frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 28 November 2019
doi: 10.3389/fmolb.2019.00130

Kinetic Modeling of the Genetic Information Processes in a Minimal Cell

*Zane R. Thornburg[1], Marcelo C. R. Melo[1,2], David Bianchi[1], Troy A. Brier[1], Cole Crotty[1], Marian Breuer[1,3], Hamilton O. Smith[4], Clyde A. Hutchison III[4], John I. Glass[4] and Zaida Luthey-Schulten[1]**

[1] Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, United States, [2] Machine Biology Group, Department of Psychiatry, Microbiology, and Bioengineering, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [3] Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, Netherlands, [4] Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, La Jolla, CA, United States

JCVI-syn3A is a minimal bacterial cell with a 543 kbp genome consisting of 493 genes. For this slow growing minimal cell with a 105 min doubling time, we recently established the essential metabolism including the transport of required nutrients from the environment, the gene map, and genome-wide proteomics. Of the 452 protein-coding genes, 143 are assigned to metabolism and 212 are assigned to genetic information processing. Using genome-wide proteomics and experimentally measured kinetic parameters from the literature we present here kinetic models for the genetic information processes of DNA replication, replication initiation, transcription, and translation which are solved stochastically and averaged over 1,000 replicates/cells. The model predicts the time required for replication initiation and DNA replication to be 8 and 50 min on average respectively and the number of proteins and ribosomal components to be approximately doubled in a cell cycle. The model of genetic information processing when combined with the essential metabolic and cell growth networks will provide a powerful platform for studying the fundamental principles of life.

Keywords: minimal cells, stochastic simulations, kinetic parameters, DNA replication, transcription, translation, mRNA production, protein production

OPEN ACCESS

Edited by:
Giulia Palermo, University of California, Riverside, United States

Reviewed by:
Juan R. Perilla, University of Delaware, United States
Ali Mohamad Farhat, University of Michigan, United States

***Correspondence:**
Zaida Luthey-Schulten
zan@illinois.edu

Specialty section:
This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences

Received: *27 August 2019*
Accepted: *07 November 2019*
Published: *28 November 2019*

Citation:
Thornburg ZR, Melo MCR, Bianchi D, Brier TA, Crotty C, Breuer M, Smith HO, Hutchison CA III, Glass JI and Luthey-Schulten Z (2019) Kinetic Modeling of the Genetic Information Processes in a Minimal Cell. Front. Mol. Biosci. 6:130. doi: 10.3389/fmolb.2019.00130

1. INTRODUCTION

JCVI-syn3A, a bacterial cell with a synthetic minimal genome of size 543 kbp and 493 genes, is an organism designed to have the fewest genes necessary for life and is therefore an ideal model organism for studying fundamental principles of life (Lachance et al., 2019). In Breuer et al. (2019), we published the flux balance analysis of the essential metabolism of JCVI-syn3A along with the gene map and the genome-wide data from essentiality and proteomics experiments. Although metabolism, including transport of nutrients into the cell, has been established, the reactions and kinetic models for genetic information processes in JCVI-syn3A are missing. The accompanying gene map in **Figure 1A** assigned all 452 protein coding genes to one of the four major functional classes: metabolism with transporters (143), genetic information processes (212), cellular processes such as cell division (6), and unclear functions (91). Accompanying the gene map is a map of the proteomics data detected for the 428 proteins in **Figure 1B**. The model presented here uses the proteomics data to guide the modeling of protein production.

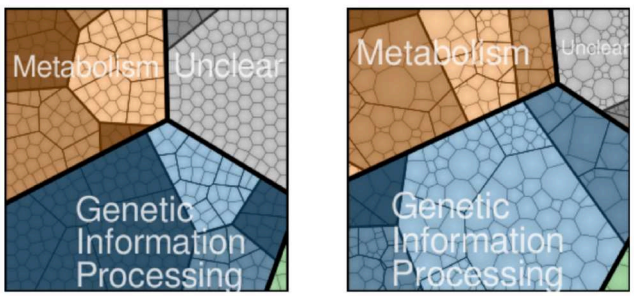

FIGURE 1 | JCVI-syn3A protein coding genes **(A)** and proteomics **(B)** distributed to four functional classes: metabolism (brown), genetic information processing (blue), cellular processes such as cell division (green), and unclear function(gray). Shading within the four functional classes indicate subsystems within the class, such as nucleotide metabolism in metabolism or transcription in genetic information processing. NCBI GenBank CP016816.2: https://www.ncbi.nlm.nih.gov/nuccore/CP016816.2 (Breuer et al., 2019).

In our previous work on ribosome biogenesis in *Escherichia coli* (Earnest et al., 2015, 2016), ribosome assembly was included along with DNA replication and transcription/translation of just the ribosomal proteins (rproteins). In this simplified model we focus on developing kinetic parameters that replicate the DNA, generate proteins comparable to the proteomics abundances, and produce sufficient numbers of rprotein and ribosomal RNA (rRNA) to generate approximately 500–700 ribosomes estimated from the biomass equation in Breuer et al. (2019). Here we introduce the construction and results of our simplified genetic information processing model for a cell 400 nm in diameter. The kinetics for initiation of DNA replication is based on a mechanism derived from the JCVI-syn3A genomic sequence, crystal structures of the initiator protein DnaA complexed with DNA and kinetics parameters from single molecule fluorescence resonance energy transfer (smFRET) experiments. Parameters for simplified kinetics describing DNA replication, transcription, mRNA degradation, translation, and protein degradation are derived from the literature and our previous studies on JCVI-syn3A (Breuer et al., 2019) and *E. coli* (Earnest et al., 2015, 2016). Within the cell cycle of 105 min, these processes duplicate the genome, generate, and translate sufficient amounts of mRNA to approximately reproduce the proteomics data, and the estimated number of ribosomes. All 452 protein coding genes and 35 genes for rRNAs and tRNAs in the genome of JCVI-syn3A are expressed. Three pseudo genes and three genes for small RNA are not expressed in this model.

2. METHODS

Each of the genetic information processing subsystems involve species that are low in population in the cell, for example one or two copies of a gene and 0–10 copies of a protein-coding mRNA. To capture the stochastic nature of genetic information processes, the kinetics were modeled with chemical master equation (CME) simulations and solved using the Gillespie algorithm as implemented in the software Lattice Microbes (Roberts et al., 2013; Hallock et al., 2014; Earnest et al., 2015, 2018) with the pyLM interface in a Python 3 Jupyter notebook. Due to the small size of JCVI-syn3A, 400 nm in diameter, we neglect the spatial location of species inside the cell in this simplified model which allows us to stochastically model the kinetics as well-stirred using CME simulations. The results of stochastic simulations were averaged over 1,000 replicates/cells. Each replicate requires a run time of one second. The Jupyter notebooks are available and are posted at GitHub (https://github.com/zanert2/Thornburg_FrontMolBiosci_2019).

2.1. Polymerization Model and Rate Forms

In our genetic information processing model, DNA replication, transcription, and translation are all reactions that involve an enzyme (DNAP, RNAP, or ribosome) catalyzing polymerization reactions based on a preexisting template polymer (the entire ssDNA, each unique gene on the ssDNA, or its corresponding mRNA). In the case of replication, the single template is the entire genomic sequence of 543 kpb. In the case of transcription, the templates are the individual 493 genes, each with a unique length and sequence. In the case of translation, the templates are the number of individual messengers for each of the proteins. We use a rate form based on Equation (33) from Hofmeyr et al. (2013) that was derived assuming polymerization from a single unique template where the enzyme is in excess and the concentration of free enzyme is constant. DNA replication, transcription, and translation all involve a situation in which the enzyme is in excess of unique templates. For DNA replication, there is a single start site, oriC, and 35 DNAP molecules in the proteomics data. In the case of transcription, there are 187 RNAP and if we consider any one gene as the template for the rate form, there are at most two copies of the gene at any point in the cell cycle. In translation, there are over 500 ribosomes available to translate the individual mRNAs which typically number <10. In each case, we assume a constant steady-state concentration of free enzymes in determining the kinetic rates, although the template concentrations will change over time. The general polymerization rate form can be written as

$$v_{poly} = \frac{k_{cat}[T]}{\left(1 + \frac{K_0}{[E]}\right)\frac{K_{D1}K_{D2}}{[M]_1[M]_2} + \sum_i \frac{n_i K_{Di}}{[M]_i} + n_{tot}} \quad (1)$$

which we modify for transcription and translation in the following sections to address that there is competition among unique templates of different lengths n_{tot} in each process. For our experimental situation, the polymerization rate is dominated by k_{cat}, n_{tot}, and template concentrations. The variation in rates based on these assumptions is discussed further below in Equation (2). The general rate form considers a mechanism starting with enzyme E (DNAP, RNAP, or ribosome) binding to a polymer template T with binding constant K_0. Once the enzyme and template have bound, the first two monomers (dNTP, NTP, or the charged aa-tRNA) M_1 and M_2 bind to the template/enzyme complex with association constants K_{D1} and K_{D2}. The monomer concentrations are determined by the pool sizes provided in Zhang and Ignatova (2009) and Breuer et al. (2019). A value of K_D has been measured for a single elongation

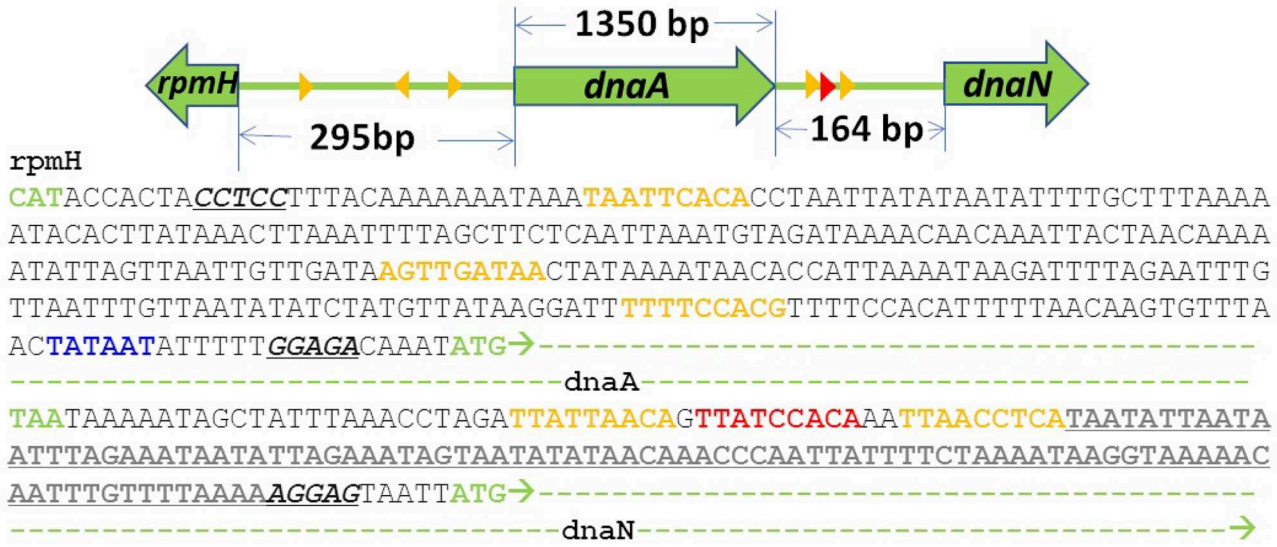

FIGURE 2 | DNA sequence near oriC of JCVI-syn3A defined by the 3 high (red) and low (yellow) affinity DnaA(IV) binding sites: The AT-rich region (bold and underlined) binds DnaA(III). The AT-rich region ends at the Shine-Dalgarno sequence (italicized and underlined) preceding the dnaN gene (green dashes). A putative promoter for the DnaA gene preceding its Shine Dalgarno sequence is shown in blue.

step of mRNA by RNAP, but not for DNAP or ribosomes (Larson et al., 2012). Values for K_D were fitted to maximize the rate of each process assuming their respective pool sizes and other experimentally measured kinetic parameters. Our fitted value for RNAP agrees well with the experimentally determined value. Monomers of type i are then added to the growing polymer by the binding with their respective association constant K_{Di} and we assume that they are the same for any one process. The growing polymer is elongated at a rate k_{cat}. The resulting polymer (DNA, rRNA, mRNA, tRNA, or protein) of length n_{tot} will consist of n_i of each respective monomer type M_i following the first two positions in the polymer.

In general, both the enzyme and template concentrations are functions of time. In evaluating the rate constant, the enzyme concentrations were held constant to the values derived from the proteomics data making the polymerization rate obey first order kinetics

$$\nu = k(n_{tot}, k_{cat})[T] \tag{2}$$

where the rate constant is defined as

$$k(n_{tot}, k_{cat}) = C \times \frac{k_{cat}}{\left(1 + \frac{K_0}{[E]}\right)\frac{K_{D1}K_{D2}}{[M]_1[M]_2} + \sum_i \frac{n_i K_{Di}}{[M]_i} + n_{tot}} \tag{3}$$

in which C represents any modifications to the rates of transcription or translation. For the kinetic parameters, pool sizes, and low enzyme concentrations assumed in the kinetic model, the denominator is dominated by the third term, the length of the new polymer n_{tot}. In analyzing the sensitivity of DNA replication, transcription, and translation to the concentration of each respective enzyme, we found that the rate constants k from Equation (3) deviated no more than 10^{-4}% as the concentration of enzyme is doubled over the cell cycle. Our above approximations hold assuming the cell is in the exponential growth phase where nutrient and pool sizes are in a steady state. The approximations no longer hold in cases such as the transition from exponential to stationary growth. As nutrients in the environment become depleted, the rate of elongation steps in DNA replication, transcription, and translation will be slowed down due to a lack of monomers M_i.

2.2. Replication Initiation

Previous treatments of replication initiation have proposed a mechanism based on *E. coli* and *B. subtillis* that began with the initiator protein DnaA binding to four 9-bp signatures of the DNA near oriC, followed by accumulation of DnaA monomers around that location until a buildup of 20–30 monomers was reached (Atlas et al., 2008; Karr et al., 2012). Our model of DNA replication initiation is based on the genomic sequence of JCVI-syn3A in **Figure 2** and a mechanism derived from crystal structures of the multi-domain DnaA binding to ds- and ssDNA shown in **Figure 3**. In the genomic sequence structure, a strong DnaA binding signature (TTATCCACA) is located near the origin matching the whole 9-bp sequence with two neighboring signatures matching 7 out of 9 bp (Schaper and Messer, 1995; Weigel et al., 1997; Speck et al., 1999). These signatures lie next to an AT-rich region 93 bp in length.

DnaA domain IV [DnaA(IV)] binds most strongly to the sequence TTATCCACA. DnaA(IV) binds to the dsDNA signatures (Erzberger et al., 2006; Duderstadt et al., 2011). DnaA domain III [DnaA(III)] binds to AT-rich ssDNA in 3 nucleotide increments forming a helical, filament-like structure (Erzberger et al., 2006; Duderstadt et al., 2011). Our mechanism assumes

that the binding of DnaA(IV) to the three neighboring dsDNA signatures near oriC opens up a small pocket of ssDNA in the neighboring AT-rich region. This mechanism is illustrated in **Figure 4A**. Once the dsDNA sites are occupied, DnaA(III) can start binding to the neighboring AT-rich region on the ssDNA. The DNA continues to be unwound until the AT-rich region is wrapped by the DnaA filament. Since DnaA(III) binds to ssDNA in 3 nt increments (Duderstadt et al., 2011; Cheng et al., 2014) the 93 bp AT-rich region shown in **Figure 2**, produces a filament with 30 DnaA. After formation of the filament, replication can be initiated.

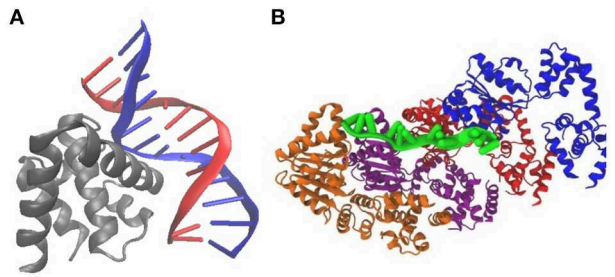

FIGURE 3 | Crystal structures of DnaA binding to *E. coli* DNA suggest a mechanism for initiation of replication: **(A)** PDB 1J1V; DnaA(IV) binds to a 9-bp signature on dsDNA. **(B)** PDB 3R8F; Four DnaA(III) bind to 3-nucleotide increments on ssDNA.

To capture the proposed mechanism, we begin with a reaction binding a DnaA to the high affinity binding signature near OriC on dsDNA, creating a bound site and the two low affinity free sites on either side of the high affinity site. The low affinity sites on dsDNA then react with one DnaA each, creating a bound site for each. The dsDNA binding rates use second order rate forms using the rate constants shown in **Table 1**. There is also a reaction in the model for DnaA binding to other high affinity sites around the chromosome. This is included since the filament length strongly depends on the number of free DnaA available. The kinetic model for the formation of the DnaA filament is based on an smFRET study on ssDNA (Cheng et al., 2014). The smFRET study in **Figure 4B** reports values for k_{on} for addition of a DnaA molecule to the growing DnaA filament bound to ssDNA and k_{off} for removal of a DnaA molecule from the filament as shown in **Figure 4C**. These kinetic parameters are presented in **Table 1** and were used for each independent binding and unbinding until a filament consisting of 30 DnaA has formed. Once the filament is formed and replication begins, the filament is assumed to be removed at the rate of the polymerization in DNA replication which models removal of DnaA by DNA helicase. The model is constructed so that only one replication initiation event occurs in a cell cycle.

2.3. Replication

The replisome, a complex containing proteins necessary for DNA replication including DNA helicase, DNAP, DNA primase, gyrase/topoisomerase, and the beta clamp, binds at oriC once

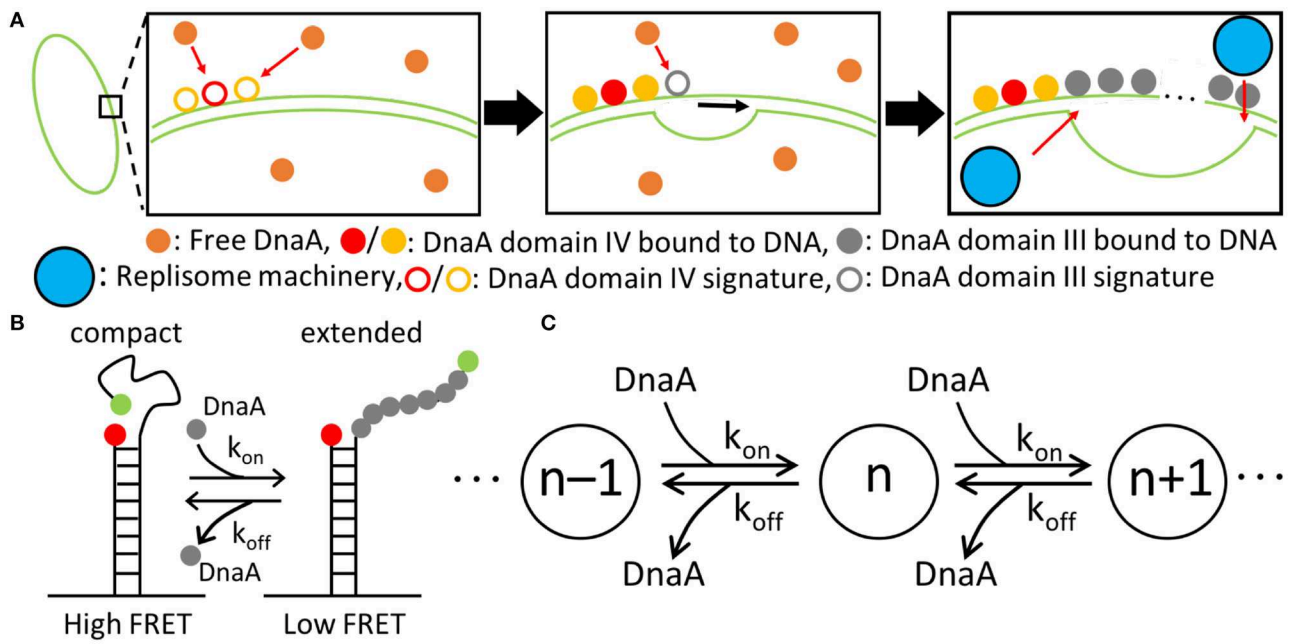

FIGURE 4 | Replication initiation mechanism. **(A)** DnaA(IV) binds to three signatures on dsDNA next to the AT-rich region near oriC. DnaA(III) subsequently binds in 3-nucleotide increments. DnaA(III) continues to bind to ssDNA until the AT-rich region is opened, allowing the replisome machinery to be loaded. **(B)** The kinetic parameters for the binding of DnaA(III) to ssDNA were obtained from a smFRET study (Cheng et al., 2014) where the FRET signal depended on the number of DnaA bound. Fewer DnaA corresponded to compact ssDNA, resulting in a high FRET signal. Increasing the number of DnaA bound to ssDNA extends the filament, lowering the FRET signal. **(C)** Schematic of the binding kinetics of DnaA(III) to ssDNA forming a DnaA filament of length n. The k_{on} and k_{off} values correspond to the kinetics measured by smFRET.

the replication initiation event has occurred and then proceeds in both directions around the chromosome, creating the two replication forks as shown in **Figure 5**. Using smFRET experiments, the replisome has been observed to assemble in just a few seconds (Downey and McHenry, 2010; Cho et al., 2014). We do not model the assembly of the replisome and assume its assembly occurs during or before replication initiation. As the replisome proceeds along the chromosome, the original chromosome shown in green is unzipped and the two new chromosomes shown in red and blue are polymerized on the original ssDNA template. Both strands of ssDNA at the replication fork are treated the same with continuous polymerization, and okazaki fragments are not modeled. The model assumes that once the replisomes reach the terminus, they fall off quickly and the two new chromosomes are instantaneously separated. The number of dATP, dTTP, dCTP, and dGTP monomers n_i appearing in the rate form (Equation 1) are calculated from the A, T, C, and G content of the genome: 203606 A, 207816 T, 67238 C, and 64720 G. Since there are no metabolic reactions to produce deoxynucleotides or ATP for the reactions to occur, constant pools for each are assumed using the pool sizes from Breuer et al. (2019) presented in **Table 2**.

Kinetic parameters for replication are given in **Table 3**. The elongation rate constant k_{cat} (Xie et al., 2008) and the association constant for DNAP to DNA K_0 (Zhang et al., 2016) were obtained from the literature for *E. coli*. In order to make a second copy of the genome within the 105 min doubling time, the choice of K_D was made in order to minimize the time to duplicate the DNA. Assuming the constant pool sizes and DNAP concentrations, the value of K_D corresponds to the value where the length of the genome is the dominant term in the denominator of k in Equation (3).

2.4. Transcription

To modify the general rate form for transcription, we incorporate two factors: the probability of an active RNAP selecting any gene $P_{gene\ selection}$ and the strength of the gene's promoter $S_{promoter}$. The fraction of active RNAP as estimated by Bremer and Dennis (2008) for a cell with a ∼100 min doubling time implies that around 29 of the 187 RNAP are actively transcribing at any time. Of the actively transcribing RNAPs, Bremer and Dennis (2008) estimate that approximately 24% are involved in making stable RNA like rRNA. Since each rRNA operon only contains the 16S, 23S, and 5S rRNAs and no tRNAs, transcription of the two drRNA genes will require four RNAP. Therefore, the probability of any other gene being selected is $P_{gene\ selection} = 25/487 = 0.05$. We estimate that each rRNA operon is always being actively

TABLE 1 | Kinetic parameters used in the model of replication initiation.

Parameter	Value	Units	References
High affinity binding rate	7,800	$mM^{-1}\ s^{-1}$	Schaper and Messer, 1995; Weigel et al., 1997
Low affinity binding rate	35	$mM^{-1}\ s^{-1}$	Schaper and Messer, 1995; Weigel et al., 1997
k_{on}	100	$mM^{-1}\ s^{-1}$	Cheng et al., 2014
k_{off}	0.55	s^{-1}	Cheng et al., 2014

TABLE 2 | Pool sizes from Breuer et al. (2019) and estimated from Zhang and Ignatova (2009) and Mackie (2013)*.

Species	Pool size (mM)
dATP	0.018
dTTP	0.022
dCTP	0.012
dGTP	0.007
ATP	1.04
UTP	0.68
CTP	0.34
GTP	0.68
tRNA*	0.0020
aa-tRNA*	0.0076

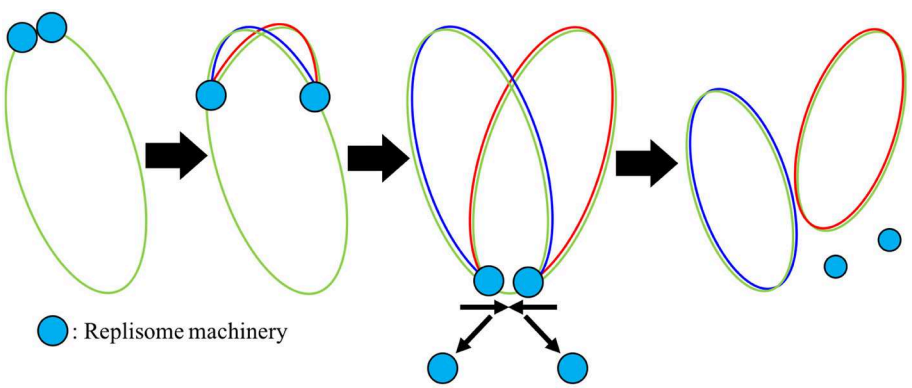

FIGURE 5 | Mechanism of DNA replication: Once the replisome machinery is loaded onto the chromosome shown in green, the machinery begins to polymerize around the DNA, elongating new DNA onto both halves of the original chromosome shown in red and blue. Once the replisome reaches the terminus, we assume that the replisomes fall off quickly.

transcribed by two RNAP, and therefore has a probability of gene selection of 1. The expression from Hofmeyr et al. (2013) did not include competition for multiple templates which is now captured with the probability of gene selection. This gives us a transcription rate

$$v_{transcription} = P_{gene\ selection} \times v_{poly} \quad (4)$$

which we use for transcription of rRNA, tRNA, and ribosomal protein-coding genes.

The the rate of transcribing a gene also depends on the strength of its promoter sequence (Jones et al., 2014), however the precise promoter sequences and their strengths have not been measured for JCVI-syn3A. In a preliminary analysis of the sequences preceding each protein-coding gene, we found that, in general, a protein is more likely to have a higher proteomics value if the start codon is preceded by both a Shine Dalgarno sequence a promoter sequence TANAAT as characterized in *Mycoplasma pneumoniae* (Lloréns-Rico et al., 2015). Using this information, to incorporate a proxy for promoter strength, $S_{promoter}$, into the kinetics, the transcription rate for each non-ribosomal protein coding gene is multiplied by the ratio of gene's proteomics count to the average proteomics count of 180

$$v_{mRNA\ transcription} = S_{promoter} \times P_{gene\ selection} \times v_{poly} \quad (5)$$

Since some ribosomal proteins were not reported in the proteomics data, this factor is not used in the transcription rates of ribosomal protein coding genes.

The model expresses the genes for all 452 protein coding genes and the genes for rRNA and tRNA. For each protein or RNA, the gene identifier from the NCBI entry (NCBI GenBank CP016816.2: https://www.ncbi.nlm.nih.gov/nuccore/CP016816.2; Breuer et al., 2019) is read and the corresponding sequence is used to determine the nucleotide stoichiometries for the formation and degradation reactions. RNA formation reactions use our modified polymerized, template-driven rate forms in Equations (4) and (5) and the degradation reactions of mRNA follow first order kinetics. The nucleotide stoichiometries are used to determine the monomer counts n_i and total polymer length n_{tot} in the rate form. Constant pools of nucleotides are assumed using the pool sizes from Breuer et al. (2019) presented in **Table 2**. For the transcription reactions, the enzyme is RNAP and the template is the total concentration of the gene in the cell as a function of time and includes the replication of DNA. This model, however, does not take into account the location of a gene on the genome during DNA elongation. The elongation rate constant k_{cat} and the association constants K_0 and K_D are listed in **Table 3**. Literature values of mRNA and tRNA elongation rates of 25 nt/s are used for k_{cat} (Chen et al., 2015). A messenger half-life of 4 min is used for all mRNA degradation. The half-life of 1 min in Breuer et al. (2019) did not result in mRNA abundances that produced proteins quickly enough to double the number of proteins in the cell cycle. The 4 min half life gives a total mRNA abundance in better agreement with the data published in Lynch and Marinov (2015). The experimentally observed rRNA operon elongation rate k_{cat} of 90 nt/s (Ryals et al., 1982) was multiplied by two for both operons to model the effect of two RNAP simultaneously transcribing each operon. The association constant for association of RNAP to DNA K_0 was calculated according to Hofmeyr et al. (2013) using the concentrations of the free and actively transcribing RNAP (Bremer and Dennis, 2008) and concentration of the gene. The association constant for nucleotides binding to the RNAP/gene complex K_D was fitted so that the rate of transcription was maximized by making transcript length the dominant term in the denominator of k in Equation (3). Our fitted value agrees with a measured experimental value of 0.14 mM (Larson et al., 2012). With no transcriptomic data available, each mRNA begins with a count of 1 and each tRNA is divided evenly at 190 each to have a total tRNA abundance of 3,750, a value scaled from *E. coli* based on differences in cell volume (Mackie, 2013).

TABLE 3 | Parameters used in kinetics for replication, transcription, translation, mRNA degradation, and protein degradation.

Subsystem	Parameter	Value	Units	References
Replication	k_{cat}	600	bp/s	Breier et al., 2005; Xie et al., 2008
	K_0	0.26	μM	Zhang et al., 2016
	K_D	1.0	μM	Fitted
Transcription	k_{cat} (mRNA and tRNA)	25	nt/s	Chen et al., 2015
	k_{cat} (rRNA)	180	nt/s	Ryals et al., 1982
	K_0	100	nM	Bremer and Dennis, 2008
	K_D	0.1	mM	Fitted; Larson et al., 2012
Translation	k_{cat}	5	aa/s	Cox, 2004
	K_0	100	nM	Bremer and Dennis, 2008
	K_D	0.01	mM	Fitted
mRNA Degradation	$t_{1/2}$	4	min	Bernstein et al., 2004; Briani et al., 2008
Protein degradation	$t_{1/2}$	25	hr	Maier et al., 2011

2.5. Translation

Since the number of total mRNA is approximately on the same order of the number of ribosomes, the probability of any mRNA being translated is near unity. The only other modification of the translation rate expression is to allow more than one ribosome (polysomes) N_{ribo} to bind to a long transcript in Equation (6).

TABLE 4 | ATP hydrolysis costs of reactions in genetic information processing subsystems (Russell and Cook, 1995; Lynch and Marinov, 2015).

Reaction	ATP cost	Units
Replication	1	ATP per bp
Transcription	1	ATP per nt
Translation	2	ATP per aa
mRNA degradation	1	ATP per nt
Protein degradation	1	ATP per aa

The cost of translation does not include charging of the tRNAs as those reactions are incorporated in the essential metabolism (Breuer et al., 2019).

This factor is an integer calculated as the length of the transcript over an estimated ribosome spacing of 300 nt in *E. coli* (Brandt et al., 2009). If the value is calculated as <1, the value of N_{ribo} is set to 1. The ribosome spacing was estimated using an observed approximate average of 4 ribosomes per polysome for an average transcript length of 1,200 nt.

$$v_{translation} = N_{ribo} \times v_{poly} \quad (6)$$

The model includes the translation and degradation of each protein made from each mRNA. The gene identifier from the NCBI entry also includes the amino acid sequence for protein coding genes which is used to determine the corresponding stoichiometries of tRNA charged with their corresponding amino acids (aa-tRNA) required to build the protein and the amino acid stoichiometries when the protein is degraded. For the translation reactions, the template in the polymerization rate form (Equation 1) is the associated mRNA. The model uses whole, intact ribosomes as the enzyme and does not model association of messengers to the 30S small subunit followed by association of the 50S large subunit. The elongation rate constant k_{cat} and the association constants K_0 and K_D are listed in **Table 3**. For *E. coli*, experimentally measured elongation rates range from 10 to 20 aa/sec (Bremer and Dennis, 2008), however slower rates have been reported in other bacteria such as *Mycobacterium bovis* with an elongation rate of 2 aa/sec (Cox, 2004). A value within the estimated range of 2–10 aa/sec of 5 aa/sec was chosen so that the number of proteins was approximately doubled in a cell cycle. The association constant of the ribosome to the mRNA K_0 was estimated using the average fraction of actively translating ribosomes (Bremer and Dennis, 2008) and an average concentration of an mRNA to be one in the cell. The association constant for aa-tRNA binding to the ribosome/mRNA complex K_D was fitted to maximize the rate of translation assuming constant aa-tRNA pool sizes and ribosome concentration. The value of K_D was computed using the length of the shortest protein, ribosomal protein L34 (40 aa), in the equation for the rate constant k (Equation 3). A half-life of 25 h was used for

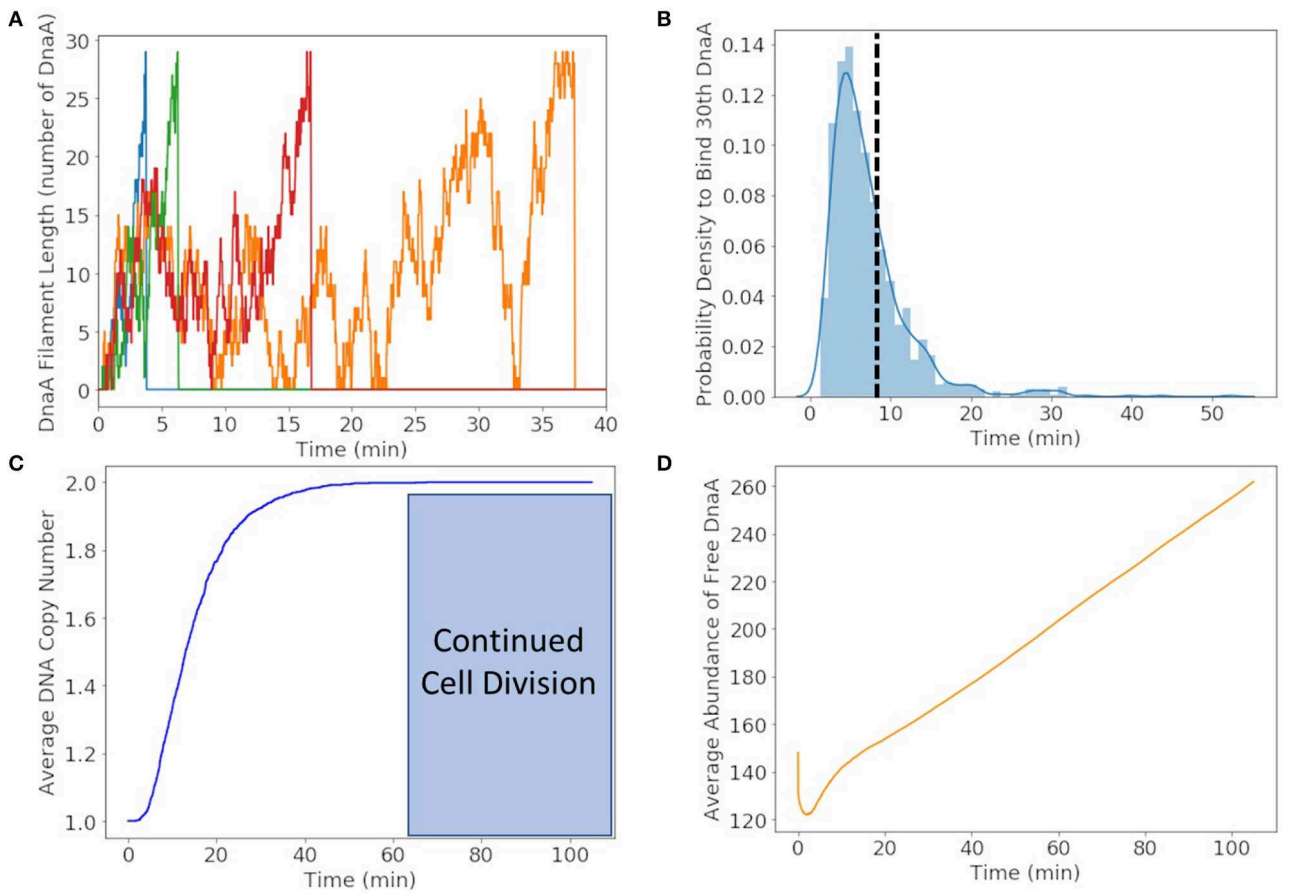

FIGURE 6 | (A) DnaA filament formation for four different replicates shown in different colors. The stochastic effects of the filamentation kinetics result in a wide range of times to form the filament from <5 to 50 min. **(B)** Probability distribution of replication initiation times when the thirtieth DnaA in the ssDNA filament binds. We predict the most probable time to form the filament to be approximately 5 min and the average time to be approximately 8 min shown with a dotted line. **(C)** Average of genome duplication over 1,000 replicates shows that on average the genome will be duplicated in 65 min of the 105 min cell cycle, leaving approximately 40 min for continued cell division. **(D)** The average abundance of DnaA not bound to DNA gets depleted by filament formation and replenished by translation and removal of the filament by DNA helicase.

protein degradation reactions (Maier et al., 2011) Degradation of the proteins in extremely slow, so the main source of dilution would be by cell division after 105 min.

2.6. ATP Energy Costs

Replication, transcription, translation, mRNA degradation, and protein degradation have associated ATP hydrolysis costs. Although the mechanism for ATP hydrolysis is not explicitly modeled, the costs are incorporated as additional time dependent reactions for each subsystem. For example, in DNA replication the DNA helicase is not explicitly modeled, but we assume that 1 ATP hydrolysis event per bp is required to unwind the dsDNA. The ATP cost of each reaction in each subsystem is determined by the length of the DNA/RNA/protein being formed or mRNA/protein being degraded (Russell and Cook, 1995; Lynch and Marinov, 2015). In transcription, we assume that the RNAP uses 1 ATP hydrolysis event per bp to unwind the dsDNA. The mRNA degradation reactions also assume that

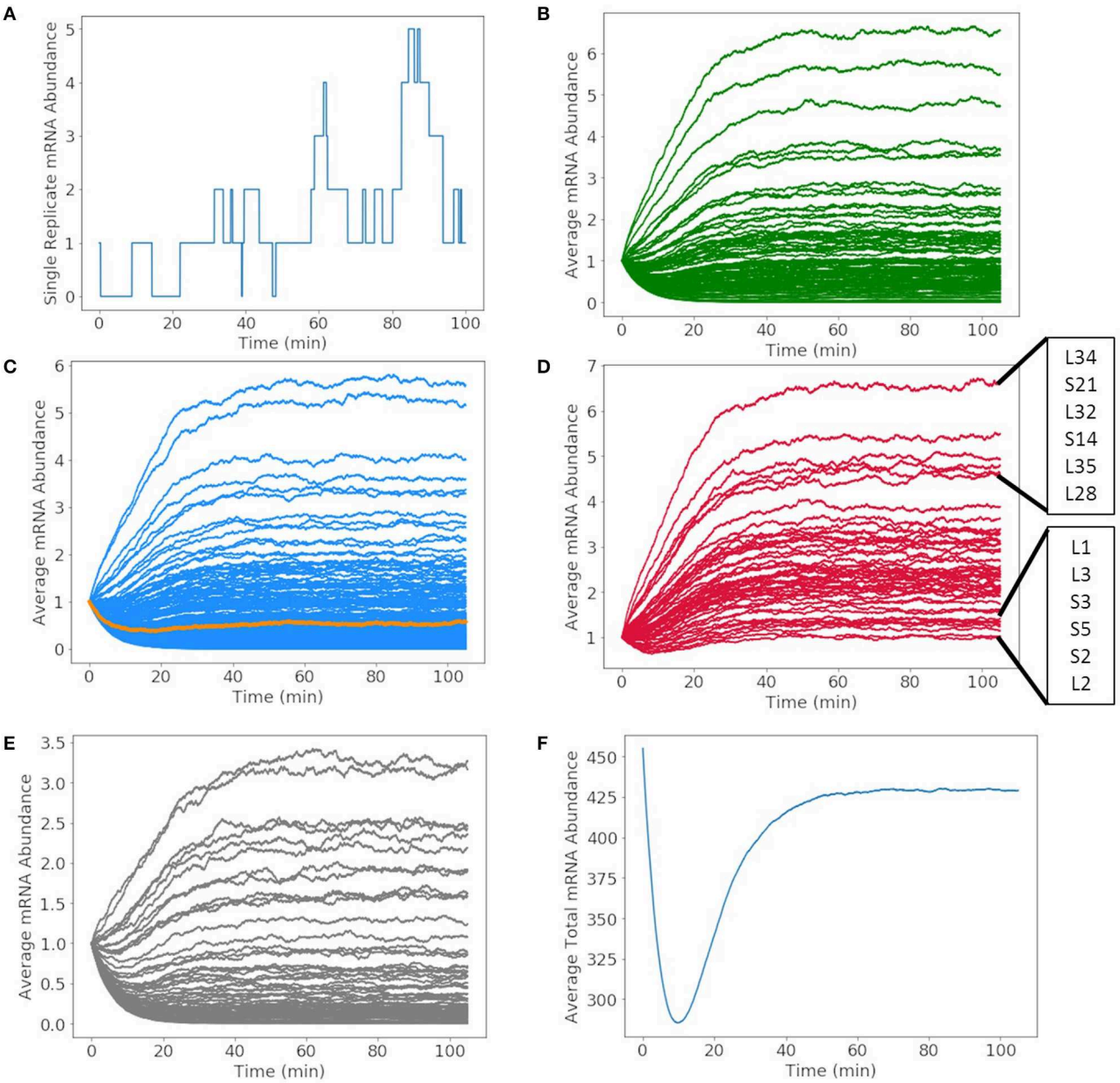

FIGURE 7 | Abundances of mRNA and tRNA transcribed in a 105 min cell cycle. **(A)** A single replicate from the stochastic simulation of the mRNA abundance for glucose-6-phosphate isomerase shows fluctuations in the average integer abundance of messengers. Fluctuations arise from competing rates of formation, degradation, and replication. The average mRNA abundances of mRNA coding for **(B)** metabolic proteins, **(C)** genetic information processing, DnaA (orange), and cell division proteins, **(D)** ribosomal proteins, and **(E)** proteins of unclear function all have average abundances between zero and seven. **(F)** The total number of all messengers during a cell cycle averaged over 1,000 replicates shows that typically there are 300–450 messengers present in the cell at any time.

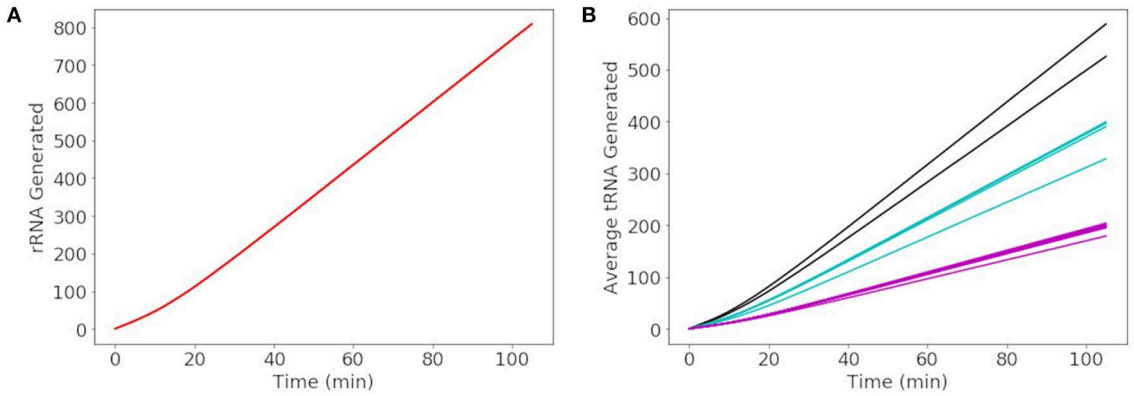

FIGURE 8 | Generated abundances of rRNA and tRNA for a cell cycle averaged over 1000 replicates. **(A)** 16S, 23S, and 5S rRNA are each generated to the same average abundance of 800 over a cell cycle. **(B)** The average abundance of each tRNA generated in a cell cycle results in a total abundance of approximately 4000 tRNAs. The tRNAs in black are methionine and leucine, in blue are threonine, tryptophan, lysine, arginine, and serine, and remaining tRNAs are shown in pink.

1 ATP hydrolysis event is required per nucleotide removed from the messenger. The transcription reactions assume 2 ATP hydrolysis events per amino acid addition. These reactions use 2 instead of 4 ATP hydrolysis events since the amino acid charging of the tRNA are already included in the essential metabolic network (Breuer et al., 2019). The costs used are also shown in **Table 4**.

3. RESULTS

3.1. Replication Initiation and Replication

We found that DnaA(IV) requires <1 min to bind to all three dsDNA signatures. The stochastic trajectories of DnaA filament formation from four representative cells are shown in **Figure 6A**. The distribution of times to form the DnaA filament in **Figure 6B** is peaked at 5 min, but on average it takes 8 min for the DnaA filament to form on ssDNA as shown with a dotted line. Once the filament is 30 DnaA in length, replication begins and the DnaA filament is removed by the polymerization of DNA, resulting in the fast drop from 30 to 0 DnaA in the filament as seen in the trajectories in **Figure 6A**. It then takes another 50 min on average for replication to reach completion in **Figure 6C**. We predict replication initiation and replication are completed by 65 min, leaving another 40 min for the cell to divide in the 105 min cell cycle.

To illustrate the time-dependent variation in protein formation, the average abundance of free DnaA is shown in **Figure 6D**. Within the first minute we see a fast drop due to DnaA(IV) binding to high affinity dsDNA binding sites around the genome. The filament formation slowly removes DnaA from the free DnaA abundance until around 8 min when replication most frequently begins. DnaA in then replenished over several minutes due to removal of the filament by DNA helicase and translation of new DnaA.

3.2. Transcription

The mRNA production in a single cell exhibits fluctuations due to competing rates of formation and degradation. A representative of the mRNA production for glucose-6-phosphate isomerase over the 105 min simulation is shown in **Figure 7A**. The abundance of the messenger fluctuates from zero to two before DNA replication occurs and then one to five once the gene has been duplicated. The time dependence of all mRNA over a cell cycle averaged over 1,000 replicates are shown in **Figures 7B–E**. The mRNA are divided by mRNA for metabolic proteins (**Figure 7B**), genetic information processing and cell division proteins (**Figure 7C**), ribosomal proteins (**Figure 7D**), and proteins of unclear function (**Figure 7E**). The resulting kinetics show each mRNA growing or depleting in population from the initial one copy until the effects of replication are fully manifested around 60 min. In the early phase, the increase or decrease of mRNA reflects the competition between mRNA decay and the length of the transcript and the strength of the gene's promoter. As the genome is duplicated, this equilibrium for each mRNA shifts once a second copy of the gene is present. As the position of the gene in the genome is not considered, the variations are proportional to change in the DNA copy number of the cell cycle and not the nearness to oriC. The total number of mRNAs in **Figure 7F** varies from its initial value of 452 (one for each of the protein-coding genes) to an equilibrium value of approximately 425.

More than 500 of each rRNA were produced in a cell cycle shown in **Figure 8A**, reaching the number required to produce 500–700 ribosomes in the cell cycle estimated by Breuer et al. (2019). The number of each tRNA produced in **Figure 8B** reveals three groupings of tRNA production. The three groupings depend on the number of genes for each tRNA present in the genome. The groups consisting of more than one gene include 3 each of methionine and leucine tRNA genes making up the tRNA grouped between 500 and 600 tRNA and 2 each of threonine, tryptophan, lysine, arginine, and serine tRNA genes making up the tRNA grouped between 300 and 400 tRNA. Overall the model produces approximately 4,000 total tRNAs over a cell cycle, in close agreement with the initial estimate of 3,750 obtained from scaling the abundances in *E. coli* (Mackie, 2013).

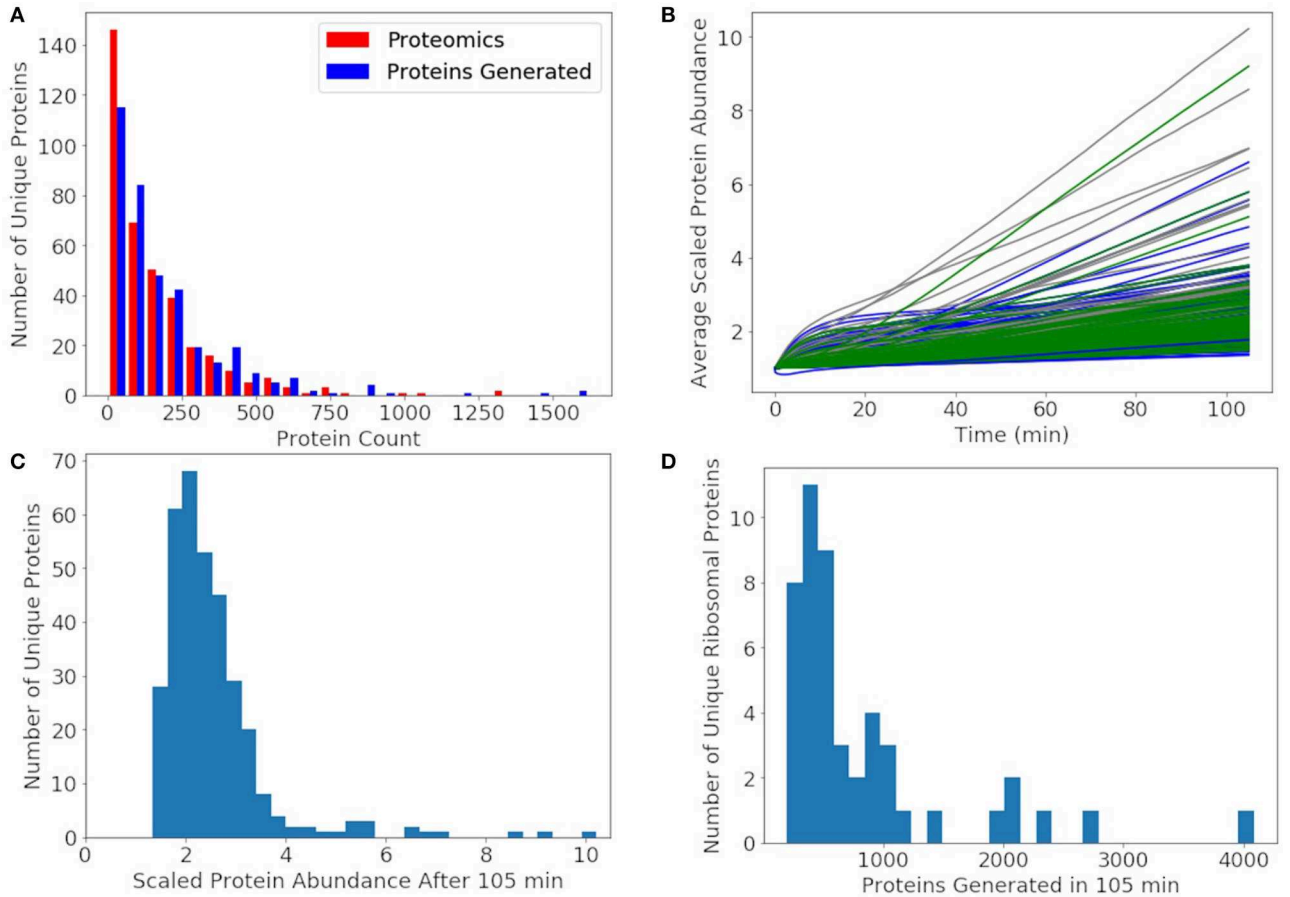

FIGURE 9 | (A) The distribution of proteins generated after 105 min averaged over 1,000 replicates approximately reproduces the experimental proteomics distribution for JCVI-syn3A. **(B)** Average protein counts scaled to the proteomics numbers for proteins with a proteomics counts >10. Metabolic proteins (green), non-ribosomal genetic information processing and cell division proteins (blue), and proteins of unclear function (gray). **(C)** A histogram of the scaled protein abundances at 105 min shows that the model doubles the abundances of most proteins with only a few outliers including mostly proteins of unclear function and thioredoxin, acyl carrier protein, transcription antitermination factor NusB, aspartyl/glutamyl-tRNA amidotransferase, and transporter ptsH. **(D)** A histogram of the number of ribosomal proteins generated shows that the model produces approximately 500 of most ribosomal proteins, enough to form the predicted 500 ribosomes. Some ribosomal proteins were produced in large excess including L34 above 4000, S21 near 3000, and L32, L35, S14, and L28 around 2000 each.

3.3. Translation

Since the protein degradation rate of 25 h is much slower than the mRNA degradation rate of 4 min, proteins will accumulate and only decay significantly by dilution through cell division. The goal of the model was to approximately reproduce the experimental proteomics distribution, double the abundance of each non-ribosomal protein, and produce 500–700 of each ribosomal protein. We compare our distribution of generated proteins over a cell cycle to the experimental proteomics in **Figure 9A**. We approximately reproduce most of the distribution with the greatest deviation being for proteins with fewer than 10 counts in the proteomics data. In the rest of our analysis of non-ribosomal proteins, we focus on proteins with experimental proteomics abundances >10. For further comparison, the number of each non-ribosomal protein generated over a cell cycle is compared to its proteomics value used to initialize the simulations (**Figure 9B**). From the histogram in **Figure 9C** we see that most non-ribosomal proteins double in number over a cell cycle with a few outliers, of which most are proteins of unclear function. The remaining outliers include thioredoxin, acyl carrier protein, transcription antitermination factor NusB, aspartyl/glutamyl-tRNA amidotransferase, and ptsH, all of which are short proteins around 100 amino acids in length or shorter. The histogram of ribosomal proteins abundances generated by the model in **Figure 9D** reveals that the model produces 500 copies for the majority of the ribosomal proteins, while the shortest are being overproduced. Ribosomal proteins overproduced include L34 above 4,000, S21 near 3,000, and L32, L35, S14, and L28 above 2,000 each. Ribosomal proteins not generated to an abundance of at least 500 include L1, L3, S3, S5, S2, and L2.

3.4. ATP Energy Costs

The model was constructed to estimate the ATP hydrolysis requirements for the genetic information processes in the minimal cell using per bp, nt, or aa usage of ATP in DNA

TABLE 5 | ATP hydrolysis costs of the deterministic model for genetic information processes.

Subsystem	ATP used in 105 min (millions)	ATP cost for a 400 nm cell (mM)
Total	77	3,800
Replication	0.54	28
Transcription	10	500
Translation	59	2,900
mRNA degradation	5.9	290
Protein degradation	1.9	93

The ATP cost for transcription reported here only includes the hydrolysis costs of the RNAP, it does not include ATP built into RNA sequences.

elongation, transcription, translation, mRNA degradation, and protein degradation. The estimates of the ATP hydrolysis cost over a 105 min simulation are presented in **Table 5** as both the total number of ATP used and the corresponding concentration of ATP required for a 400 nm cell. The model predicts that the total ATP hydrolysis cost over a cell cycle to be approximately 3,800 mM for JCVI-syn3A. This estimate does not suggest that 3,800 mM of ATP needs to be present in the cell, but provides an estimate for how quickly the metabolism will need to convert ADP into ATP. The most significant of the ATP hydrolysis costs in the genetic information processes comes from translation requiring 2,900 mM and the smallest of the costs is for DNA replication at 28 mM. The cost for translation will be higher once the genetic information processes are paired with the metabolism, as this cost did not account for the two ATP hydrolysis events to charge each tRNA which are included in the essential metabolism (Breuer et al., 2019). The cost for transcription of 500 mM does not include the ATP built into RNA sequences, it only includes the ATP hydrolysis costs of the RNAP. The predicts ATP requirements for mRNA degradation and protein degradation are predicted to be 290 and 90 mM, respectively. The cost for protein degradation is smaller due to the long protein have-life of 25 h relative to the 4 min half-life of messengers.

4. DISCUSSION

Our detailed model for the initiation of DNA replication builds upon observations from crystal structures of the initiator protein DnaA bound to signatures on ds-and ssDNA found near the oriC and smFRET measurements of the DnaA filament formation on ssDNA. The time taken for DNA replication initiation is predicted to vary from <5 min up to 50 min. We predict a total time of 65 min on average for the formation of the second copy of the genome, which means at least one copy of the DNA can be generated in a cell cycle.

The average number of any mRNA is within the expected range from zero to ten as reported in *E. coli* (Milo and Phillips, 2015) and can be used as predictions for mRNA counts in JCVI-syn3A until transcriptomic data or smFISH experiments are available for validation. We predict that approximately 450 messengers will be present in the cell on average, agreeing with the extrapolated number for a 400 nm diameter cell from Lynch and Marinov (2015). In our previous treatments of replication and transcription of a given gene in *E. coli* (Peterson et al., 2015; Cole and Luthey-Schulten, 2017) we showed how the variation in DNA copy number and position of the gene in circular DNA can broaden the mRNA distribution. We are likely underestimating the distributions for genes close to oriC and overestimating the distributions for genes near the terminus. In the case of rRNA, a higher transcription rate generated a sufficient number of rRNA to form 500–700 ribosomes in a cell cycle. A higher transcription rate was justified from the greater promoter strength of the rRNA operon observed in *E. coli* and other bacteria (Maeda et al., 2015) as well as the presence of multiple RNAPs estimated to be reading the operon (Bremer and Dennis, 2008). While the model produces over 500 rRNAs, there is variation in the number of ribosomal proteins. For the majority of the ribosomal proteins, approximately 500 of each were generated. However, the long ribosomal proteins were not generated quickly enough and the shorter ribosomal proteins occurred in much higher numbers. This is likely due to no promoter strength being assigned to the transcription of genes coding for ribosomal proteins. In the case of non-ribosomal proteins where we assigned promoter strengths based on proteomics counts, our model, to the most part, approximately doubles the number of proteins over a cell cycle. Identification of the promoter sequences and operonal structures for genes in JCVI-syn3A would help assign variation in promoter strengths and transcription rates on the basis of genomic information rather than proteomics values.

The simplified kinetic models for the genetic information processing reactions in the minimal cell JCVI-syn3A neglected the explicit assembly of the protein complexes that replicate DNA (replisome), transcribe the genes, and translate the mRNA and instead focused on the "polymerization" reactions that replicated the DNA, transcribed the genes into mRNAs, and translated them into proteins and how they are coupled. In some cases, this neglect can be justified by assumed timescale separation of the processes, but in general more experimental measurements of the assembly reactions would help to establish to what degree the association of the complexes are captured in the kinetic parameters given in the literature for the fundamental processes of replication, transcription, and translation. As the next step, the results from the genetic information processes will first be connected to uptake reactions that transport nucleobases, nucleosides, and amino acids into the minimal cell. Coupling genetic information processes with the essential metabolism and cell growth should result in a complete whole cell kinetic model of JCVI-syn3A.

DATA AVAILABILITY STATEMENT

The jupyter notebooks containing the models in this study can be found at https://github.com/zanert2/Thornburg_FrontMolBiosci_2019.

AUTHOR CONTRIBUTIONS

ZT and ZL-S: developed models for genetic information processes, data curation, writing—original draft, and writing—reviewing and editing. MM: assistance in writing of Jupyter notebooks. DB and TB: advised Lattice Microbes interface for the stochastic model. HS: assisted in development of DNA replication initiation model. CC: constructed initial stochastic model of replication initiation. MB: data curation. CH and JG: reviewing.

FUNDING

Partial support from NSF MCB 1818344 and 1840320, The Center for the Physics of Living Cells NSF PHY 1430124, NSF PHY 1505008, and NSF REU 1659598.

ACKNOWLEDGMENTS

The authors thank Tyler Earnest for help with the gene map software.

REFERENCES

Atlas, J., Nikolaev, E., Browning, S., and Shuler, M. (2008). Incorporating genome-wide DNA sequence information into a dynamic whole-cell model of *Escherichia coli*: application to dna replication. *IET Syst. Biol.* 2, 369–382. doi: 10.1049/iet-syb:20070079

Bernstein, J. A., Lin, P.-H., Cohen, S. N., and Lin-Chao, S. (2004). Global analysis of *Escherichia coli* RNA degradosome function using dna microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2758–2763. doi: 10.1073/pnas.0308747101

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. doi: 10.1016/j.cell.2008.11.016

Breier, A. M., Weier, H.-U. G., and Cozzarelli, N. R. (2005). Independence of replisomes in *Escherichia coli* chromosomal replication. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3942–3947. doi: 10.1073/pnas.0500812102

Bremer, H., and Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* 3, 1–49. doi: 10.1128/ecosal.5.2.3

Breuer, M., Earnest, T. M., Merryman, C., Wise, K. S., Sun, L., Lynott, M. R., et al. (2019). Essential metabolism for a minimal cell. *eLife* 8:e36842. doi: 10.7554/eLife.36842

Briani, F., Curti, S., Rossi, F., Carzaniga, T., Mauri, P., and Dehò, G. (2008). Polynucleotide phosphorylase hinders mrna degradation upon ribosomal protein S1 overexpression in *Escherichia coli*. *RNA* 14, 2417–2429. doi: 10.1261/rna.1123908

Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* 11:808. doi: 10.15252/msb.20159000

Cheng, H.-M., Gröger, P., Hartmann, A., and Schlierf, M. (2014). Bacterial initiators form dynamic filaments on single-stranded dna monomer by monomer. *Nucleic Acids Res.* 43, 396–405. doi: 10.1093/nar/gku1284

Cho, W.-K., Jergic, S., Kim, D., Dixon, N. E., and Lee, J.-B. (2014). Loading dynamics of a sliding DNA clamp. *Angew. Chem. Int. Edn.* 53, 6768–6771. doi: 10.1002/anie.201403063

Cole, J. A., and Luthey-Schulten, Z. (2017). Careful accounting of extrinsic noise in protein expression reveals correlations among its sources. *Phys. Rev. E* 95:062418. doi: 10.1103/PhysRevE.95.062418

Cox, R. A. (2004). Quantitative relationships for specific growth rates and macromolecular compositions of *Mycobacterium tuberculosis*, *Streptomyces coelicolor* A3 (2) and *Escherichia coli* B/r: an integrative theoretical approach. *Microbiology* 150, 1413–1426. doi: 10.1099/mic.0.26560-0

Downey, C. D., and McHenry, C. S. (2010). Chaperoning of a replicative polymerase onto a newly assembled dna-bound sliding clamp by the clamp loader. *Mol. Cell* 37, 481–491. doi: 10.1016/j.molcel.2010.01.013

Duderstadt, K. E., Chuang, K., and Berger, J. M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478:209. doi: 10.1038/nature10455

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. doi: 10.1088/1361-6633/aaae2c

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. doi: 10.1002/bip.22892

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. doi: 10.1016/j.bpj.2015.07.030

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. doi: 10.1038/nsmb1115

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. doi: 10.1016/j.parco.2014.03.009

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. doi: 10.1016/j.febslet.2013.07.011

Jones, D. L., Brewster, R. C., and Phillips, R. (2014). Promoter architecture dictates cell-to-cell variability in gene expression. *Science* 346, 1533–1536. doi: 10.1126/science.1255301

Karr, J. R., Sanghvi, J. C., MacKlin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B. Jr., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401. doi: 10.1016/j.cell.2012.05.044

Lachance, J.-C., Rodrigue, S., and Palsson, B. O. (2019). Synthetic biology: minimal cells, maximal knowledge. *eLife* 8:e45379. doi: 10.7554/eLife.45379

Larson, M. H., Zhou, J., Kaplan, C. D., Palangat, M., Kornberg, R. D., Landick, R., et al. (2012). Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6555–6560. doi: 10.1073/pnas.1200939109

Lloréns-Rico, V., Lluch-Senar, M., and Serrano, L. (2015). Distinguishing between productive and abortive promoters using a random forest classifier in *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 43, 3442–3453. doi: 10.1093/nar/gkv170

Lynch, M., and Marinov, G. K. (2015). The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15690–15695. doi: 10.1073/pnas.1514974112

Mackie, G. A. (2013). RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* 11, 45–47. doi: 10.1038/nrmicro2930

Maeda, M., Shimada, T., and Ishihama, A. (2015). Strength and regulation of seven rRNA promoters in escherichia coli. *PloS ONE* 10:e0144697. doi: 10.1371/journal.pone.0144697

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R., and Serrano, L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol. Syst. Biol.* 7:511. doi: 10.1038/msb.2011.38

Milo, R. and Phillips, R. (2015). *Cell Biology by the Numbers.* Garland Science.

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. U.S.A.* 112, 15886–15891. doi: 10.1073/pnas.1516246112

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. doi: 10.1002/jcc.23130

Russell, J. B., and Cook, G. M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Mol. Biol. Rev.* 59, 48–62.

Ryals, J., Little, R., and Bremer, H. (1982). Temperature dependence of rna synthesis parameters in *Escherichia coli*. *J. Bacteriol.* 151, 879–887.

Schaper, S., and Messer, W. (1995). Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* 270, 17622–17626. doi: 10.1074/jbc.270.29.17622

Speck, C., Weigel, C., and Messer, W. (1999). ATP–and ADP–dnaA protein, a molecular switch in gene regulation. *EMBO J.* 18, 6169–6176. doi: 10.1093/emboj/18.21.6169

Weigel, C., Schmidt, A., Rückert, B., Lurz, R., and Messer, W. (1997). DnaA protein binding to individual dnaa boxes in the *Escherichia coli* replication origin, oric. *EMBO J.* 16, 6574–6583. doi: 10.1093/emboj/16.21.6574

Xie, X. S., Choi, P. J., Li, G.-W., Lee, N. K., and Lia, G. (2008). Single-molecule approach to molecular biology in living bacterial cells. *Annu. Rev. Biophys.* 37, 417–444. doi: 10.1146/annurev.biophys.37.092607.174640

Zhang, G., and Ignatova, Z. (2009). Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS ONE* 4:e5036. doi: 10.1371/journal.pone.0005036

Zhang, H., Tang, Y., Lee, S.-J., Wei, Z., Cao, J., and Richardson, C. C. (2016). Binding affinities among DNA helicase-primase, DNA polymerase, and replication intermediates in the replisome of bacteriophage T7. *J. Biol. Chem.* 291, 1472–1480. doi: 10.1074/jbc.M115.698233

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.