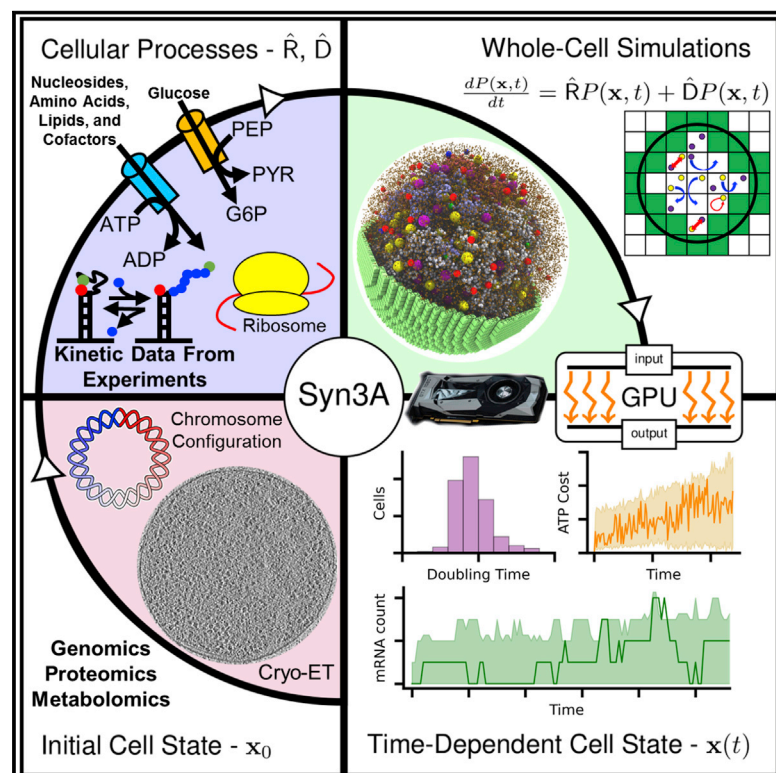


# Fundamental behaviors emerge from simulations of a living minimal cell

## Graphical abstract



## Authors

Zane R. Thornburg, David M. Bianchi, Troy A. Brier, ..., Hamilton O. Smith, John I. Glass, Zaida Luthey-Schulten

## Correspondence

zan@illinois.edu

## In brief

A kinetic model for a minimal bacterial cell offers quantitative insight into how the cell balances processes from metabolism to gene expression to growth.

## Highlights

- 3D spatial resolution of a fully dynamical whole-cell kinetic model
- Detailed single-reaction, single-cell accounting of time-dependent ATP costs
- Genome-wide mRNA half-lives emerge from length-dependent kinetics and diffusion
- Connections among metabolism, genetic information, and cell growth are revealed



## Article

# Fundamental behaviors emerge from simulations of a living minimal cell

Zane R. Thornburg,<sup>1</sup> David M. Bianchi,<sup>1</sup> Troy A. Brier,<sup>1</sup> Benjamin R. Gilbert,<sup>1</sup> Tyler M. Earnest,<sup>1</sup> Marcelo C.R. Melo,<sup>1</sup> Nataliya Safronova,<sup>2</sup> James P. Sáenz,<sup>2</sup> András T. Cook,<sup>3</sup> Kim S. Wise,<sup>3</sup> Clyde A. Hutchison III,<sup>3</sup> Hamilton O. Smith,<sup>3</sup> John I. Glass,<sup>3</sup> and Zaida Luthey-Schulten<sup>1,4,5,6,\*</sup>

<sup>1</sup>Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

<sup>2</sup>Technische Universität Dresden, B CUBE Center for Molecular Bioengineering, 01307 Dresden, Germany

<sup>3</sup>J. Craig Venter Institute, La Jolla, CA 92037, USA

<sup>4</sup>NSF Center for the Physics of Living Cells, Urbana, IL 61801, USA

<sup>5</sup>NIH Center for Macromolecular Modeling and Bioinformatics, Urbana, IL 61801, USA

<sup>6</sup>Lead contact

\*Correspondence: [zan@illinois.edu](mailto:zan@illinois.edu)

<https://doi.org/10.1016/j.cell.2021.12.025>

## SUMMARY

We present a whole-cell fully dynamical kinetic model (WCM) of JCVI-syn3A, a minimal cell with a reduced genome of 493 genes that has retained few regulatory proteins or small RNAs. Cryo-electron tomograms provide the cell geometry and ribosome distributions. Time-dependent behaviors of concentrations and reaction fluxes from stochastic-deterministic simulations over a cell cycle reveal how the cell balances demands of its metabolism, genetic information processes, and growth, and offer insight into the principles of life for this minimal cell. The energy economy of each process including active transport of amino acids, nucleosides, and ions is analyzed. WCM reveals how emergent imbalances lead to slowdowns in the rates of transcription and translation. Integration of experimental data is critical in building a kinetic model from which emerges a genome-wide distribution of mRNA half-lives, multiple DNA replication events that can be compared to qPCR results, and the experimentally observed doubling behavior.

## INTRODUCTION

An overarching goal of molecular biology is to explain the basic processes of life in terms of the laws of physics and chemistry. In 1984, Morowitz proposed the study of the simplest living cells, the Mycoplasmas, as models for understanding the fundamental principles of life (Morowitz, 1984). Just as the study of hydrogen, the simplest atom, led to the understanding of more complex atoms, it seems plausible that the study of the simplest living cells will reveal principles that apply to all living systems. For this reason, we have been interested in studying “minimal cells” by designing and building cellular genomes that do not include genes that are non-essential in the laboratory. We have been able to produce living cells with fewer genes than any known naturally occurring cell (Hutchison et al., 2016; Breuer et al., 2019). Such cells should be easier to completely describe than any known naturally occurring cells. To approach the question “what is life?” using our minimal cell model, we are testing whether the combined functions of the minimal cell genes can inform a computer model that correctly predicts the behavior of the cell. This will provide a test of our understanding of the minimal requirements for life. Such a model will also provide a design tool for predicting the effects of changes in the genome; for example, when a reaction pathway is added.

A complete description of the state of the cell requires knowledge of its size, shape, components, intracellular reactions, and interactions with its environment, all of these as a function of time and cell growth. Adding to this list is the need for theoretical models and simulations that interpret and integrate this daunting amount of experimental data. Due to large numbers of genes with unknown function and the complexity in model systems such as *Escherichia coli*, along with the broad range of concentrations and timescales that need to be considered, simulating a complete description of the state of a cell has been challenging. The development of whole-cell models (WCMs) and how they have progressed from genome-scale metabolic models (GSMs) (Varma and Palsson, 1994) and the calculation of their steady-state fluxes has been recently reviewed (Goldberg et al., 2018; Marucci et al., 2020). The most comprehensive models have been developed for *Mycoplasma genitalium* and *E. coli* (Karr et al., 2012; Macklin et al., 2020), where the subsystems were treated in terms of ordinary differential equations, flux balance analysis, and stochastic simulations. Common challenges are establishing the reaction networks and the availability of kinetic parameters and -omics data such as metabolomics and proteomics to use as initial conditions. No one bacterium has a complete set of parameters and -omics data, so the development of a WCM relies upon synthesizing information from other



well-studied organisms. Unlike the previous WCMs, the simulations we present here are based on fully dynamical kinetic models where subsystem networks and chemical species are interconnected continuously over time on a single-cell basis.

An ideal system for such a whole-cell model would be a minimal cell consisting of as few genes and reactions as possible for the cell to grow and divide (Luthey-Schulten, 2021). JCVI-syn3A is a genetically minimal bacterial cell, consisting of only of 493 genes on a single 543-kbp circular chromosome with 452 genes coding for proteins (Breuer et al., 2019), some of which are subunits of multi-domain complexes (NCBI GenBank: CP016816.2). Syn3A's genome and physical size are approximately one-tenth those of the model bacterial organism *E. coli*. Syn3A has a smaller fraction of genes with unclear function (87/452, 20%) than *E. coli* (1780/4637, 38%) and *Mycoplasma pneumoniae* (311/688, 45%) (Breuer et al., 2019). The reduction in complexity and scale of Syn3A presents a unique opportunity to develop a near-complete whole-cell kinetic model. The Syn3A genome was synthesized based on the known genome sequence of the natural parent Gram-positive organism *Mycoplasma mycoides subsp. capri* str. GM12 (GenBank: CP001621.1) and has been synthetically reduced to achieve a minimal genome producing living cells that grow, divide in about 100 min, and have consistent spherical morphologies with 400–500-nm diameters (Breuer et al., 2019; Pelletier et al., 2021; Hutchison et al., 2016; Gibson et al., 2010).

In Breuer et al. (2019), we established the essential GSMM for Syn3A along with genome-wide gene essentiality and proteomics. The protein products of 155 genes involved in 175 metabolic reactions were organized into seven subsystems: central, nucleotide, lipid, cofactors, amino acid, ions, and macromolecule metabolism, providing the starting point for the kinetic metabolic model presented here. The reactions in macromolecule metabolism are now kinetically modeled (Thornburg et al., 2019) through approximately 2,000 reactions involving the 251 genes in the genetic information processes of DNA replication, transcription of all 493 genes, translation and degradation of all 452 mRNA, tRNA charging, and cell growth. In addition to biochemical reactions, whole-cell, 3D spatial models require cellular architecture, including spatial distributions of ribosomes and configurations of the circular chromosome. The cellular architectures (Figure 1) are reconstructed at the single-cell level directly from cryo-electron tomograms (cryo-ET) that reveal a near-random distribution of ribosomes throughout the cell with a few present in polysomes (Gilbert et al., 2021).

Kinetic parameters for the majority of the cellular reactions have been measured in related organisms through decades of biochemical, single-molecule (sm) FRET, and spectroscopic studies reported in the literature and kinetic databases like Bremer and Dennis (2008), BRENDA (Chang et al., 2021), and equilibrium constants reported in NIST's TECRdb (Goldberg et al., 2004) and Equilibrator (Flamholz et al., 2012). Comparative proteomics analyses to *Mesoplasma florum* (Matteau et al., 2020; Lachance et al., 2021), *B. subtilis* (Wang et al., 2015), and *E. coli* (Taniguchi et al., 2010) were used to approximate missing or questionable information regarding a few of the Syn3A enzymes. At the moment, only relative metabolomics data on Syn3A is available, so the metabolite concentrations used to initialize the

simulations of the Syn3A WCM were estimated from a scaling of the comprehensive study done on *E. coli* (Park et al., 2016) and a limited list from *M. pneumoniae* (Yus et al., 2009). For completeness, all the modifications to the metabolic map, genetic information processing, and kinetic parameters are provided in Figures 2, S1, and S2; Tables S1 and S2; and STAR Methods (Metabolic rates and parameterization).

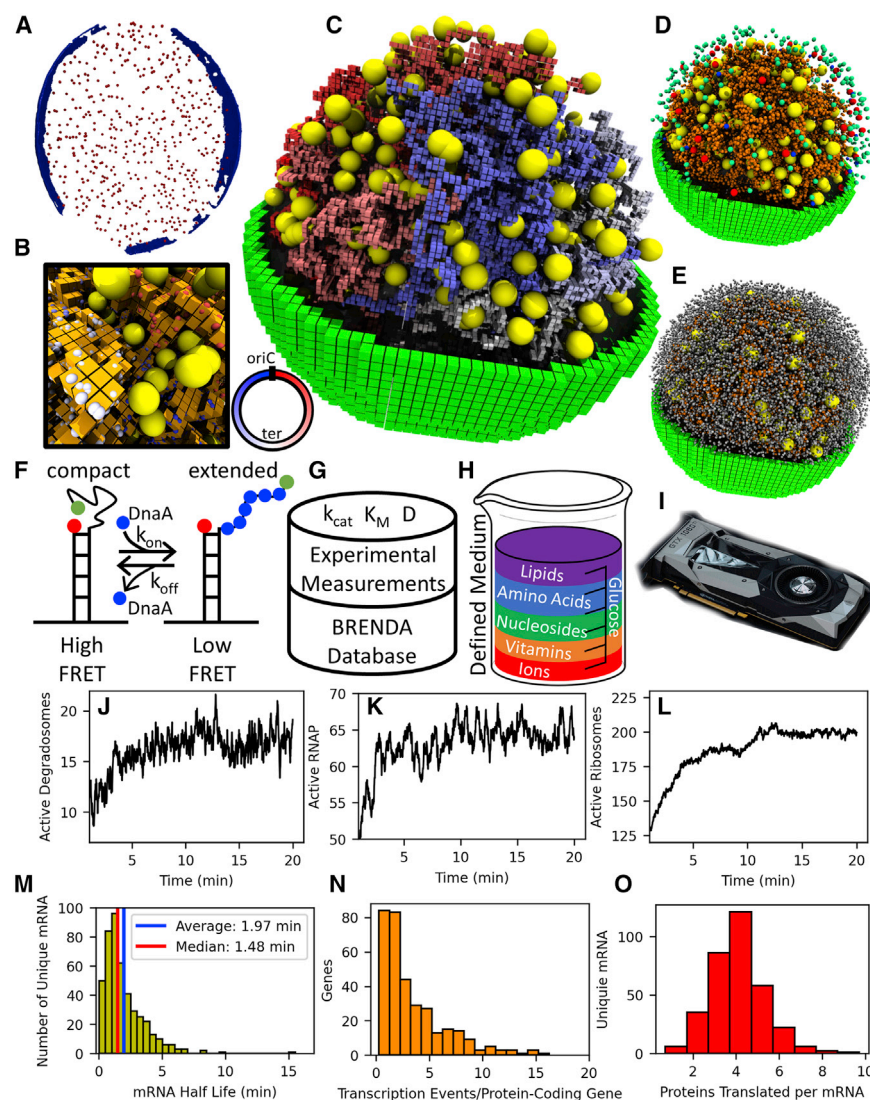
With the background data now available for Syn3A, we were able to develop a whole-cell kinetic model of this minimal cell. Because of the large variation in timescales and concentrations, developing a whole-cell model that treats metabolism, genetic information processes, and growth can, at the moment, only be achieved by hybrid stochastic and deterministic simulations. Kinetics of the essential metabolic network (Breuer et al., 2019) are handled deterministically via ordinary differential equations (ODEs), and the kinetics of the genetic information processes are handled with stochastic simulations. We consider here two models of the stochastic simulations (Figure S3): a chemical master equation (CME) description that assumes that the whole cell is well stirred and implicitly includes the effects of diffusion in the rates of transcription, mRNA degradation, and translation; and a reaction-diffusion master equation (RDME) description that requires macromolecules to diffuse to each other for reactions to take place in the spatially heterogeneous environment of the cell. In our simulations, we record time-dependent particle counts of each molecule and intermediate, fluxes of all metabolic reactions, and in the spatial model, the position of each macromolecule within the cell. We present results of the well-stirred model over a complete cell cycle for 174 healthy replicate cells out of a total population of 207 cells. Unhealthy cells within the simulations run out of phosphoenol pyruvate (pep), halting glycolysis.

For the spatial model, which is computationally expensive, we use our graphics processing unit (GPU)-based Lattice Microbes software (Roberts et al., 2013; Hallock et al., 2014; Earnest et al., 2018) to simulate replicates over the first 20 min of the cell cycle, before DNA replication and substantial cell growth have occurred. These relatively short-term simulations are critical in calculating binding probabilities and estimating mRNA half-lives in the well-stirred model. Being computationally less expensive, the well-stirred model runs on CPUs and provides a whole-cell model that can be easily complexified to allow addition of more pathways and testing. More importantly, it allows us to quantitatively understand the “principles of life” for a minimal cell growing with little regulation.

## RESULTS

### 3D spatial whole-cell simulations incorporate experimental data and inform homogeneous well-stirred simulations

The general workflow of constructing an initial cell state is shown in Figure 1. Starting from the coordinates of the 503 ribosomes and cell boundary of a small cell with ~ 400-nm diameter from cryo-ET (Figure 1A), the DNA is folded around ribosomes as a circular self-avoiding polymer on a lattice in such a way that the sequence order and gene positions are maintained (Figures 1B and 1C) (Gilbert et al., 2021). Experimental 3C maps showed



**Figure 1. Workflow for whole-cell simulations**

(A) Ribosome coordinates and cell boundaries are obtained from cryo-electron tomograms.

(B) The self-avoiding lattice DNA (red, white, and blue spheres) is folded around the ribosomes (yellow spheres).

(C) The membrane (green cubes) surrounds the ribosomes, DNA, and 200-nm radius cytoplasmic space.

(D) A representative set of membrane complexes and proteins (degradosomes—red spheres, SecY—blue spheres, and PtsG—green spheres) are randomly distributed in the peripheral membrane and transmembrane space.

(E) All other macromolecules are randomly distributed throughout the cytoplasm as shown in all gray spheres.

(F) Some rates have been reported from single-molecule experiments such as the DnaA filament formation rate.

(G) Otherwise, we used the BRENDA and other databases for kinetic rates.

(H) The defined medium composition is used to determine nutrient uptake in our simulations.

(I) Spatial simulations require GPU acceleration.

(J–L) The spatial simulations predict numbers of (J) active degradosomes breaking down mRNA, (K) transcribing RNAP, and (L) translating ribosomes.

(M) The length-dependent kinetics of mRNA decay and requiring mRNA to diffuse to degradosomes results in a distribution of mRNA half-lives.

(N) The average number of times each gene is transcribed over the course of the 20-min spatial simulations.

(O) A distribution of the average number of times each mRNA type is translated in its lifetime shows that every mRNA is translated at least once in its lifetime on average.

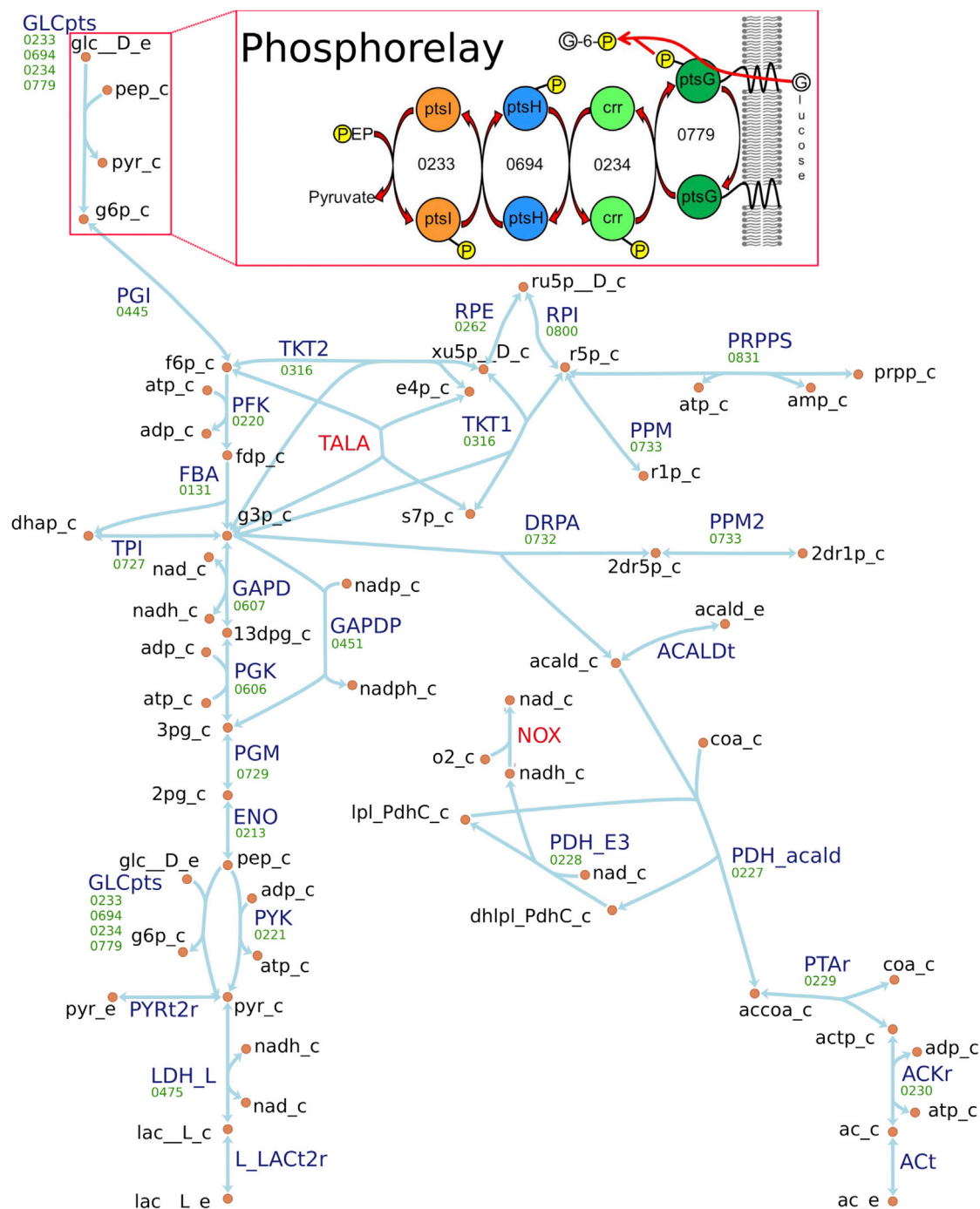
3D visualization done with VMD (Humphrey et al., 1996).

no significant features of persistent supercoiled domains or loops, so the chromosome is assumed to be in a relaxed state (Gilbert et al., 2021). Each replicate cell uses the same ribosome coordinates from cryo-ET, but a different chromosome configuration unique among the ensemble. The top of the membrane is cut away in Figure 1C to reveal the ribosomes and DNA. According to the 3C-seq maps and the proteomics of nucleoid-associated proteins (NAPs) (Gilbert et al., 2021; Breuer et al., 2019), the DNA configurations are assumed to be relaxed with no supercoiling so that the genes are easily accessible. Figure 1D shows 120 degradosome complexes in red, 66 SecY proteins in blue, and 831 PtsG proteins in green as three examples. The remainder of the proteome consisting of over 77,000 proteins, 200 mRNA, and 5,800 tRNA are then randomly distributed throughout the cytoplasm and membrane, resulting in the crowded environment (Figure 1E).

Our spatial model includes a total of 7,765 unique molecules and intermediates and over 7,200 reactions including binding re-

actions such as RNAP binding to a gene start site. Where possible, kinetic parameters are obtained from single-molecule experiments, such as the smFRET experiment for the formation of the DnaA filament along the AT-rich single-stranded DNA (Figure 1F) near the origin. Otherwise, as described in STAR Methods (Metabolic rates and parameterization), kinetic parameters are developed from a targeted survey of the primary literature or kinetic databases (Figure 1G) as discussed above. Simulations of a spatially resolved cell are computationally expensive and require GPU (Figure 1I) acceleration to make them possible on a human timescale (Hallock et al., 2014). The GPUs used for spatial simulations included NVIDIA Titan V and NVIDIA Tesla Volta V100 GPUs, which took 10 h and 8 h to simulate 20 min of cell time, respectively. Because of this computational expense, we simulated the first 20 min of the cell cycle for only 8 cells, a limited time frame during which we assume no substantial growth or DNA replication has occurred. The simulations provide insight into the numbers of active degradosomes, RNAP,





**Figure 2. The JCVI-syn3A central metabolic network**

Central metabolism starts with glucose uptake via the phosphorelay mechanism shown in the inset including protein names and gene numbers. Red reaction names indicate reactions that do not have a gene annotated but are assumed to be performed by one of the uncharacterized proteins. ACALDt, PYRt2r, L\_Lact2r, and ACT are all assumed to be non-enzymatic passive transport reactions. The reaction map was generated using Escher (King et al., 2015). See also Figures S1 and S2.

and ribosomes (Figures 1J–1L). The early increases are due to the initial conditions of the spatial simulations. The cell is initialized with no active complexes (no RNAP are on the chromosome and no mRNA are on ribosomes and degradosomes), and the

transient behavior reflects the time required for RNAP to diffuse to genes and mRNA to be translated or degraded.

Bremer and Dennis (2008) calculated that anywhere from 15.5% to 36.2% of RNAP are active at any one time in *E. coli*

depending on the doubling time, with slower-growing cells having a smaller fraction of active RNAP. The spatial model predicts that Syn3A will have an average of 63 of its 187 RNAP active (34%) early in the cell cycle, falling within the calculated range. For fast-growing *E. coli*, approximately 80% of the ribosomes are active (Bremer and Dennis, 2008; Dai et al., 2016), but for slow-growing *E. coli*, this number drops between 20%–50% depending on the growth rate (Dai et al., 2016). The spatial model predicts that, on average, 220 of the 503 ribosomes, roughly 45% of ribosomes, are active.

We calculated the mRNA half-lives, the number of times a gene is transcribed, and the number of times an mRNA is translated in its lifetime for all the 452 protein-coding genes (Figures 1M–1O). The average and median half-lives are in reasonable agreement with the 2-min average half-life experimentally measured in *Mycoplasma gallisepticum* (Kirk and Morowitz, 1969). The broad distribution of half-lives, including the long tail out to 15 min, has been observed in a genome-wide study of mRNA half-lives in *B. subtilis*, Hambraeus et al. (2003). Each gene is transcribed at least once within the ensemble of simulations, but not in every simulation. The number of times each gene is transcribed reflects both its length and, more importantly, its promoter strength, which is weighted relative to proteomics counts. Lastly, the genome-wide average translations per mRNA is four times, but several factors impact this number including gene length and how many times a mRNA can be read by a polysome using a polysome spacing of 120 nt estimated from a distribution of polysome sizes in *E. coli* (Brandt et al., 2009).

### The kinetic model is influenced by the defined medium composition and new genome annotations

The time-dependent metabolite concentrations within the cell are determined by the metabolic reactions that depend on transport of key metabolites like glucose, nucleosides, fatty acids, amino acids, and cofactors. With a defined growth medium, exact uptake kinetics can be simulated using the external metabolite concentrations and the numbers of transporters. The metabolic maps in Syn3A here have been revised to be consistent with the defined growth medium, updated gene annotations, and experimental measurements such as lipidomics. To refer to genes in JCVI-syn3A, we simplify the locus tags from the NCBI entry from JCVISYN3A\_xxxx to xxxx. For example, JCVISYN3A\_0527 is referred to as gene 0527.

The only sugar source in the defined medium is glucose, so the revised map for central metabolism (Figure 2) starts with the phosphorelay relay system (Rohwer et al., 2000; Meadow et al., 2005a, 2005b), which is responsible for the uptake and phosphorylation of glucose to glucose-6-phosphate (g6p) and is shown in the inset. Each phosphate exchange reaction of the phosphorelay is simulated independently, and the overall kinetics for the phosphorelay predict that Syn3A takes up 15,000 glucose molecules per second for a cell with a radius of 200 nm. Syn3A does not have the proteins to perform oxidative phosphorylation, so all ATP in Syn3A is generated by the central metabolism (Breuer et al., 2019). Pyruvate kinase (*pyk*/0221) (PYK) converts ADP to ATP and pep to pyruvate, competing for pep molecules with the glucose transport reaction, so the

fluxes between these two reactions need to be carefully balanced throughout the cell cycle. Because the fructose-1,6-bisphosphate aldolase (*fbaA*/0131) (FBA) reaction splits fdp into two molecules, the rate of lower glycolysis is twice that of upper glycolysis. Therefore, a maximum of 45,000 ATP can be generated per second assuming no other NTPs are being formed: 30,000 ATP per second can be generated by the phosphoglycerate kinase (*pgk*/0606) (PGK) and 15,000 by pyruvate kinase (PYK) where pep is split between the glucose uptake reactions and PYKs. A much smaller amount of ATP is generated through acetate kinase (*ackA*/0230) (ACKr) upon the secretion of acetate. In our model, the cell was able to survive off of the ATP generated by central metabolism. Detailed discussions of the amino acid, cofactor, nucleotide, and lipid metabolic subsystems are included in STAR Methods.

The kinetic model of the genetic information processing reactions including DNA replication initiation and elongation, mRNA degradation, transcription, and translation are based on the kinetic model by Thornburg et al. (2019) with a few modifications. Each of the elongation reactions are treated using a polymerization rate dependent on the respective monomer concentrations (dNTPs, NTPs, aa:tRNAs). For full details on these rate forms, see STAR Methods (Genetic information processes).

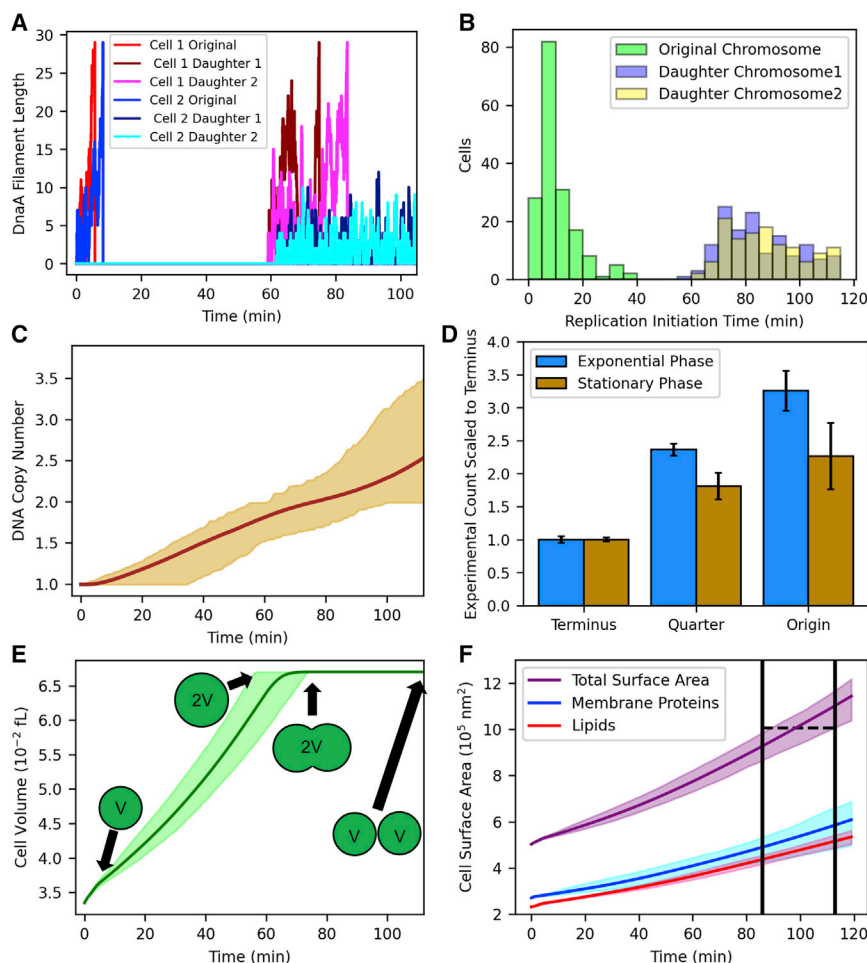
### Timing of the cell cycle and cell growth are determined by dynamics of DNA replication and surface area growth

Unlike the spatial RDME-CME-ODE model, cells are simulated for whole cell cycles in the well-stirred CME-ODE model. To highlight the key features defining the cell cycle in Syn3A, we examine in Figure 3 the time dependence for the initiation of DNA replication, chromosome duplication, and the doubling of both the cell volume and surface area in the well-stirred simulations. In our previous work (Thornburg et al., 2019), only one replication event was allowed to occur per cell cycle, but here we complexify this model and allow for multiple replication initiation events based solely on the kinetics. We use the same equation for the rate of DnaA(III) filament formation on ssDNA for each independent origin

$$\nu = k_{on}[DnaA] - k_{off} = (k_{on} / N_A V(t)) N_{DnaA} - k_{off}. \quad (\text{Equation 1})$$

Both  $k_{on}$  and  $k_{off}$  for the binding of DnaA(III) were measured using smFRET (Cheng et al., 2015). We allow replication initiation events on the daughter chromosomes in our model after they separate following the first replication cycle. The initial cell volume is used for the first replication initiation event, and the doubled cell volume is used for the daughter chromosomes late in the cell cycle. Based on these kinetics, if the number of DnaA doubles faster than the volume, it is likely that another replication initiation event will occur and the cell will have more than two chromosomes at the end of its cell cycle.

Two representative cells were selected to demonstrate the stochastic nature of replication initiation, and their DnaA filament lengths are shown as functions of time (Figure 3A). The first cell had completed its first replication initiation event around 5 min into the cell cycle. Both daughter chromosomes then initiate events late in the cell cycle. The second cell did not translate enough DnaA for it to be favorable for either daughter



**Figure 3. Processes determining the cell cycle progression**

(A) DNA replication initiation in two individual cells with and without subsequent replication initiation on the daughter chromosomes.

(B) Distributions of replication initiation times for 174 original chromosomes of which 141 of the daughter chromosomes had further replication initiation events.

(C) The DNA copy number as a function of time among 174 cells. The solid line is the average and the shaded region represents the full range among the population.

(D) Relative quantities of origins, quarter positions, and termini of the chromosome from qPCR for exponential and stationary phase cells. Error bars represent the standard deviation among six biological replicates. Exponential phase standard deviations are 0.05 (Terminus), 0.09 (Quarter), and 0.3 (Origin). Stationary phase standard deviations are 0.03 (Terminus), 0.2 (Quarter), and 0.5 (Origin).

(E) The cell doubles in volume between 50 to 70 min.

(F) The cell surface area doubles between 88 to 112 min. Cells maintain a 55:45 ratio of protein surface area to lipid surface area over the course of a cell cycle.

See also Figure S5.

chromosome to complete a full DnaA filament. The distribution of replication initiation times for 174 cells (Figure 3B) for the original chromosome ranges from 3 min up to 36 min with an average time of 10 min and most probable time of 6 min. Of the 174 cells, 33 had only one replication event occur in their cell cycle. The daughter chromosomes in cells with multiple replication events have replication initiation times ranging from 58 min up to 110 min with an average time of 82 min.

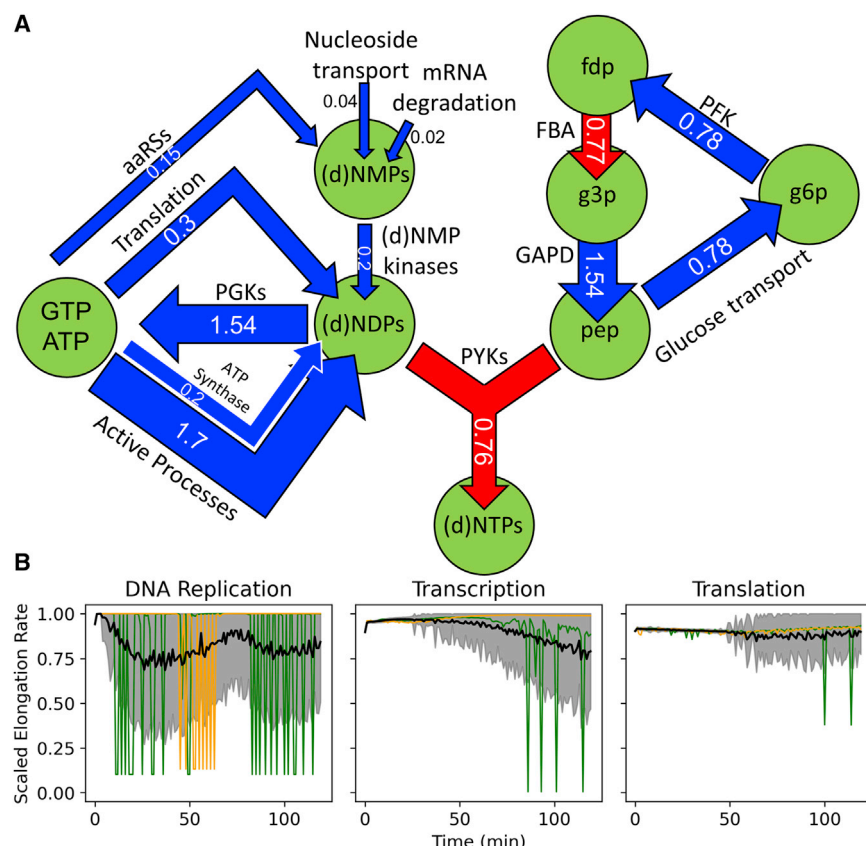
Each cell is initialized with a single chromosome, and from Figure 3C we find that the average duplication time for the original chromosome is around 70 min with the earliest being completed at 56 min and the latest at 90 min. The average chromosome number grows to 2.8 at the end of the cell cycle, which reflects that either one or both of the daughter chromosomes have partially completed another chromosome. Cells with multiple replication events can have chromosome copy numbers as high as 3.8, indicating that they have nearly duplicated both daughter chromosomes in their cell cycle.

Although other bacteria have been shown to undergo multiple replication events per cell cycle (Helmstetter and Cooper, 1968; Nielsen et al., 2007), it had not yet been directly observed in Syn3A. From quantitative polymerase chain reaction (qPCR) experiments, we have determined the relative quantities of origins,

cells, after the first replication initiation event occurs, another replication initiation event will occur on the same chromosome before the first replication cycle completes. In our current model, a maximum ratio of 2 would occur, as we do not allow multiple initiation events to occur until the first replication cycle completes.

Critical to determining the length of the cell cycle are the times required to double the volume and surface area of the cell. The kinetics for cell growth are characterized by the formation of lipids and insertion of membrane proteins to determine the cell surface area and volume in the model. The cell volume is calculated from the surface area assuming the cell maintains spherical morphology until the onset of division. The volume doubles in the simulations anywhere from 56 to 72 min with an average of 64 min (Figure 3E). Without explicit kinetics for cell division by FtsZ/FtsA filaments, division is assumed to begin when the volume has doubled, during which the volume stays constant and the surface area continues to grow until two separate cells are formed. The surface area will double anywhere from 88 to 112 min with an average doubling time of 97 min (Figure 3F). Syn3A has an experimentally measured doubling time of 105 min in rich growth medium (Breuer et al., 2019). We report a simulated doubling time (Figures 3E and 3F) based only on

quarter positions, and termini in Syn3A cells in both exponential and stationary phases (Figure 3D). The results are presented with all the quantities scaled to the number of termini. In the exponential phase cells, there are more than three times the number of origins than termini on average. This indicates that in many



**Figure 4. The dominant connections between gene expression and cellular metabolism**

(A) A simplified diagram of the ATP and GTP use, nucleotide metabolism, and glycolysis pathway show connections among the networks. Arrow width corresponds to rate through the reaction given in mM/s on the arrow. Red arrows indicate the rate-limiting steps of glycolysis in the simulations.

(B) If a cell runs low on dNTPs, NTPs, or charged tRNAs, the rates of the corresponding genetic information processing reactions are reduced (*dnaA/0001*). The black trace shows the average rate among the population with the full range in gray. The orange and green traces represent two individual cells. See also Figure S4.

membrane proteins, which are known to be underreported by proteomics when only a trypsin digest is used (D. Gonzalez, personal communication).

### Balance of genetic information processes and metabolism

Due to reduction in its minimal genome, Syn3A has few remaining transcription/translation/transport regulatory proteins and must adjust the fluxes through the cellular subsystems to maintain stable growth. The simplified map of the reaction network with fluxes from a representative cell from the well-stirred model early in its cell cycle demonstrates the balance among use of ATP and GTP, nucleotide metabolism, and glycolysis (Figure 4A). The glycolysis pathway and nucleotide metabolism are connected through the PYK reactions converting all (d)NDPs to (d)NTPs, which results in the shared usage of pep with glucose uptake. Syn3A exclusively makes pep by the action of enolase at the end of glycolysis, generating two pep per glucose taken up. Because the glucokinase was removed in genome reduction (Breuer et al., 2019), we assume that the only way Syn3A can phosphorylate glucose is by PtsG in the phosphorelay. According to our kinetic model, if the cell runs out of pep, there is no way to continue glycolysis or phosphorylate NDPs except ADP, which can be converted to ATP by the reversible ATP synthase. On average, if PYK reactions use more than half of the pep formed, less glucose will gradually be taken up and the cell will run out of pep and cease to take up glucose. Roughly 16% of the cells in a total ensemble of 207 cells in the well-stirred simulations experienced pep's shortage, leaving 174 cells that could successfully complete a cell cycle.

Previous studies have indicated the possibility of PYK, PFK, and GAPD being rate-limiting reactions of glycolysis (Iwami and Yamada, 1980; Boscá and Corredor, 1984); however, fructose-1,6-bisphosphate aldolase (FBA) appears to control the overall flux of glycolysis in Syn3A according to our simulations, which agrees with the findings of Kitamura et al. (2021). FBA

healthy cells, whereas the experimentally measured doubling time includes a whole population, which may include unhealthy cells, reducing the average doubling time of the colony.

To determine the cell surface area, each lipid and membrane protein has an assigned surface area contribution (STAR Methods, Lipid metabolism). The contributions of proteins and lipids to the surface area are separated in Figure 3F where the two contributions are seen to maintain a rough 55% to 45% surface area contribution ratio (Sáenz et al., 2012). Because the translation and insertion of membrane proteins are both totally stochastic in the simulations, there is more variation in the surface area contribution from membrane proteins. The variation in lipids increases with simulation time as lipid synthesis genes are stochastically expressed in each individual cell.

The liponucleotide synthase *cdsA/0304* catalyzing reaction DASYN (Figure S2) that adds CDP to a phospholipid precursor (phosphatidic acid, PA) was identified as a "choke point" in the phospholipid biosynthesis, in agreement with a recent whole-cell model of *E. coli* (Macklin et al., 2020). Additionally, acyltransferase *plsY/0117*, which catalyzes the conjugation of fatty acid and glycerol moieties at the membrane was found to limit the production of the downstream intermediate PA. We attribute both of these effects to their low counts in the reported proteomics of Syn3A (Breuer et al., 2019) and adjusted the counts to values similar to those observed in other bacterial species (Table S2 and STAR Methods, Lipid metabolism). Their low counts are likely due to the fact that both are multiple domain



has a low experimental proteomics count of 227 relative to the other glycolytic enzymes in Syn3A, having counts of 400 or greater. Relative to other bacteria, Syn3A appears to have a lower FBA count (Table S1), so in parameterizing the simulations, the count of FBA was scaled to 775 based on the enzyme's concentration in *E. coli* so that kinetic parameters would better match the known equilibrium constant for the reaction. The FBA reaction is almost always at its maximum rate and is the slowest step in upper glycolysis in our models. The rates of the reactions in lower glycolysis are dictated by how much flux goes through the FBA reaction.

The balance of metabolic reactions also affects the rates of genetic information reactions through the monomer-dependent polymerization reaction rates discussed in STAR Methods (Genetic information processing). The time-dependent rates of each elongation reaction (DNA replication, transcription, and translation) scaled to its maximum rate for *dnaA*/0001 expression are shown in Figure 4B. The single cell in orange has only one complete DNA replication event occur in its cell cycle, and the green cell has a second. Their replication rates are at the maximum until the cell runs low on dNTPs (< 0.01 mM) (Figure S4), in this case dATP (data not shown). The slowdown gives the cell time to import more deoxynucleosides and generate more dNTPs. The rate of replication will fluctuate from minute to minute as long as the cell runs low on a particular dNTP. In general, DNA replication rate is the most frequently affected of the genetic information processes with its average (black) going as low as 75% of the maximum.

There are no significant deviations in the transcription and translation rates for the cell with a single replication event (orange) (Figure 4B), as the instantaneous pool sizes of NTPs and charged tRNA remain high enough to not slow down any rates. For cells with multiple replication events (green), more RNA are transcribed, potentially depleting concentrations of NTPs, thereby reducing its transcription rate. This slowdown in transcription leads to pauses during which nucleosides can be taken up and phosphorylated.

Translation is infrequently altered because the amount of charged tRNA (aa:tRNA) depends directly on the uptake of amino acids, which are present in millimolar concentrations in the medium. So even though cells will sometimes run low on amino acids, a brief slowdown is enough for them to import more amino acids and recover their charged tRNA levels ( $>10^2$  of each aa:tRNA).

### Time-dependent ATP costs in the minimal cell quantify the cell's significant energy costs

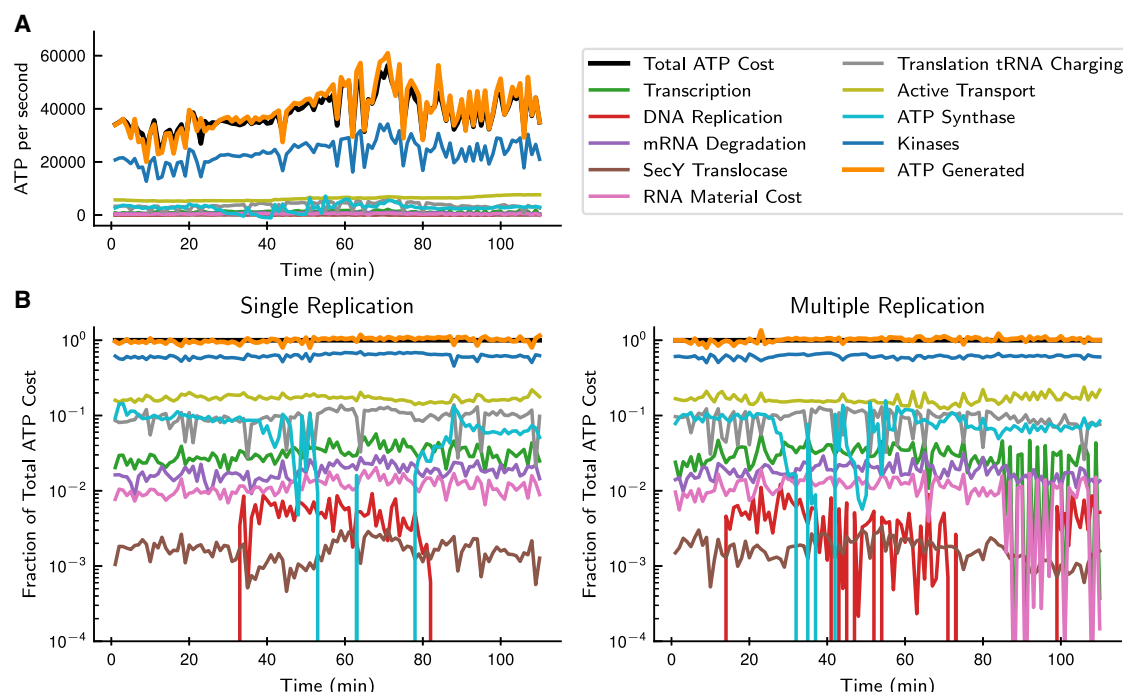
A key feature of our model is the explicit tracking of every energy molecule used by activated reactions in both metabolism and genetic information processes. Already in 1973, Stouthamer (1973) comprehensively calculated the amount of ATP required for the formation of a microbial cell based on the studies available on energetic costs at the time. He broke down the ATP demands of a cell into requirements for formation of polysaccharides, proteins, lipids, RNA, and DNA, uptake of amino acids, phosphate, and ions, and turnover of RNA. He notes that exactly accounting for transport reactions is difficult and little information on their ATP costs were available at the time, so only a few transport reactions are included in his calculations. More recently, a comprehensive

review of the literature and calculations for both bacteria and eukaryotes (Lynch and Marinov, 2015) assigned costs for the synthesis of macromolecules and determined the total ATP costs for DNA replication, transcription, and translation.

While the previous reviews calculated ATP costs for overall cell formation, we advance ATP cost calculations to single-cell, single-reaction resolution as a function of time for the cellular networks of Syn3A including both metabolism and genetic information processing, which is made possible by fully dynamical kinetic modeling. Figure 5A shows the ATP generated and used as functions of time for a representative cell from the well-stirred simulations. Note that these plots do not represent the number of phosphate bonds made and broken, but the number of ATP molecules being used. Another notable difference is that tRNA charging is only counted as one ATP, whereas it is typically counted as two phosphate bonds (Lynch and Marinov, 2015). The charging reactions still convert ATP to AMP and pyrophosphate in the simulations (Figure S3), but because the metabolic network explicitly accounts for the ATP cost of regenerating an AMP to an ADP through the ADK1 reaction (Figure S1), tRNA charging is counted as a single ATP cost in Figure 5. Additionally, because translation elongation uses GTP instead of ATP, it is not shown in the cost plot, but it is twice what we defined as the ATP cost of charging tRNAs from two GTP per amino acid: one during the loading of an aa:tRNA into the A site of the ribosome and a second during translocation to the next step on the mRNA (Lynch and Marinov, 2015).

The total ATP generated at each time step is close to or slightly greater than the total ATP used. As discussed earlier, the maximum ATP production is 45,000 ATP per second assuming no other NTPs are being made by PGK or PYK. Roughly 35,000 ATP are made per second initially (Figure 5A). As the cell grows, the number of proteins and associated rates of metabolic reactions increase, giving rise to the overall increase in both the ATP production and cost over the cell cycle. The growth is not perfectly smooth or linear because the protein counts and reaction rates depend on stochastic gene expression and timing of DNA replication events. To better compare the relative ATP cost of each activated process, we plot their fraction of the total ATP cost for a cell with a single replication event (Figure 5B, left) and a cell with multiple replication events (Figure 5B). The highest cost in Syn3A is for metabolic reactions, in particular the PFK reaction in upper glycolysis using 75% of the total metabolic cost.

Quantifying the exact cost of activated transport reactions has been a difficult challenge in both the energy calculations by Stouthamer (1973) and Lynch and Marinov (2015), as well as in other recent whole-cell models (Karr et al., 2012; Macklin et al., 2020). Because each transport reaction is simulated independently, we know their exact ATP costs by recording the fluxes through each reaction. Unlike most organisms, which have synthesis pathways for most of its building blocks, Syn3A has been reduced to the point where it relies on having to transport them in. From crystal structures of related transporters, it is clear that the majority of them contain ATPase domains that require, on average, at least one ATP for every nutrient molecule imported (Santos et al., 2018). It is assumed here that one ATP is used for every molecule taken up. The cost of active transport



**Figure 5. A detailed accounting of cellular ATP costs**

Time-dependent ATP costs over the course of a cell cycle show the distribution of ATP costs among individual processes and a balance of ATP generation and usage. ATP is generated only in the central metabolism through the PGK, PYK, and ACKr reactions in Figure 2. ATP synthase uses ATP but is reversible in the kinetic model and can switch direction as seen in its drop in the single replication cell. Translation elongation is not included because it uses GTP rather than ATP (Lynch and Marinov, 2015).

is approximately twice the ATP cost of tRNA charging (roughly 5,000:2,500 ATP per second), making it one of the largest energetic costs in the minimal cell. This is the most exact calculation of the costs of transport presented to date and reveals how transport reactions are critical for a minimal cell, which relies almost entirely on nutrients from its environment.

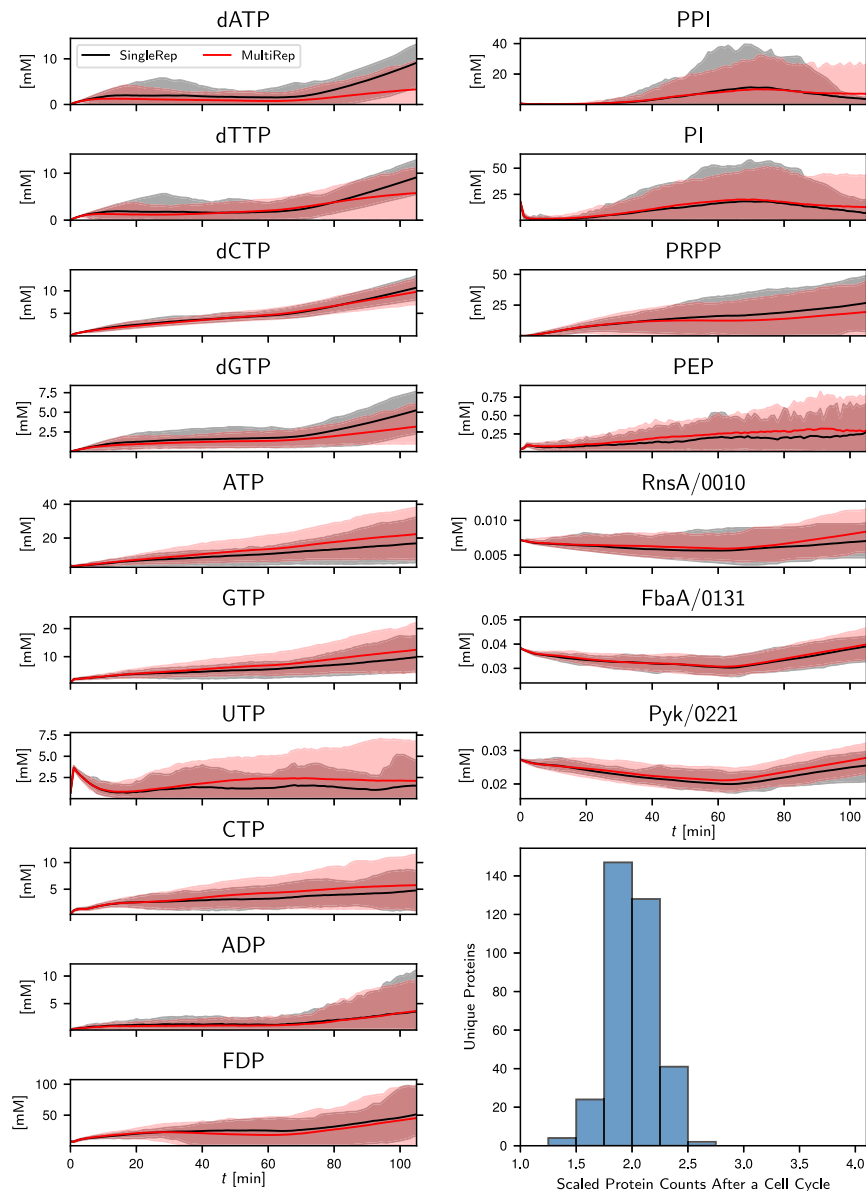
In Mycoplasmas, the ATP synthase typically breaks down ATP and pumps out protons to maintain a basic intracellular environment (Benyoucef et al., 1981a, 1981b). The reversible kinetics for ATP synthase depend on the ATP, ADP, and phosphate levels inside the cell. While ATP synthase accounts for roughly 10% of the overall ATP costs for the majority of the time, in response to low ATP levels in the cell, it can momentarily change direction.

Below the cost of ATP synthase come in descending order tRNA charging, transcription elongation, mRNA degradation, RNA material cost, DNA replication elongation, and membrane protein insertion via translocation by SecY. DNA replication only takes place during part of the cell cycle. Its ATP costs typically occur early in the cell cycle (Figure 5B, left) and again late in the cell cycle (Figure 5B, right) when initiation of another replication event occurs. The fluctuations reflect changes in the rate of DNA replication.

### Time-dependent concentrations show consistent average behavior and large population variability

Homeostasis is a property of a normal cell to maintain constant intracellular concentrations over a cell cycle suggesting that

upon experiencing a perturbation whether from the environment or an intracellular reaction, it responds by adjusting its biochemical pathways to bring the concentrations back into an acceptable range for stable functioning of its networks (Agozzino et al., 2020). The prevailing wisdom is that some degree of regulation is required to control any large fluctuations. The time traces for a representative selection of metabolites and macromolecules from the well-stirred simulations are shown in Figure 6. The complete time traces of all chemical species and reaction fluxes are provided in Data S1. The dNTP concentrations decrease late in the cell cycle for cells where multiple replication events have occurred because they are being incorporated into new chromosomes. For dATP and dTTP, it appears that the same cells maintain average concentrations that are relatively constant over the whole cell cycle. The concentrations of dGTP and dCTP both continually increase over the cell cycle, which calls for investigating the possibility of any regulation on the uptake of their precursor deoxynucleosides or the thioredoxin reactions converting GDP and CDP to dGDP and dCDP in nucleotide metabolism. The concentrations of UTP and CTP are fairly constant throughout most of the cell cycle, likely because the only annotated way to make new CTP in Syn3A is by converting UTP to CTP through a CTP synthase reaction CTPS2 in nucleotide metabolism (Figure S1). ATP and GTP, on the other hand, continue to increase over the cell cycle in both cells with single and multiple replication events.



**Figure 6. Simulated traces of key cellular metabolite and enzyme concentrations**

Concentrations for 141 cells with multiple replication events (red) and 33 cells with single replication events (black) show a wide range of intracellular concentrations over the population: solid lines (average) and the shading (10<sup>th</sup> to 90<sup>th</sup> percentiles). A proteome-wide distribution of scaled protein counts shows that most proteins have their counts accurately doubled over the course of a cell cycle on average. The scaled protein counts represent the count of a protein at 105 min (experimental end of the cell cycle) divided by its initial count from the proteomics.

candidate to be included in a complexified model that includes regulation.

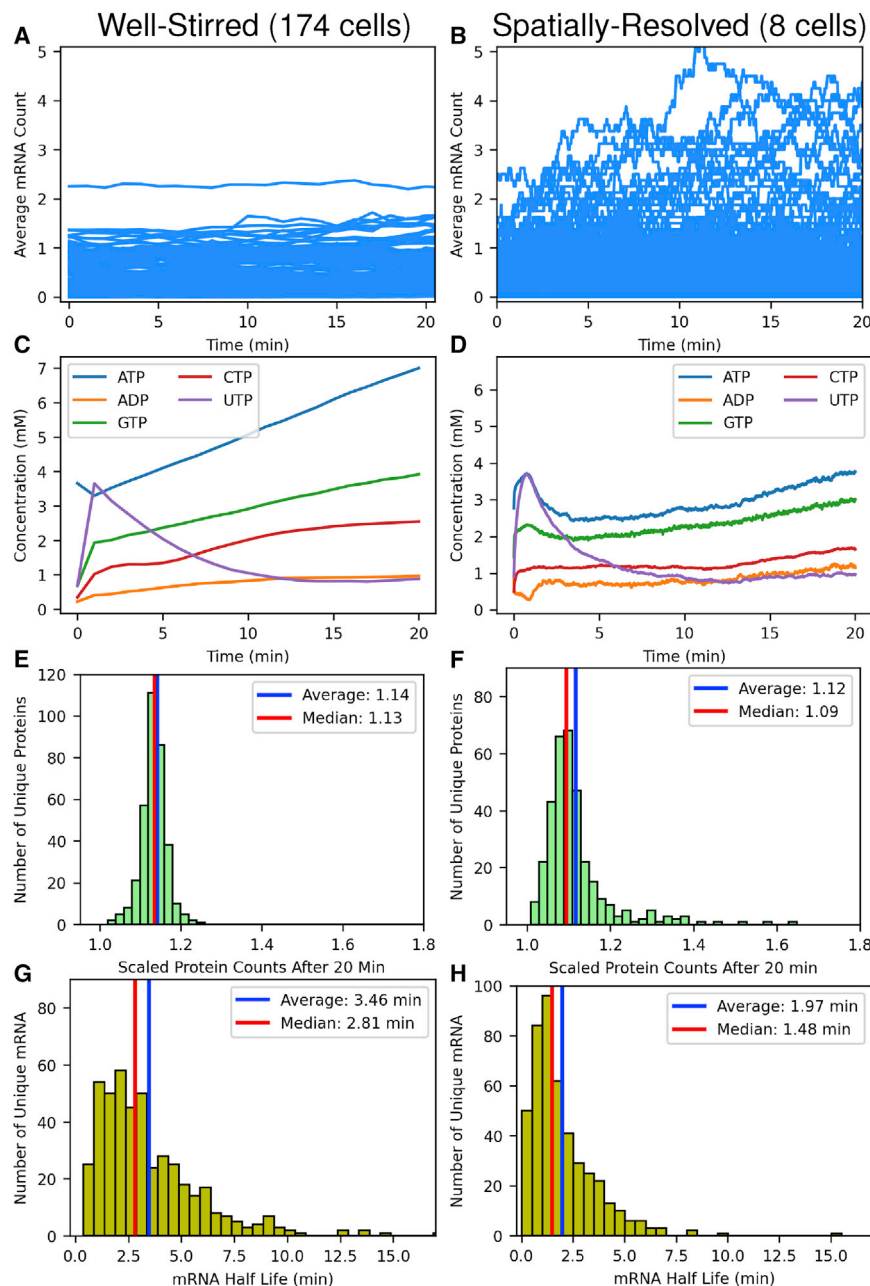
As discussed earlier, the FBA reaction limits the rate of glycolysis in our kinetics. As long as the enzyme performing the reaction, fructose-1,6-bisphosphate aldolase (FbaA), is at its average concentration or lower, FBA will be going slower than upper glycolysis, and its substrate, fdp, can build up by tens of millimolars. In cells that generate more FbaA enzymes or take up less glucose, there is less significant buildup or even no buildup resulting in the fdp pool being lower, even less than 10 mM.

Examples of three proteins, the nucleoside transporter ATP-binding protein (*rnsA/0010*), pyruvate kinase (*pyk/0221*), and fructose-1,6-bisphosphate aldolase (*fbaA/0131*), are provided in Figure 6. The concentrations of RnsA, FbaA, and Pyk all share similar behavior, being slightly diluted over the first part of the cell cycle and then increasing until the end of the cell cycle. Reactions for protein degradation are not included, so this is purely a volume effect where the number of proteins is not increasing to match the

Even though homeostasis can be observed for a population, there can be significant variation among individual cells due to stochastic fluctuations in gene expression. Some of the largest variations in our simulations were for phosphate (PI), pyrophosphate (PPI), and fructose bisphosphate (fdp). Cells with multiple replication events have a wide range of phosphate levels at the end of the cell cycle. In contrast, cells with only a single replication event return to consistently lower phosphate levels after the first replication event is complete. PPI is given off from DNA replication reactions, so cells with multiple replication events will see a broader range of PPI and therefore PI levels. While such high concentrations have been reported in yeast (Park et al., 2016), they may be inaccurate for Syn3A. Syn3A still has the gene *phoU/0428* coding for a phosphate regulator, so it is a primary

increased volume. The farther away a gene is from the origin, the more exaggerated this effect becomes because of the delay between the start of replication and when a gene gets doubled.

To gauge proteome-wide homeostasis, the scaled protein counts for all proteins reported in the proteomics data (Breuer et al., 2019) with counts of 10 or greater excluding ribosomal proteins, a total of 350/452 proteins, at the end of a cell cycle are shown in the histogram in Figure 6. Ribosomal proteins are excluded from this plot because their counts do not reflect the 503 ribosomes observed in cryo-ET, with many having counts fewer than 300. For a protein to maintain a near-constant concentration, its count must double over a cell cycle as the volume doubles. On average, the overwhelming majority of proteins end the cell cycle with 1.75 to 2.25 times their initial protein counts. Outliers include proteins whose genes are longer than 4,000 nt



**Figure 7. A comparison of gene expression between well-stirred and spatially resolved simulations**

(A and B) Counts of mRNA for genes coding for genetic information processing proteins.

(C and D) Intracellular concentrations (pools) of NTPs and ADP.

(E and F) Genome-wide scaled protein counts after the first 20 min of the cell cycle.

(G and H) Genome-wide mRNA half-life distributions. The 1.97-min mRNA half-life in the spatial model is in better agreement with a measured average half-life in *M. gallisepticum*.

the difference in transcription rates between the two models (see [STAR Methods](#), Transcription). The spatial model has lower concentrations of nucleotides ([Figures 7C and 7D](#)) than the well-stirred model, which can be tied to two factors: first, more nucleotides are being incorporated into mRNA, and second, the current spatial model does not include DNA replication, which initiates, on average, around 10 min in the well-stirred model ([Figure 3B](#)). With the genes for the nucleoside transporters being close to the origin, their genes would be duplicated early in the cell cycle. Consequently, the spatial model should have fewer nucleoside transporters, which would result in slightly lower uptake rates of nucleosides. Scaled protein counts are compared in [Figures 7E and 7F](#). The slightly higher counts produced in the well-stirred simulations reflect that partial replication events have taken place in a few of the cells within the first 20 min. Finally, mRNA half-lives are compared between the two methods ([Figures 7G and 7H](#)), where the well-stirred half-lives depend on the active degradosome statistics in the spatial model ([STAR Methods](#), mRNA Degradation). The well-stirred model has longer half-lives on average than the half-lives in the spatial

on the low end and *priB*/0026 on the high end, which has a transcript only 441 nt long.

### Agreement between hybrid well-stirred and spatial simulations

To gauge the quality of our parameterized well-stirred hybrid CME-ODE model to reproduce the results of the 3D spatially resolved hybrid RDME-gCME-ODE model with diffusion (see [Video S1](#)), we compare counts of mRNAs for genes involved in genetic information processes, nucleotide pools, protein distributions, and mRNA half-life distributions ([Figure 7](#)). The mRNA counts in the spatial model are higher on average likely due to

model, with average half-lives of 3.4 min and 1.97 min, respectively.

### DISCUSSION

We report here the results from fully dynamical kinetic models, both for the well-stirred homogeneous (CME-ODE) and 3D (RDME-CME-ODE) spatially resolved scenarios, for a living minimal bacterial cell. We provide the time-dependent information about the dynamic rates of genetic information processes, the 148 known metabolites, 452 proteins and mRNAs, 29 tRNAs, 503 ribosomes, and DNA undergoing over 7,000 reactions.



With its reduced genome of 543 kbp and 493 genes, the minimal cell JCVI-syn3A has retained only a few genes for regulatory proteins and functional small RNAs. In our present kinetic models, regulation can only occur through gene expression and the rate forms for the various genetic information processes and metabolic reactions. We have not included explicitly known regulatory proteins like PhoU (*phoU/0428*) and the riboswitches TPP and SAM, as the kinetic parameters and time-scales of conformational changes in the riboswitches are still being investigated (Scull et al., 2021). The simulations based on the hybrid well-stirred (CME-ODE) whole-cell kinetic model have already given us quantitative insight into how the cell balances the demands of metabolism, genetic information, and growth over a cell cycle. From the emergent behaviors arising from the well-stirred and spatially resolved stochastic-deterministic simulations presented here, we can begin to understand the principles of life for this minimal cell when little regulation is present.

By emergent, we specifically mean behaviors defining the state of the cell (time-dependent concentrations, patterns, reaction rates, and correlations) that arise from simulations of the kinetic models and are not imposed. Such a behavior is the relationship among stochastic gene expression, cell growth, and progression of the cell cycle (Figure 3). Formation of a complete DnaA filament along the single-stranded DNA near the origin determines the timing of initiation of DNA replication. Cell growth as measured by increasing surface area is controlled in our model solely from lipid metabolism and translation of the mRNAs for lipid enzymes and membrane proteins. In an earlier work (Peterson et al., 2015), we showed experimentally, theoretically, and computationally the effects of DNA replication or gene copy number on the variance in mRNA distributions and ultimately the protein distributions, and this prior study guided the development of the kinetic model. As most growth studies are carried out on a population of cells, the results in this figure suggest a range of doubling times to be expected at a single-cell level. Importantly, given the dependence of the DnaA binding rate to its abundance and inverse dependence to the cell volume, the expected number of replication initiation events should exhibit a distribution as it does here. Lipid metabolism based on the lipidomics study in Figure S2 leads to cell volume/surface growth that doubles the volume shortly after the first replication event in approximately 65 min and doubles the surface area in a range from 88 to 112 min.

Based on average time-dependent fluxes emerging from the simulations, the cells react to depletion of either the nucleotide or amino acid pools by slowing down replication, transcription, or translation (Figures 4 and S4A–S4F). While, on average, the cells are able to maintain these pools, the change in rates allows them to recover from the imbalance. Syn3A monitors its internal state through the metabolic network and modifies its rate of replication, transcription, and translation in response to the concentrations of dNTPs, NTPs, and aa:tRNAs, respectively. The rate of glycolysis in our kinetics is limited by the fructose-1,6-bisphosphate aldolase (*fbaA/0131*) and the PYK reactions (*pyk/0221*), which require pep to convert any (d)NDP to (d)NTP. With the removal of glucokinase in the genome reduction, pep is essential for the transport of glucose into the cell and its phosphorylation to g6p. An imbalance between upper and lower

glycolysis and nucleotide metabolism does occur in a small fraction of the cells, which stop growing due to a depletion of pep.

The kinetic model allows the cell-wide uses of ATP to be monitored (Figure 5), which revealed that the costs of active transport of ions, amino acids, and nucleosides in Syn3A are comparable to other significant costs such as translation, as first suggested by Stouthamer (1973). This increased cost reflects the simplicity of this organism and its dependence on communication with its environment. Because Syn3A lacks oxidative phosphorylation, the cell depends almost entirely on glycolysis for formation of ATP. This results in significant sensitivity to the levels of glycolytic enzymes and therefore the stochastic genetic information processing reactions that express the enzymes.

In general, average metabolite concentrations generated from the model (Figure 6) show reasonable agreement to the values reported by Park et al. (2016) in *E. coli*, but the cell-to-cell variation in the kinetic models over a cell cycle is broader than the predicted range. This discrepancy certainly calls for regulation in some cases such as the uptake of (deoxy)nucleosides, uptake of inorganic phosphates, and formation of some metabolites that build up to large concentrations in some cells such as prpp and fdp.

In summary, the emergent behaviors are validated by several experimental results. The simulations accurately double the protein counts from experimental genome-wide proteomics. While our results are not in perfect agreement with qPCR origin-to-terminus ratios, our kinetics also reflect that many Syn3A cells experience multiple replication events per cell cycle. Our model predicts an average surface area doubling time of roughly 98 min, which is close to the experimentally measured doubling time (Breuer et al., 2019) in rich medium of 105 min. Also emergent are fractions of active complexes and the distribution of mRNA half-lives from the spatial model. The fraction of active RNAP is similar to the fraction reported by Bremer and Dennis (2008). The average of the distribution of mRNA half-lives is similar to the average reported for *M. gallisepticum* (Kirk and Morowitz, 1969), and the distribution having a long tail out to 15 min has also been observed in *B. subtilis* (Hambraeus et al., 2003). At the moment, no experiments have been done to measure the number of active degradosomes in Syn3A and is a prediction to be validated in future experiments. While the predicted distribution of half-lives is in agreement with genome-wide studies carried out on related organisms, it awaits confirmation from ongoing transcriptomics studies. The hybrid CME-ODE whole-cell kinetic model requires roughly 4 h to calculate the behavior of a cell over a maximum cell cycle of 120 min. The hybrid model is straightforward to complexify by adding additional reactions before proceeding to the full spatial hybrid RDME-CME-ODE model, which is computationally more demanding. For example, the reintroduction of the glucokinase into the kinetic model restored those unhealthy cells, which halted glycolysis due to the shortage of pep. The genetic reintroduction still needs to be carried out to validate this prediction.

### Limitations of the study

Current simulations predict a lower ratio of origins to termini formed in DNA replication than observed in the qPCR experiments, which comes from restricting the formation of replication

forks on the daughter chromosomes to occur only after complete replication of the mother chromosome and from starting each simulation with a single circular chromosome. In the future, we will consider multiple cell cycles starting from daughter cells each possibly containing a chromosome with multiple replication forks. Extending the simulations over several cell cycles would allow us to obtain statistics about cell divisions and multiple initiations of DNA replication events and DnaA filament formation.

The RDME-CME-ODE simulations are currently limited by having the degradosome, RNAP, and ribosome complexes all initialized in inactive states. Starting from an inactive state results in the initial transience in Figures 1M–1O, where the first few minutes of simulation are dominated by the complexes reaching steady-state processing of mRNAs, which may be overshadowing other interesting phenomena. In the absence of experimental knowledge of the average active complexes in Syn3A, the initial transience emphasizes importance of diffusion and how the ensemble-averaged results of the spatial model could be used to parameterize probabilistic factors in the well-stirred CME-ODE kinetic model, which account for diffusion. Future simulations with averaged occupation states will address this limitation. Besides the lack of explicit regulatory factors discussed above, the reactions to modify nucleobases in DNA and rRNA have been neglected and the time-dependent assembly mechanisms of protein complexes and ribosomes have been absorbed into the overall rates. Regulation is important and will be included in future models. No kinetic model for formation of FtsZ/FtsA filament and septum formation prior to formation of the daughter cells is considered. In the future, these processes will be addressed as we modify the spatially resolved kinetic model to allow a change in cell morphology, DNA replication, and ribosome biogenesis similar to the rule-based model we created for DNA replication in a slow-growing and dividing *E. coli* cell (Earnest et al., 2016).

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Bacterial strains
  - Defined medium
- **METHOD DETAILS**
  - Background for stochastic cell simulations
  - LM interface for creating spatial simulations: jLM
  - Overview of deterministic and stochastic simulation methods
  - Construction of JCVI-syn3A cell geometry and DNA configurations
  - Initialization of proteins, mRNAs, and tRNAs
  - Genetic information processing
  - Metabolism

- Lipidomics
- Quantitative PCR protocol
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Analysis of cell simulations
  - Lipidomics analysis and post-processing

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2021.12.025>.

## ACKNOWLEDGMENTS

We thank Julio Maia for significant assistance in modifying jLM and John Stone for making it possible to visualize the RDME simulations and the cryo-electron tomogram reconstructions in VMD. All authors received partial support from NSF MCB 1818344 and 1840320. Z.R.T., D.M.B., T.A.B., B.R.G., and Z.L.-S. were supported by the Center for the Physics of Living Cells (NSF PHY 1430124) and NSF Physics of Living Systems (PHY 1505008). N.S. and J.P.S. were supported by German Federal Ministry of Education and Research BMBF grant (to J.P.S., project 03Z22EN12) and VW Foundation “Life” grant (to J.P.S., project 93090). Z.L.-S. is partially supported by the NIH Center for Macromolecular Modeling and Bioinformatics (P41-GM104601).

## AUTHOR CONTRIBUTIONS

Z.R.T., D.M.B., T.A.B., B.R.G., and Z.L.-S.: creation of models and writing – original draft. Z.R.T., D.M.B., T.M.E., and M.C.R.M.: programming and software development. Z.L.-S.: supervision and funding acquisition. A.T.C. and N.S.: data/evidence collection. K.S.W.: provision of materials. J.P.S., C.A.H., H.O.S., and J.I.G.: supervision and writing – review and editing.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 9, 2021

Revised: November 1, 2021

Accepted: December 17, 2021

Published: January 20, 2022

## REFERENCES

- Agazzino, L., Balázs, G., Wang, J., and Dill, K.A. (2020). How do cells adapt? stories told in landscapes. *Annu. Rev. Chem. Biomol. Eng.* **11**, 155–182.
- Alberty, R.A. (2003). *Thermodynamics of Biochemical Reactions* (John Wiley & Sons, Inc.).
- Benyoucef, M., Rigaud, J.-L., and Leblanc, G. (1981a). The electrochemical proton gradient in *Mycoplasma* cells. *Eur. J. Biochem.* **113**, 491–498.
- Benyoucef, M., Rigaud, J.-L., and Leblanc, G. (1981b). Gradation of the magnitude of the electrochemical proton gradient in *Mycoplasma* cells. *Eur. J. Biochem.* **113**, 499–506.
- Bianchi, D.M., Peterson, J.R., Earnest, T.M., Hallock, M.J., and Luthey-Schulten, Z. (2018). Hybrid CME-ODE method for efficient simulation of the galactose switch in yeast. *IET Syst. Biol.* **12**, 170–176.
- Blötz, C., and Stülke, J. (2017). Glycerol metabolism and its implication in virulence in *Mycoplasma*. *FEMS Microbiol. Rev.* **41**, 640–652.
- Boscá, L., and Corredor, C. (1984). Is phosphofructokinase the rate-limiting step of glycolysis? *Trends Biochem. Sci.* **9**, 372–373.
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell* **136**, 261–271.

- Bremer, H., and Dennis, P.P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *Ecosol Plus* 3. <https://doi.org/10.1128/ecosal.5.2.3>.
- Breuer, M., Earnest, T.M., Merryman, C., Wise, K.S., Sun, L., Lynott, M.R., Hutchison, C.A., Smith, H.O., Lapek, J.D., Gonzalez, D.J., et al. (2019). Essential metabolism for a minimal cell. *eLife* 8, e36842.
- Castellanos, M., Kushiro, K., Lai, S.K., and Shuler, M.L. (2007). A genomically/chemically complete module for synthesis of lipid membrane in a minimal cell. *Biotechnol. Bioeng.* 97, 397–409.
- Catipovic, M.A., and Rapoport, T.A. (2020). Protease protection assays show polypeptide movement into the SecY channel by power strokes of the SecA ATPase. *EMBO Rep.* 21, e50905.
- Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblitz, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., and Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.* 49 (D1), D498–D508.
- Chen, Y., Bauer, B.W., Rapoport, T.A., and Gumbart, J.C. (2015). Conformational changes of the clamp of the protein translocation ATPase SecA. *J. Mol. Biol.* 427, 2348–2359.
- Cheng, H.-M., Gröger, P., Hartmann, A., and Schlierf, M. (2015). Bacterial initiators form dynamic filaments on single-stranded DNA monomer by monomer. *Nucleic Acids Res.* 43, 396–405.
- Cho, K.H. (2017). The structure and function of the gram-positive bacterial RNA degradosome. *Front. Microbiol.* 8, 154.
- Clejan, S., and Bittman, R. (1984). Decreases in rates of lipid exchange between *Mycoplasma gallisepticum* cells and unilamellar vesicles by incorporation of sphingomyelin. *J. Biol. Chem.* 259, 10823–10826.
- Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J.L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423.
- Dai, X., Zhu, M., Warren, M., Balakrishnan, R., Patsalo, V., Okano, H., Williamson, J.R., Fredrick, K., Wang, Y.-P., and Hwa, T. (2016). Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nat. Microbiol.* 2, 16231.
- Daubenspeck, J.M., Jordan, D.S., Simmons, W., Renfrow, M.B., and Dybvig, K. (2015). General n- and o-linked glycosylation of lipoproteins in mycoplasmas and role of exogenous oligosaccharide. *PLoS ONE* 10, e0143362.
- Dillon, D.A., Wu, W.-I., Riedel, B., Wissing, J.B., Dowhan, W., and Carman, G.M. (1996). The *Escherichia coli* *pgpB* gene encodes for a diacylglycerol pyrophosphate phosphatase activity. *J. Biol. Chem.* 271, 30548–30553.
- Duderstadt, K.E., Chuang, K., and Berger, J.M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478, 209–213.
- Earnest, T.M., Lai, J., Chen, K., Hallock, M.J., Williamson, J.R., and Luthey-Schulten, Z. (2015). Toward a whole-cell model of ribosome biogenesis: Kinetic modeling of SSU assembly. *Biophys. J.* 109, 1117–1135.
- Earnest, T.M., Cole, J.A., Peterson, J.R., Hallock, M.J., Kuhlman, T.E., and Luthey-Schulten, Z. (2016). Ribosome biogenesis in replicating cells: Integration of experiment and theory. *Biopolymers* 105, 735–751.
- Earnest, T.M., Cole, J.A., and Luthey-Schulten, Z. (2018). Simulating biological processes: stochastic physics from whole cells to colonies. *Rep. Prog. Phys.* 81, 052601.
- Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: COntstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7, 74.
- Ejsing, C.S., Sampaio, J.L., Surendranath, V., Duchoslav, E., Ekroos, K., Klemm, R.W., Simons, K., and Shevchenko, A. (2009). Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl. Acad. Sci. USA* 106, 2136–2141.
- Erzberger, J.P., Mott, M.L., and Berger, J.M. (2006). Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* 13, 676–683.
- Facchetti, G., Chang, F., and Howard, M. (2017). Controlling cell size through sizer mechanisms. *Curr. Opin. Syst. Biol.* 5, 86–92.
- Fazal, F.M., Koslover, D.J., Luisi, B.F., and Block, S.M. (2015). Direct observation of processive exoribonuclease motion using optical tweezers. *Proc. Natl. Acad. Sci. USA* 112, 15101–15106.
- Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012). eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res.* 40 (Database issue, D1), D770–D775.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. USA* 110, 10039–10044.
- Garcia-Gonzalo, F.R., and Izpisua Belmonte, J.C. (2008). Albumin-associated lipids regulate human embryonic stem cell self-renewal. *PLoS ONE* 3, e1384.
- Gaspari, E., Malachowski, A., Garcia-Morales, L., Burgos, R., Serrano, L., dos Santos, V.A.P.M., and Suarez-Diez, M. (2020). Model-driven design allows growth of *Mycoplasma pneumoniae* on serum-free media. *npj Systems Biology and Applications* 6, 33.
- Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56.
- Gilbert, B.R., Thornburg, Z.R., Lam, V., Rashid, F.M., Glass, J.I., Villa, E., Dame, R.T., and Luthey-Schulten, Z. (2021). Generating chromosome geometries in a minimal cell from cryo-electron tomograms and chromosome conformation capture maps. *Front. Mol. Biosci.* 8, 644133.
- Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361.
- Goldberg, R.N., Tewari, Y.B., and Bhat, T.N. (2004). Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics* 20, 2874–2877.
- Goldberg, A.P., Szigeti, B., Chew, Y.H., Sekar, J.A., Roth, Y.D., and Karr, J.R. (2018). Emerging whole-cell modeling principles and methods. *Curr. Opin. Biotechnol.* 51, 97–102.
- Golding, I., and Cox, E.C. (2004). RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. USA* 101, 11310–11315.
- Haas, D., Thamm, A.M., Sun, J., Huang, L., Sun, L., Beaudoin, G.A.W., Wise, K.S., Lerma-Ortiz, C., Bruner, S.D., Breuer, M., et al. (2021). Metabolite Damage and Damage-Control in a Minimal Genome. *bioRxiv*. <https://doi.org/10.1101/2021.12.01.470718>.
- Hallock, M.J., Stone, J.E., Roberts, E., Fry, C., and Luthey-Schulten, Z. (2014). Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations. *Parallel Comput.* 40, 86–99.
- Hambraeus, G., von Wachenfeldt, C., and Hederstedt, L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Genet. Genomics* 269, 706–714.
- Helmstetter, C.E., and Cooper, S. (1968). DNA synthesis during the division cycle of rapidly growing *Escherichia coli* B/r. *J. Mol. Biol.* 31, 507–518.
- Herzog, R., Schwudke, D., Schuhmann, K., Sampaio, J.L., Bornstein, S.R., Schroeder, M., and Shevchenko, A. (2011). A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome Biol.* 12, R8.
- Herzog, R., Schuhmann, K., Schwudke, D., Sampaio, J.L., Bornstein, S.R., Schroeder, M., and Shevchenko, A. (2012). LipidXplorer: a software for consensual cross-platform lipidomics. *PLoS ONE* 7, e29851.
- Hindmarsh, A. (1983). Odepack, a systematized collection of ode solvers. *Scientific Computing* 1, 55–64.
- Hofmeyr, J.-H.S., Gqwaka, O.P., and Rohwer, J.M. (2013). A generic rate equation for catalysed, template-directed polymerisation. *FEBS Lett.* 587, 2868–2875.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38, 27–28.

- Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science* 351. <https://doi.org/10.1126/science.aad6253>.
- Ibba, M., Francklyn, C., and Cusack, S. (2005). The Aminoacyl-tRNA Synthetases (Bantam).
- Iwami, Y., and Yamada, T. (1980). Rate-limiting steps of the glycolytic pathway in the oral bacteria *Streptococcus mutans* and *Streptococcus sanguis* and the influence of acidic pH on the glucose metabolism. *Arch. Oral Biol.* 25, 163–169.
- Jeckelmann, J.-M., and Erni, B. (2019). Carbohydrate transport by group translocation: the bacterial phosphoenolpyruvate: sugar phosphotransferase system. In *Bacterial cell walls and membranes*, A. Kuhn, ed. (Springer), pp. 223–274.
- Jo, S., Lim, J.B., Klauda, J.B., and Im, W. (2009). CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* 97, 50–58.
- Johnson, G.E., Lalanne, J.-B., Peters, M.L., and Li, G.-W. (2020). Functionally uncoupled transcription-translation in *Bacillus subtilis*. *Nature* 585, 124–128.
- Jordan, D.S., Daubenspeck, J.M., Laube, A.H., Renfrow, M.B., and Dybvig, K. (2013). O-linked protein glycosylation in *Mycoplasma*. *Mol. Microbiol.* 90, 1046–1053.
- Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401.
- King, Z.A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., and Palsson, B.O. (2015). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* 11, e1004321.
- Kirk, R.G., and Morowitz, H.J. (1969). Ribonucleic acids of *Mycoplasma gallisepticum* strain A5969. *Am. J. Vet. Res.* 30, 287–293.
- Kitamura, S., Shimizu, H., and Toya, Y. (2021). Identification of a rate-limiting step in a metabolic pathway using the kinetic model and in vitro experiment. *J. Biosci. Bioeng.* 131, 271–276.
- Kornspan, J.D., and Rottem, S. (2012). The phospholipid profile of mycoplasmas. *J. Lipids* 2012, 640762.
- Lachance, J.-C., Matteau, D., Brodeur, J., Lloyd, C.J., Mih, N., King, Z.A., Knight, T.F., Feist, A.M., Monk, J.M., Palsson, B.O., et al. (2021). Genome-scale metabolic modeling reveals key features of a minimal gene set. *Mol. Syst. Biol.* 17, e10099.
- Liebermeister, W., and Klipp, E. (2006). Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.* 3, 41.
- Liebermeister, W., Uhlenhof, J., and Klipp, E. (2010). Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* 26, 1528–1534.
- Liebsch, G., Binder, M., Schifferer, R., Langmann, T., Schulz, B., and Schmitz, G. (2006). High throughput quantification of cholesterol and cholesteryl ester by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *Biochim. Biophys. Acta* 1761, 121–128.
- Lönnfors, M., Doux, J.P., Killian, J.A., Nyholm, T.K., and Slotte, J.P. (2011). Sterols have higher affinity for sphingomyelin than for phosphatidylcholine bilayers even at equal acyl-chain order. *Biophys. J.* 100, 2633–2641.
- Lubitz, T., and Liebermeister, W. (2019). Parameter balancing: consistent parameter sets for kinetic metabolic models. *Bioinformatics* 35, 3857–3858.
- Lubitz, T., Schulz, M., Klipp, E., and Liebermeister, W. (2010). Parameter balancing in kinetic models of cell metabolism. *J. Phys. Chem. B* 114, 16298–16303.
- Luthey-Schulten, Z. (2021). Integrating experiments, theory and simulations into whole-cell models. *Nat. Methods* 18, 446–447.
- Lynch, M., and Marinov, G.K. (2015). The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. USA* 112, 15690–15695.
- Mackie, G.A. (2013). RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* 11, 45–57.
- Macklin, D.N., Ahn-Horst, T.A., Choi, H., Ruggero, N.A., Carrera, J., Mason, J.C., Sun, G., Agmon, E., DeFelice, M.M., Maayan, I., et al. (2020). Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation. *Science* 369. <https://doi.org/10.1126/science.aav3751>.
- Marucci, L., Barberis, M., Karr, J., Ray, O., Race, P.R., de Souza Andrade, M., Grierson, C., Hoffmann, S.A., Landon, S., Rech, E., et al. (2020). Computer-aided whole-cell design: taking a holistic approach by integrating synthetic with systems biology. *Front. Bioeng. Biotechnol.* 8, 942.
- Matteau, D., Lachance, J.-C., Grenier, F., Gauthier, S., Daubenspeck, J.M., Dybvig, K., Garneau, D., Knight, T.F., Jacques, P.-É., and Rodrigue, S. (2020). Integrative characterization of the near-minimal bacterium *Mesoplasma florum*. *Mol. Syst. Biol.* 16, e9844.
- McElwain, M.C., and Pollack, J.D. (1987). Synthesis of deoxyribonucleotides in Mollicutes: dependence on deoxyribose-1-phosphate and PPI. *J. Bacteriol.* 169, 3647–3653.
- Meadow, N.D., Mattoo, R.L., Savtchenko, R.S., and Roseman, S. (2005a). Transient state kinetics of Enzyme I of the phosphoenolpyruvate:glucose phosphotransferase system of *Escherichia coli*: equilibrium and second-order rate constants for the phosphotransfer reactions with phosphoenolpyruvate and HPr. *Biochemistry* 44, 12790–12796.
- Meadow, N.D., Savtchenko, R.S., Nezami, A., and Roseman, S. (2005b). Transient state kinetics of enzyme IICBGl, a glucose transporter of the phosphoenolpyruvate phosphotransferase system of *Escherichia coli*: equilibrium and second order rate constants for the glucose binding and phosphotransfer reactions. *J. Biol. Chem.* 280, 41872–41880.
- Morowitz, H.J. (1984). The completeness of molecular biology. *Isr. J. Med. Sci.* 20, 750–753.
- Neale, G.A., Mitchell, A., and Finch, L.R. (1983). Enzymes of pyrimidine deoxyribonucleotide metabolism in *Mycoplasma mycoides subsp. mycoides*. *J. Bacteriol.* 156, 1001–1005.
- Neidhardt, F.C., and Umbarger, H.E. (1996). Chemical Composition of *Escherichia coli*. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 1, F.C. Neidhardt, ed. (ASM Press), pp. 2–4.
- Nielsen, H.J., Youngren, B., Hansen, F.G., and Austin, S. (2007). Dynamics of *Escherichia coli* chromosome segregation during multifork replication. *J. Bacteriol.* 189, 8660–8666.
- Noor, E., Flamholz, A., Bar-Even, A., Davidi, D., Milo, R., and Liebermeister, W. (2016). The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLoS Comput. Biol.* 12, e1005167.
- Pappas, A., Park, T.-S., and Carman, G.M. (1999). Characterization of a novel dUTP-dependent activity of CTP synthetase from *Saccharomyces cerevisiae*. *Biochemistry* 38, 16671–16677.
- Park, E., Ménétret, J.-F., Gumbart, J.C., Ludtke, S.J., Li, W., Whynot, A., Rapoport, T.A., and Akey, C.W. (2014). Structure of the SecY channel during initiation of protein translocation. *Nature* 506, 102–106.
- Park, J.O., Rubin, S.A., Xu, Y.-F., Amador-Noguez, D., Fan, J., Shlomi, T., and Rabinowitz, J.D. (2016). Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* 12, 482–489.
- Pelletier, J.F., Sun, L., Wise, K.S., Assad-Garcia, N., Karas, B.J., Deerinck, T.J., Ellisman, M.H., Mershin, A., Gershenfeld, N., Chuang, R.-Y., et al. (2021). Genetic requirements for cell division in a genomically minimal cell. *Cell* 184, 2430–2440.e16.
- Peterson, J.R., Cole, J.A., Fei, J., Ha, T., and Luthey-Schulten, Z.A. (2015). Effects of DNA replication on mRNA noise. *Proc. Natl. Acad. Sci. USA* 112, 15886–15891.
- Petzold, L. (1983). Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM Journal on Scientific and Statistical Computing* 4, 136–148.
- Plackett, P. (1967). The glycerolipids of *Mycoplasma mycoides*. *Biochemistry* 6, 2746–2754.



- Pollack, J.D., Myers, M.A., Dandekar, T., and Herrmann, R. (2002). Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: the replacement of the missing “household” nucleoside diphosphate kinase gene and activity by glycolytic kinases. *OMICS* 6, 247–258.
- Quentin, Y., Fichant, G., and Denizot, F. (1999). Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. *J. Mol. Biol.* 287, 467–484.
- Razin, S., and Tully, J.G. (1970). Cholesterol requirement of mycoplasmas. *J. Bacteriol.* 102, 306–310.
- Roberts, E., Stone, J.E., and Luthey-Schulten, Z. (2013). Lattice Microbes: high-performance stochastic simulation method for the reaction-diffusion master equation. *J. Comput. Chem.* 34, 245–255.
- Rodwell, A.W. (1983). Defined and partly defined media. In *Methods in mycoplasmaology*, S. Razin and J.G. Tully, eds. (Academic Press), pp. 163–172.
- Rohwer, J.M., Meadow, N.D., Roseman, S., Westerhoff, H.V., and Postma, P.W. (2000). Understanding glucose transport by the bacterial phosphoenolpyruvate:glucose phosphotransferase system on the basis of kinetic measurements in vitro. *J. Biol. Chem.* 275, 34909–34921.
- Ryals, J., Little, R., and Bremer, H. (1982). Temperature dependence of RNA synthesis parameters in *Escherichia coli*. *J. Bacteriol.* 151, 879–887.
- Sáenz, J.P., Sezgin, E., Schwillie, P., and Simons, K. (2012). Functional convergence of hopanoids and sterols in membrane ordering. *Proc. Natl. Acad. Sci. USA* 109, 14236–14240.
- Sampaio, J.L., Gerl, M.J., Klose, C., Ejsing, C.S., Beug, H., Simons, K., and Shevchenko, A. (2011). Membrane lipidome of an epithelial cell line. *Proc. Natl. Acad. Sci. USA* 108, 1903–1907.
- Santos, J.A., Rempel, S., Mous, S.T., Pereira, C.T., Ter Beek, J., de Gier, J.-W., Guskov, A., and Slotboom, D.J. (2018). Functional and structural characterization of an ECF-type ABC transporter for vitamin B12. *eLife* 7. <https://doi.org/10.7554/eLife.35828>.
- Schieck, E., Lartigue, C., Frey, J., Voza, N., Heggermann, J., Miller, R.A., Valguarnera, E., Muriuki, C., Meens, J., Nene, V., et al. (2016). Galactofuranose in *Mycoplasma mycoides* is important for membrane integrity and conceals adhesins but does not contribute to serum resistance. *Mol. Microbiol.* 99, 55–70.
- Scull, C.E., Dandpat, S.S., Romero, R.A., and Walter, N.G. (2021). Transcriptional riboswitches integrate timescales for bacterial gene expression control. *Front. Mol. Biosci.* 7, 607158.
- Serdiuk, T., Steudle, A., Mari, S.A., Manioglou, S., Kaback, H.R., Kuhn, A., and Müller, D.J. (2019). Insertion and folding pathways of single membrane proteins guided by translocases and insertases. *Science advances* 5. <https://doi.org/10.1126/sciadv.aau6824>.
- Stouthamer, A.H. (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek* 39, 545–565.
- Surma, M.A., Herzog, R., Vasilj, A., Klose, C., Christinat, N., Morin-Rivron, D., Simons, K., Masoodi, M., and Sampaio, J.L. (2015). An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. *Eur. J. Lipid Sci. Technol.* 117, 1540–1549.
- Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538.
- Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* 16, 1146–1152.
- Thornburg, Z.R., Melo, M.C.R., Bianchi, D., Brier, T.A., Crotty, C., Breuer, M., Smith, H.O., Hutchison, C.A., 3rd, Glass, J.I., and Luthey-Schulten, Z. (2019). Kinetic modeling of the genetic information processes in a minimal cell. *Front. Mol. Biosci.* 6, 130.
- Torres, M., Balada, J.-M., Zellars, M., Squires, C., and Squires, C.L. (2004). *In vivo* effect of NusB and NusG on rRNA transcription antitermination. *J. Bacteriol.* 186, 1304–1310.
- Varma, A., and Palsson, B.O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60, 3724–3731.
- Wallden, M., Fange, D., Lundius, E.G., Baltekin, Ö., and Elf, J. (2016). The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* 166, 729–739.
- Wang, M., Herrmann, C.J., Simonovic, M., Szklarczyk, D., and von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* 15, 3163–3168.
- Webb, A.J., and Hosie, A.H.F. (2006). A member of the second carbohydrate uptake subfamily of ATP-binding cassette transporters is responsible for ribonucleoside uptake in *Streptococcus mutans*. *J. Bacteriol.* 188, 8005–8012.
- Werner, A. (2011). Predicting translational diffusion of evolutionary conserved RNA structures by the nucleotide number. *Nucleic Acids Res.* 39, e17.
- Willemoës, M., and Sigurskjöld, B.W. (2002). Steady-state kinetics of the glutamine reaction of CTP synthase from *Lactococcus lactis*. The role of the allosteric activator GTP incoupling between glutamine hydrolysis and CTP synthesis. *Eur. J. Biochem.* 269, 4772–4779.
- Williamson, D.L., and Whitcomb, R.F. (1975). Plant mycoplasmas: a cultivable spiroplasma causes corn stunt disease. *Science* 188, 1018–1020.
- Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., et al. (2012). SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res.* 40 (Database issue, D1), D790–D796.
- Yus, E., Maier, T., Michalodimitrakakis, K., van Noort, V., Yamada, T., Chen, W.-H., Wodke, J.A., Güell, M., Martínez, S., Bourgeois, R., et al. (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326, 1263–1268.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
JCVI-syn3A	<a href="#">Breuer et al., 2019</a>	GenBank: CP016816.2
JCVI-syn3B	<a href="#">Breuer et al., 2019</a>	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
C5-CMRL Defined Medium	This study	N/A (Ingredients below)
Alanine	Sigma-Aldrich	05129
Arginine . HCl	Sigma-Aldrich	A5131
Asparagine	Sigma-Aldrich	A4159
Cysteine .HCL.H2O	Sigma-Aldrich	C7880
Glutamate Na.H2O	Sigma-Aldrich	G1626
Glutamine	Sigma-Aldrich	G3126
Glycine	Sigma-Aldrich	G8895
Histidine	Sigma-Aldrich	H6034
Isoleucine	Sigma-Aldrich	I2752
Leucine	Sigma-Aldrich	L8000
Lysine HCL	Sigma-Aldrich	L5626
Methionine	Sigma-Aldrich	M9625
Phenylalanine	Sigma-Aldrich	P2126
Proline	Sigma-Aldrich	81709
Serine	Sigma-Aldrich	S4500
Threonine	Sigma-Aldrich	T8625
Tryptophan	Sigma-Aldrich	T8941
Tyrosine	Sigma-Aldrich	T3754
Valine	Sigma-Aldrich	V0500
Na2HPO4 (Sodium Phosphate dibasic)	Sigma-Aldrich	S0876
NaH2PO4 (Sodium Phosphate monobasic)	Sigma-Aldrich	S5011
KCl (Potassium Chloride)	Sigma-Aldrich	P3911
MgSO4 (Magnesium Sulfate)	Sigma-Aldrich	246972
glycerol	Sigma-Aldrich	G5516
Spermine tetraHCL	Sigma-Aldrich	S2876
Nicotinic acid	Sigma-Aldrich	72309
Thiamine HCL	Sigma-Aldrich	T1270
DL Lipoic (thioctic) acid ((+)-alpha-Lipoic acid)	Sigma-Aldrich	T1395
CoenzymeA, sodium salt hydrate	Sigma-Aldrich	C3144
Adenine hemisulfate salt	Sigma-Aldrich	A3159
Guanine HCL	Sigma-Aldrich	51030
Uracil	Sigma-Aldrich	U0750
Thymine	Sigma-Aldrich	T0376
Adenosine	Sigma-Aldrich	A4036
Guanosine	Sigma-Aldrich	G6264
Uridine	Sigma-Aldrich	U3750
Thymidine	Sigma-Aldrich	T9250
Cytidine	Sigma-Aldrich	C122106
KnockOut™ (serum replacement)	GIBCO	10828-028

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Riboflavin	Sigma-Aldrich	R9504
Glucose	Sigma-Aldrich	DX0145
choline chloride	Sigma-Aldrich	C7017
5-Formyl-5,6,7,8-tetrahydrofolic acid calcium salt (aka folinic acid calcium salt hydrate)	Sigma-Aldrich	F7878
dTMP (thymidylic acid); Thymidine 5'-monophosphate disodium salt hydrate	Sigma-Aldrich	T7004
Thiamine diphosphate (Thiamine pyrophosphate)	Sigma-Aldrich	C8754
pyridoxal 5'-phosphate.H <sub>2</sub> O	Sigma-Aldrich	P3657
cholesterol	Sigma-Aldrich	C8667
palmitic acid	Sigma-Aldrich	P0500
oleic acid	Sigma-Aldrich	O1008
Phenol red (0.5%w/v, sterile)	Sigma-Aldrich	P0290
Penicillin G sodium	Sigma-Aldrich	P3032
CMRL 1066 10x (-) phenol red, (-) L-Glutamine, (-) NaHCO <sub>3</sub>	GIBCO	ME19150L1 (ordered with phenol red removed)
SP4 Medium	<a href="#">Williamson and Whitcomb, 1975</a>	N/A (Ingredients below)
Mycoplasma Broth Base	BD Biosciences	211458
Bacto Tryptone	BD Biosciences	211705
Bacto Peptone	BD Biosciences	211677
Bacto Yeastolate	BD Biosciences	255772
Yeast Extract Solution	GIBCO	18180-059
Serum (heat inactivated, fetal bovine)	GIBCO	16140-071
KnockOut™ (serum replacement)	GIBCO	10828-028
<b>Critical commercial assays</b>		
Applied Biosystems PowerUp SYBR Green Master Mix	ThermoFisher	Cat# A25741
QIAquick Gel Extraction Kit	QIAGEN	Cat# 28704
<b>Deposited data</b>		
Aligned cryo-ET tilt-series	<a href="#">Gilbert et al., 2021</a>	EMPIAR – Accession Number: 10686
Cryo-ET reconstruction	<a href="#">Gilbert et al., 2021</a>	EMDB – EMD: 23661
JCVI-syn3A reference genome	<a href="#">Breuer et al., 2019</a>	GenBank: CP016816.2
Mass spectrometry data of Syn3A	<a href="#">Breuer et al., 2019</a>	MassIVE – Accession Number: 000081687
Proteomics of Syn3A	<a href="#">Breuer et al., 2019</a>	ProteomeXchange – Accession Number: PXD008159
qPCR of Syn3A origin:quarter:terminus ratios	This study	Mendeley Data: <a href="https://doi.org/10.17632/nprw2h5tx6.1">https://doi.org/10.17632/ nprw2h5tx6.1</a>
BRENDA	<a href="#">Chang et al., 2021</a>	<a href="https://www.brenda-enzymes.org/">https://www.brenda-enzymes.org/</a>
<b>Oligonucleotides</b>		
Forward qPCR primer for Syn3A origin of replication AATCGGTGCAAGTAATGAACAAG	Integrated DNA Technologies	N/A
Reverse qPCR primer for Syn3A origin of replication TCCCATTCAGATTACCATATAA	Integrated DNA Technologies	N/A
Forward qPCR primer for the quarter way point around the Syn3A genome GCTGACATAGGTGAAGGTCTAAC	Integrated DNA Technologies	N/A
Reverse qPCR primer for the quarter way point around the Syn3A genome GCCACTGGAGCAGGTATTT	Integrated DNA Technologies	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Forward qPCR primer for the terminus point around the Syn3A genome GCATTAGGCATTGTTGGCATAA	Integrated DNA Technologies	N/A
Reverse qPCR primer for the terminus point around the Syn3A genome ACATCGTTTCGTGCTGGATTGA	Integrated DNA Technologies	N/A
Forward qPCR primers for the qPCR control molecule AAAAATTTGTAATCGGTGC	Integrated DNA Technologies	N/A
Reverse qPCR primer for the qPCR control molecule CTGCAATTTTCCAGCCAC	Integrated DNA Technologies	N/A
319 bp qPCR control molecule AAAAATTTGTAATCGGTGCAAGT AATGAACAAGCTTTTATAGCAGT TCAAACAGTAAGTAAAAATCCTG GGATTTCTTATAATCCATTGTTTA TTTATGGTGAATCTGGAATGGGA AAATTAGCATTAGGCATTGTTGG CATAAATCCAGCACGAACGATGT TATTTTGCTGACATAGGTGAAGG TCTAACAGAAGGAACAGTCGCT GAAGTTTTAGTTAAAGTTGGTGA TGTTGTTAAAGAAGGACAATCAT TATACTTTGTTGAAACTGATAAA GTAAACAGTGAAATACCTGCTCC AGTGGCTGGAAAAATTGCAG	Integrated DNA Technologies	N/A
<b>Software and algorithms</b>		
Code to reconstruct Syn3A cells	Gilbert et al., 2021	<a href="https://github.com/brg4/SAP_chromosome">https://github.com/brg4/SAP_chromosome</a>
Code for genetic information processing	Thornburg et al., 2019	<a href="https://github.com/zanert2/Thornburg_FrontMolBiosci_2019">https://github.com/zanert2/Thornburg_FrontMolBiosci_2019</a>
Code for hybrid CME-ODE model	This paper	<a href="https://github.com/Luthey-Schulten-Lab/Minimal_Cell">https://github.com/Luthey-Schulten-Lab/Minimal_Cell</a> <a href="https://doi.org/10.5281/zenodo.5780120">https://doi.org/10.5281/zenodo.5780120</a>
Code for hybrid RDME-CME-ODE model	This paper	<a href="https://github.com/Luthey-Schulten-Lab/Minimal_Cell">https://github.com/Luthey-Schulten-Lab/Minimal_Cell</a> <a href="https://doi.org/10.5281/zenodo.5780120">https://doi.org/10.5281/zenodo.5780120</a>
Lattice Microbes – v2.4	This paper	<a href="https://github.com/Luthey-Schulten-Lab/Lattice_Microbes">https://github.com/Luthey-Schulten-Lab/Lattice_Microbes</a> <a href="https://doi.org/10.5281/zenodo.5780138">https://doi.org/10.5281/zenodo.5780138</a>
OdeCELL – v1.0	This paper	<a href="https://github.com/Luthey-Schulten-Lab/Minimal_Cell">https://github.com/Luthey-Schulten-Lab/Minimal_Cell</a> <a href="https://doi.org/10.5281/zenodo.5780120">https://doi.org/10.5281/zenodo.5780120</a>
Escher – v1.7.3	King et al., 2015	<a href="https://escher.github.io/#/">https://escher.github.io/#/</a>
VMD - v1.9.4a53	Humphrey et al., 1996	<a href="http://www.ks.uiuc.edu/Research/vmd/">http://www.ks.uiuc.edu/Research/vmd/</a>
Biopython – v1.78	Cock et al., 2009	<a href="https://biopython.org/">https://biopython.org/</a>
QuantStudio Software for qPCR	Applied Biosystems Inc.	<a href="https://www.thermofisher.com/us/en/home/global/forms/life-science/quantstudio-6-7-pro-software.html">https://www.thermofisher.com/us/en/home/global/forms/life-science/quantstudio-6-7-pro-software.html</a>
IDT PrimerQuest tool	Integrated DNA Technologies	<a href="https://www.idtdna.com/PrimerQuest/Home/Index">https://www.idtdna.com/PrimerQuest/Home/Index</a>
<b>Other</b>		
QuantStudio 6 qPCR instrument	Applied Biosystems, Inc.	Cat# A43159
QExactive mass spectrometer	Thermo Scientific	N/A
TriVersa NanoMate ion source	Advion Biosciences	N/A



## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Zaida Luthey-Schulten ([zan@illinois.edu](mailto:zan@illinois.edu)).

### Materials availability

This study did not generate any new unique cell strains or reagents, however the JCVI-syn3A and JCVI-syn3B bacterial strains are available to researchers through the JCVI and Codex DNA, Inc under a material transfer agreement through John Glass ([jglass@jcv.org](mailto:jglass@jcv.org)). Note that to handle JCVI-syn3A and JCVI-syn3B, United States scientists must obtain a United States Veterinary Permit for Importation and Transportation of Controlled Materials and Organisms and Vectors from the U.S. Department of Agriculture Animal and Plant Health Inspection Service. The organisms require Biosafety Level 2 containment.

### Data and code availability

- qPCR data have been deposited at Mendeley Data: (<https://doi.org/10.17632/nprw2h5tx6.1>) and are publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- All original code has been deposited at [https://github.com/Luthey-Schulten-Lab/Lattice\\_Microbes](https://github.com/Luthey-Schulten-Lab/Lattice_Microbes) and [https://github.com/Luthey-Schulten-Lab/Minimal\\_Cell](https://github.com/Luthey-Schulten-Lab/Minimal_Cell) and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Bacterial strains

The principal mycoplasmal strains used in this study are JCVI-syn3A (GenBank: CP016816.2) and JCVI-syn3B, which is genetically identical to JCVI-syn3A except is has a synthetic DNA landing pad to allow for easy introduction of new genes. These strains were propagated in the SP4 and defined medium compositions listed in the Key Resources Table. JCVI-syn3A and JCVI-syn3B were handled in labs with permits for Biosafety Level 2 containment.

### Defined medium

In a continuing effort to completely define the minimal external environment sufficient to support growth of Syn3A, increasingly defined media have been developed. Our current iteration is based on the defined medium developed and fully described by Rodwell (Rodwell, 1983) which supports growth of *Mycoplasma mycoides* subspecies including the natural precursor of JCVI-syn1.0, *Mycoplasma mycoides subsp. capri str. GM12* (Rodwell medium designation C5). This medium did not support growth of Syn3A. Defined components were therefore empirically added based on i) predictions from the analysis of metabolic pathways for Syn3A (Breuer et al., 2019), ii) components of defined media reported for growth of *M. pneumoniae* (Yus et al., 2009) and iii) a defined component of SP4 medium (CMRL-1066). The complete description of the medium is given in the Key Resources Table and [Table S1](#). In all cases lipid delivery was provided using KnockOut (KO) serum replacement as a source of albumin (Garcia-Gonzalo and Izpissúa Belmonte, 2008). Concentrations of assorted constituents in this medium served as the basis for simulated parameters in the model. A subset of the components of the defined medium were used to simulate transport reactions in metabolic reactions. The subset was selected as each component present in a transport reaction in the metabolic reactions.

Adaptation of Syn3A was accomplished by progressive dilution of SP4 cultures into the defined medium. Cultures were maintained as previously described (Hutchison et al., 2016) in static liquid culture at 37 °C, monitoring growth with the pH indicator phenol red for acid production. Mycoplasma growth was confirmed by plating dilutions of adapted cultures onto solid agar SP4 medium to identify and quantify colonies.

## METHOD DETAILS

### Background for stochastic cell simulations

The computational methods in this study employ both deterministic and stochastic reaction solvers and communication between them. Before discussing specific construction of the hybrid simulations of Syn3A, we must first introduce the stochastic simulation program Lattice Microbes (LM) (Roberts et al., 2013; Hallock et al., 2014; Earnest et al., 2018). Stochastic simulations are necessary when reactions involving few particles such as those in genetic information processing and the charging of the tRNAs can lead to large variations in the state of the cell (Taniguchi et al., 2010). LM stochastically simulates reactions without (in the well-stirred homogeneous scenario) and with the explicit inclusion of particle diffusion to account for the spatially-heterogeneous intracellular environment. For example, ribosome and nucleoid regions are distinct from cytoplasmic, membrane, and peripheral membrane regions, and we allow different reactions  $r$  to take place in each with a different propensity  $a_r$  and stoichiometry  $S_r$ . In our spatially-resolved

reaction-diffusion master equation (RDME) simulations, a cubic lattice is imposed on the cell and the system is divided into subvolumes  $v$  centered about the lattice points. Within each subvolume the reactions are assumed to be well-stirred and simulated using the Gillespie stochastic algorithm (Gillespie, 1977). Diffusion of particles from one subvolume to another is described by a diffusion operator, and the overall evolution of the state of the cell is given by the combined RDME.

$$\begin{aligned} \frac{dP(\mathbf{x}, t)}{dt} = & \mathbf{R}P(\mathbf{x}, t) + \mathbf{D}P(\mathbf{x}, t) = \sum_v \sum_r^R [-a_r(\mathbf{x}_v)P(\mathbf{x}_v, t) + a_r(\mathbf{x}_v - \mathbf{S}_r)P(\mathbf{x}_v - \mathbf{S}_r, t)] \\ & + \sum_v \sum_{\xi}^{i,j,k} \sum_{\alpha}^N [-d^{\alpha} \mathbf{x}_v^{\alpha} \times P(\mathbf{x}, t) + d^{\alpha} (\mathbf{x}_{v+\xi}^{\alpha} + 1)P(\mathbf{x} + \mathbf{1}_{v+\xi}^{\alpha} - \mathbf{1}_v^{\alpha}, t)]. \end{aligned} \quad (\text{Equation 2})$$

The state of the cell  $\mathbf{x}$  represents all the species (genes, RNAs, and proteins) present at any instant of time. The first term is the chemical master equation (CME) probabilistic description of the subvolume-localized reactions for every subvolume in the system and the second term is the diffusion operator  $\mathbf{D}$  for each particle type  $\alpha$  in the  $x$ ,  $y$ , and  $z$  directions specified by  $i$ ,  $j$ , and  $k$ . For clarity, within the context of the spatially-resolved simulations, we will use the term *local CME* to describe the modeling of subvolume-localized stochastic reactions and the term *global CME* to describe the modeling of cell-wide stochastic reactions between species assumed to be well-stirred in the full cellular volume. In the whole-cell simulations presented in this study, we combine the above simulation methods with ordinary differential equation (ODE) solvers in hybrid methods that use the results of each method to update the particle counts of the subsequent method in a backward-updating fashion (Bianchi et al., 2018), which we present schematically in Figure S3. LM allows for periodic communication between its stochastic CME and RDME solvers and other simulation methods, such as an ordinary differential equations solver for metabolic reactions. The time step for communication,  $\tau$ , is a parameter based on general timescale separation between the reactions in the metabolism, which are modeled by ODEs, and the reactions involved in the gene expression, which are modeled by a CME/RDME, that have significantly longer times between individual reactions. The communication times for linking these two methodologies were proven for similar metabolic and gene expression reactions (where regime separation was again chosen based on varying reaction propensities) in a genetic switch in yeast (Bianchi et al., 2018). The accuracy and efficiency for a series of communication times was reported there, even to the numerical limit of a computationally “exact” implementation of the algorithm, where communication occurs after the firing of each stochastic reaction. The choice of this parameter does affect simulations (Figure 2 in Bianchi et al. (2018)), however the size of these effects diminishes substantially as the communication time step is decreased from the minute to single second scale. In these simulations, we used a communication time step of 1 s.

In the spatial RDME model, the genetic information processes are simulated as reactions and diffusion into and out of each local subvolume. Specifically, RNAP diffuses to the location of the start of the gene and binds to it. Later it is released along with the transcript at the end of the gene. mRNA can diffuse and bind to either the ribosomes or degradasomes. These reactions communicate with a global CME model of transcription elongation kinetics and tRNA charging kinetics, which are assumed to take place anywhere within the cell volume. Information about the stochastic reactions that occurred during the time interval of length  $\tau$  are communicated to the ODE kinetic model for metabolism, the ODE solver is then run for an identical time interval, and the final state of the ODE model is used to update the state of the stochastic reaction system. The entirety of this procedure is shown in Figure S3.

Below we discuss more thoroughly the RDME, CME, and ODE solvers, and the combined hybrid algorithms. After introducing how the cellular architecture and DNA configurations are created from cryo-electron tomograms using the theory of a circular self-avoiding polymers (Gilbert et al., 2021), we discuss the construction of the spatial model for Syn3A that includes the full kinetic model of metabolism. A detailed description of lipid metabolism requires use of information coming from the lipidomics studies by the Saenz group which we provide at the end of the methods. The minimal cell simulation programs are available at [https://github.com/Luthey-Schulten-Lab/Minimal\\_Cell](https://github.com/Luthey-Schulten-Lab/Minimal_Cell)

### LM interface for creating spatial simulations: jLM

The algorithms discussed in the subsections below use the new release of LM, Lattice Microbes v2.4 ([https://github.com/Luthey-Schulten-Lab/Lattice\\_Microbes](https://github.com/Luthey-Schulten-Lab/Lattice_Microbes)), which comes with a new user interface package designed to be used in Jupyter Python notebooks: jLM. LM v2.4 has a more user-friendly installation. (The programs and user manuals are available through Github address in Resources and our website). jLM has an improved interface for generating initial conditions for spatially resolved simulations, including constructing cell geometries such as those shown in Figures 1B–1E, defining diffusion rates and rules, and defining reactions. To check the initial setup of the physical cell and the simulations, jLM includes visualization of cell geometries using active rendering, table visualization for particle diffusion coefficients, and table visualization of reaction details such as subvolume localization and the particles involved in the reaction. jLM allows for incorporation of data such as cryo-electron tomograms where we can include cell features directly into cell architecture in simulations as long as positional data about the cellular features is annotated. For diffusion, the simulations require diffusion probabilities for each particle between each region defined in the simulation, for example between the membrane and cytoplasm, as well as diffusion probabilities within the region. While this sounds daunting at first to define so many diffusion probabilities for many species, jLM allows the user to define global diffusion rules that will be set for all particles in the simulation which can be later modified. To define any one diffusion rule, jLM takes the particle name, subvolume region (meaning the region in which the particle currently resides), destination subvolume region, and diffusion coefficient as inputs.

To define reactions, jLM takes a list of substrates, list of products, and rate constant as inputs where the rate constant is earlier defined giving a value and reaction order as inputs.

### Overview of deterministic and stochastic simulation methods

Here, we describe individually the three simulation methods used to model the chemical kinetics in order of increasing temporal and spatial complexity. At the lowest level of complexity, we model the time-evolution of metabolite concentrations using ordinary differential equations (ODEs). We solved the deterministic initial value problem for the ODE system modeling the metabolic reactions using the LSODA solver within the ODEPACK software suite (Hindmarsh, 1983; Petzold, 1983). The LSODA solver has implementations of multistep methods via the Adams or BDF methods for both stiff and non-stiff systems of ODEs. We used the backward differentiation formula method, with order varying between 1 and 5, to solve the stiff ODE system of metabolic reactions. The initial conditions for the enzyme counts and metabolite concentrations for the metabolic reactions are in Table S1 and are based on the proteomics data from (Breuer et al., 2019) and metabolomics data primarily from (Park et al., 2016) and (Yus et al., 2009). Based on the cryo-electron tomograms for Syn3A (Gilbert et al., 2021), we assumed the initial radius of Syn3A of 200 nm and cell volume of 0.0335 fL. Missing or questionable data was supplemented by comparative analysis to *E. coli*, *M. florum*, and *B. subtilis*. The kinetic parameters are listed in Table S2.

To write the metabolic reactions into a system of ODEs that can be numerically evaluated by the ODE solver, we use a custom python package named odecell. This package was designed to have a simple application programming interface (API) for defining rate forms, specifying kinetic parameters, and assigning reactants and products for individual reactions. The package odecell comes with some predefined simple rate forms such as first and second order kinetics, but most reactions in the whole-cell model have a custom rate form defined for the random binding model of enzyme kinetics. Once all reactions have been defined and parameters, reactants, and products have been assigned, odecell is used to pass the time-evolution equation for every metabolite to the ODE solver discussed above.

The chemical master equation (CME) is an equation describing the time-evolution of a well-stirred system of reacting particles. Like other master equations, it models the probability of the system being found in a set of different states and the transitions between those states. In the case of the chemical master equation, the states are different combinations of discrete counts of particles and the system transitions between states as a result of chemical reactions. Under the well-stirred condition, the chemical reactions are equally probable between any reactant particles in the system volume. The system is assumed to be Markovian and the transition probabilities are conditionally dependent on the current state and the parameterization of the transition probabilities. Given this assumption, realizations of the system's time-evolution can be simulated using the Gillespie algorithm and its variants. Ensembles of those realizations can then be used to reconstruct the time-evolution of the probability distribution of system states. The advantage of a stochastic CME model of chemical kinetics versus a deterministic ODE model is that the CME model reports the distribution of system states (and fluctuations), rather than only the mean concentration. This is especially relevant for systems with low copy numbers, such as models of genetic information processing, where the fluctuations and mean are of a similar order of magnitude.

At the greatest level of complexity, we use the reaction-diffusion master equation (RDME) to model diffusion and reactions within Syn3A. The RDME is a spatially-resolved version of the chemical master equation, where the system is now a set of connected subvolumes and the system state is the distribution of particles across that set of subvolumes. The system state changes by particles diffusing between subvolumes or reacting within subvolumes. The reactions among particles in a subvolume are handled as well-stirred and can be simulated using the same methods as the CME.

We sampled the stochastic initial value problems for the CME and RDME systems using the newly-released version 2.4 of Lattice Microbes (LM) (Roberts et al., 2013; Hallock et al., 2014; Earnest et al., 2018), a GPU-based stochastic simulation software for chemical kinetics. The CME system was sampled using the implementation of the Gillespie direct algorithm and the RDME system was sampled using the implementation of the multi-particle diffusion algorithm. The reaction network for the CME system was constructed using the pyLM subpackage, a problem-solving environment that provides Python bindings to access the underlying LM code written in C++. The spatial model and reaction network for the RDME system were constructed using the new jLM subpackage, which extends the functionality of pyLM to include real-time visualization and interrogation of the system using ipywidgets within Jupyter notebooks. The simulations and analyses presented below were performed using Python scripts and Jupyter notebooks that are listed within the Key Resource Table.

### Construction of JCVI-syn3A cell geometry and DNA configurations

We model the spatially-resolved kinetics using a RDME formalism and simulate the system using LM, which necessitates recreating the geometry of cells in a cubic lattice representation. Centered about the lattice sites are cubic reaction subvolumes. Particles within a common subvolume are assumed to be well-stirred and the subvolume-localized reactions proceed in a manner identical to the CME. Particles may also diffuse between subvolumes that are directly adjacent and share a face. The selection of the lattice size is an essential step in the creation of the spatial model. Through a constraint introduced by the greatest diffusion rate, the lattice size and maximum timestep are interdependent, i.e., a greater lattice size permits greater maximum timesteps (Roberts et al., 2013). Ultimately, the balance between increased spatial-resolution and computationally-achievable timescales is a choice on the part of the modeler. Previous LM simulations of ribosome biogenesis in the model Gram-negative bacterium *E. coli* used a lattice with dimensions of 32 × 32 × 192 sites (196,608 total sites) and a lattice size of 32 nm, with timesteps of 25 μs (Earnest et al.,

2016). Syn3A is approximately one-tenth of the physical size of *E. coli* and we chose a lattice size of 8 nm to realize the effects of spatial heterogeneities in the small system, while allowing for simulations over biologically-relevant timescales. The dimensions of the lattice used for the Syn3A model is  $64 \times 64 \times 64$  sites (262,144 total sites). The location of reactions and diffusion are controlled by manipulating the types of individual lattice sites, which in turn associate the surrounding subvolume with a region of the cell. For example, translation reactions are prohibited from occurring in subvolumes associated with the chromosome. In another example, proteins may diffuse from subvolumes associated with the ribosomes into neighboring cytoplasmic subvolumes, but the inverse is prohibited. Constructing the spatial-model can be decomposed into three essential steps: 1) creating the cell architecture on a cubic lattice by manipulating the site types, 2) specifying the rates for reactive and diffusive events, along with their region-based rules, and 3) placing the particles within the model.

We created single-cell architectures of Syn3A cells, including the ribosome distribution and chromosome configurations, using the cryo-ET and self-avoiding polymer model described in our previous work (Gilbert et al., 2021). In summary, there are four steps to the process. 1) Syn3A cells were imaged using cryo-ET by the lab of Elizabeth Villa at UCSD and ribosome coordinates were determined by applying an iterative binary classification procedure (Tegunov and Cramer, 2019) to tomographic reconstructions of Syn3A cells. A cell observed to have roughly a 200 nm radius contained 503 ribosomes distributed nearly-randomly throughout the cell. In a few cases, neighboring ribosomes were so close that possible polysomes ranging in size of 2-5 ribosomes could be identified, and we include them in our treatment of translation (Figure 1A). For comparison, *M. florum* contains 1,600 to 2,100 ribosomes (Matteau et al., 2020) which corresponds to 650 to 850 ribosomes when scaled to the volume of Syn3A. A larger ribosome count is not surprising because of the faster doubling time of *M. florum*.

2) Assuming the ribosomes to be spherical with a radius of 10 nm, the ribosomes were placed in the lattice representation using seven 8 nm lattice sites arranged as a star. 3) We modeled the 543 kbp circular chromosome of Syn3A as a lattice polymer on a 4 nm lattice cubic lattice (11.8 bp per monomer). This lattice polymer model was constrained to be self-avoiding and circular, a type of model also known as a self-avoiding polygon (SAP). The 4 nm lattice was made to be coincident with the 8 nm cubic lattice, and the chromosome model was also constrained to avoid the ribosomes and remain within the membrane.

Ensembles of chromosome configurations constrained by the transformed tomogram data (ribosomes and membrane) were then generated using a Monte Carlo algorithm that alternated between freely-growing the SAP through the insertion of monomers and then equilibrating the lattice model while subject to a simplified Hamiltonian containing a nearest-neighbor excluded volume term and a bending potential based on the assumed persistence length of dsDNA.

4) Finally, a coarse-graining procedure was applied to the configurations to localize up to eight of the 4 nm monomers within the 8 nm lattice sites, this is shown as the spheres embedded in the orange lattice sites in Figure 1B. Upon completion, 46,188 sequence-specific DNA particles, each representing an 11.8 bp portion of Syn3A's chromosome, were distributed among the 8 nm lattice sites according to the final positions of the monomers in the 4 nm lattice polymer model of the chromosome. An example of a Syn3A cell geometry with the ribosomes and DNA particles is shown Figure 1C. There were few intrachromosomal interactions in the preliminary 3C-Seq map, and only a small number of possible DNA looping interactions were identified (Gilbert et al., 2021). The lack of intrachromosomal interactions was hypothesized to be caused by a lack of persistent supercoiling, which may result from two factors present in Syn3A: 1) a low abundance of the nucleoid-associated protein HU, which can stabilize plectonemic loops resulting from supercoiling, and 2) a relatively high abundance of topoisomerases and gyrases, which can relax translation-induced supercoiling (Gilbert et al., 2021). The limited number of possible DNA looping interactions were assumed to be unsupercoiled loops created by SMC protein complexes. Due to the uncertainty about nature of the interactions in the preliminary 3C-seq map, the whole-cell simulations in this study used chromosome configurations without DNA looping present. The DNA configurations for the independent RDME simulations were chosen from an ensemble of over 100 possible configurations, so that no two simulations contained the same configuration. Within the spatially-resolved model, the lifetime of a gene's mRNA transcript is strongly-dependent on the proximity of the gene's to ribosomes and the membrane-associated degradosomes. We evaluated the uniqueness of the DNA configurations in our study by comparing the radial distances from the center of the cell for all 493 gene end sites. There were significant variations between the 8 configurations we selected with an average difference in radial distances of approximately 50 nm.

### Initialization of proteins, mRNAs, and tRNAs

Having placed the ribosomes and circular DNA, we create the rest of the cellular architecture by specifying regions of the cell. Within spatially-resolved LM simulations, this is done by manipulating the lattice site types that the individual reaction subvolumes are centered about. To do this, we use new functionality in the jLM subpackage, in which the user can specify the parametric form of select three-dimensional shapes and jLM will generate matching lattice representations. Additionally, jLM allows for logical set operators to be used to compose multiple lattice representations to create more complex geometries. If a particle restricted to a region may freely diffuse to every subvolume within that region during the course of an RDME simulation, then we refer to that region as being contiguous. All lattice sites in the simulation space are initialized as belonging to the *extracellular* region. We first create a lattice representation of a sphere with a radius of 200 nm and define those lattice sites as belonging to the *cytoplasm* region. Within the spatially-resolved reaction model, the outermost layer of the cytoplasm is defined as a separate *outer-cytoplasm* region, so that peripheral membrane complexes, such as the degradosome, can be treated independently from transmembrane complexes. We next



create this *outer-cytoplasm* region by using jLM to compose a spherical shell that is exterior and adjacent to the *cytoplasm* region, a minimum of one subvolume in thickness, and contiguous. In the final step, we create the *membrane* region by using jLM to compose a spherical shell that is exterior and adjacent to the *outer-cytoplasm* region, a minimum of one subvolume in thickness, and contiguous.

Once the cellular architecture has been constructed, we place other macromolecular complexes such as degradosomes and transporters, mRNA, and proteins as shown in Figures 1D and 1E. The degradosome is a complex that consists of a membrane-bound endoribonuclease scaffold (RNase Y), two metabolic enzymes (enolase and phosphofructokinase), an RNA helicase, a 3' to 5' exoribonuclease (RNase R or a putative exoribonuclease), and two 5' to 3' exoribonucleases (RNases J1 and J2) (Cho, 2017). The 3' to 5' exoribonuclease that typically binds most favorably in Gram-positive bacteria is PNPase. However, since PNPase isn't present in JCVI-syn3A, we assume one of the other two 3' to 5' exoribonucleases present in the cell can take its place. The degradosome breaks down mRNA by first cleaving messengers with RNase Y, unwinding any dsRNA using the helicase, and then degrading the fragments from end to end using the exoribonucleases. We assume the stoichiometry of all degradosome components to be 1, so we take the minimum proteomics count among the components to determine the total number of complete degradosomes: 120 shared by RNase J1 (*rnjA*/0600) and the putative degradosome helicase (0410).

Proteins are initialized to their counts from the genome-wide proteomics for Syn3A (Breuer et al., 2019). The 93 membrane proteins making up roughly 9,000 of the 77,000 proteins in the proteomics are randomly distributed in the membrane. Proteins not reported in the proteomics or with counts less than 10 are initialized with a count of 10. 10 was chosen as the default protein count because the average uncertainty among all proteins over the three replicate mass spec experiments performed by Breuer et al. (2019) was  $\pm 5$ . See Table S1 for all initial counts used in the simulations.

In the absence of transcriptomics data, mRNA are initialized using information from previous simulations. Well-stirred simulations were run to long times without replication reactions (1 h cell time) for 100 cells to get average mRNA counts that will be accurate for early times in the cell cycle when only one chromosome is present. From these simulations, the time-average count of each mRNA was calculated over the 100 cells. The initial mRNA counts for every gene were independently sampled from Poisson distributions with mean parameters equal to their respective time-averaged mRNA counts in the steady-state simulations. Following mRNA initialization, the cells have 150 total mRNA particles on average. tRNA are initialized to 200 tRNA per isoform in the genome. The counts of tRNAs were determined by scaling a total count of roughly 200,000 tRNAs in *E. coli* (Neidhardt and Umbarger, 1996), which corresponds to roughly 6,000 total tRNAs in Syn3A when scaled by cell volume. Syn3A has a total of 29 tRNA isoforms, and the total count was partitioned evenly among all species and rounded to counts of 200 tRNAs per isoform. Each protein, mRNA, and tRNA are initialized as randomly distributed in their respective region in the cell. For comparison, *M. florum* contains roughly 18,000 total tRNAs and 420 total mRNAs (Matteau et al., 2020) which correspond to roughly 7,000 tRNAs and 170 mRNAs when scaled to the volume of Syn3A.

Ribosomes and DNA are assumed to be stationary in this present model. Diffusion coefficients of proteins were set to  $1.0 \mu\text{m}^2/\text{s}$  (Earnest et al., 2015, 2016). Diffusion coefficients of all RNA were calculated using a modified Stokes-Einstein equation (Werner, 2011)

$$D = \frac{k_B T}{6\pi\eta R_H} \quad (\text{Equation 3})$$

where the hydrodynamic radius  $R_H$  was calculated using the equation

$$R_H = \left( \frac{3m_m L_{RNA}}{4\pi N_A \rho} \right)^{1/3}, \quad (\text{Equation 4})$$

which assumes the globular RNA to be approximately spherical. To calculate the hydrodynamic radii, a molar mass  $m_m$  of  $337 \text{ g mol}^{-1}$  per nucleotide (the average molar mass of AMP, UMP, GMP, and CMP) and density  $\rho$  of  $1.8 \text{ g cm}^{-3}$  (Werner, 2011) were used along with the length of the RNA  $L_{RNA}$  in nucleotides and Avogadro's number  $N_A$ . The viscosity  $\eta$  was calculated using a measured mRNA diffusion coefficient of  $0.03 \mu\text{m}^2/\text{s}$  in *E. coli* for a mRNA 4,000 nt in length (Golding and Cox, 2004). From the measured diffusion coefficient and mRNA length, the viscosity was back calculated to be  $1.2 \text{ Pa s}$  using the hydrodynamic radius and modified Stokes-Einstein equations. The temperature was assumed to be 310 K. mRNAs are allowed to diffuse into ribosome subvolumes, but proteins were excluded once they leave the ribosome after translation. From the fastest diffusion rate, the diffusion coefficient used for protein diffusion, and the lattice spacing of 8 nm, the time step of the simulation was set to  $30 \mu\text{s}$ . The diffusion coefficients can all be found in Table S2.

### Genetic information processing

The set of genetic information processing reactions improve upon the model by Thornburg et al. (2019) that simulates DNA replication initiation and elongation, transcription of all 493 genes, degradation of all 452 mRNA, and translation of all 452 proteins in the CME model. As in the recent studies on *B. subtilis* (Johnson et al., 2020), we continue to assume that transcription is decoupled from translation. We again simulate the elongation reactions of each process under the well-stirred assumption of the chemical master equation using the Gillespie. The genetic information processes now also include directly effects of diffusion of RNAP and mRNA and their binding to the DNA, ribosomes, and degradosome including diffusion and spatial heterogeneity using the reaction-diffusion master

equation (RDME). In the hybrid CME-ODE simulations, there are no binding reactions because the whole cell is assumed to be well-stirred, but the effects of diffusion are included as probabilistic prefactors multiplying  $k_{cat}$  involving the active numbers of RNAP, ribosomes, and degradosomes. In the RDME-gCME-ODE (RDME-ODE), particles must diffuse to one another to undergo a chemical reaction. For example, mRNA must diffuse to degradosomes to get degraded and to ribosomes to be translated. Counts of all proteins, mRNA, and tRNA are initialized in the same way as in the spatially resolved model discussed above. In the CME-ODE model, the DNA replication, transcription, and translation elongation reactions are all modeled using an enzyme-catalyzed, template-driven polymerization mechanism based on the general rate form (Thornburg et al., 2019)

$$\nu_{polymerization}^{CME-ODE} = \frac{k_{cat}}{\left(1 + \frac{K_0}{[Enzyme]}\right) \frac{K_{D1}K_{D2}}{[Monomer_1][Monomer_2]} + \sum_i \frac{n_i K_{Di}}{[Monomer_i]} + L_{polymer} - 1} N_{Template} \quad (\text{Equation 5})$$

This equation was first derived in Hofmeyr et al. (2013) for the case where the enzyme is more abundant than its template polymer. Each reaction type requires the DNA, RNA, or protein sequence to determine the polymer length  $L_{polymer}$  and the number of each monomer type in the sequence  $n_i$ . This rate form includes the binding of the enzyme (DNAP, RNAP, or ribosome) to the template polymer (ssDNA or mRNA) because of the assumption that the templates and enzymes are well-stirred when using the chemical master equation. The binding step is accounted for through  $K_0/[Enzyme]$ . Dissociation constants of monomers  $K_D$  (dNTPs, NTPs, or aa:tRNAs respectively) to the respective enzyme enter in the rate form for the first two monomers in the sequence  $Monomer_1$  and  $Monomer_2$  as well as in the summation over all monomer types  $i$  in the polymer. Since the reactions are simulated stochastically, these rates are calculated using the time-dependent counts of their respective templates  $N_{Template}$  rather than their concentrations.  $N_{Template}$  is typically less than 6 and enzyme counts are typically greater than 100. Thornburg et al. (2019) assumed that the pools of monomers would remain constant over the cell cycle and that the polymerization rates are dominated by the total length of the polymer product. As long as the monomer pools remain high, this assumption is reasonable, but not all cells have large nucleotide and amino acid pools. Here we use time-dependent pool sizes updated once per minute so that the genetic information processing reactions can respond dynamically to the metabolism. The kinetic parameters appear in Table S2.

In our spatially-resolved model, we do not simulate the reactions for DNA replication since we only simulate the first 20 min of the cell cycle. The model still includes reactions to transcribe all 493 genes, degrade all 452 mRNA, and translate all 452 mRNA into proteins. The polymerization rates in the spatial model do not include the binding step of the enzyme to the template and instead simulate separate binding reactions that require the RNAP to diffuse to genes and mRNA to diffuse to degradosomes and ribosomes. The general polymerization rate form using this assumption was derived to be

$$\nu_{polymerization}^{RDME-ODE} = \frac{k_{cat}}{\frac{K_{D1}K_{D2}}{[Monomer_1][Monomer_2]} + \sum_i \frac{n_i K_{Di}}{[Monomer_i]} + L_{polymer} - 1} N_{Enzyme:Template} \quad (\text{Equation 6})$$

Rather than depending on the time-dependent template counts, the rates are now dependent on the count of the enzyme-template bound state  $N_{Enzyme:Template}$ , for example  $N_{RNAP:gene,i}$  once the enzyme and/or template has diffused to the other and bind. All sequences were read from the Syn3A NCBI entry (CP016816.2) using Biopython (Cock et al., 2009)

### DNA replication initiation

Thornburg et al. (2019) identified the three neighboring 9 bp affinity sites where domain DnaA(IV) is known to bind to dsDNA based on the crystallographic study (PDB: 1j1v and PDB: 3r8f) from Erzberger et al. (2006); Duderstadt et al. (2011). We simulate reactions for DnaA(IV) binding to the 9 bp high affinity site next to the origin, TTATCCACA, and subsequent binding to two neighboring low affinity sites (7/9 bp matches 2 and 1 bp away). The binding of the three DnaA(IV) molecules to the 27 bp opens the dsDNA is assumed to allow the further binding of domain DnaA(III) to the neighboring AT rich region on the genome. The binding of DnaA(III) requires 3 nt along the ssDNA (Duderstadt et al., 2011), so the AT region of 90 nt would create a filament of 30 DnaA proteins to stabilize a bubble large enough for the replicating machinery represented by DNA polymerase to enter and begin replication. The kinetics of replication initiation depend directly on the instantaneous cell volume and number of the multi-domain protein DnaA.

$$\nu = k_{on}[DnaA] - k_{off} = (k_{on} / N_A V(t)) N_{DnaA} - k_{off} \quad (\text{Equation 7})$$

This equation represents the formation rate of a DnaA filament along single-stranded DNA near the origin. Both  $k_{on}$  and  $k_{off}$  for the binding of DnaA(III) were measured using single molecule FRET (Cheng et al., 2015). The on and off rates compete with each other depending on the available DnaA and the concentration dependence on the cell volume  $V$ . Because the replication initiation reactions are treated stochastically, the equation using the number of DnaA ( $N_{DnaA}$ ) is used. The on rate is calculated using Avogadro's number ( $N_A$ ) and the initial cell volume for the first replication initiation event. Once the chromosome is doubled late in the cell cycle, it is assumed based on Figure 3D that the volume has doubled. The on rate for replication initiation events on the two daughter chromosomes later in the cell cycle are then recalculated using the doubled volume. Based on these kinetics, if the number of DnaA doubles faster than the volume, it is likely that another replication initiation event will occur and the cell will have more than two chromosomes at the end of its cell cycle.

### DNA replication elongation

Once a filament 30 DnaA long has formed, we initiate replication forks proceeding in both directions around the circular chromosome. Thornburg et al. (2019) had DNA replication duplicate all genes at the same time after the whole chromosome was copied. Here, we duplicate genes one at a time in both directions in order from origin to terminus using the elongation rate

$$v_{DNA\ replication}^{CME-ODE} = \frac{100\ bp/s}{\left(1 + \frac{K_0}{[DNAP]}\right) \frac{K_{D1}K_{D2}}{[dNTP_1][dNTP_2]} + \sum_i \frac{n_i K_{D1}}{[dNTP_i]} + L_{gene} + L_{intergenic} - 1} N_{replication\ fork:gene} \quad (\text{Equation 8})$$

As the replication fork proceeds along the chromosome, we update which gene it is at to be used as the template polymer  $N_{replication\ fork:gene}$  in the polymerization rate. While we simulate the translation of all proteins involved in the DNAP, we use the count of the  $\alpha$  subunit of DNAP as a representative count based on stoichiometry of the complex. In the duplication reaction for each gene, we include the intergenic region following the gene in the rate of elongation to accurately simulate the time of elongation for the whole chromosome, making the total length of polymerization  $L_{gene} + L_{intergenic}$ . We do not treat Okazaki fragments and assume the lagging strand is duplicated at the same time as the leading strand without any time delays. The dNTP costs of the leading and lagging strands are both accounted for by communication with the instantaneous pools in the metabolism every minute. DNA replication uses 1 ATP as an energy molecule per bp.

### Transcription

We transcribe each protein-coding gene individually using the polymerization rate

$$v_{transcription,mRNA}^{CME-ODE} = \frac{S_{promoter} \times (N_{active\ RNAP}/493\ genes) \times 20\ nt/s}{\left(1 + \frac{K_0}{[RNAP]}\right) \frac{K_{D1}K_{D2}}{[NTP_1][NTP_2]} + \sum_i \frac{n_i K_{D1}}{[NTP_i]} + L_{mRNA} - 1} N_{gene} \quad (\text{Equation 9})$$

Since the RNAP could read any of the 493 genes, we include the probabilistic factor  $N_{active\ RNAP}/493\ genes$  which is the probability of an active RNAP reading one of the 493 genes. To achieve a doubling time range in the well-stirred simulations that is close to the experimentally observed doubling time (Breuer et al., 2019), a lower fraction of active RNAP is assumed than was calculated in the spatial simulations. Rather than the 33% calculated, 16% of RNAP was selected to be active in the well-stirred simulations (Bremer and Dennis, 2008). The higher fraction of active RNAP results in larger counts of mRNA and therefore also larger counts of proteins, including the membrane proteins growing the membrane surface area and lipid synthesis proteins, increasing the growth rate of the cell (data not shown). We estimate the number of active RNAP  $N_{active\ RNAP}$  using the count of 187 RNAP in Syn3A (Breuer et al., 2019), which together with the active fraction predicts 30 active RNAP in Syn3A. Each gene is not equally likely to be read, though, so we include a factor to simulate promoter strength  $S_{promoter}$ . We use the assumption of Thornburg et al. (2019) that the genome-wide proteomics can be used to calculate proxies for the promoter strength of each gene as the proteomics count of the protein product of the gene divided by the average proteomics count of 180. An upper limit is enforced equal to the rRNA operon elongation rate of 2x85 nt/s, one factor of 85 nt/s per operon. We do not simulate any operonal structures for protein-coding genes and transcribe each gene individually. Once transcriptomics data becomes available for Syn3A to determine operonal structures, we can incorporate transcription of operons into this model. Transcription reactions use 1 ATP as an energy molecule per nt.

Transcription of tRNA also includes the probabilistic factor of an active RNAP to be reading a particular gene, but does not include a promoter strength. No operons are assumed for transcription of tRNA. This results in the tRNA transcription rate taking the form

$$v_{transcription,tRNA}^{CME-ODE} = \frac{(N_{active\ RNAP}/493\ genes) \times 20\ nt/s}{\left(1 + \frac{K_0}{[RNAP]}\right) \frac{K_{D1}K_{D2}}{[NTP_1][NTP_2]} + \sum_i \frac{n_i K_{D1}}{[NTP_i]} + L_{tRNA} - 1} N_{gene} \quad (\text{Equation 10})$$

Because Syn3A still has genes *nusG*/0840 and *nusB*/0107, both of which are factors required for rRNA transcription antitermination (Torres et al., 2004), we use an accelerated rate of 85 nt/s for transcription of the RNAP on the rRNA operons (Ryals et al., 1982). Thornburg et al. (2019) assumed that 2 RNAP will always be transcribing each of the rRNA operons based on counts of RNAP reading the rRNA operons in *E. coli* reported by Bremer and Dennis (2008). We assume that if a RNAP is transcribing a rRNA it will transcribe the whole operon for both rRNA operons in Syn3A. So in transcription of rRNA we assume that the polymer length entering into the transcription rate is the length of the whole rRNA operon  $L_{rRNA\ operon}$  including intergenic space, making the rRNA-specific polymerization rate

$$v_{transcription,rRNA}^{CME-ODE} = \frac{2 \times 85\ nt/s}{\left(1 + \frac{K_0}{[RNAP]}\right) \frac{K_{D1}K_{D2}}{[NTP_1][NTP_2]} + \sum_i \frac{n_i K_{D1}}{[NTP_i]} + L_{rRNA\ operon} - 1} N_{gene} \quad (\text{Equation 11})$$

Since reactions are not included for post-transcriptional modifications, the products of this reaction are separated 16S, 23S, and 5S rRNA.

In the spatial model, the RNAP have to diffuse to a gene, which are in fixed positions in these simulations, to transcribe it. The transcription rate for mRNA still includes the proxy promoter strength defined above, but takes on the modified form

$$v_{transcription, mRNA}^{RDME-ODE} = \frac{S_{promoter} \times 20 \text{ nt/s}}{\frac{K_{D1}K_{D2}}{[NTP_1][NTP_2]} + \sum_i \frac{n_i K_{Dj}}{[NTP_i]} + L_{mRNA} - 1} N_{RNAP:gene} \quad (\text{Equation 12})$$

The rate of transcription has a set maximum of 85 nt/s to match the elongation rate of the rRNA operon and a lower limit of 10 nt/s. Since each gene is unique and distinguishable in the cell, these reactions are simulated in the global CME once the RNAP has bound to a gene using the time-dependent bound state  $N_{RNAP:gene}$ . In moving these reactions to the global CME, which is reinitialized at every communication time of 1 s, the transcription rates are calculated using the instantaneous NTP pool sizes. Because gene order or position in the chromosome is known for each DNA configuration, we place the products of transcription at the end of the gene being transcribed, thereby taking into account gene directionality.

The elongation rates of transcription for tRNA and rRNA are similarly modified in the spatial model. Rather than multiplying the rRNA operon elongation rate by 2 to mimic the effect of having 2 RNAP on the operon, the spatial model allows for 2 RNAP to bind to and read the operon at any one time.

$$v_{transcription, tRNA}^{RDME-ODE} = \frac{20 \text{ nt/s}}{\frac{K_{D1}K_{D2}}{[NTP_1][NTP_2]} + \sum_i \frac{n_i K_{Dj}}{[NTP_i]} + L_{tRNA} - 1} N_{RNAP:gene} \quad (\text{Equation 13})$$

$$v_{transcription, rRNA}^{RDME-ODE} = \frac{85 \text{ nt/s}}{\frac{K_{D1}K_{D2}}{[NTP_1][NTP_2]} + \sum_i \frac{n_i K_{Dj}}{[NTP_i]} + L_{rRNA \text{ operon}} - 1} N_{RNAP:gene} \quad (\text{Equation 14})$$

### mRNA degradation

The rate of the mRNA degradation reactions are influenced by the results of the spatially-resolved simulations by using the predicted number of active degradosomes. The rate form used for mRNA degradation is

$$v_{mRNA \text{ degradation}}^{CME-ODE} = \left( \frac{N \text{ active degradosomes}}{452 \text{ mRNA types}} \right) \left( \frac{88 \text{ nt/s}}{L_{mRNA}} \right) N_{mRNA} \quad (\text{Equation 15})$$

where  $N_{mRNA}$  is the number of one unique type of mRNA and the probability of a degradosome selecting one type of mRNA is the number active degradosomes, 18 degradosomes predicted by the spatial model, divided by 452 mRNA types from the genome. The rate of degradation is then determined by a velocity of 88 nt/s measured for the exoribonuclease RNase R using optical tweezers (Fazal et al., 2015) divided by the length of the mRNA. mRNA degradation uses 1 ATP as an energy molecule per nt broken down.

To simulate the binding reaction of an mRNA to a degradosome in the spatial model, a binding rate of mRNA to RNase Y was calculated using a measured 11 cleaving events per protein per cell for Rnase E in *E. coli* Mackie (2013), an protein analogous to Rnase Y in Syn3A. Once an mRNA diffuses and binds to a degradosome, which is fixed to the membrane in these simulations, it is assumed that the mRNA has been cleaved and cannot unbind as a functional mRNA. The mRNA is then decayed at a rate equal to a velocity of 88 nt/s measured for RNase R Fazal et al. (2015) divided by the length of the mRNA.

### Translation

The polymerization rate of translation is modified by a model for polysomes both in the hybrid well-stirred CME-ODE and spatially-resolved RDME-ODE simulations. We improved upon the polysome model assumed by Thornburg et al. (2019) using polysome density estimated by Gilbert et al. (2021) and the predicted fraction of active ribosomes in the spatial model presented in this study (Thornburg et al., 2019). A cell roughly 200 nm in radius consisting of 503 ribosomes was estimated to have 25% of ribosomes participating in polysomes from cryo-ET. Our spatial model predicts that roughly 45% of ribosomes will be actively translating at any one time in Syn3A. Of the 45% of ribosomes that are active, in our well-stirred model we assume all 25% of ribosomes in polysomes are active ( $P_{active \text{ polysome}}$ ), leaving 20% of active ribosomes as single ribosomes ( $P_{active \text{ ribosome}}$ ). These prefactors make our overall rate of translation

$$v_{translation}^{CME-ODE} = \frac{(P_{active \text{ polysome}} \times L_{max \text{ polysome}} + P_{active \text{ ribosome}}) \times 12 \text{ aa/s}}{\left( 1 + \frac{K_0}{[Ribosome]} \right) \left( \frac{K_{D1}K_{D2}}{[AA:tRNA_1][AA:tRNA_2]} + \sum_i \frac{n_i K_{Dj}}{[AA:tRNA_i]} + L_{protein} - 1 \right)} N_{mRNA} \quad (\text{Equation 16})$$

If the mRNA is read by a polysome, we assume a polysome spacing of 120 nt to determine the size of polysome  $L_{max \text{ polysome}}$  that can be reading the mRNA estimated from a distribution of polysome sizes in *E. coli* (Brandt et al., 2009). We enforce the polysome size to be one fewer ribosome than the spacing allows so that the whole mRNA can be read by each ribosome in the polysome simultaneously. Polysome sizes estimated from the two cryo-ET methods applied to Syn3A (Gilbert et al., 2021) range from 2 to 13, but in our well-stirred CME-ODE model we enforce an upper bound of 15 ribosomes in a polysome which was the largest polysome size re-



ported in experimental distributions from cryo-ET by Brandt et al. (2009). Translation uses 2 GTP as energy molecules per amino acid: one during the loading of an aa:tRNA into the A site of the ribosome and a second during translocation to the next step on the mRNA (Lynch and Marinov, 2015).

The translation rate in the spatial model does not include any prefactors for polysomes, making the rate of translation for the full length of a protein

$$v_{\text{translation}}^{\text{RDME-ODE}} = \frac{12 \text{ aa/s}}{\frac{K_{D1}K_{D2}}{[AA : tRNA_1][AA : tRNA_2]} + \sum_i \frac{n_i K_{Di}}{[AA : tRNA_i]} + L_{\text{protein}} - 1} N_{\text{Ribosome:mRNA}} \quad (\text{Equation 17})$$

Since ribosomes are in fixed positions in these simulations, it is required that the mRNA diffuse to a ribosome and bind before translation reactions occur. Therefore the rate of translation is dependent on the count of mRNA bound to a ribosome  $N_{\text{Ribosome:mRNA}}$ . If two of the same type of mRNA, for example coding for DnaA, were bound to ribosomes at different locations in the cell, they would be indistinguishable in the global CME method. Because the two would be indistinguishable, the simulation would not know where to put the products of translation inside the cell, so these reactions are locally evaluated in their own subvolumes in the RDME portion of the simulation.

The spatial RDME-ODE simulations take into account polysomes from the ribosome positions in cryo-ET. Polysome clusters are assigned using a center-to-center distance of 20 nm between ribosomes, and we assigned polysome clusters ranging from 2 ribosomes to 5 ribosomes for the small cell (200 nm radius) used in the spatial simulations. For clusters of 2 ribosomes, order of the polysomes in the ribosomes is assigned randomly. For clusters of 3 or more ribosomes in a polysome, the two ribosomes at the ends of the polysome are determined as the ribosomes with the greatest total distance from each ribosome in the polysome. From those two ribosomes, the start and end are randomly assigned. Then, from the start ribosome, order is assigned as the next nearest-neighbor until reaching the ribosome at the end of the polysome.

In the translation reactions for polysomes, the mRNA is allowed to bind to any ribosome in the polysome. Once a polysome translation reaction is complete, the simulation again checks if the mRNA will be released or passed to the next ribosome in the polysome. The first ribosome reading the mRNA reads the full length of the mRNA using the translation rate for a single ribosome producing a protein of length  $L_{\text{protein}}$ . Once the first translation reaction is complete, the simulation checks if the next ribosome in the polysome order is occupied. If the ribosome is occupied or if it is at the end of the polysome, the mRNA and protein come off and diffuse away. If the next ribosome is free, then the mRNA is passed to that ribosome which will undergo a translation reaction at the rate

$$v_{\text{polysome translation}}^{\text{RDME-ODE}} = \frac{12 \text{ aa/s}}{\frac{K_{D1}K_{D2}}{[AA : tRNA_1][AA : tRNA_2]} + \sum_i \frac{n_i K_{Di}}{[AA : tRNA_i]} + 40 \text{ aa} - 1} N_{\text{Ribosome:mRNA}} \quad (\text{Equation 18})$$

only reading the last 40 amino acids of the protein sequence, corresponding to the polysome spacing of 120 nt used in the well-stirred model. This treatment of polysomes results in accurate timing of proteins coming off of polysomes for the metabolic reactions while reducing the number of intermediate states that need to be tracked over the course of the simulation.

### Membrane Protein Insertion

In Syn3A, membrane proteins are inserted via a translocation reaction by the SecA-SecY complex. In the CME-ODE model, we assume that newly translated membrane proteins in the cytoplasm will find a SecY under the well-stirred assumption. The rate of the insertion reaction is then the elongation rate of SecY (50 aa/s (Serdiuk et al., 2019)) divided by the length of the protein multiplied by the cytoplasmic number of that protein

$$v_{\text{insertion}} = \frac{50 \text{ aa/s}}{L_{\text{protein}}} \times N_{\text{protein cytoplasmic}} \quad (\text{Equation 19})$$

The elongation rate of SecY was calculated from the insertion ratio of LacY measured in phospholipid liposomes using atomic force microscopy in units of membrane inserted segments per second (Serdiuk et al., 2019). The SecA-SecY complex has an associated ATP cost for the power stroke motion used to insert membrane proteins. From FRET measurements and molecular dynamics simulations, it has been estimated that the complex moves 10 amino acids per stroke using 1 ATP per stroke (Park et al., 2014; Chen et al., 2015; Catipovic and Rapoport, 2020), so we assume a cost of 1 ATP per 10 amino acids in our insertion reactions.

For a membrane protein to be inserted into the membrane in the spatial model, it must first diffuse to a SecY, which are in fixed, but random positions in the membrane in these simulations. We assume that the rate of binding of a membrane protein to the SecY is fast using a rate of  $10^6 \text{ M}^{-1}\text{s}^{-1}$  once the membrane protein has diffused to the SecY. The protein is then inserted at a rate equal to a velocity of 50 aa/s divided by the length of the protein.

### Sensitivity of polymerization reactions

The most critical parameter to assess the sensitivity in the genetic information processing polymerization reactions is the ratio of  $K_D$  for each respective enzyme associating with their respective monomers to the concentration of the monomers themselves ( $K_D/[M]$ ), for example NTPs associating with an RNAP before their incorporation into the nascent transcript. The value of these  $K_D$ 's affects the monomer concentration at which the rate of polymerization is reduced from its maximum, and we demonstrate how this ratio affects

the overall rates of polymerization in Figure S4. Rates in all plots of Figure 4 are calculated relative to the maximum polymerization rates. The plot for general polymerization (Figure S4G) shows the case when all  $K_D/[M]$  ratios are being changed simultaneously, which would show the same trend regardless of reaction type because of the identical rate law. When the monomer concentration is in equilibrium with its binding site, i.e.,  $K_D/M = 1$ , the overall rate of polymerization is halved. We also see that for the polymerization reactions to be at their maximum rate, the monomer concentration needs to be roughly 100 times greater than the  $K_D$ . The reaction will stop almost entirely once the monomer concentration is 100 times less than the  $K_D$ .

Figures S4H–S4J used the sequences of the gene, mRNA, and protein for *dnaA*/0001 because it is representative of the average monomer contents in Syn3A. For replication, we show the rate's dependence on the  $K_D/[M]$  ratio (Figure S4H) as well as the sensitivity to the dATP and dNTP concentrations using the  $K_D$  used in the simulations (Figure S4A and Figure S4B). The dependence on the  $K_D/[M]$  ratio only shows what happens as the ratio is varied for a single dNTP, dATP, because varying all of them simultaneously would result in the general plot discussed above. When the dATP concentration is equal to the  $K_D$ , the overall rate of DNA replication elongation is reduced by roughly 25%. The rate does not decrease by 50% because dATP only makes up roughly 35% of the summation over monomer types. The black line in the plot shows that for the initial dATP concentration and  $K_D$  used in the simulation are in the regime at which the DNA elongation rate is at its maximum, but is close to being reduced. The other two plots show the dependence on the replication elongation rate on the concentration of dATP and all dNTPs using the  $K_D$  from the simulations (0.001 mM). If all dNTP's drop below 0.1 mM or if dATP goes below 0.01 mM, the rate of replication elongation will decrease.

Transcription shows a similar trend looking at the  $K_D/[M]$  ratio for ATP (Figure S4I) with the transcription rate being reduced roughly 25% when the ATP concentration is equal to the  $K_D$ . The trend is similar because the A content of the mRNA is similar to that of the gene. The transcription rate is also plotted as a function of ATP and all NTP concentrations using the  $K_D$  from the simulations (0.1 mM) (Figure S4C and Figure S4D). These plots show that if all NTPs go below ~5 mM or ATP goes below 1 mM the transcription rate will decrease. Because there are 20 amino acids contributing to the composition of proteins, the translation rate is less sensitive to individual levels of charged tRNAs. This is reflected in the  $K_D/[M]$  ratio plot (Figure S4J) where the charged isoleucine tRNA was used as the example. When the concentration of charged lysine tRNA is equal to the  $K_D$ , the overall rate of transcription is reduced by less than 10%. Because charged tRNA counts are small, the translation rate is plotted as a function of their count rather than concentration, still using the  $K_D$  used in the simulation (Figure S4E and Figure S4F). These plots show that the transcription rate is already on the cusp of decreasing with 200 tRNA per isoform and won't start decreasing until fewer than 100 tRNA are charged for charged isoleucine tRNA.

## Metabolism

### Metabolic rates and initial parameterization

To simulate metabolic reactions, we use the rate form for a random binding model for enzymatic reactions

$$v_{\text{Random Binding}} = \frac{k_f[E] \prod_i^{\text{substrates}} ([i]/K_{M,i}) - k_r[E] \prod_j^{\text{products}} ([j]/K_{M,j})}{\prod_i^{\text{substrates}} (1 + [i]/K_{M,i}) + \prod_j^{\text{products}} (1 + [j]/K_{M,j}) - 1} \quad (\text{Equation 20})$$

In this rate form,  $[E]$  is the concentration of enzyme,  $i$  and  $j$  are the reactants and products, and  $K_{M,i} = [E_{-i}][i]/[E_{+i}]$ , where  $[E_{-i}]$  is the enzyme without substrate/product  $i$  bound and  $[E_{+i}]$  is the enzyme complexed with substrate/product  $i$  (Alberty, 2003; Liebermeister and Klipp, 2006; Liebermeister et al., 2010; Noor et al., 2016).

Initial estimates for every kinetic parameter and metabolite concentration in the metabolic reactions were determined using Parameter Balancing (Liebermeister et al., 2010; Lubitz and Liebermeister, 2019), a software that estimates parameters within biochemical reaction networks using Bayesian estimation and prior distributions derived from user-inputted surveys of kinetic data. Our survey of reported experimental measurements for forward and reverse catalytic rate constants ( $k_f$  and  $k_r$ ), equilibrium constants ( $K_{eq}$ ), and binding constants ( $K_{M,i}$ ) for each reaction were obtained from the online databases BRENDA (Chang et al., 2021) and TECRdb (NIST) (Goldberg et al., 2004). Parameters reported for mutants were excluded from the survey. When no published results for the  $K_{eq}$  were available from TECRdb, they were generated from the thermodynamic analysis program Equilibrator (Flamholz et al., 2012) at biologically relevant pH and temperatures. Values of intracellular small molecule concentrations were collected from experimental metabolomics studies in *E. coli* and *M. pneumoniae* (Park et al., 2016; Yus et al., 2009). This survey of experimental measurements serves as input for the Parameter Balancing software, which produces as output estimates of the kinetic parameters and metabolite concentrations. These estimates can additionally be made to satisfy thermodynamic constraints, namely the Haldane relationship between the kinetic ( $k_f$ ,  $k_r$ ,  $K_M$ ) and thermodynamic ( $K_{eq}$ ) quantities. To summarize briefly, the Bayesian estimation within the Parameter Balancing procedure uses prior normal or log-normal distributions with specified means and standard deviations. When possible, the means and standard deviations of the prior distributions are equal to the sample statistics calculated from the inputted survey of values from the literature and databases, but are set to default values (for example 0.1 mM for means of metabolite concentrations) in the absence of reported values. Even in the absence of reported values, the Haldane relationships still impose the thermodynamic constraint of the equilibrium constant  $K_{eq}$  on the predicted  $k_f$ ,  $k_r$ , and  $K_M$ .

One can go to [https://www.parameterbalancing.net/pb/static/css/css\\_template/documentation.html](https://www.parameterbalancing.net/pb/static/css/css_template/documentation.html) for further details on the default prior distributions. The equilibrium constants are not explicitly present in the random binding Michaelis-Menten like rate laws (Equation 20) that are utilized for ODE simulation of metabolic reactions in the final kinetic model, but they were accounted

for in generating the values for  $k_f$ ,  $k_r$ , and  $K_M$ . The Parameter Balancing procedure calculates standard deviations for each estimated parameter. These standard deviations are dependent on the inputted survey of experimental measurements (i.e., concentrations and kinetic parameters from the literature and databases) and inform the modeler of uncertainty in the estimated parameters. For the most part, the standard deviations of the estimated parameters are relatively small fractions of the mean values used as parameters in kinetic simulations.

After the initial estimation of the parameters and concentrations in the metabolic network, two methods were used to further improve the selection of parameters and initial concentrations: 1) time-averaged reaction fluxes in the kinetic model of metabolism were compared to steady-state fluxes predicted by flux balance analysis and 2) Gibbs free energies of the reactions in the kinetic model were compared to experimental measurements and free energies predicted on the basis of equilibrium thermodynamics.

### Comparison to steady-state flux balance analysis

The flux-balance analysis model (FBAm) by (Breuer et al., 2019) was modified to match the reaction network simulated in the kinetic model. Modifications include the updated lipid composition in the biomass expression based on the lipidomics experiments (Figure S2), removal of the “non-quantifiable” ATP cost, and any added/removed reactions in individual subsystems detailed below in Central Metabolism (Figure 2), Nucleotide Metabolism (Figure S1), and Lipid Metabolism (Figure S2). All other model details, such as uptake reaction flux constraints, were maintained from the model given in Breuer et al. (2019). FBAm solutions were evaluated using the publicly available program COBRApy (Ebrahim et al., 2013), with JCVI-syn3A biomass production, and therefore BIOMASS reaction flux, being the objective function. The doubling time obtained via the optimization process is approximately 96 min, closely matching the mostly commonly observed kinetic model doubling time of 97 min and the experimentally observed doubling time of 105 min. The final set of concentrations and kinetic parameters used to initialize the simulations are provided in Tables S1 and S2, respectively.

The fluxes from the FBAm are directly compared to the average CME-ODE fluxes over three periods of the cell cycle in Table S3, specifically 0 to 20 min when replication is initiating, 20 to 60 min during DNA replication elongation and volume growth, and 60 to 100 min once the volume has doubled and the cell is dividing, possibly with more DNA replication elongation events. While most of the fluxes are comparable, there are some significant differences between the steady-state fluxes and the ODE fluxes even after adjusting parameters, some due to the nature of the FBAm. The FBAm is designed to optimize production of “biomass” at the fastest rate possible, meaning production of the components necessary to double the cell as quickly as possible. Because of this optimization, several reactions end up unused as they would slow down or not contribute to biomass production, some by consuming ATP (for example through activated transport) that does not contribute directly to necessary biomass components. An example of a reaction that slows down biomass production is amino acid uptake. Amino acids can be taken up through either a proton symporter or active transport. While there are reactions to take up amino acids through an active transporter, the FBAm steady-state solution generates no flux through those uptake reactions, indicated by their fluxes being 0 mM/s in the FBAm in Table S3, because it can take up amino acids without using any ATP through the symporter reactions. Using only the symporters minimizes the ATP used for amino acid uptake, allowing the FBAm to use the ATP elsewhere to maximize the rate of production for other cell components. An example of reactions that don’t contribute directly to biomass production are the RNDR1 and RNDR2 reactions in nucleotide metabolism that convert ADP to dADP and GDP to dGDP, respectively. Because both dADP and dGDP can be made by directly taking up their deoxynucleoside precursors, it is not necessary to convert ADP or GDP to dADP or dGDP, hence the FBAm has no flux through the RNDR1 and RNDR2 reactions.

A reaction that does not directly use ATP, but can reduce the overall ATP formed is GAPDP in central metabolism. GAPDP converts g3p to 3pg through a redox reaction with NADP to NADPH. Although this reaction does not use ATP, any flux that passes through this reaction is flux that does not go through PGK to produce ATP because GAPDP bypasses PGK in the main glycolysis pathway. Due to this ATP opportunity cost, no flux is observed through GAPDP in the FBAm solution.

Another significant difference comes with the correlation between nucleoside uptake and acetate secretion. Because Syn3A lacks the pyruvate dehydrogenase E1 genes *pdhA* and *pdhB* and therefore its reaction, the only way for Syn3A to secrete acetate is by the connection to nucleotide metabolism through 2-deoxyribose-1-phosphate (2dr1p). Once a deoxyadenosine or deoxyguanosine is taken up, it can either be phosphorylated to its dNMP form or the deoxyribose can be removed to make the respective nucleobase through a PUNP reaction, giving off a 2dr1p molecule. The 2dr1p molecule is then converted to acetate through a series of reactions in central metabolism, converting one ADP to ATP in the process. Therefore, for each deoxyadenosine and deoxyguanosine converted to a nucleobase, the cell can net 1 ATP. This causes the FBAm to optimize to only take up the deoxynucleosides for the adenine and guanine pathways in nucleotide metabolism, because if the cell takes up the regular nucleoside, it costs one ATP to take up the molecule and no ATP are recovered whereas the ATP for transport is recovered if the cell takes up a deoxynucleoside and removes the deoxyribose. In Table S3 this is reflected by the ADNabc and GSNabc fluxes being 0 mM/s for the FBAm. This also causes the fluxes through the DADNabc and DGSNabc reactions to be higher in the FBAm than in the kinetic model. The lower fluxes through DADNabc and DGSNabc in the kinetic model then result in lower fluxes through the acetate secretion pathway, some of the reversible reactions even running the opposite direction among the population of cells once the fluxes are averaged over time.

The total combined uptake of nucleosides and deoxynucleosides is higher in the kinetic model than in the FBAm. The FBAm accounts for cellular generation of only one additional copy of the chromosome over the cell cycle in its biomass expression (Breuer et al., 2019). In reality, the DNA is only duplicated for part of the cell cycle, so the cell needs to take up deoxynucleosides quickly enough to make the first new chromosome after roughly 70 min (Figure 3). The FBAm assumes a uniform mRNA half-life of 1 min

among all mRNA (Breuer et al., 2019) and that most of the nucleotide costs for transcription are in making rRNA. In our kinetic model, there is a distribution of mRNA half-lives, resulting in some mRNA being more stable, having more counts on average, and therefore requiring more nucleotides. This results in the need for higher rates of nucleoside uptake to meet the demands of transcription.

### Comparison of free energy calculations

The Gibbs free energy change of any reaction in the system is given by

$$\Delta G = \Delta G_0 + RT \log Q, \quad (\text{Equation 21})$$

where  $Q$  is the reaction quotient

$$Q = \frac{\prod_i^{\text{products}} [j]}{\prod_j^{\text{substrates}} [j]} \quad (\text{Equation 22})$$

and  $\Delta G_0$  is the standard Gibbs free energy of the reaction

$$\Delta G_0 = \sum_i^{\text{products}} \nu_i \mu_i^0 - \sum_j^{\text{substrates}} \nu_j \mu_j^0, \quad (\text{Equation 23})$$

as calculated using the component-contribution method for each of the reacting species (Flamholz et al., 2012). Due to the relation between the Gibbs free energy change and the forward and backward fluxes,  $\Delta G = -RT \log(J^+/J^-)$ , the standard Gibbs free energy may alternatively be calculated from the rate law (Equation 20) and kinetic parameters as

$$\Delta G_0 = -RT \log \left[ \frac{k_f \prod_i^{\text{products}} K_{M,i}}{k_r \prod_j^{\text{substrates}} K_{M,j}} \right]. \quad (\text{Equation 24})$$

As discussed above, equilibrium constants for metabolic reactions were collected from Equilibrator and TECRdb. From these equilibrium constants,  $\Delta G_0$  was calculated.  $\Delta G_0$  was also calculated using Equation 24 and the final set of kinetic parameters ( $k_{cat}$ 's and  $K_M$ 's). Agreement between the standard Gibbs free energy,  $\Delta G_0$ , calculated from equilibrium thermodynamics (equilibrium constants) versus kinetics (rate law and kinetic parameters) served as a guiding criteria in the refinement of the final set of kinetic parameters. The time-averaged metabolite concentrations from the well-stirred model were used to calculate the reaction quotient,  $Q$ , of each reaction in the central metabolism. The free energy changes of the reactions,  $\Delta G$ , were then calculated using Equation 21 with the common  $Q$  and either the  $\Delta G_0$  from Equilibrator or the kinetic parameters. These two cases are plotted together in Figure S6A. For each reaction, the kinetic parameters were adjusted such that the differences in the free energies calculated from kinetic parameters and those generated by Equilibrator were minimized. While the free energies for most reactions are similar using the final model parameters, there are a few notable deviations. First, the fructose-1,6-bisphosphate aldolase (FBA) reaction has a deviation of roughly 20 kJ/mol due to a difference in the equilibrium constant from Equilibrator and the resulting equilibrium constant from the kinetic parameters. Although there is a possible set of parameters from BRENDA to bring the free energies into closer agreement, the parameters selected were required to make the reaction go in the forward direction at a rate matching the flux in the FBAm. This is the same case for the DRPA and TPI reactions. The transaldolase reaction (TALA) will be better parameterized when the gene catalyzing this reaction is annotated as it is currently subject to the default protein count value in the absence of a proteomic value related to a specific gene.

The main glycolysis pathway carries the greatest flux in the entire metabolic network of Syn3A. Others have discussed the cumulative free energy change throughout the main glycolysis pathway to indicate the strong favorability of the reactions (Flamholz et al., 2013; Park et al., 2016). The cumulative free energy change through the glycolysis pathway of our model of Syn3A is plotted in Figure S6B where the free energy of each reaction is summed proceeding in order down the glycolysis reactions. The plot shows an overall free energy change of roughly  $-80$  kJ/mol from the start to end of glycolysis. Other studies have reported overall free energy changes from  $-30$  to  $-70$  kJ/mol (Flamholz et al., 2013; Park et al., 2016), indicating that the kinetic parameters selected for the glycolysis reactions approximately reproduce the thermodynamics observed in other organisms. The main deviation here is again with the FBA reaction where the kinetic parameters result in a free energy drop of roughly 25 kJ/mol which has been reported to be less than 10 kJ/mol (Flamholz et al., 2013; Park et al., 2016).

Finally, to get a sense of reversibility the forward, reverse, and net reaction rates along the glycolysis pathway are plotted in Figure S6C. These rates were calculated by separating the forward and reverse rates in Equation 20 using time-averaged enzyme and metabolite concentrations from the well-stirred model. The larger the gap between the forward and reverse rate, the more irreversible the reaction is in our model. The PFK and PYK reactions were already annotated as irreversible, but the gaps in FBA and PGK also indicate that the forward direction is significantly more favorable in the model. This is due to the concentrations of the products of both reactions being significantly lower than the concentrations of the substrates, resulting in significantly slower reverse reaction rates than forward.



### Sensitivity analysis for external metabolite concentrations

Sensitivity analysis for external metabolite concentrations in the defined medium (Table S1) was also carried out. In doing this, we found that almost all the external metabolite concentrations (for example ions, phosphate, glycerol, cholesterol, etc.) used in this formulation were not growth limiting (when compared to the FBA fluxes required to achieve optimized biomass formation), except when the phosphate concentration was set too low ( $< 1$  mM). This phosphate growth medium adjustment was tested experimentally in an attempt to observe the efficacy of an alternative buffer composition, but this modification resulted in no significant cellular growth, confirming the kinetic model prediction. Other metabolites present in near growth limiting medium concentrations are spermidine (which is involved in DNA charge stabilization) and the amino acid aspartate (primarily used in protein translation). The program and the results of the sensitivity analysis are provided in the analysis programs for the well-stirred model (CME-ODE) on Github. The plot of the phosphate example is provided in Figure S7.

### Central metabolism

The kinetic model of central metabolism begins with a detailed mechanism of the glucose uptake phosphorelay as shown in Figure 2. The mechanism starts with the reversible phosphorylation of protein encoded by *ptsI*/0233 with a phosphate group from phosphoenolpyruvate (pep). This phosphate group is then passed to *ptsH*/0694, then to *crr*/0234, to membrane-bound *ptsG*/0779, and finally to the glucose taken up by *ptsG*/0779. The genes for each of the phosphorelay enzymes present in JCVI-syn3A are considered essential in the genome-wide transposon mutagenesis essentiality study and structures are known for each component in other bacteria (Breuer et al., 2019; Jeckelmann and Erni, 2019). Second-order rate constants for each of the reactions in the mechanism, including the reverse reactions, were measured using transient-state kinetic methods in work pioneered by Meadow and coworkers (Rohwer et al., 2000; Meadow et al., 2005a, 2005b). Their rate constants were implemented for each step of the phosphorelay mechanism except for the final reaction where glucose is transported through the membrane by phosphorylated PtsG and the phosphate is transferred from the *ptsG*/0779 to the glucose to form glucose-6-phosphate (g6p). For that reaction, a second-order rate constant was estimated from a glucose uptake rate measured using C13 labeling, a proteomics count of *ptsG*/0779, and a media concentration of external glucose all reported in a single study in *M. pneumoniae* (Yus et al., 2009). Experimental proteomics counts were used for each of the phosphorelay enzymes and the extracellular glucose concentration was fixed at the concentration in the growth medium (Table S1).

Before further testing of kinetic parameters, we removed the non-quantifiable ATP cost included in the steady-state flux balance analysis model by Breuer et al. (2019), accounting for approximately 41% of the ATP expenses in their model (Breuer et al., 2019). This cost was included assuming there are other significant ATP costs not accounted for in the metabolic network or gene expression and was estimated using a non-quantifiable ATP cost calculated in a model of *M. pneumoniae*. We construct our model for the metabolic reactions assuming the majority of the ATP costs are being accounted for through known metabolic reactions and genetic information processes.

When removing the non-quantifiable ATP cost from the network, we simultaneously removed the pyruvate dehydrogenase PDH\_E1 and PDH\_E2 reactions leading to the acetate secretion pathway. The genes coding for the E1 subunits of pyruvate dehydrogenase (*pdhA* and *pdhB*) were removed during genome reduction after being classified as unessential (Breuer et al., 2019). We removed the PDH\_E2 reaction because it requires the modifications made to the pyruvate dehydrogenase complex (PdhC) after the PDH\_E1 reaction, which has been genetically removed. The E2 (*pdhC*/0227) and E3 (*pdhD*/0228) components of pyruvate dehydrogenase still remain in the genome and are part of another acetate secretion pathway that begins with 2-deoxyribose-1-phosphate (2dr1p), a side product from nucleotide metabolism reactions. The E3 component is involved in reaction PDH\_acald that adds coenzyme A (coa) to acetaldehyde. The flux through this pathway is two orders of magnitude smaller than when PDH\_E1 was present, therefore significantly less ADP will be converted to ATP through this pathway and the secretion of acetate is much reduced. However, the amount of ATP conversion lost by removing this pathway is approximately the same flux as the non-quantifiable ATP cost in the steady-state flux model.

With the defined medium in which glucose is the only sugar present, we made modifications to the set of possible uptake reactions. The defined medium contains no mannose, glucosamine, or acetyl-D-mannosamine, so we removed the respective transport reactions for those metabolites and left only the main glucose transporter. The defined medium contains no mannose, glucosamine, or acetyl-D-mannosamine, so we removed the transport reactions for each of these metabolites. Without their uptake, we removed all pathways processing these sugars, which all lead to f6p generation (Breuer et al., 2019).

We also tried to remove the NAD oxidase reaction (NOX) since the gene coding for the enzyme was removed during genome reduction. However, NOX appears to be critical in maintaining the balance of NAD and NADH in Syn3A. GAPD and lactate dehydrogenase (LDH\_L) achieve redox/charge balance via NAD/NADH, but because we assume Syn3A can secrete pyruvate, not all of the flux through glycolysis will pass through LDH\_L going to lactate secretion. This leads to a slight imbalance between GAPD and LDH\_L and requires another reaction to convert NADH to NAD, so we leave the NOX reaction for this function, assuming one of the proteins of unknown function fulfills this task. The default protein count of 10 proteins is used to simulate the kinetics for NOX.

Finally, when comparing the fluxes to the steady-state FBA and the kinetic parameters to the known equilibrium constants, one adjustment was made in the protein counts along the main glycolysis pathway. The enzyme for the fructose-1,6-bisphosphate aldolase (FBA) reaction coded by gene *fbaA*/0131 has a lower protein count than any other protein along the main glycolysis pathway in the reported proteomics for Syn3A (227 proteins compared to roughly 400 for most other enzymes). The lower count made it such that the required parameters to have the reaction go in the forward direction resulted in an equilibrium constant that made the free

energy deviate by 50 kJ/mol from the known free energy calculated from the known equilibrium constant (see Figure S6). The comparative proteomics study (Table S1) was used to scale the count of the FBA enzyme to the same concentration as the FBA enzyme in *E. coli*, resulting in an initial FBA count of 775 proteins. The increased protein count allowed for the net flux per enzyme to be reduced while still maintaining the required flux through the FBA reaction, thereby enabling a change in the kinetic parameters to those with a lesser disparity in the forward versus reverse rates and a standard free energy with a reduced deviation of 20 kJ/mol.

### Amino acid metabolism

As Syn3A only contains salvage reactions for the cellular building blocks, it does not synthesize amino acids and requires them to be taken up by two transport mechanisms: AA permeases (0876, 0878, and 0886) and the ATP-dependent Opp (0165-0169). Breuer et al. (2019) assumed the Opp transporter to take up tetrapeptides, however the defined growth medium does not contain polypeptides, only individual amino acids. Given the contents of our defined medium, we assume that the Opp transporter can also take up amino acids and the ATP costs are proportional to the flux through this competing pathway. The AA permeases take up the amino acids, but with different affinities. Once amino acids are taken up in Syn3A, some are used in reactions in Nucleotide and Cofactor metabolisms, but the primary use of the amino acid is protein synthesis. For all amino acids except glutamine, we use the same mechanism for charging their respective tRNAs, even though it is well-known that there are two classes of aminoacyl-tRNA synthetases (for review see Ibbá et al. (2005)) that can function as either monomers or dimers. We first bind the tRNA synthetase with an ATP, then its respective amino acid, and finally its uncharged tRNA with the anticodon for the amino acid. Once all three substrates are associated, we simulate a conversion step where all products leave the synthetase; the tRNA is charged with the amino acid, and ATP is converted to AMP and PP<sub>i</sub>. Amino acid uptake is treated in the deterministic ODE kinetics, but we treat the charging mechanism stochastically since the reactions involve changing the states of two macromolecules with small counts: the tRNA synthetases and the tRNAs. The stochastic reaction scheme is shown in Figure S3.

In the mischarging mechanism for GLN tRNA, we first simulate the reactions described above charging a GLN tRNA with GLU using glutamyl-tRNA synthetase. The mischarged tRNA is then handled by the glutamyl-tRNA amidotransferase. The transferase is first bound with the mischarged tRNA, then ATP, and finally either a GLN or ASN amino acid. The complex then undergoes a conversion step giving off ADP, phosphate, GLU if GLN was bound, ASP if ASN was bound, and GLN-charged GLN tRNA. The amidotransferase had a lower affinity for the ASN according to parameters found in BRENDA (Chang et al., 2021).

### Nucleotide metabolism

The nucleotide metabolic network of Syn3A has previously been established as a series of salvage reactions (Breuer et al., 2019), which rely on the import of nucleotide precursors in the form of ribo- and deoxy-nucleosides. The nucleotide metabolism serves as the connection between the genetic information processing and the rest of the cellular metabolism by providing the necessary NTPs and dNTPs for use in the construction of the RNA and DNA. The reactions can be separated into two distinct reaction types: transport reactions (all reactions responsible for the uptake of the precursors) and salvage reactions (all non-transport reactions). A schematic diagram of the updated nucleotide metabolic network is presented in Figure S1.

Nucleotide transport is performed by the ribonucleoside ATP binding cassette (*rnsABC*) transporter protein, which contains four subunits: two permease domains (*rnsD*/0008 and *rnsC*/0009), an ATP binding domain (*rnsA*/0010), and a substrate-binding domain (*rnsB*/0011). The *rnsABC* system takes up all ribo- and deoxy-nucleosides needed by the cell. Breuer et al. (2019) postulated that Syn3A could uptake nucleobases because the parent organism has this ability, however no gene has been assigned to the function nor has the uptake been confirmed via experiment. Therefore, uptake of nucleobases is not simulated in the present kinetic model. Based on growth medium components (Table S1), uptake of cytidine, uridine, adenosine, guanosine, 2-deoxycytidine, 2-deoxyadenosine, 2-deoxyguanosine, and thymidine is possible by the cell. However, cytidine is not used by the cell (see Figure S1), so its uptake is not simulated. Nucleotide transport reactions were modeled using the random binding model with parameters from the literature. Webb and Hsieh (2006) characterized the *rnsABC* system of *Streptococcus mutans* and measured the binding constants for cytidine, uridine, and adenosine. Since only the ribonucleosides were measured, the values for the deoxy-forms were assumed to be the same. In the case of guanine nucleoside derivatives the  $K_m$  for adenosine was used, and similarly the uridine  $K_m$  was also used for thymidine. All kinetic parameters used to simulate the transport of nucleotides into the cell can be found in Table S2.

Pyruvate kinase (*pyk*/0221) and phosphoglycerate kinase (*pgk*/0606) share similar enzymatic function in Syn3A. Their primary function is to generate energy (ATP) for the cell via central metabolism using ADP, but the additional activity of these enzymes provides formation of other NTP/dNTP from NDP/dNDP. In many bacteria this process is performed by nucleoside diphosphate kinases (*ndk*), however Mycoplasmas and Syn3A do not have *ndk* (Pollack et al., 2002). Parameters derived from activities measured in Mycoplasmas by Pollack et al. (2002) were used to develop the kinetic parameters. The activity relative to ADP for the two enzymes has been measured under saturated substrate conditions in several Mycoplasmas (Pollack et al., 2002). Using the parameters for ADP as the substrate, the catalytic rate constants for the PGK and PYK reactions were modified using the relative activities for the other substrates as measured by Pollack et al. (2002) and the affinity constants for all other substrates were assumed to be equal to that of ADP for the corresponding enzyme.

CTP synthase (*pyrG*/0129) was found in Syn3A forming CTP from UTP (reaction CTPS2) using an amino donor glutamine catalyzed by ATP Breuer et al. (2019). Additional activity for CTP synthase converting dUMP to dCMP (reaction CTPSDUMP) was predicted for Syn3A resulting from the fact that dUMP was a dead-end in the network and this secondary activity being the most probable solution (Breuer et al., 2019). In other organisms such as *S. cerevisiae* and *Lactococcus lactis*, their respective CTP synthases have activity toward dUTP Pappas et al. (1999); Willemoës and Sigurskjöld (2002). Validation of this secondary activity has yet to be confirmed in

Syn3A, however broader substrate specificity is common in *Mycoplasma* (Breuer et al., 2019; Pollack et al., 2002). Parameters for the CTPS2 reaction were curated from the literature pipeline with a minor adjustment in the product catalytic rate constant for better agreement with the FBA steady state fluxes reported in (Breuer et al., 2019). In the case of the secondary functionality, CTPSDUMP parameters were used from CTPS2 for analogous metabolites due to a lack of parameters in the literature.

Phosphatase activity has been observed experimentally in *Mycoplasmas* and was predicted to be in the nucleotide metabolism of Syn3A (Breuer et al., 2019). This activity was proposed for the deoxymononucleotides (dAMP, dGMP, dUMP, and dTMP), however no gene was assigned for these reactions. Recently, using mass spectrometry and single gene deletion studies of Syn3A on wild-type and mutant Syn3A, a gene was identified responsible for the hydrolase activity against the mononucleotides (Haas et al., 2021). Gene 0066 was identified as dUMP phosphatase primarily acting on dUMP and dTMP, which agrees with previous experimental results from the literature (Neale et al., 1983). The dUMPase was also observed to act on dGMP and dAMP albeit less than dUMP, which is agreement with experimental data from other *Mycoplasma* (Neale et al., 1983). As a result reactions NTD1, NTD5, NTD6, and NTD8 have been assigned to be catalyzed by the protein coded by gene 0066 and simulated in the kinetic model. Parameters were gathered and tested using the process outlined above (Table S2). Although other hydrolase activity is predicted, no other hydrolase activity is accounted for in the kinetic model due to a lack of gene assignment(s).

Syn3A relies on purine-nucleoside phosphorylase (*punA*/0747) activity for conversion of ribo- and deoxy-nucleosides to their respective nucleobase form (Breuer et al., 2019). Previously the PunA activity was only expected toward adenine and guanine nucleosides (Breuer et al., 2019). With no uracil transport (due to no gene with the function) or deoxyuridine transport (due to no deoxyuridine in the external media), the cell can only bring in uracil nucleotide derivatives via uridine transport. McElwain and Pollack (1987) provided evidence of *punA*/0747 activity toward uridine in related *Mycoplasma*. Parameters for the uridine to uracil formation catalyzed via *punA*/0747 (reaction PUNP5) were obtained and implemented in the kinetic model as described above.

### Lipid metabolism

Typically there are three main processes by which *Mycoplasmas* can build lipid pools: direct incorporation of lipids, lipase activity, and lipid synthesis (Figure S2) (Gaspari et al., 2020). However, in Syn3A all identified lipases were removed in the genome reduction from *Mycoplasma mycoides subsp. capri* str GM12 to JCVI-syn1.0 and then to Syn3A. For this reason, incorporation of lipids such as phosphatidylcholine (PC) or sphingomyelin (SM) and cholesterol (Chol) from the media remains the key mechanism for rapid alteration of cellular lipid biomass as shown by the lipidomics analysis from the Sáenz laboratory in Figure S2B. A phospholipid synthesis pathway from glycerol and fatty acids (palmitate and oleate) taken up from the media passively, to produce phosphatidylglycerol (PG) and cardiolipin (CLPN) remains in place. Inspiration for the design of the main pathway (Breuer et al., 2019) from fatty acids to phosphatidylglycerol was available from a previous attempt to create a lipid synthesis model for a minimal cell (Castellanos et al., 2007) as well as other mycoplasma lipid synthesis pathways (Blötz and Stülke, 2017). In addition, synthesis of glycolipids via the sugar moieties from the UDP-glucose pathway (connected to the Central Metabolism by reaction PGMT) and with the lipid intermediate diacylglycerol to produce galactosyl diacylglycerol (GAL-DAG/galfur12 gdr) is present. It was assumed that *pgpB*/0214 could catalyze the PAPA reaction linking the phospholipid and glycolipid synthesis paths via phosphatidate phosphatase side activity in addition to its role in synthesis of phosphatidylglycerol. This phosphatase side-activity by PgpB was speculated in Breuer et al. (2019) and observed in *E. coli* by Dillon et al. (1996).

The model was parametrized via a similar procedure to that described for the other metabolic modules by inputting data from Equilibrator (Flamholz et al., 2012), BRENDA (Chang et al., 2021), and a manual search of the SABIO-RK kinetic database (Wittig et al., 2012) (for a few parameters, such as the forward catalytic rate constant for the AGPAT reaction) and Parameter Balancing (Lubitz et al., 2010; Lubitz and Liebermeister, 2019) in Metabolic Rates and Initial Parameterization. Since the majority of the enzymes involved in lipid synthesis are membrane embedded or peripheral membrane proteins, these were more likely to suffer from poor proteomics coverage, as was previously discussed, so adjustments were made to obtain agreement to FBA fluxes using values selected from the comparison to other organisms (Table S1).

Important to note are the lipid related processes that have not been modeled in this work. While lipids in the model are incorporated into the membrane directly (PC, SM and Chol) or synthesized at the membrane (PG, CL, etc.) we do not include the activity of flippases (*ywjA*/0371 and *ywjA*/0372) (Quentin et al., 1999) that transition lipids between the inner and outer leaflet of the membrane. Without a detailed composition of the inner and outer leaflet relative to each other for each of the modeled lipid classes and with the expectation that the energy cost of this process is likely small compared to more costly gene expression processes, flippase activity is not included in the present model. In addition, we do not include potential activity of phospholipase A1 or A2 which allow for acyl chain scavenging from various phospholipid species because these enzymes have not yet been genetically identified in Syn3A.

Crucially, due to the completion of further experimental studies suggested in Breuer et al. (2019) we have increased our knowledge about the lipid composition of Syn3A and have thus updated the lipid biomass and metabolic pathways from those previously presented. In addition to a membrane glycolipid component, Breuer et al. (2019) also identified the production of a lipid bound capsular polysaccharide (CPS, hereafter referred to as capsule) as a component of the Syn3A membrane architecture. *M. mycoides*, the parent organism of Syn3A produces a galactofuranose based capsule (Schieck et al., 2016). However, upon studying both JCVI-Syn1.0 and Syn3A, by Gas Chromatography/Mass Spectrometry (GC/MS) it appears that JCVI-syn1.0 produces a capsule, while Syn3A does not (K. Dybvig, personal communication). While leading to a simplification of the UDP-Glucose pathway to be responsible only for glycolipid production and not for production of CPS, this information does bring into question the function of gene 0113 which was previously speculated to be an *epsG*-like capsule producing polymerase (Breuer et al., 2019). This gene could still be

active in the process of protein glycosylation, an important process in Mycoplasmas (Jordan et al., 2013; Daubenspeck et al., 2015), but this is yet to be determined.

Beyond demonstrating the absence of a capsule in Syn3A we conducted lipidomics studies to gain an increased understanding of the lipid makeup of the cell (Sáenz Group TU-Dresden). The updated lipid species molar percentages are presented in Figure S2. A crucial caveat to this data is that due to the heavy dependence on lipid incorporation for Syn3A that the lipidomic makeup observed is likely substantially dependent upon the growth medium that is used. However, this is the clearest picture of Syn3A lipid makeup to date, and it is therefore used to parameterize the lipid biomass and growth model used in our simulations.

Interestingly, cholesterol which is typically absent from many bacterial membranes, but is present in those of Mycoplasma (Razin and Tully, 1970) in order to help maintain membrane integrity and fluidity, is present in a relatively higher proportion in Syn3A compared to its parent organism *M. mycoides*. While not previously predicted in the Lipid Biomass given in Breuer et al. (2019) uptake of SM and PC by Mycoplasmas has been previously shown by Kornspan and Rottem (2012) and was confirmed by lipidomic analysis for Syn3A. If we consider the difference between the SP4-FBS serum that was used to grow the cells analyzed in Figure S2B and the SP4-KO media typically used to grow Syn3A cells, it seems reasonable that this amount of SM and cholesterol are present. In addition, cholesterol is highly correlated with high sphingomyelin concentration and lower PC concentration, since sphingolipids attract cholesterol while PC repels it (Clejan and Bittman, 1984; Lönnfors et al., 2011; Gaspari et al., 2020; Kornspan and Rottem, 2012). Uptake rates for cholesterol, PC, and SM are fitted to approximately obtain (at cell doubling) the measured lipidomic biomass mole fractions observed in the lipidomics studies (Figure S2B). The relative ratio of the main phospholipids (CL and PG) observed increased relative to what had been observed in earlier studies of Mycoplasma (Plackett, 1967) and reported in Breuer et al. (2019).

The molar percentages of each lipid species obtained via lipidomic analysis were then used to rescale the lipid biomass of Syn3A from the values reported previously, by assuming that the cellular dry weight fraction of non-capsule lipid compounds to be the same fraction given in Breuer et al. (2019). This mass fraction of approximately 7.5% or 0.8 fg was then divided between the lipid species observed in the lipidomic analysis according to their molar percentages in Figure S2. The additional assumption of a constant glycolipid component of the lipid dry weight fraction (which was subtracted out from the total lipid biomass) was maintained. The notion that the glycolipid biomass fraction stays approximately constant for the growth medium and membrane has been previously observed in *M. pneumoniae* (Gaspari et al., 2020) and agrees with our findings. A flux balance analysis study of the metabolic model containing the rescaled biomass and updated metabolic pathways given in this work gave a growth rate of 97 min which falls within the range of simulated doubling times of this study (88–112 min) and is close to the experimentally measured doubling time of 105 min (Breuer et al., 2019).

With the experimentally refined metabolic pathways and lipid biomass in place we developed a simulated cell growth model based on synthesis and incorporation of each particle making up the membrane. Lipids and membrane proteins, each with a specific surface area contribution, are dynamically added to the membrane via synthesis, incorporation, and translocation over the duration of the cell cycle. Surface area contributions per lipid species were obtained from monolayer experiments (Sáenz et al., 2012) and molecular dynamics (MD) studies (Jo et al., 2009) (Table S2). Cell growth occurs via a process shown in Figure 3D where the cell initially experiences simultaneous membrane surface area and cell volume growth until the cell volume is doubled. Translocation of integral membrane proteins was simulated by a process in which each protein is translocated at the cost of 1 ATP per 10 amino acids as was observed in Catipovic and Rapoport (2020). After volume doubling, additional membrane surface area is produced to pinch off and separate the two spherical Syn3A daughter cells of volume equal to the initial simulation conditions. Such a constant increase of cell volume per generation follows the “sizer principle” that has been observed experimentally for slow-growing Gram-negative and Gram-positive bacterial cells (Wallden et al., 2016; Facchetti et al., 2017).

## Lipidomics

### Lipid extraction for mass spectrometry lipidomics

Mass spectrometry-based lipid analysis was performed by Lipotype GmbH (Dresden, Germany) as described by Sampaio et al. (2011). Lipids were extracted using a two-step chloroform/methanol procedure (Ejsing et al., 2009). Samples were spiked with internal lipid standard mixture containing: cardiolipin 14:0/14:0/14:0/14:0 (CL), ceramide 18:1/2/17:0 (Cer), diacylglycerol 17:0/17:0 (DAG), hexosylceramide 18:1/2/12:0 (HexCer), lyso-phosphatidate 17:0 (LPA), lyso-phosphatidylcholine 12:0 (LPC), lyso-phosphatidylethanolamine 17:1 (LPE), lyso-phosphatidylglycerol 17:1 (LPG), lyso-phosphatidylinositol 17:1 (LPI), lyso-phosphatidylserine 17:1 (LPS), phosphatidate 17:0/17:0 (PA), phosphatidylcholine 17:0/17:0 (PC), phosphatidylethanolamine 17:0/17:0 (PE), phosphatidylglycerol 17:0/17:0 (PG), phosphatidylinositol 16:0/16:0 (PI), phosphatidylserine 17:0/17:0 (PS), cholesterol ester 20:0 (CE), sphingomyelin 18:1/2/12:0/0 (SM), triacylglycerol 17:0/17:0/17:0 (TAG) and cholesterol D6 (Chol). After extraction, the organic phase was transferred to an infusion plate and dried in a speed vacuum concentrator. 1st step dry extract was re-suspended in 7.5 mM ammonium acetate in chloroform/methanol/propanol (1:2:4, V:V:V) and 2nd step dry extract in 33% ethanol solution of methylamine in chloroform/methanol (0.003:5:1; V:V:V). All liquid handling steps were performed using Hamilton Robotics STARlet robotic platform with the Anti Droplet Control feature for organic solvents pipetting.

### MS data acquisition

Samples were analyzed by direct infusion on a QExactive mass spectrometer (Thermo Scientific) equipped with a TriVersa NanoMate ion source (Advion Biosciences). Samples were analyzed in both positive and negative ion modes with a resolution of  $Rm/z = 200 = 280000$  for MS and  $Rm/z = 200 = 17500$  for MSMS experiments, in a single acquisition. MSMS was triggered by an inclusion list en-



compassing corresponding MS mass ranges scanned in 1 Da increments (Surma et al., 2015). Both MS and MSMS data were combined to monitor CE, DAG and TAG ions as ammonium adducts; PC, PC O<sup>-</sup>, as acetate adducts; and CL, PA, PE, PE O<sup>-</sup>, PG, PI and PS as deprotonated anions. MS only was used to monitor LPA, LPE, LPE O<sup>-</sup>, LPI and LPS as deprotonated anions; Cer, HexCer, SM, LPC and LPC O<sup>-</sup> as acetate adducts and cholesterol as ammonium adduct of an acetylated derivative (Liebisch et al., 2006).

### Quantitative PCR protocol

The relative amounts of the different regions of the genome present in syn3A cells were measured by quantitative PCR (qPCR). PCR primer pairs were designed that amplify at the origin, at the terminus and halfway between the origin and the terminus of the Syn3A genome. That last point is hereafter called the quarter point. The sequence for the approximate Origin of replication qPCR 106 bp amplicon comprising Syn3A basepairs 354–459, which is in the gene *dnaA/0001*, is **GCATTAGGCATTGTTGGCATAAATCCAGCACGAACGATGT** with reverse primer TCCCATTCAGATTCACCATAAA. The sequence for the approximate Quarter way around the genome 137 bp amplicon comprising Syn3A basepairs 138,342–138,478, which is in the gene *pdhC/0227*, is **GCTGACATAGGTGAAGGTCTAACAGAAGGAACAGTCGCTGAAGTTTTAGTTAAAGTTGGTGATGTTGTTAAAGAAGGACAATCATTATACTTTGTTGAACTGATAAAGTAAACAGTGAAATACCTGCTCCAGTGGC** with reverse primer GCCACTGGAGCAGGTATTT. The sequence of the approximate Terminus of replication qPCR 39 bp amplicon comprising Syn3A basepairs 271,774–271,783, which is in the gene *plsX/0419*, is **GCATTAGGCATTGTTGGCATAAATCCAGCACGAACGATGT** with reverse primer ACATCGTTCGTGCTGGATTTA. The bolded sequences are the forward primer locations and the underlined sequences the reverse primer locations. Note that in the Terminus amplicon, the primers overlap by 3 bp. Primers were designed using The Integrated DNA Technologies PrimerQuest Tool (<https://www.idtdna.com/PrimerQuest/Home/Index>) set for a maximum amplicon length of 150 base pairs.

To enable normalization of the three different qPCRs, we designed and had Integrated DNA Technologies synthesize a 319 base pair synthetic DNA molecule that contained one copy of each of the three qPCR amplicons. This qPCR standard molecule contains each of the 3 amplicons plus 3 bases at the ends of each amplicon and either 10 or 14 bases at the ends of the molecule:

AAAATTTTGT**AATCGGTGCAAGTAATGAACAAGCTTTATAGCAGTTCAAACAGTAAGTAAAAATCCTGGGATTTCCTATAATCCA**  
**TTGTTATTTATGGTGAATCTGGAATGGGAAAATTAGCATTAGGCATTGTTGGCATAAATCCAGCACGAACGATGTTATTTGCTG**  
**ACATAGGTGAAGGTCTAACAGAAGGAACAGTCGCTGAAGTTTTAGTTAAAGTTGGTGATGTTGTTAAAGAAGGACAATCATTATA**  
**CTTTGTTGAAACTGATAAAGTAAACAGTGAAATACCTGCTCCAGTGGCTGGAAAAATTGCAG**

In the molecule, all qPCR amplicon sequences are italicized, the forward primers are bolded, and the reverse primers are underlined. The inter-amplicon and flanking bases are not italicized. To prepare the qPCR standard molecule for use in qPCRs, the molecule was amplified using the primers AAAATTTTGT**AATCGGTGC** as the standard forward and CTGCAATTTTCCAGCCAC as the standard reverse. The amplicon were electrophoresed on agarose gels and the correct size band was excised and purified using a QIAquick Gel Extraction Kit according to the manufacturer's instructions. All qPCR sequences and their locations on the genome are in Table S4.

To prepare Syn3A cells for qPCR analysis, a 100  $\mu$ L aliquot of cells grown in SP4 media (Williamson and Whitcomb, 1975) supplemented with 17% KnockOut Serum Replacement (hereafter called SP4KO) was mixed with 900  $\mu$ L of fresh SP4KO. Multiple two fold dilutions of this cell suspension were then made by serially transferring 500  $\mu$ L aliquots to tubes containing 500  $\mu$ L of fresh SP4KO. A total of 15 dilutions were made and these tubes were capped and incubated at 37° C overnight. The stage of growth was determined by the color of the phenol red pH indicator, which is a component of SP4KO. As the cells metabolize glucose, they release acid, which causes the normally red SP4KO to turn orange and then yellow. The next day the highest dilution tube that was yellow was selected as the stationary phase culture. The most concentrated tube that had not begun transitioning from red to orange was designated exponential phase culture.

To prepare the Syn3A genomic DNA for qPCR analysis, 10  $\mu$ L of the exponential and stationary phase cultures were diluted 1:100 in water, incubated at 98° C for 10 min in a thermocycler with a heated lid and immediately added to the qPCRs. To prepare qPCR control DNA templates, a set of 6 two-fold dilutions was made using the gel purified DNA. The qPCRs were set up as follows: 5  $\mu$ L PowerUp SYBR Green Master Mix (Applied Biosystems), 2  $\mu$ L DNA template, 0.55  $\mu$ L 10  $\mu$ M forward primer, 0.55  $\mu$ L 10  $\mu$ M reverse primer, 1.9  $\mu$ L H<sub>2</sub>O for a total reaction volume of 10  $\mu$ L.

Six replicates were made for each sample. Reactions were run in a QuantStudio 6 (Applied Biosystems) qPCR machine with the following program: Hotstart polymerase activation 50° C 120 s, ramp 2.05° C/s, 95° C 120 s; amplification 45 cycles of ramp 2.05° C/s, 95° C 15 s, ramp –1.71° C/s, 52.9° C 15 s, ramp 2.05° C/s, 72° C 60 s; melting curve 95° C 15 s, ramp –1.71° C/s, 52.9° C 60 s, ramp 0.05° C/s up to 95° C. The Threshold Cycle (Ct) values were calculated automatically using the QuantStudio Software (Applied Biosystems) with the default parameters. Amplification plots and standard curves generated for each of the three regions of the genome are shown in Figure S5. Note that our standard curves are unitless. We did not quantify the amount of qPCR standard in the gel purified material. We simply used the qPCR standard to normalize the amounts of standards in the origin, quarter, and terminus reactions.

The amounts of origin, quarter, and terminus were each averaged among the six replicates separately for the exponential and stationary phase samples. Next, we set the average amount of exponential and stationary phase DNA in the terminus reactions to be one, and then determined the ratios of the origin and quarter samples to the terminus samples. The standard deviations were calculated as the ratio of the standard deviations for origin and quarter samples to the average of the terminus. The qPCR data is available on Mendeley Data (<https://doi.org/10.17632/nprw2h5tx6.1>).

## QUANTIFICATION AND STATISTICAL ANALYSIS

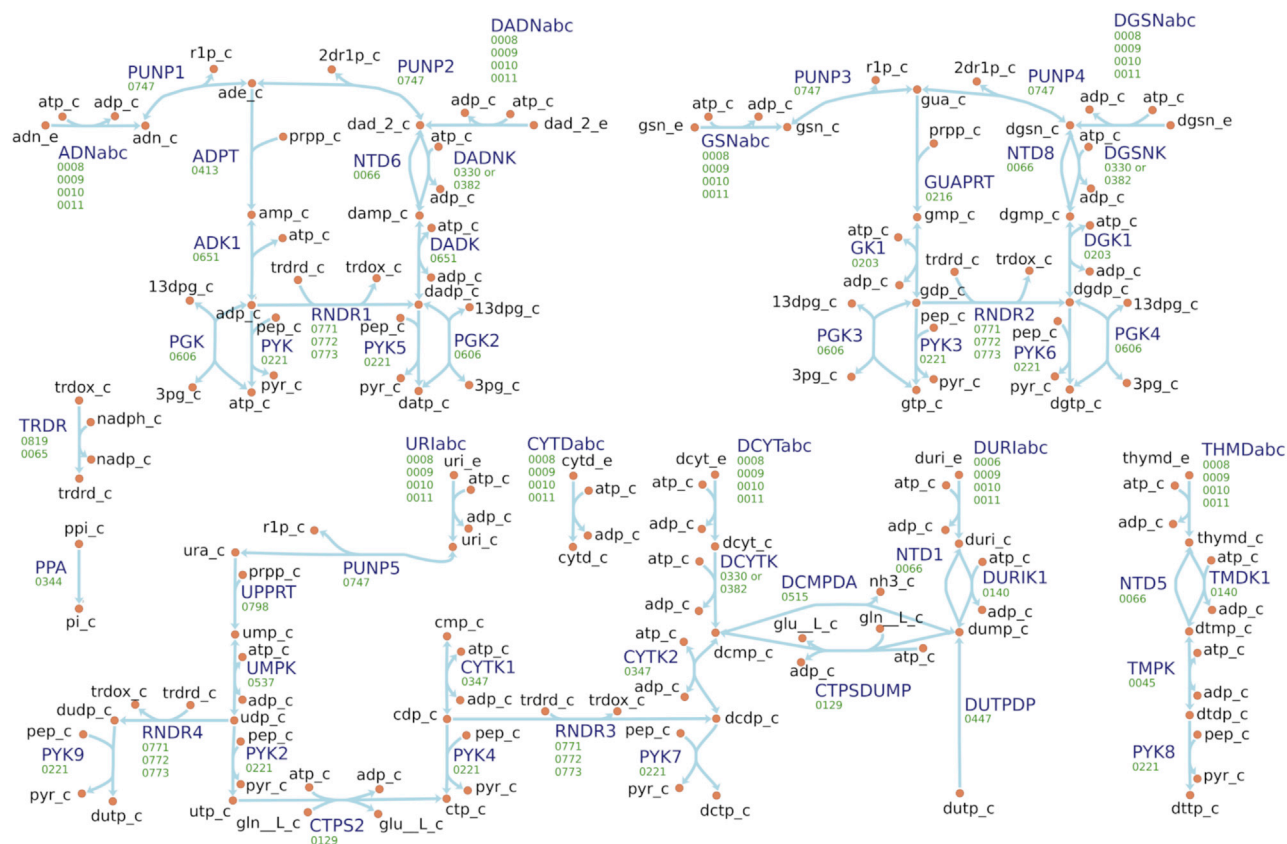
### Analysis of cell simulations

The jupyter Python notebooks used to analyze cell simulations are included with the programs in the Minimal Cell Github repository listed in the Key Resources Table ([https://github.com/Luthey-Schulten-Lab/Minimal\\_Cell](https://github.com/Luthey-Schulten-Lab/Minimal_Cell)).

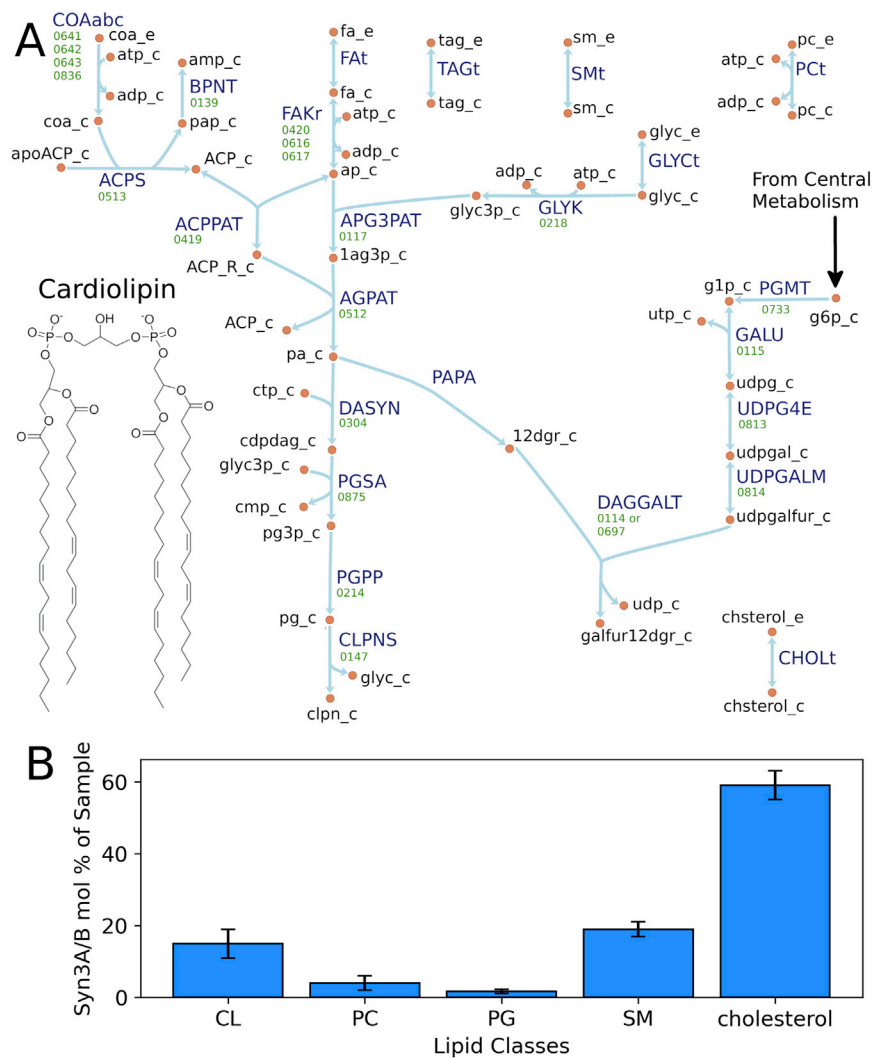
### Lipidomics analysis and post-processing

Data were analyzed with Lipotype GmbH developed lipid identification software based on LipidXplorer (Herzog et al. (2011), Herzog et al. (2012)). Data post-processing and normalization were performed using an in-house developed data management system. Only lipid identifications with a signal-to-noise ratio > 5, and a signal intensity 5-fold higher than in corresponding blank samples were considered for further data analysis. For data visualization, filtering threshold of 0.2 mol% was applied. Only those lipid species that are present in 3 biological replicates were considered. For class distribution plot (Figure S2), lipid species were summarized to lipid class and classes plotted as averages across 3 biological replicates. Cholesterol ester (CE) was excluded from the lipidome, as it was shown that it is not a membrane lipid.

# Supplemental figures



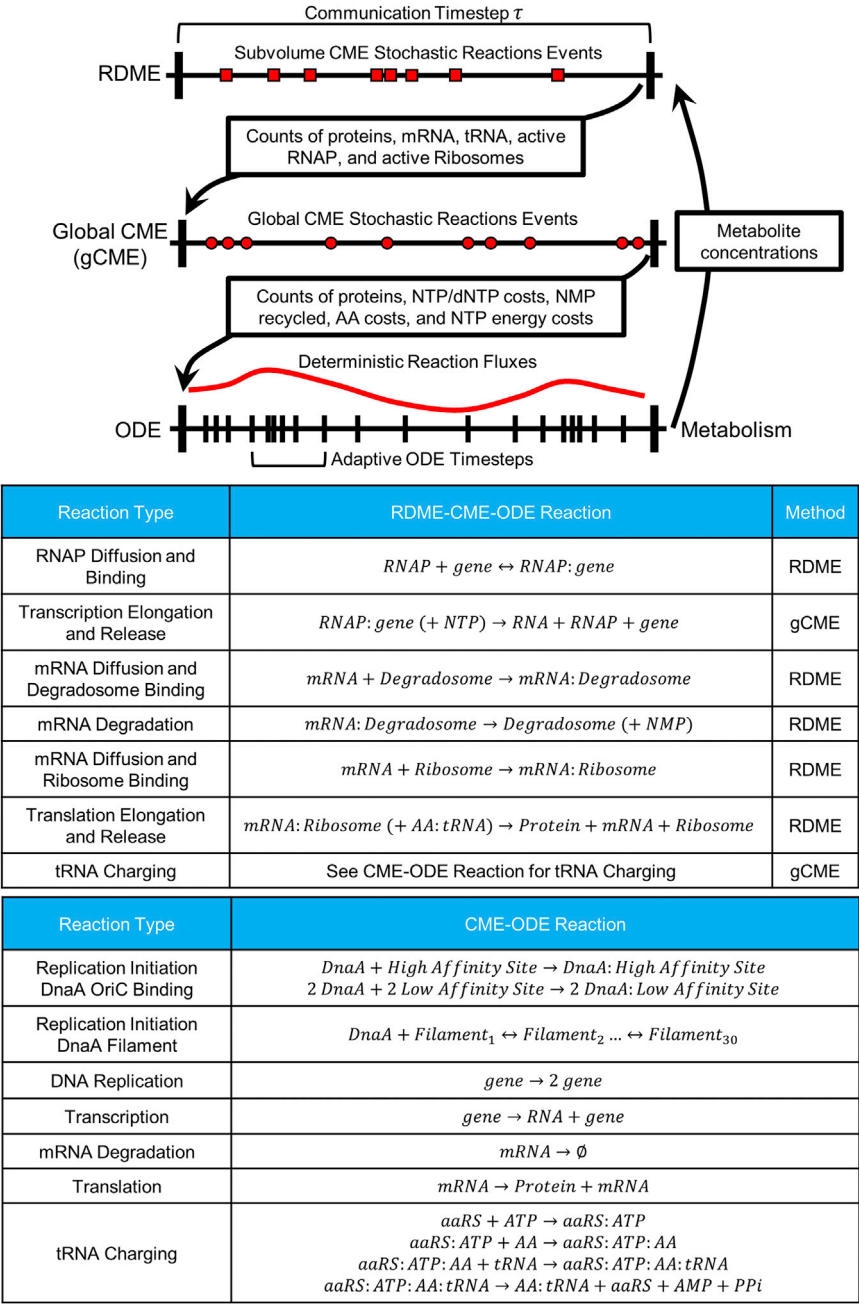
**Figure S1. Nucleotide metabolism reaction network, related to Figure 2**  
Updated reaction network of nucleotide metabolism in Syn3A.



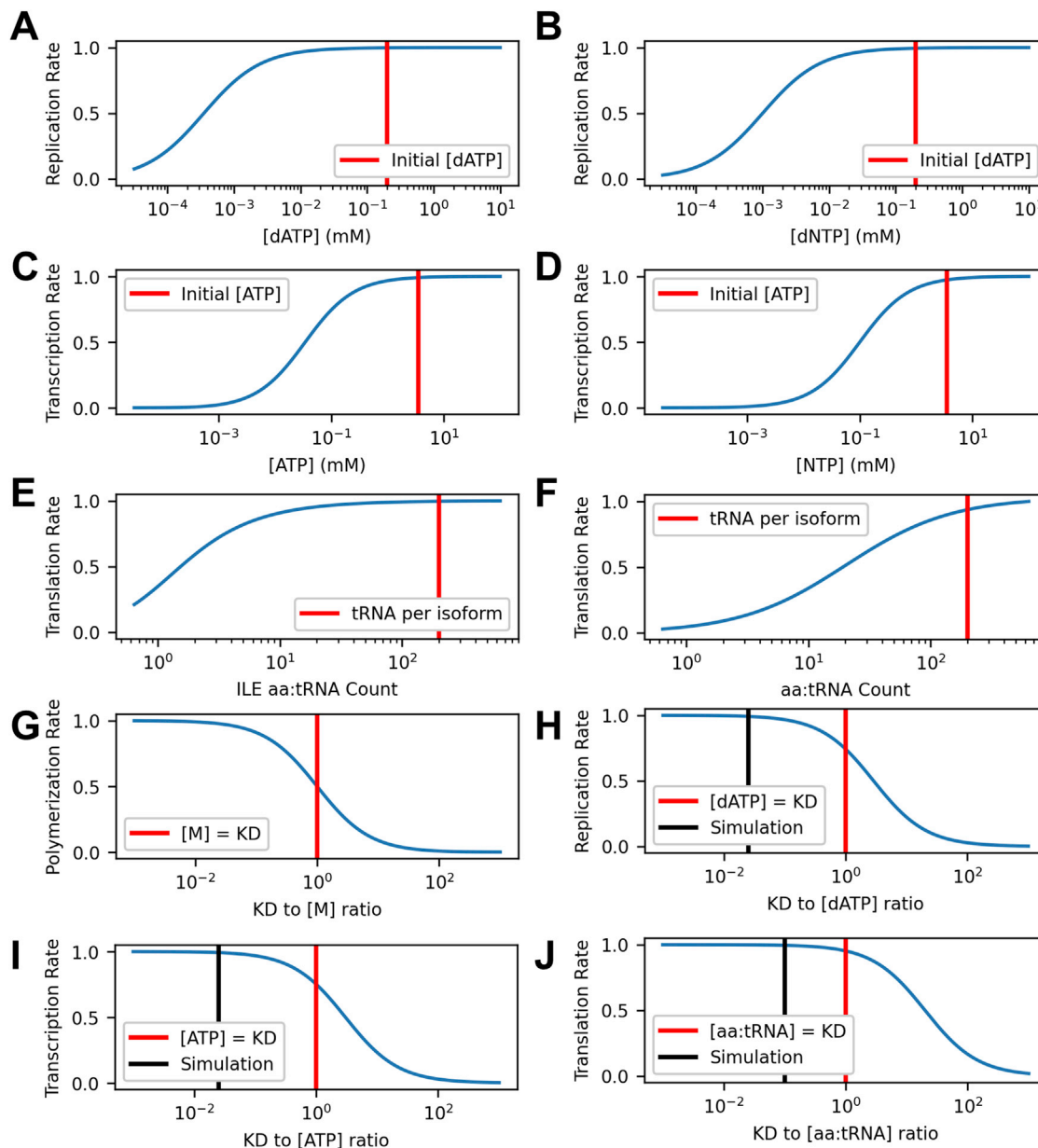
**Figure S2. Lipid metabolism reaction network and lipidomics data, related to Figure 2**

(A) Updated reaction network of lipid metabolism in Syn3A. The pathway leading to capsule formation has been removed and reactions for uptake of sphingomyelin (SM) and phosphatidylcholine (PC) have been added. (B) Lipidomics study of the lipid composition of JCVI-syn3B (Syn3A with a synthetic DNA landing pad present, but genetically identical to Syn3A) from mass spectrometry. Error bars represent the standard deviations of three biological replicates.



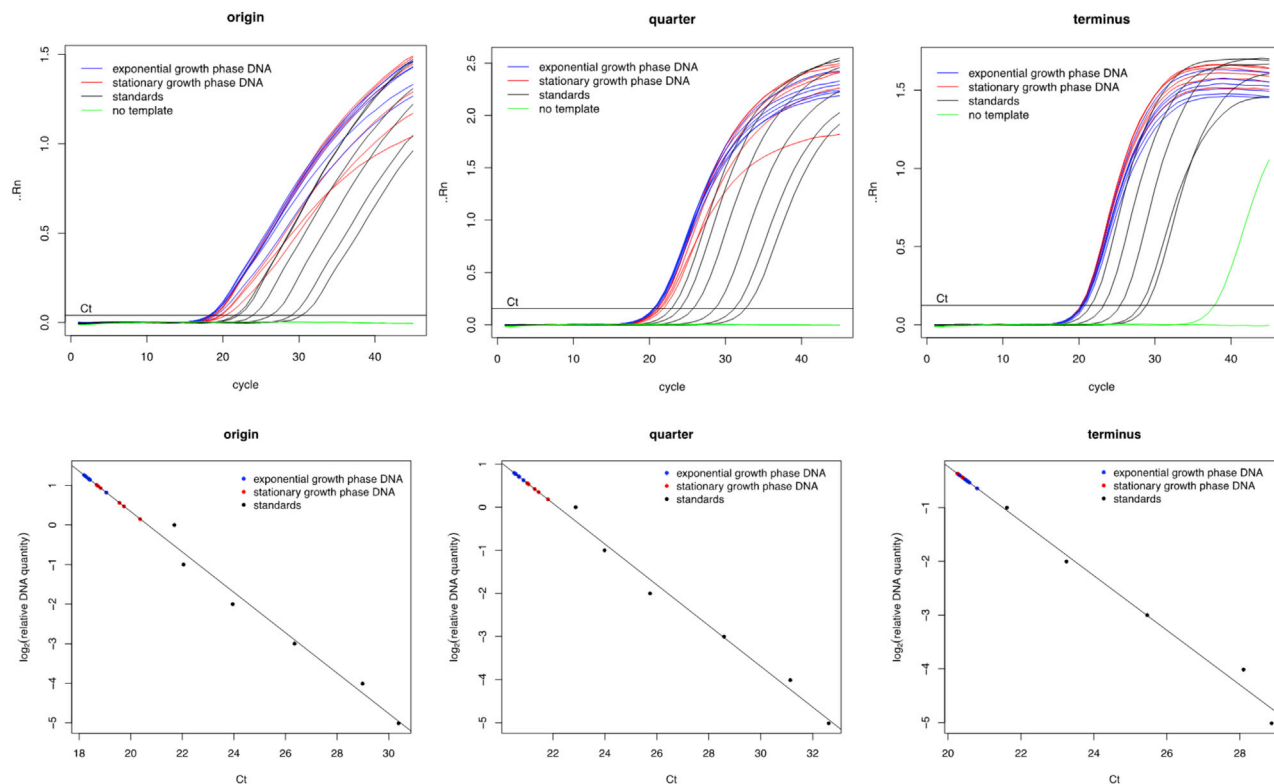


**Figure S3. Pictorial RDME-gCME-ODE algorithm and stochastic reaction table, related to STAR Methods**  
Pictorial algorithm for the hybrid RDME-gCME-ODE simulations and formulations of the stochastic reactions. The first table designates the handling of each reaction type in the spatially-resolved simulations. The second table gives the reactions used in the totally well-stirred simulations.



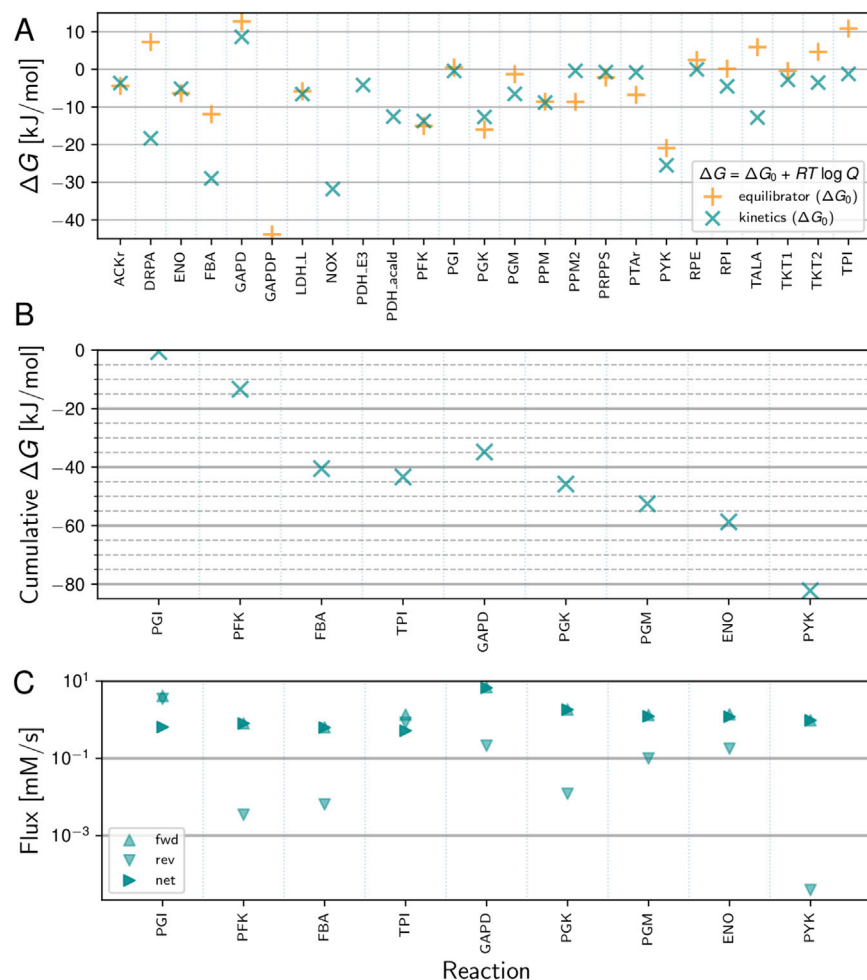
**Figure S4. Sensitivity of genetic information processing reactions to metabolite concentrations, related to Figure 4**

Each rate is scaled to their maximum rate for an average gene/mRNA/protein. Replication elongation rate as a function of dATP concentration (A) and of all dNTPs (B). Transcription elongation rate as a function of ATP concentration (C) and of all NTPs (D). Translation elongation rate as a function of charged (aa:tRNA) tRNA counts for isoleucine (ILE) (E) and of all aa:tRNAs (F). (G-J) The polymerization reactions are sensitive to the ratio of their  $K_D$  to monomer concentrations. (G)  $K_D$  to monomer concentration ratio plot for the general polymerization rate. When the monomer concentration is equal to the  $K_D$ , the overall rate is halved. (H-J)  $K_D$  to monomer concentration ratio plots varying the dATP, ATP, and ILE aa:tRNA ratios for DNA replication elongation, transcription, and translation respectively for DnaA. Black vertical lines (Simulation) indicate the ratio of the  $K_D$  used in the simulations to the initial monomer concentrations.



**Figure S5. qPCR amplification and standard curve plots, related to Figure 3D and STAR Methods**

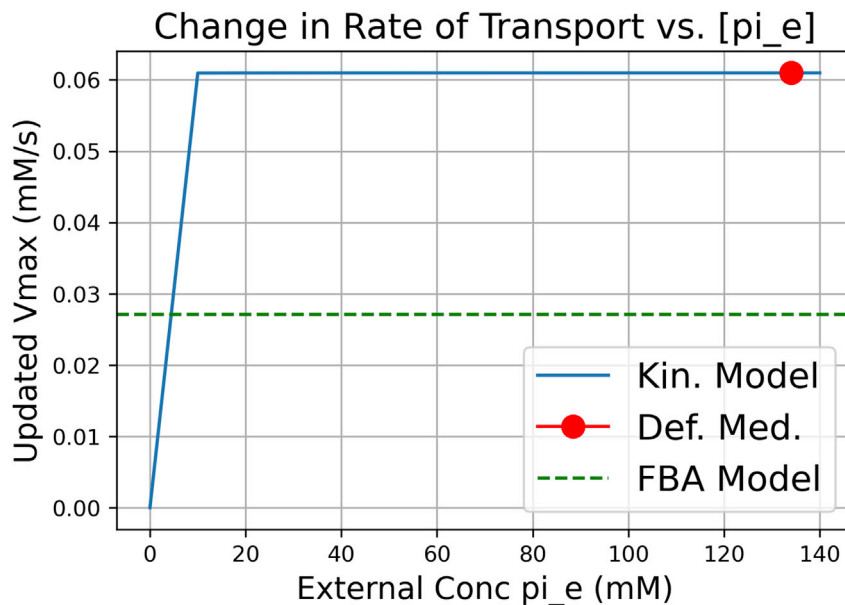
Amplification plots for the origin, quarter, and terminus qPCRs are shown on the top row. The amounts of PCR amplicon accumulated (vertical axis listed as *Rn*) after each PCR cycle (horizontal axis) are plotted. The threshold cycle (*Ct*) is the cycle at which each sample crosses the threshold cycle line, which is determined by the qPCR instrument QuantStudio software. The lower row shows standard curves generated using the qPCR standard DNA to calculate the relative amounts of DNA present in each sample.



**Figure S6. Free energies of central metabolism reactions, related to STAR Methods**

(A) Gibbs free energy change of reactions in central metabolism using reaction quotients calculated with time-averaged metabolite concentrations. GAPDP is irreversible in the kinetic model ( $k_r = 0$ ) and  $\Delta G_0$  cannot be evaluated,  $\Delta G_0$  from Equilibrator for NOX is extremely negative ( $\sim 450$  kJ/mol) and not displayed. Equilibrator cannot calculate  $\Delta G_0$  for PDH\_E3 and PDH\_acald reactions due to the enzyme itself participating in the reactions. (B) Cumulative Gibbs free energy change along the Embden-Meyerhof-Parnas (EMP) glycolytic pathway of Syn3A. (C) Forward, reverse, and net fluxes of reactions along EMP glycolytic pathway of Syn3A calculated with time-averaged metabolite and enzyme concentrations.





**Figure S7. Sensitivity of phosphate transport rate to external phosphate concentration, related to STAR Methods**

Of special interest is phosphate transport, as at one point in the defined medium growth formulation (see [Table S1](#)) the concentration of phosphate salt buffer was decreased from 140 mM to 0.5 mM. This resulted in negative effects on cell growth and the inability to produce viable cell colonies (as the kinetic model predicts). After this, the phosphate buffer component was re-adjusted to a concentration of 134 mM, well within the acceptable range predicted by the computational model.