 5' domain, forming state 200. This intermediate is the bifurcation point at which both assembly pathways begin to diverge. The majority of trajectories from the major pathway complete the central

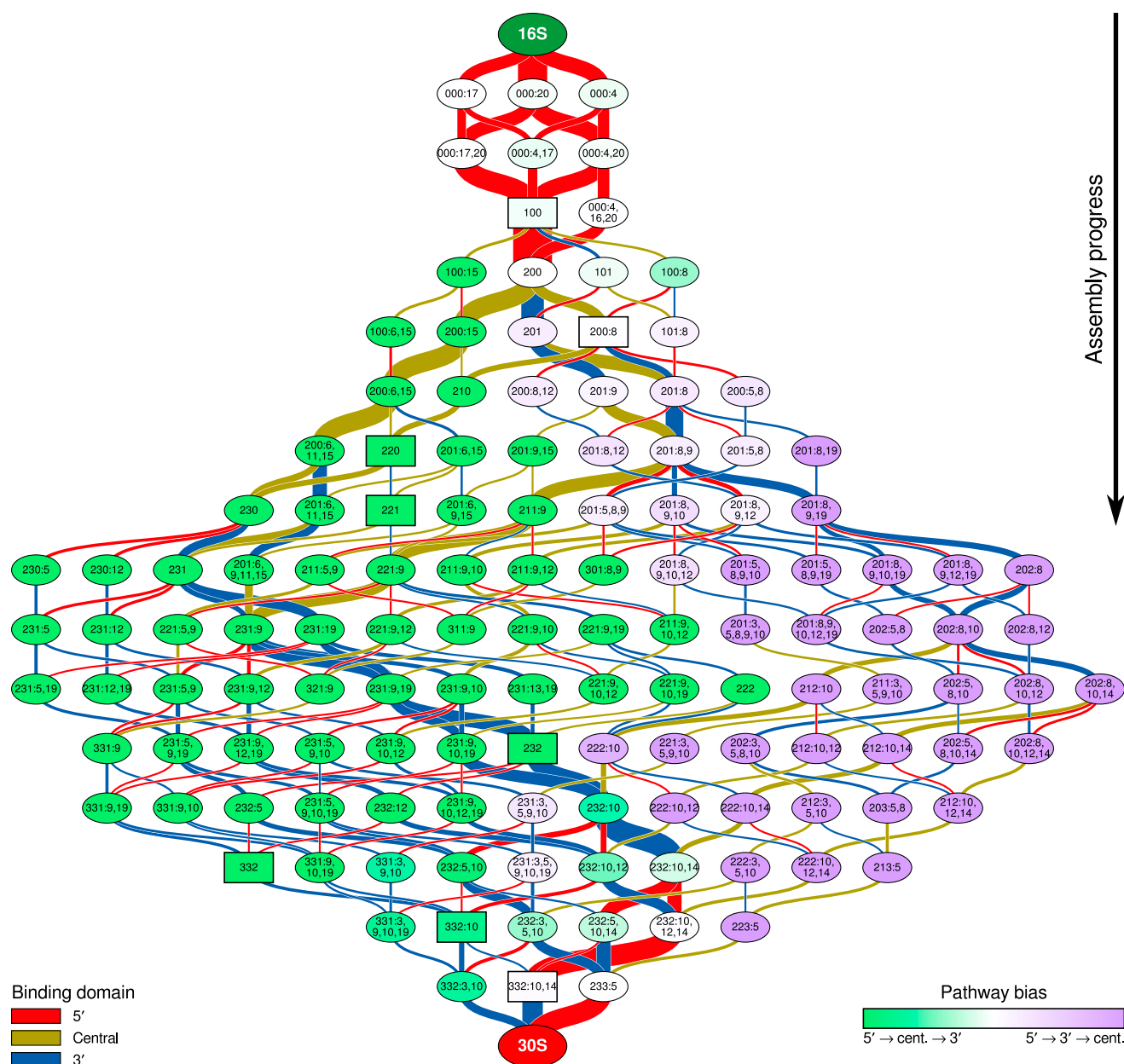


FIGURE 3 Reduced network for 30S assembly at 15°C. Each node is an assembly intermediate, labeled according to which proteins are bound. A three-digit number describes the set of r-proteins bound to each domain (5', central, and 3', respectively), and all remaining r-proteins are listed after the three-digit number (see Analysis of low-temperature binding-reaction network). The edges connecting the intermediates represent the r-protein binding reactions. The width represents the total amount of intermediate converted by that reaction, and the color indicates the binding domain of that protein (5', red; central, yellow; 3', blue). The color of each node indicates its bias toward its use of the two assembly pathways. Green indicates that clustering of protein binding-order trajectories have indicated that this species is more likely to take part in the 5' → cent. → 3' pathway. Predicted assembly intermediates from pulse/chase qMS and cryoEM (14) are represented using rectangles. To see this figure in color, go online.

domain before starting the 3' domain, but the minor pathway switches between binding 5'- and central-domain proteins until it reaches state 201:8. This is another branch point at which the minor path can either rejoin the major pathway or continue finishing the 3' domain. With the exception of state 200:8, no intermediates predicted using cryoEM and P/C qMS are present on the minor pathway. State 200:8 feeds about half of the reaction flux from that species

back into the major pathway. The majority of the remaining flux ends up at state 201:8,9, from which half of the flux flows back to the major pathway as well. Although the clustering analysis identified state 200:8 as a minor pathway species, it contributes equally to each path. Finally, both pathways converge in the vicinity of state 232:10, from which the remaining tertiary 5'- and 3'-domain proteins bind to complete the 30S.

MD simulations to probe network bifurcation and structural barriers at 15°C

The minor pathway in the kinetic model has not been experimentally observed; however, the proteins bound to the *in vitro* states 100 and 200:8, appearing before and after the bifurcation point, have been predicted using cryoEM and P/C qMS (14). Using MD simulations, we probed the ensemble of conformations of states 201, 200:8, and 200:15 near the bifurcation point at state 200 (Table S2). All states contain the intact 16S rRNA and are prebound with uS4, uS17, bS20, and bS16, whereas states 201, 200:8, and 200:15 have bound, in addition, uS7, uS8, and uS15, respectively. To observe the maximum fluctuations in the nucleic acid conformations, we prepared the MD simulations with a neutralizing concentration of sodium ions with no magnesium ions present.

In our previous MD simulations and experiments (10,39,40) on the motions of the 5' domain under similar conditions, we saw that the dominant role of uS4 in state 100 and 200 is to bring together helices h16 and h18, whereas r-proteins uS17, bS20, and bS16 tighten helices in their binding sites on the 5' and central domains. Because the central domain is already partially formed in state 200, it is expected that the main role of uS8 and uS15 is to add rigidity to the central domain. uS7 binds to the partially formed 3' domain, whereas uS8 and uS15 bind to regions in the central domain already formed (see Figs. S3 and S4).

In the 3' domain, all four simulations showed similar motions. These fluctuations are dominated by the partial unfolding of the 3' domain. Helices in the lower four-way junction (h29, h30, h41–h43) separate from helices in the upper three-way junction (h34–h40) (Fig. 4 a). Time traces of the centers of mass for the different junctions in all four MD simulations show that the helices separate from 40 Å to over 60 Å after 140 ns (Fig. 4 b). Simultaneously, the structural signature (2) h33 separates from h31 and h32 and becomes more solvent exposed. This is expected, since h33 is connected to these junctions. Similar results are seen in simulations of the *Thermus thermophilus* SSU (Fig S2), suggesting that these motions are probably common to all bacterial organisms. The fact that states 200, 201, 200:8, and 200:15 all have similar motions suggests that there is no strong bias to binding either uS7, uS8, or uS15 and that the next major assembly barrier, the folding of the 3' domain, occurs further along in the assembly pathway.

Because the binding of uS7 and uS8 have a minimal effect globally on the structure of the ribosome-assembly intermediates, we probed the effect of adding the 3'-domain binding r-proteins uS9 and uS19. In the folded ribosomal SSU, uS9 binds to both the lower four-way and upper three-way junction, whereas uS19 binds to the structural signature h33 (Fig. S4). As the uS19 binding site is more local than uS9, we investigated the binding of uS19 first

(Fig. S4). Adding uS19 to the simulations (moving from state 200:8 to 201:8,19) tightens the structural signature in h33 and keeps h33 packed against h31–h32, and as in the four previous simulations, state 201:8,19 also shows similar unfolding of the 3' domain (Fig. 4 b). State 201:8,9,19, on the other hand, does not have the separation in the 3' domain (Fig. 4 b). Interestingly, all six MD simulations showed the 3' domain rotating away from the five-way junction in the 5' domain, suggesting that there is another folding barrier further along in the assembly pathway. This motion might only be arrested upon the addition of uS5.

Construction of the *in vitro* high-temperature kinetic model of SSU assembly

The previously described model fits the experimental data well over many different initial intermediate configurations and has predictive power, but it is not adequate for use in an *in vivo* model of *E. coli*, since it describes the reconstitution of the 30S at a temperature much lower than that required for optimal *E. coli* growth. Since the rates of binding for each protein will vary independently with temperature in ways that are difficult to predict, it is not sufficient to simply scale the rates of the low-temperature model to match the observed assembly time *in vivo*. To prepare a kinetic model of SSU assembly at physiologically optimal temperatures, we constructed a model based on *in vitro* reconstitution experiments performed at 40°C (14). These experiments were performed at concentrations lower than those in the low-temperature model, 0.02 μM 16S rRNA and 0.04 μM labeled r-proteins, but the fivefold molar excess of the chase unlabeled proteins was the same as before. Since only the control protein binding curves were measured in this work, we are not able to include the effect of cooperative binding. Due to the lack of these reactions, the high-temperature model does not fit the experimental data as well as the low-temperature model (Fig. 5). However, the correct protein binding order is represented, and protein abundance half-lives are reproduced within 6%.

The reduced network assembly model at 40°C contains 145 unique intermediates and 325 protein binding reactions. The number of intermediates was set to focus on the core binding network and to allow efficient RDME simulations of the *in vivo* model discussed below. Although Fig. 5 shows that the reduced set captures the binding kinetics well, we carried out additional simulations to investigate whether important assembly pathways are being removed. Reducing the full high-temperature model from 1612 to 638 states, we repeated the previous analysis of the assembly network. It was observed (data not shown) that there is a minor partitioning of protein binding order trajectories into the two pathways seen in the 15°C data. However, the 5' → central → 3' trajectories occur >90% of the time, compared to the 70% seen in the low-temperature network. The dominance of the 5' → central → 3' pathways is likely due to the effects of the higher temperature, which increases the rates of binding

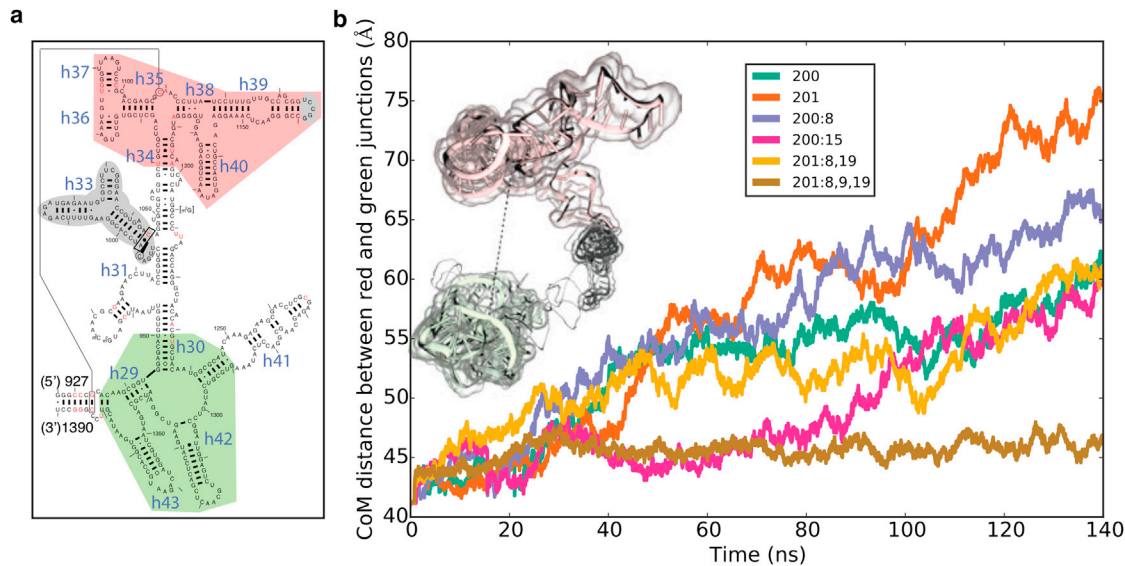


FIGURE 4 (a) Secondary structure diagram of the 3' domain (41). Centers of mass are computed from the lower four-way junction helices h29, h30, h41–h43 (green region), and the upper three-way junction helices h34–h40 (red region). These centers are separated by the structural signature h33 (gray region) (2). (b) Time traces of center-of-mass distances in the 3' domain. The r-protein binding sites in the folded SSU for each domain are provided in Figs. S3 and S4. To see this figure in color, go online.

in the primary proteins and diminishes the differences previously observed between the secondary and tertiary proteins. Since the rate constants have changed significantly with respect to the low-temperature model, the reduced network structure has changed as well. The problems we experienced with uS3 and uS5 were not repeated here, since the experimental binding order of these proteins was consistent with the Nomura map. The assembly pathway is much less

directed, i.e., for most states, there are many binding reactions that occur at similar reaction rates (Fig. 6). It is evident that the temperature has had a large effect on the utilization of assembly pathways. The bifurcation into two distinct pathways seen in the low-temperature model is absent in the high-temperature model (Fig. 6). Although the binding order is less well defined at higher temperatures, the assembly still progresses in a 5'→central→3' directionality, with

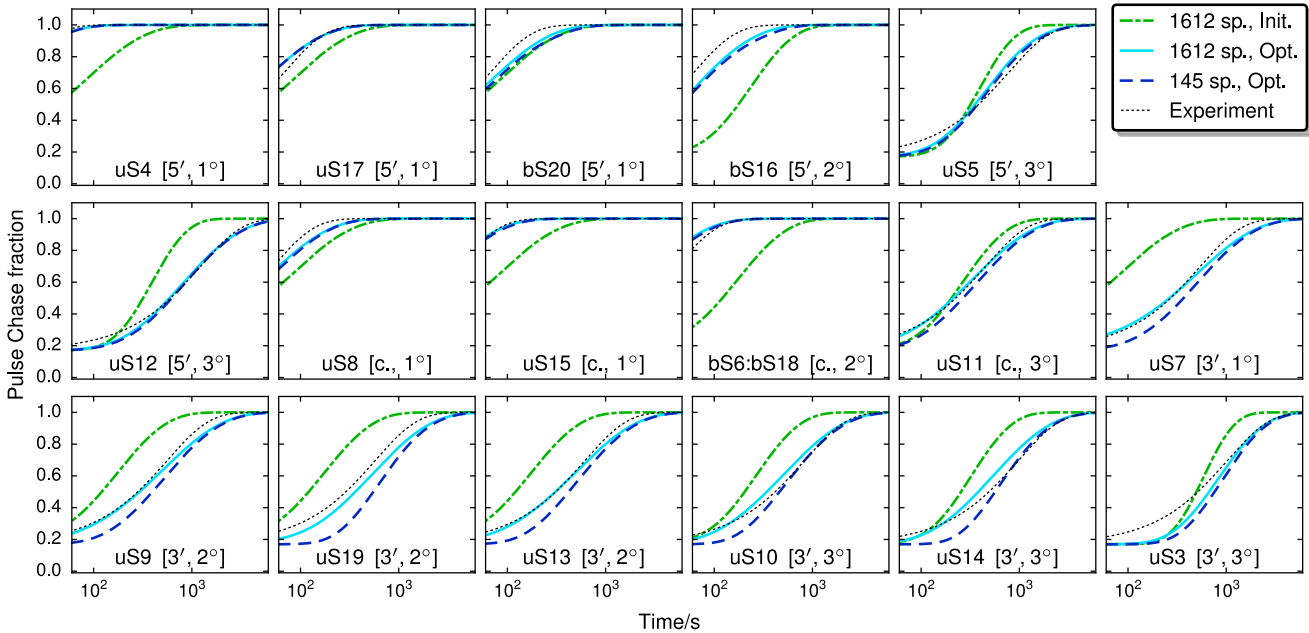


FIGURE 5 Fitting of protein binding curves from the high-temperature in vitro model to the curves measured from the 40°C experiment. Deviations of the reduced model with respect to the full model tend to only impact the 3'-domain binding proteins. To see this figure in color, go online.

Binding domain

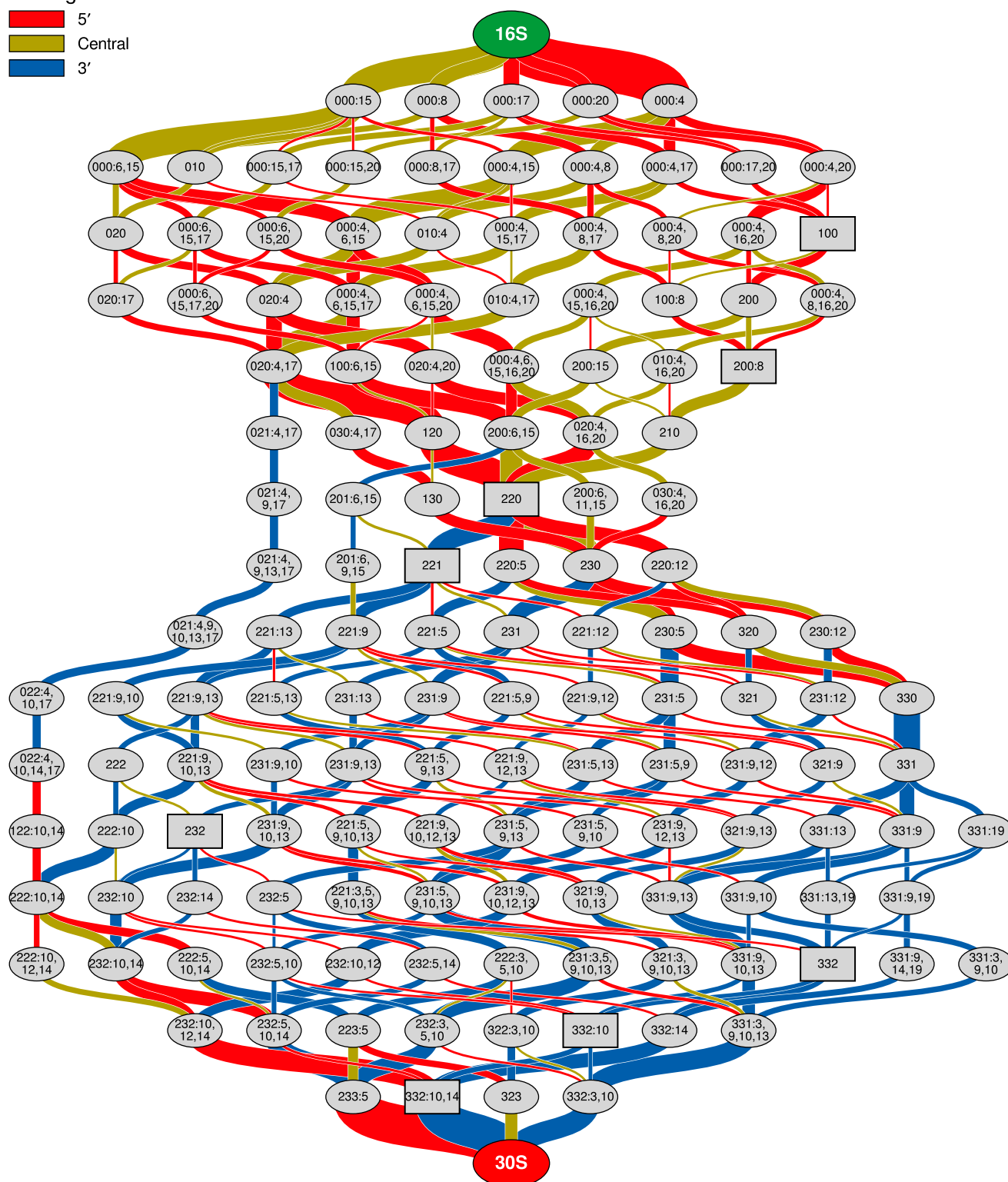


FIGURE 6 Reduced network for 30S assembly at 40°C. The 5'- and central-domain proteins bind simultaneously, leading to state 220. From here, two weakly defined paths emerge: either the 5' and central domains are completed simultaneously, followed by the 3' domain, or vice versa, ending in the formation of the 30S. To see this figure in color, go online.

the 5'- and central-domain proteins binding in parallel, followed by the 3'-domain proteins and, finally, the remaining tertiary proteins from the 5' domain.

Binding of the primary proteins uS4 and uS15 to the 5' and central domains, respectively, dominates the nucleation of the nascent 30S. The most highly traversed intermediates seen at low temperatures, states 100 and 200, appear less prominent at high temperatures. State 100 appears 1 min into the assembly process in both the proposed mechanism (14) and our kinetic model. The state 220 acts as a central hub for most assembly paths in our network and is also predicted as an intermediate in the proposed mechanism. It reaches its peak concentration at 2.2 min which is comparable to the time of 3 min inferred from P/C qMS and cryo-EM. The following state, 221, appears in both our model and the predicted mechanism as well, but the timings are different. It was predicted to bind 8 min into the assembly process, but we are observing the intermediate 221 coming in ~6 s after state 220. The next predicted assembly intermediate is state 232, which is less prominent in our model than what would be expected from the P/C qMS and cryo-EM data. The maximum concentration of state 232 is reached much sooner than expected from the proposed mechanism, coming in at 5 min instead of the 12 min predicted. The latest predicted intermediate, 332:10,14, which is missing only uS3, comes in at 20 min instead of the 70 min predicted. The timing discrepancies between the experiments and our results is likely due to the lack of kinetic cooperativity in our model. Though there are differences between these times, the P/C qMS study did not identify exact intermediates experimentally, instead they are inferred from the data. The relative ordering of intermediates suggests that this model and the published mechanism are in agreement.

Modeling in vivo ribosome biogenesis

Construction of the ribosome biogenesis model

In addition to the hierarchical assembly of the SSU described above, the process of ribosome biogenesis in the cell must also include the transcription of rRNA and mRNA coding for r-proteins, the translation of r-protein, and the degradation of mRNA. The high-temperature in vitro model of SSU assembly developed from kinetic experiments with well-mixed solutions of rRNA and r-proteins is now applied to biogenesis in the heterogeneous cellular environment. For the full ribosome biogenesis model, we control the birth rate of the LSU to match that of the SSU without explicitly including LSU assembly, and we include 70S formation and dissociation reactions, with rates taken from the literature (42–44).

We present a spatially resolved model of the process in a simulation of a slow-growing *E. coli* cell, of dimensions $4.0 \times 0.9 \times 0.9 \mu\text{m}^3$ and initially containing ~3000 ribosomes (16,21). Using our LM 2.2.1 software, we monitor

the stochastic changes in the number of species in a cell over its doubling time of 120 min. The capsule-shaped cell is discretized onto a lattice with 32 nm spacing between lattice sites, allowing us to neglect excluded-volume effects from the 20-nm-diameter 70S particles. The nucleoid region of dimensions $3.1 \times 0.45 \times 0.45 \mu\text{m}^3$ is centered within the cell volume (Fig. 7 a). At each lattice site, we assume the well-stirred approximation to evaluate the reaction time course using the Gillespie algorithm (45).

The protein diffusion constants are estimated based on their mass using a scaling relation between the diffusion constant in water versus that in cytosol (46) leading to diffusion constants in the range $8\text{--}20 \mu\text{m}^2 \text{s}^{-1}$. The maximum time step, Δt , that can be used in the MPD-RDME simulation is determined by the fastest-diffusing species, which in this case is bS18. To ensure that no particles diffuse more than a single lattice site per step, the maximum time step is chosen to ensure that the RMS displacement of a Brownian particle, $\sqrt{6D\Delta t}$, is shorter than the lattice spacing. To speed up the simulation, the protein diffusion constants were all scaled by a factor of 0.3 to allow for longer time steps, resulting in a maximum time step of 25 μs . This should not have a significant effect on the outcome of the simulation, since the slowest protein diffuses at a rate nearly an order of magnitude faster than the rate of the fastest nonprotein species.

mRNA diffuses at $0.3 \mu\text{m}^2 \text{s}^{-1}$, as measured in the literature (47). The diffusion constant for rRNA is computed from the radius of gyration (48) using the same scaling relationship to account for diffusion in cytosol as for r-protein. Assembly intermediate diffusion rates are assigned by counting the number of proteins bound and using this number to linearly interpolate between the diffusion constants of 16S and 30S species. Transition rates between compartments are computed from the geometric mean of the diffusion rates for each compartment.

Single-particle tracking experiments on individual SSUs and LSUs, as well as complete ribosomes, have shown that ribosomes are partially excluded from the nucleoid region and diffuse at a rate 10-fold slower than the rate for individual subunits (20). From this study, we take the rates of $0.4 \mu\text{m}^2 \text{s}^{-1}$ (20) for both SSUs and LSUs and $0.055 \mu\text{m}^2 \text{s}^{-1}$ (18,20) for full 70S ribosomes. We decrease the diffusion constant of ribosomes, ribosomal subunits, and assembly intermediates within the nucleoid region by a factor of 10 to account for the increase in molecular crowding due to the presence of a compacted chromosome. The 70S particles are observed to be partially excluded from the nucleoid region. The reason for this is not well understood (18,20), but it most likely is a result of the excluded-volume interactions between the ribosomes and DNA. To account for ribosome exclusion without explicitly simulating the chromosome, we bias the transition rates between the nucleoid and cytoplasm by a factor of 4.0. A summary of the diffusion parameters is given in Table 3, and the complete list can be found in Table S3.

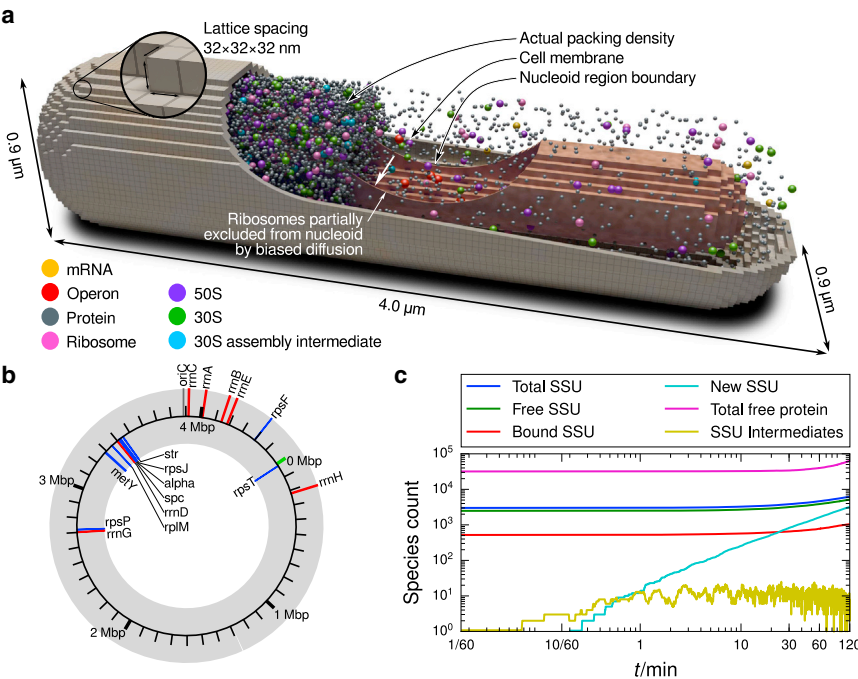


FIGURE 7 (a) Cutaway of a representative simulated cell configuration. Operon locations (red) are fixed within the nucleoid region. Messengers (yellow) are transcribed from these sites and diffuse to find 30S particles (green), upon which a 50S subunit (purple) joins the complex, forming a translating ribosome (pink). The ribosome emits r-proteins (gray), which diffuse away and bind to SSU intermediates (cyan). Translating 70S particles are excluded from the nucleoid region through a bias in their intercompartmental transition rates. (b) Genome diagram of the operons transcribed in the in vivo biogenesis model. (c) Species counts for a single replicate during a full 120-min cell cycle. The initial species counts are set to their mean values from a well-stirred simulation at steady state. The counts of 16S rRNA and assembly intermediates are set to zero to investigate the formation of new intermediates. Dilution reactions are omitted from this simulation to investigate the change in particle count over a cell cycle. The curve Bound SSU measures the total count of 30S particles that are not bound to other species in the cell, i.e., all translating ribosomes and 30S/mRNA complexes. Total SSU measures all 30S particles in the cell, including both free species and bound. To see this figure in color, go online.

The initial species counts (see Table S4) are determined from the mean copy numbers at the steady state of a well-stirred stochastic simulation of the in vivo network within a volume equal to the cell volume (2.37 fL) using Lattice Microbes. The freely diffusing species are placed uniformly throughout the cell, the translating ribosomes are placed outside the nucleoid uniformly in the cytoplasm, and the operons are placed based on their genetic loci. These seven rRNA operons and nine r-protein operon species are placed in the nucleoid region at random about the central axis. Assuming that the origin of replication is at the center of the cell and the chromosome is linearly organized (49), operons are placed along the cell axis at positions relative to

their distance from *oriC* (Fig. 7 b). Subsequent simulations are initialized from random time steps taken from a long-running simulation approaching steady state (Fig. 7 c).

The next step toward a spatially resolved model of ribosomal biogenesis is to provide constant and balanced production of rRNA and r-protein through transcription, translation, and degradation in the cell. Transcription is modeled as a simple birth process localized at operon sites within the nucleoid region. Transcription of 16S rRNA occurs from seven ribosomal operons (rrnABCDEGH) at a birth rate resulting in a mean count of 4500 ribosomes at steady state. This number is chosen to approximate a cell that initially contains 3000 ribosomes immediately after cell division, and the number of ribosomes doubles to 6000 over the 120-min cell cycle. Transcription of mRNA from the nine r-protein operons is modeled similarly to rRNA. Since mRNA is actively degraded by RNase E at various rates depending on the content of the transcript, we use data from a genome-wide microarray study of *E. coli* mRNA half-lives (50) to estimate the decay rate for each messenger species individually. In lieu of explicit gene regulation, we tune the mRNA birth rates such that the steady-state copy numbers are roughly equal for each r-protein species. Since the volume of the cell does not change in our simulations, dilution reactions (modeled as a first-order death process) are added to account for the effect of increasing cell volume as the cell grows. Dilution reactions in addition to the mRNA degradation reactions are added for all species with the exception of the operons. These reactions occur at a rate of $\ln 2/120$ min, approximating a slow-growing cell with a doubling time of 2 h.

TABLE 3 Summary of diffusion constants for the in vivo ribosome biogenesis model

Species	Compartment	D ($\mu\text{m}^2 \text{s}^{-1}$)
Ribosome	cytoplasm	0.055
	nucleoid	0.0055
	cytoplasm \rightarrow nucleoid	0.0043
	nucleoid \rightarrow cytoplasm	0.0017
Subunit	cytoplasm	0.4
	nucleoid	0.04
	cytoplasm \leftrightarrow nucleoid	0.126
Protein	cytoplasm, nucleoid	2.6–6.4
mRNA	cytoplasm, nucleoid	0.3
Intermediate	cytoplasm	0.15–0.39
	nucleoid	0.015–0.039
	cytoplasm \leftrightarrow nucleoid	0.047–0.122

The dilution rate is simply $\ln 2/120$ min, the cell doubling time. Transition rates between compartments are computed from the geometric mean of their diffusion constants.

Our model of transcription and translation takes the operon structure in the mRNA transcripts into account and allows for multiple gene products to be produced from a single mRNA molecule. Translation is modeled in three stages. First, initiation occurs by the association of the messenger and SSU, followed by the association of the LSU to this complex to form a translating ribosome. Since a model of LSU assembly has yet to be developed, we simply add 50S species to the system at a rate that matches the production rate of 30S SSUs. Second, translation of the ribosome along the mRNA strand is simulated by assuming that once a 50S species associates to the 30S/mRNA complex, the ribosome translates with a constant speed until it dissociates from the end of the transcript. Each SSU r-protein is made sequentially at a rate k_{ti}/N_i where k_{ti} is the translation rate per amino acid (10 aa/s, estimated from Bremer and Dennis (5)) and N_i is the number of codons between the stop codons of the previous and current SSU r-protein genes, including the length of any intervening genes not represented in the model (e.g., LSU r-protein). This extrapolation to 10 aa/s is not well justified, but it is sufficient for our work, since protein production is limited by the number of available mRNAs. Genomic data are taken from the *E. coli* K-12 MG1655 genome (GenBank accession number U00096 (25)). Finally, termination occurs after translation past any remaining genes not considered in the model, by the simultaneous dissociation of the ribosome into mRNA, 30S, and 50S subunits. An example of the derivation of the translation reactions from genomic data is given in the [Supporting Material](#) (Section S4) for the *spc* operon. No postprocessing is assumed to occur for the protein. However, bS6 and bS18 dimerize before associating with rRNA at an assumed rate of $1.0 \mu\text{M}^{-1} \text{s}^{-1}$ (51) and dissociate at a rate of $8.7 \times 10^{-3} \text{s}^{-1}$, computed from the dissociation constant reported in Recht and Williamson (35). A summary of the in vivo reactions, rate constants, and diffusion parameters is presented in [Tables 2 and 3](#). All parameters are reported in [Table S3](#).

Simulation results of the ribosome biogenesis model

We start with the initial conditions derived from the steady-state well-stirred simulation. Since these initial conditions describe the mean of a growing cell—starting at 3000 ribosomes and ending at 6000 ribosomes—we scale all species counts by 2/3 to approximate the initial conditions of a newly divided cell. The initial rRNA and 30S intermediate counts are set to zero so that the birth of new ribosomes over the 120-min cell cycle can be monitored. The first new 30S begins to appear after 17 s ([Fig. 7 c](#)), and the cell quickly reaches a stable-state bulk 30S production rate of 27/min (from the slope of the production line), with new SSUs appearing uniformly within the cell. The production rate is accelerated with respect to the in vitro simulations and is due to the greater r-protein concentration in the in vivo simulations. The total ribosome count, using the sum of 30S, 30S:mRNA, and 70S particles, increases from 3000 to 6000 over the 120-min cell doubling time.

The assembly intermediate counts fluctuate significantly over the course of the cell cycle, with a mean count of 9.7 ± 3.8 (mean \pm SD; coefficient of variation, 0.39). All 145 intermediates appear with nonzero counts at some point during the cell cycle. Intermediate 233:5 (30S missing uS12) had a maximum copy number of 12, which is greater than that of any other intermediate during the cell cycle. None of the other final intermediates ([Fig. 6](#)) were found in such high quantities.

To gather more statistics on the formation times of the intermediates and new subunits, we designed simulations based on the previous cell-cycle-long simulation ([Fig. 7 c](#)) to measure the delay between the appearance of rRNA and the formation of intermediate species. Since the assembly time of the 30S is of the order of a few minutes, we performed 5 min of simulation time over 64 replicates to collect sufficient data to compute distributions of assembly times. The initial conditions for each replicate are selected from random time points during the cell cycle simulated previously and have been modified to remove all assembly intermediates. The rRNA operons are removed and 100 rRNA molecules are distributed uniformly throughout the cell, allowing for measurement of the time interval between the formation of 16S rRNA and the subsequent intermediates. Since the protein count ([Fig. 7 c](#)) is much higher than the initial rRNA count, the results from these simulations will be comparable to the full cell cycle. From these formation-time simulations, we measure the birth times of the species of interest from the start of the simulation. The results of this process are equivalent to computing the species birth times by following the fate of each rRNA in the cell-cycle simulation.

To investigate the spatial distribution of assembly intermediates, we perform clustering in time to partition the set of intermediates into classes of species that are correlated in time. We use the data from the formation-time simulations to compute mean copy number versus time curves for each intermediate. The curves from each intermediate are scaled to unit amplitude to treat each species equally with respect to its maximum concentration and are then compared using an RMS difference metric. Hierarchical clustering is used to partition the intermediates into six classes (T0–T5), where each class contains species that are formed at similar times. The fraction of the total rRNA that contributes to each temporal class (derived from the formation-time simulations) is provided in [Fig. 7 a](#), and the membership of all intermediates in each cluster is provided in [Fig. S6](#). To achieve adequate sampling of the spatial distribution of all intermediates, we performed 128 short (5 min) simulations from multiple starting conditions sampled randomly from the cell-cycle simulation. Using the temporal clustering, we computed mean intermediate distributions over the whole cell volume and projected the distribution onto the *xz* plane, leading to a measurement of density qualitatively similar to one performed using an optical microscope.

The first class, designated T0, contains the 16S rRNA and 40 early intermediates and is formed at the sites of the rRNA operons. These intermediates are localized because the timescale of the protein binding reactions of the primary and secondary proteins of the 5' and central domains is of the same order as the rRNA diffusion time (Fig. 8 *b*). In the next class, T1, the 3' primary and secondary proteins uS7 and uS9 bind (Fig. 8 *c*), and the distribution of intermediates in this class begins to leave the nucleoid region. T2 contains the main bottleneck species 200 and includes intermediates as late as 220:10. Because of this, there is a path through the network that can skip over T3 entirely. T3 consists of less common intermediates undergoing the binding of 3'-domain proteins and later binding 5'-domain proteins. This is the last cluster where any spatial heterogeneity is evident. T4 consists of more common late-stage intermediates undergoing binding similar to that observed for T3. The distribution of T4 is effectively uniform over the cell. Finally, T5 contains species missing tertiary proteins and is distributed uniformly. This leads to production of new 30S occurring uniformly throughout the cell. The temporal class membership of all intermediates is given in Fig. S6.

The complex formed from the binding of mRNA to the SSU is found either in the cytoplasm or close to the messenger's originating operon. The mRNA cannot diffuse far from its originating transcription site because of the high concentration of 30S particles throughout the cell. Once the trans-

lating complex is formed by binding a 50S particle to the 30S/mRNA complex, the particle will diffuse out of the nucleoid. Its diffusion back into the nucleoid is hampered by the biased intercompartmental transition rates. Once translation is complete, the 70S dissociates, leaving 30S, 50S, and mRNA species free outside the nucleoid region. This leads to a distribution where the 30S/mRNA binding events are localized around their originating operons and in the cytoplasm compartment. The termination of translation appears to occur almost entirely outside of the nucleoid region, since the translation process is slow enough to allow the ribosome to completely diffuse out of the nucleoid (Fig. 8 *e*).

The mean assembly time for individual subunits was measured to be 30 s. The distribution of assembly times is approximately gamma distributed, with a scale parameter of 2.35 s and a shape parameter of 0.208 (Fig. 8 *d*). This mean assembly time is similar to the experimentally measured *in vivo* maturation time for 30S of 1.3–3.5 min at a cell doubling time of 100 min (52).

Performance of Lattice Microbe software

To our knowledge, our simplified model, with its 251 unique species and 1336 reactions (676 within the nucleoid region and 660 in the cytoplasm), is the largest time-dependent simulation of *in vivo* ribosome biogenesis to date. The cell model tests the limits of LM 2.2.1 (the current version of Lattice Microbes) with regard to its handling of the number of species

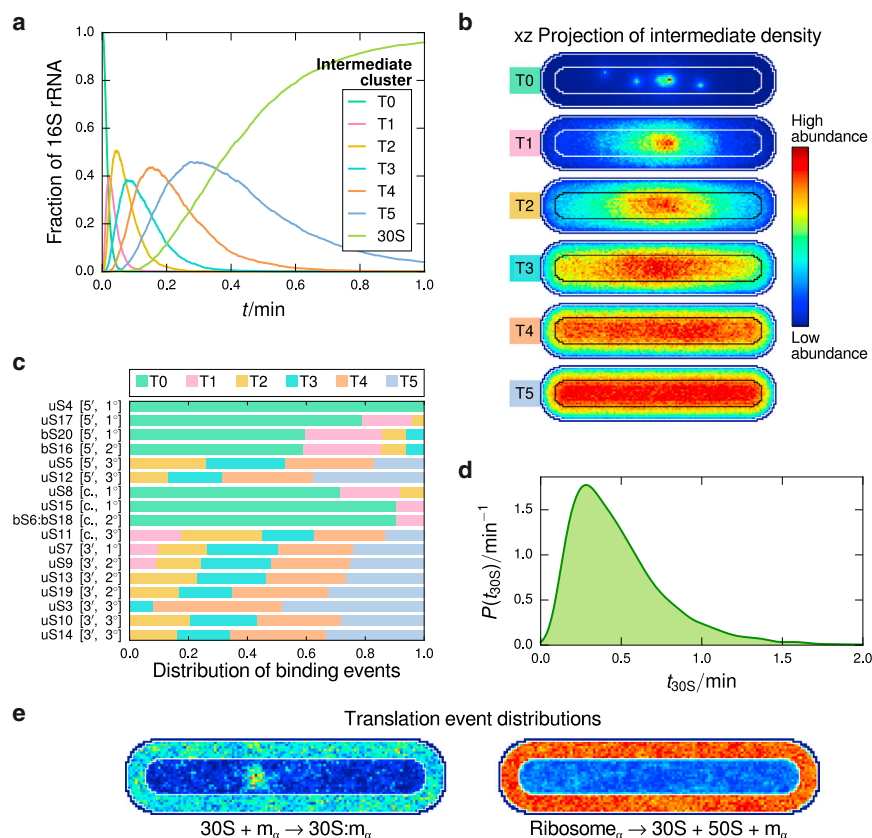


FIGURE 8 The assembly process of the 30S particle is spatially dependent. (*a*) Fraction of intermediate temporal clusters present as a function of time. Temporal clustering groups the 145 intermediate species into mutually exclusive groups based on their order of appearance in the assembly process. The precise assignment of intermediates to clusters is provided in Fig. S6. (*b*) Projections of the intermediate spatial probability distributions for the six temporal classes (T0–T5) onto the xz axis. The distribution of individual intermediates is reported in Fig. S5. (*c*) Distribution of protein binding events in each temporal class, providing a timeline of protein binding reactions. For example, all uS4 binding reactions occur in group T0 and all uS15 and bS6:bS18 binding reactions occur in T0 and T1. (*d*) Distribution of assembly times for the SSU. The birth-time distribution, measured as the time from birth of 16S to birth of 30S, is approximately gamma distributed. (*e*) Translation is spatially dependent. Central y -slices of the 3D probability density of binding events showing 30S associating with mRNA from the α -operon (*left*) and dissociation events of ribosomes translating α -mRNA (*right*). Binding of messenger to SSU appears to happen in two locations: outside the nucleoid region, and inside the nucleoid region localized near the originating operon. From the dissociation events, it is clear that the translating ribosomes are correctly excluded from the nucleoid region, as intended. To see this figure in color, go online.

and reactions. Two major data structures used by Lattice Microbes are the stoichiometric matrix, S , with dimensions of $N_{\text{reactions}} \times N_{\text{species}}$, and the reaction-location matrix, RL , with dimensions $N_{\text{reactions}} \times N_{\text{compartments}}$, specifying the reactions that can occur in a given compartment. Both of these structures are typically stored in the graphics processing unit (GPU) constant memory, which is limited to 48 kB in size in most GPUs. The size requirements of S and RL are 16 kB and 10 kB, respectively, so for the species count required for the ribosome biogenesis model, only 64 reactions could have been supported. In LM 2.2.1, we added the functionality to relocate S and RL to GPU global memory and access them via the read-only data-cache path added to the Kepler class GPUs. Current GPU constant memory usage now only handles the remaining data structures, allowing simulations of 2400 reactions without any additional changes.

The performance of the MPD-RDME simulations is determined by the wall time required for particle diffusion, reaction evaluations, and handling of input/output and simulation overflows. The scaling of computational time of a single time step is consistent with the previous version developed for multi-GPU simulations (27), where the evaluation of reactions is a linear time operation in the number of reactions, since the reaction list must be traversed for every nonempty lattice site. Because of this, a single time step on Kepler class NVIDIA GPUs (K20X; CUDA 6.5) on the NCSA Blue Waters supercomputer takes ~18 ms. At a time step of 25 μs , 1 h of simulation time requires 21 days of wall time. On Maxwell class central processing units (GTX 980; CUDA 6.5) in a desktop computer, the time step is ~6 ms and 1 h of simulation time will finish within a week.

To further accelerate the reaction-kernel runtime, we investigated specialization and employed code generation techniques to write a reaction kernel to solve the specific model being simulated. This has the benefit of requiring even fewer data structures to be accessed in constant memory, as memory references are now replaced with immediate value loads and loops that could not be unrolled at compilation time are flattened before compilation. Using this technique, run times on GTX 980 GPUs and the K20X accelerators was reduced to 1.9 ms and 4.0 ms per time step, respectively, allowing 1 h of simulation time to be completed in ~3–6 days. Simulations of the full 120-min cell cycle would require 6–12 days, depending on the GPU used. The enormous improvement in performance is achieved by applying algorithms that exploit the newest features in the rapidly developing field of GPU computing. These improvements will allow us to add more species and reactions to a simplified model describing regulation and coupling to the metabolic network.

DISCUSSION

Here, we report on the progress in developing a simplified RDME description of the transcription, translation, and pro-

tein/rRNA association events comprising ribosome biogenesis in whole cells. We have constructed an assembly model of the SSU, which is, to our knowledge, the most detailed description to date. Our whole-cell model accurately reproduces the assembly timescales of the SSU and predicts both the identity of major assembly intermediates and their spatial distributions throughout the cell. By tuning the formation rate of the LSU to match the formation rate of the SSU, we capture the increase of the ribosome count from 3000 to 6000 over the full 120-min cell cycle. Nevertheless, there are several important features and reactions that are required for a more complete model of ribosome biogenesis.

The low-temperature assembly model predicts a heretofore unrecognized assembly pathway through which the SSU is assembled in a $5' \rightarrow 3' \rightarrow \text{central}$ directionality. However, it is unlikely that this assembly pathway is biologically relevant due to the conditions from which it emerges. It appears to be an artifact of the low-temperature (15°C) *in vitro* conditions. This pathway is not seen in the reduced high-temperature (40°C) network, used as the basis of the whole-cell RDME simulations. In addition, if *in vivo* assembly occurs cotranscriptionally, the proteins will bind in the order $5' \rightarrow \text{central} \rightarrow 3'$ as the transcript leaves the polymerase. Although not directly relevant to ribosome biogenesis *in vivo*, this alternate pathway illustrates the sensitivity of coordinated assembly networks to varying conditions such as temperature.

The spatially resolved simulations exhibit strong localization of early SSU intermediates within the nucleoid region, even without explicitly treating cotranscriptional assembly. Our model predicts that 50% of the SSUs will be assembled within 42 s, which is faster than the accepted 30S maturation time of 30–90 s in rich media or 78–150 s in minimal media (52). The two main contributions to the assembly time difference are the lack of uS2 and bS21 in our model and the omission of rRNA processing. These remaining tertiary proteins would be expected to have slow binding rates, on the order of those for uS3 and uS5, and could add 10–15 s to the assembly time.

An important additional feature to consider is rRNA processing and maturation reactions. We assume in the simplified model that the 16S is emitted from the ribosomal operons completely processed, but the transcript is actually polycistronic and includes the 16S, 5S, and 23S rRNA, and tRNA as well. Each gene in the transcript has to be processed individually. The processing of the rRNA involves a number of enzymes and is considered to take place primarily in the nucleoid region, although there are suggestions in the literature that some processing may occur at the inner membrane. The maturation processes are still being investigated, but as soon as a consistent understanding emerges, these reactions can be included (53–55).

Another feature missing in our model is the action of assembly cofactors. Though the ribosome is capable of being reconstituted *in vitro* from only rRNA and r-protein, in

living cells, the process is aided by RNA chaperones, RNA helicases, ribosome-dependent GTPases, and other maturation factors (4). These species act to improve the speed and efficiency of assembly by minimizing the misfolding of nascent subunits into kinetic dead ends. P/C qMS experiments have shown that the assembly cofactors RimM, RimP, and Era significantly increase the binding rates of particular r-protein during the in vitro assembly of the 30S (56). However, kinetic data with varying cofactor concentrations are unavailable, limiting the applicability of P/C qMS to our model. KsgA is an assembly cofactor that appears to have its greatest effect during in vivo assembly. Inclusion of this cofactor could significantly change the assembly landscape as well, since it functions as a checkpoint that blocks binding sites until the intermediate reaches the correct conformation to continue assembly (57). However, the kinetics are likely difficult to measure, since they must be measured in vivo.

The actual distribution of messengers in bacteria and their diffusive behavior is not well-understood, and conflicting reports have been published stating that mRNA are freely diffusing throughout the cell (18), mRNA are addressed to certain subcellular areas in a sequence-specific way (58), and mRNA is localized near its originating operon (59). Though we assume that the mRNA can diffuse freely, we see that the regions with the largest density of 30S/mRNA association reactions are found near the originating operon of the messenger and outside of the nucleoid region. This distribution arises due to two effects. First, the new messenger is created at the location of its operon and cannot diffuse far before association with an SSU. Second, translating ribosomes are excluded from the nucleoid region, which leads to an accumulation of mRNA outside the nucleoid region from the dissociation into 30S, 50S, and mRNA.

In our whole-cell simulations, the ribosomes are distributed such that only 7% are found in the nucleoid region. In fast-growing *E. coli*, 12% are found in the nucleoid region (18). This is a reasonable result, since we are modeling slow-growing *E. coli*, where the chromosome is assumed to be densely packed into a single copy of the genome. It has been proposed that the segregation arises from maximizing the conformational entropy of the chromosome and the translational entropy of the ribosomes (60), but this alone does not explain the compaction of the chromosome seen in the stationary phase and translationally arrested cells. Our method for imposing a difference in ribosome densities between the two compartments is rather simplistic, but since the exact reason ribosomes are excluded from the nucleoid region is not clear, implementing a more physically realistic segregation mechanism may be premature. In the future, we will include the full DNA in our model in the form of a biased random walk, as used in our previous work (16).

It is known that in living *E. coli* cells, 15% of the ribosomes are not actively engaged in translation (61). Only

~25% of the 30S subunits are found in translating ribosome complexes in our simulations. This seems problematic, but in this model, only messengers that code for the SSU r-proteins uS3–bS20 are transcribed. This leads to overexpression of the r-protein, as well as underutilization of the available ribosomes. Transcription of mRNA that does not code for the r-proteins used in this model could restore the correct balance of free/transcribing ribosomes and could also correct the steady-state levels of protein and free messenger.

The number of ribosomes in a bacterial cell is observed to be roughly linearly correlated with the cell's growth rate. Such a relationship is captured by the empirical growth law (62,63), which draws parallels between growth rates of bacterial cells and how they allocate resources to protein synthesis and metabolic functions. However, the cell's effort to enforce such balance between metabolism and macromolecular synthesis is yet to be understood. This SSU assembly model can be combined with genome-scale models of metabolism and protein expression (64,65). Through network reduction methods and parameter space searches, these models could be integrated into our RDME simulations to simulate living cells.

The integration of metabolism with the model of ribosomal biogenesis would require the explicit regulation of rRNA and r-protein expression. Currently, we prescribe a constant transcription rate for each operon such that all r-protein is produced at approximately the same rate. Introducing gene regulation would alleviate the necessity of fine-tuning these rates. The two most important modes of regulation to model are the autoregulation of translation of r-protein mRNA and the regulation of transcription by ppGpp (4). In the autoregulation mechanism, certain free r-proteins can bind to their own transcripts, although at an affinity lower than that with which they bind to rRNA, inactivating the mRNA by blocking its translation. Any excess of r-protein will downregulate its own expression, leading to a small free r-protein pool. Most r-protein operons are regulated this way. The other mode of regulation is transcription deactivation via the global regulator, ppGpp, which is produced through the stringent response, i.e., during amino acid starvation conditions. The molecule binds to RNA polymerase, affecting its affinity to specific promoters. This effect depends on the sequence of the promoter, downregulating most of the genes necessary for growth, including r-protein and rRNA, and upregulating various stress-regulation genes and genes necessary for amino acid synthesis.

In summary, we have presented the first steps toward a whole-cell-level model of ribosome biogenesis in *E. coli*, starting with the assembly of the SSU. Our low-temperature in vitro assembly model fits the experimental kinetic data extraordinarily well and predicts previously unobserved assembly pathways. The high-temperature model reproduces the same binding timescales for all proteins measured in

in vitro studies and predicts key assembly intermediates in agreement with the cryoEM data. The high-temperature model was used to construct a spatially resolved, whole-cell model of ribosome biogenesis taking transcription and translation into account. The cellular environment was constructed to approximate slow-growing *E. coli* with a densely packed nucleoid region that excludes ribosomes. Although the assembly model was developed from experiments performed in vitro, with the increased cellular concentrations of r-protein, it yielded 30S assembly times comparable to those observed in experiments performed in vivo. The RDME model predicted nonuniform spatial distributions of mRNA and early 30S intermediates. Although simplified, this model has real predictive power and will be used as the basis for more complete models of ribosome biogenesis and cellular metabolism. SBML versions of the well-stirred simulation and LM 2.2.1 input files of the whole-cell simulations will be made available on our web site: http://www.scs.illinois.edu/schulten/research/ribosome_biogenesis_2015/. A tutorial describing the use of Lattice Microbes is available on our web site as well.

SUPPORTING MATERIAL

Supporting Materials and Methods, six figures, and four tables are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(15\)00765-1](http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00765-1).

AUTHOR CONTRIBUTIONS

T.M.E. designed the in vitro and in vivo models, developed the network generation and optimization code, analyzed and interpreted modeling results, and wrote the article. J.L. performed MD studies and in vivo cluster analysis, analyzed and interpreted MD trajectories, and wrote the article. K.C. started the initial development of the in vitro and in vivo models, analyzed simulation results, and wrote the article. M.J.H. modified Lattice Microbes for this work, performed timings on the code, and wrote the article. J.R.W. contributed the high-temperature data. Z.L.S. supervised the project and wrote and edited the article.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation grants PHY1026550 and MCB-1244570, and it is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. This work also used the Extreme Science and Discovery Environment (XSEDE), which is supported by National Science Foundation grant No. ACI-1053575. A portion of this research was sponsored by the DOE/BER (ORNL 4000134575) as part of the Adaptive Biosystems Imaging Focus at ORNL.

REFERENCES

1. Fox, G. E., L. J. Magrum, ..., C. R. Woese. 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. USA*. 74:4537–4541.
2. Roberts, E., A. Sethi, ..., Z. Luthey-Schulten. 2008. Molecular signatures of ribosomal evolution. *Proc. Natl. Acad. Sci. USA*. 105:13953–13958.
3. Ban, N., R. Beckmann, ..., M. Yusupov. 2014. A new system for naming ribosomal proteins. *Curr. Opin. Struct. Biol.* 24:165–169.
4. Kaczanowska, M., and M. Rydén-Aulin. 2007. Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 71:477–494.
5. Bremer, H., and P. P. Dennis. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd ed. F. C. Neidhardt, R. Curtiss, III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, editors. ASM Press, Washington, DC, pp. 1553–1569.
6. Liebermeister, W., E. Noor, ..., R. Milo. 2014. Visual account of protein investment in cellular functions. *Proc. Natl. Acad. Sci. USA*. 111: 8488–8493.
7. Held, W. A., B. Ballou, ..., M. Nomura. 1974. Assembly mapping of 30 S ribosomal proteins from *Escherichia coli*. Further studies. *J. Biol. Chem.* 249:3103–3111.
8. Adilakshmi, T., P. Ramaswamy, and S. A. Woodson. 2005. Protein-independent folding pathway of the 16S rRNA 5' domain. *J. Mol. Biol.* 351:508–519.
9. Adilakshmi, T., D. L. Bellur, and S. A. Woodson. 2008. Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature*. 455:1268–1272.
10. Kim, H., S. C. Abeyirigunawardene, ..., S. A. Woodson. 2014. Protein-guided RNA dynamics during early ribosome assembly. *Nature*. 506:334–338.
11. Talkington, M. W., G. Siuzdak, and J. R. Williamson. 2005. An assembly landscape for the 30S ribosomal subunit. *Nature*. 438:628–632.
12. Sykes, M. T., and J. R. Williamson. 2009. A complex assembly landscape for the 30S ribosomal subunit. *Annu. Rev. Biophys.* 38:197–215.
13. Bunner, A. E., A. H. Beck, and J. R. Williamson. 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.
14. Mulder, A. M., C. Yoshioka, ..., J. R. Williamson. 2010. Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science*. 330:673–677.
15. Sashital, D. G., C. A. Greeman, ..., J. R. Williamson. 2014. A combined quantitative mass spectrometry and electron microscopy analysis of ribosomal 30S subunit assembly in *E. coli*. *eLife*. 3. <http://dx.doi.org/10.7554/eLife.04491>.
16. Roberts, E., A. Magis, ..., Z. Luthey-Schulten. 2011. Noise contributions in an inducible genetic switch: a whole-cell simulation study. *PLOS Comput. Biol.* 7:e1002010.
17. Wang, W., G.-W. Li, ..., X. Zhuang. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*. 333: 1445–1449.
18. Bakshi, S., A. Sityaporn, ..., J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* 85:21–38.
19. Bakshi, S., H. Choi, ..., J. C. Weisshaar. 2014. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. *Mol. Microbiol.* 94:871–887.
20. Sanamrad, A., F. Persson, ..., J. Elf. 2014. Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. *Proc. Natl. Acad. Sci. USA*. 111:11413–11418.
21. Taniguchi, Y., P. J. Choi, ..., X. S. Xie. 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. 329:533–538.
22. Mahmutovic, A., D. Fange, ..., J. Elf. 2012. Lost in presumption: stochastic reactions in spatial models. *Nat. Methods*. 9:1163–1166.

23. Serban, R., and A. C. Hindmarsh. 2005. CVODES: the sensitivity-enabled ODE solver in SUNDIALS. *ASME 5th Int. Conf. Multibody Syst. Nonlinear Dynamics Control*. 6:257–269.
24. Liu, D. C., and J. Nocedal. 1989. On the limited memory BFGS method for large scale optimization. *Math. Program.* 45:503–528.
25. Benson, D. A., M. Cavanaugh, ..., E. W. Sayers. 2013. GenBank. *Nucleic Acids Res.* 41:D36–D42.
26. Cock, P. J. A., T. Antao, ..., M. J. L. de Hoon. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 25:1422–1423.
27. Hallock, M. J., J. E. Stone, ..., Z. Luthey-Schulten. 2014. Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations. *Parallel Comput.* 40:86–99.
28. Rodríguez, J. V., J. A. Kaandorp, ..., J. G. Blom. 2006. Spatial stochastic modelling of the phosphoenolpyruvate-dependent phosphotransferase (PTS) pathway in *Escherichia coli*. *Bioinformatics*. 22:1895–1901.
29. Lampoudi, S., D. T. Gillespie, and L. R. Petzold. 2009. The multinomial simulation algorithm for discrete stochastic simulation of reaction-diffusion systems. *J. Chem. Phys.* 130:094104.
30. Roberts, E., J. Stone, ..., Z. Luthey-Schulten. 2009. Long time-scale simulations of in vivo diffusion using GPU hardware. *Proc. IEEE Int. Symp. Parallel Distrib.* 2009:1–8.
31. Berk, V., W. Zhang, ..., J. H. D. Cate. 2006. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl. Acad. Sci. USA*. 103:15830–15834.
32. Best, R. B., X. Zhu, ..., A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.
33. Denning, E. J., U. D. Priyakumar, ..., A. D. Mackerell, Jr. 2011. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* 32:1929–1943.
34. Phillips, J. C., R. Braun, ..., K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
35. Recht, M. I., and J. R. Williamson. 2001. Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex. *J. Mol. Biol.* 313:35–48.
36. Zimmermann, R. A., A. Muto, ..., C. Branlant. 1972. Location of ribosomal protein binding sites on 16S ribosomal RNA. *Proc. Natl. Acad. Sci. USA*. 69:1282–1286.
37. de Narvaez, C. C., and H. W. Schaup. 1979. In vivo transcriptionally coupled assembly of *Escherichia coli* ribosomal subunits. *J. Mol. Biol.* 134:1–22.
38. Powers, T., G. Daubresse, and H. F. Noller. 1993. Dynamics of in vitro assembly of 16 S rRNA into 30 S ribosomal subunits. *J. Mol. Biol.* 232:362–374.
39. Chen, K., J. Eargle, ..., Z. Luthey-Schulten. 2012. Assembly of the five-way junction in the ribosomal small subunit using hybrid MD-Gō simulations. *J. Phys. Chem. B*. 116:6819–6831.
40. Lai, J., K. Chen, and Z. Luthey-Schulten. 2013. Structural intermediates and folding events in the early assembly of the ribosomal small subunit. *J. Phys. Chem. B*. 117:13335–13345.
41. Cannone, J. J., S. Subramanian, ..., R. R. Gutell. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*. 3:2.
42. Goss, D. J., L. J. Parkhurst, and A. J. Wahba. 1980. Kinetics of ribosome dissociation and subunit association. The role of initiation factor IF3 as an effector. *J. Biol. Chem.* 255:225–229.
43. Studer, S. M., and S. Joseph. 2006. Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol. Cell*. 22:105–115.
44. Milon, P., A. L. Konevega, ..., M. V. Rodnina. 2007. Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol.* 430:1–30.
45. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.
46. Kalwarczyk, T., M. Tabaka, and R. Holyst. 2012. Biologistics—diffusion coefficients for complete proteome of *Escherichia coli*. *Bioinformatics*. 28:2971–2978.
47. Golding, I., and E. C. Cox. 2004. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA*. 101:11310–11315.
48. Mandiyan, V., S. J. Tumminia, ..., M. Boublik. 1991. Assembly of the *Escherichia coli* 30S ribosomal subunit reveals protein-dependent folding of the 16S rRNA domains. *Proc. Natl. Acad. Sci. USA*. 88:8174–8178.
49. Wiggins, P. A., K. C. Cheveralls, ..., J. Kondev. 2010. Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc. Natl. Acad. Sci. USA*. 107:4991–4995.
50. Selinger, D. W., R. M. Saxena, ..., C. Rosenow. 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* 13:216–223.
51. Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA*. 89:3338–3342.
52. Lindahl, L. 1975. Intermediates and time kinetics of the in vivo assembly of *Escherichia coli* ribosomes. *J. Mol. Biol.* 92:15–37.
53. Shajani, Z., M. T. Sykes, and J. R. Williamson. 2011. Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* 80:501–526.
54. Taghbalout, A., and L. Rothfield. 2007. RNaseE and the other constituents of the RNA degradosome are components of the bacterial cytoskeleton. *Proc. Natl. Acad. Sci. USA*. 104:1667–1672.
55. Khemici, V., L. Poljak, ..., A. J. Carpousis. 2008. The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol. Microbiol.* 70:799–813.
56. Bunner, A. E., S. Nord, ..., J. R. Williamson. 2010. The effect of ribosome assembly cofactors on in vitro 30S subunit reconstitution. *J. Mol. Biol.* 398:1–7.
57. Connolly, K., J. P. Rife, and G. Culver. 2008. Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Mol. Microbiol.* 70:1062–1075.
58. Nevo-Dinur, K., A. Nussbaum-Shochat, ..., O. Amster-Choder. 2011. Translation-independent localization of mRNA in *E. coli*. *Science*. 331:1081–1084.
59. Montero Llopis, P., A. F. Jackson, ..., C. Jacobs-Wagner. 2010. Spatial organization of the flow of genetic information in bacteria. *Nature*. 466:77–81.
60. Mondal, J., B. P. Bratton, ..., J. C. Weisshaar. 2011. Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells. *Biophys. J.* 100:2605–2613.
61. Forchhammer, J., and L. Lindahl. 1971. Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of *Escherichia coli* 15. *J. Mol. Biol.* 55:563–568.
62. Scott, M., C. W. Gunderson, ..., T. Hwa. 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science*. 330:1099–1102.
63. Scott, M., and T. Hwa. 2011. Bacterial growth laws and their applications. *Curr. Opin. Biotechnol.* 22:559–565.
64. Lerman, J. A., D. R. Hyde, ..., B. O. Palsson. 2012. In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* 3:929.
65. Liu, J. K., E. J. O'Brien, ..., A. M. Feist. 2014. Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. *BMC Syst. Biol.* 8:110.