Towards a whole-cell modeling approach for synthetic biology

Oliver Purcell, Bonny Jain, Jonathan R. Karr, Markus W. Covert, and Timothy K. Lu

Citation: Chaos: An Interdisciplinary Journal of Nonlinear Science **23**, 025112 (2013); doi: 10.1063/1.4811182
View online: http://dx.doi.org/10.1063/1.4811182
View Table of Contents: http://scitation.aip.org/content/aip/journal/chaos/23/2?ver=pdfcov
Published by the AIP Publishing

Towards a whole-cell modeling approach for synthetic biology

Oliver Purcell,[2,a)] Bonny Jain,[1,a)] Jonathan R. Karr,[3] Markus W. Covert,[4] and Timothy K. Lu[1,2,b)]
[1]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[2]*Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[3]*Graduate Program in Biophysics, Stanford University, 318 Campus Drive, Stanford, California 94305, USA*
[4]*Department of Bioengineering, Stanford University, 318 Campus Drive, Stanford, California 94305, USA*

(Received 22 January 2013; accepted 28 May 2013; published online 13 June 2013)

Despite rapid advances over the last decade, synthetic biology lacks the predictive tools needed to enable rational design. Unlike established engineering disciplines, the engineering of synthetic gene circuits still relies heavily on experimental trial-and-error, a time-consuming and inefficient process that slows down the biological design cycle. This reliance on experimental tuning is because current modeling approaches are unable to make reliable predictions about the *in vivo* behavior of synthetic circuits. A major reason for this lack of predictability is that current models view circuits in isolation, ignoring the vast number of complex cellular processes that impinge on the dynamics of the synthetic circuit and vice versa. To address this problem, we present a modeling approach for the design of synthetic circuits in the context of cellular networks. Using the recently published whole-cell model of *Mycoplasma genitalium*, we examined the effect of adding genes into the host genome. We also investigated how codon usage correlates with gene expression and find agreement with existing experimental results. Finally, we successfully implemented a synthetic Goodwin oscillator in the whole-cell model. We provide an updated software framework for the whole-cell model that lays the foundation for the integration of whole-cell models with synthetic gene circuit models. This software framework is made freely available to the community to enable future extensions. We envision that this approach will be critical to transforming the field of synthetic biology into a rational and predictive engineering discipline.
 [http://dx.doi.org/10.1063/1.4811182]

Synthetic biology aims to engineer synthetic genetic circuits to endow cells and organisms with the ability to address new applications, including the production of drugs and industrial products, the design of new therapeutics for human diseases, and the study of basic biological processes. Despite significant advances in the field over the last decade, the synthetic biology design cycle has been hampered by the lack of predictive models. In contrast, more established engineering disciplines, such as civil and electrical engineering can rely on rational design by using models that can capture the behavior of real-world systems with a high degree of accuracy. This is currently not the case for synthetic biology, as models are generally only able to make qualitative predictions about system behavior. As a result, producing a functional engineered biological system usually requires extensive manual tuning by trial-and-error, a time-consuming process that significantly slows the biological design process. The lack of predictive power in current models is partly because these models account for only the synthetic circuits themselves, but not other ongoing processes in the cells in which they reside. However, the complex dynamics of synthetic circuits may affect the host cell and vice versa, leading to divergence between current models and experimental results. Recently, a whole-cell model of a small and simple bacterium was developed and shown to capture numerous aspects of this organism's behavior. Here, we adapt this model to enable the easy incorporation of synthetic circuits and investigate its use for design in synthetic biology. We show that synthetic circuits can be integrated into the model and demonstrate the effects that synthetic genes can have on the host cell. We anticipate that this whole-cell modeling approach for synthetic gene circuits will enable more predictive and rational design for the field of synthetic biology.

I. INTRODUCTION

Synthetic biology promises to revolutionize many areas of technology, including bio-manufacturing, therapeutics and diagnostics.[1,2] To meet this promise, synthetic biology must become a rigorous engineering discipline based on rational and predictive design.[3–6] Modeling efforts within synthetic biology have so far focused on synthetic components and their interactions, where more detailed models have tried to refine these descriptions.[7] For example, Tigges *et al.*[8] considered how opposing RNA polymerases interact, while Stricker *et al.*[9] incorporated DNA looping into their

[a)]O. Purcell and B. Jain contributed equally to this work.
[b)]Electronic mail: timlu@mit.edu

model of transcriptional regulation. The increasing use of stochastic models as a means of describing synthetic gene circuits further intends to improve these network-level descriptions.[9–12]

Interactions between the host cell and synthetic circuits have been recognized since the observation that oscillations in the repressilator stopped upon entry of the cell into stationary phase.[10] More recently, interactions between the host cell and synthetic circuits have been used to uncover novel aspects of host physiology[13] and population-level altruism[14,15] and to guide design. These interactions are also an important consideration when porting circuits between organisms.[16] However, modeling could be further used to accelerate synthetic biological design if it can account for the critical role of the cellular host in a comprehensive and systematic way, the state (e.g., mass, volume, energy status, metabolite concentrations) of which governs the parameters of the synthetic circuit. Consequently, biological circuit design has primarily relied on *ad hoc* engineering, requiring extensive experimental tuning, with *post hoc* modeling rather than predictive models and systematic design rules.

Many synthetic circuits have been modeled in isolation or only in the context of simplified host parameters due to the lack of comprehensive host models. This is a highly simplified approach for designing synthetic gene circuits that are expected to function inside of living cells. However, this situation is beginning to change. Recently, Karr *et al.*[17] developed the first "whole-cell" model of *Mycoplasma genitalium*, which attempts to describe this organism in its entirety. *M. genitalium* is an ideal candidate for a whole-cell model because it has one of the smallest known genomes, containing only 525 genes.[17,18] The model comprises 28 sub-models of distinct cellular processes including transcription, DNA repair, metabolism, and cell division. Each sub-model is implemented using the most appropriate mathematical representation. For example, metabolism is modeled using flux-balance analysis,[19] whereas RNA degradation and protein degradation are modeled as Poisson processes.[17] The model independently executes the sub-models for 1 s of real time and integrates the inputs and outputs of the sub-models at longer time-scales. In total, the whole cell model comprises 1423 chemical reactions and processes (e.g., formation of protein complexes) and approximately 1800 parameters. Each simulation starts at the beginning of the cell cycle and continues until the *in silico* cell "divides." The whole-cell model captures the intracellular dynamics of *M. genitalium* remarkably well; it predicts previously observed gene essentiality with 79% accuracy and qualitatively agrees with measured metabolite concentrations, reaction fluxes, and RNA expression levels.[17] Here, we adapted the whole-cell model of *M. genitalium* to enable the straightforward inclusion of artificial synthetic gene constructs into the *in silico* simulation. This updated software framework is made freely available to the community.

Towards a whole-cell modeling approach for synthetic biology, we investigate a critically important question: How do synthetic gene circuits affect the host cell? This question is relevant not only to the synthetic biology community but to the wider biological community because expression of non-native genes and genetic engineering are ubiquitously used throughout basic biological research, medicine, and biological engineering.

We used the *M. genitalium* whole-cell model to investigate the interactions between synthetic gene circuits and their hosts. First, we added *lacI* genes to the whole-cell model to study the effects of synthetic genes on the host cell and found a clear relationship between the number of genes added and the length of the cell cycle. We found that a substantial portion of this effect was due to the increased time required to replicate the DNA, and that another probable factor was the effect on the production of mRNAs from the native genome. We then used the whole-cell model to examine the impact of codon usage on gene expression. Codon optimization is a simple and commonly used approach to attempt to increase protein expression. We chose to investigate codon usage to test the predictive capacity of whole-cell models. Consistent with experimental findings,[20] the model predicts little difference in the protein expression levels of codon-optimized and non-optimized genes but captures growth differences in the host cell. Finally, we implemented a Goodwin oscillator in the whole-cell model. We found that the oscillator displayed the characteristic noisy and irregular dynamics of a Goodwin oscillator.

II. THE EFFECTS OF SYNTHETIC GENES ON THE HOST CELL

We sought to investigate how synthetic circuits might affect the dynamics and behavior of the host cells in which they reside. Furthermore, we wanted to see how these effects vary with circuit size. As a proxy for complex gene circuits, we added varying numbers of *lacI* genes, with identical promoters (promoters having the exact same DNA sequence), in tandem. Terminator sequences (sequences that halt the process of transcription and demarcate the "end" of the transcriptional unit) were not included into the genome (see Appendix for details). We used *lacI* as our model gene, as it is extensively studied[21] and a commonly used gene in the implementation of synthetic gene networks.[9,10,22,23] The most compelling effect was on the length of the cell cycle. Figure 1(a) (solid line) shows the relationship between the number of *lacI* genes added (which correlates approximately linearly with the relative fraction of LacI protein, see supplementary information[24] Figure 1) and the length of the cell cycle. The length of the cell cycle increases quickly with the addition of the first few genes and continues to increase up to 100 added genes. We saw similar results for the genes *araC*, *tetR*, and *gfp* (also well-understood genes which are commonly used in synthetic biology[10,25–27]) (see supplementary information[24] Figure 2) and similar relationships (both quantitatively and qualitatively) between the number of genes added and the relative fraction of the protein (see supplementary information[24] Figure 1). To ease comparison between results for *lacI* and *araC, tetR*, and *gfp,* we artificially set all proteins to form tetramers (a complex of four proteins) in all the studies reported here.

We then determined the contribution that the replication of additional DNA made to the increased cell cycle time.

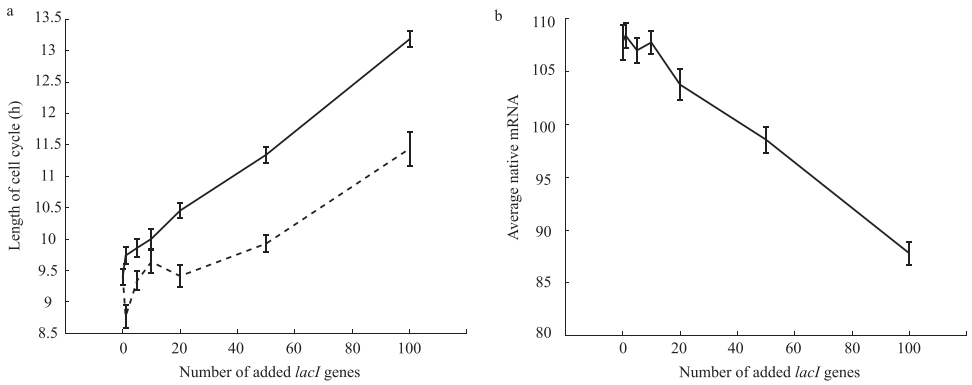

FIG. 1. *Additional genes increase the length of the cell cycle*. (a) The relationship between the number of additional expressed (solid line) and non-expressed (dashed line) *lacI* genes and the cell cycle length is shown. Cell cycle length (hours) is plotted against the number of added genes, for 1, 5, 10, 20, 50, and 100 genes added. (b) The mean (per time point, averaged over the length of the cell cycle) number of mRNAs produced from the native genome is plotted against the number of added expressed *lacI* genes for 1, 5, 10, 20, 50, and 100 genes added. 25 *in silico* cell simulations were performed for each circuit size. Bars indicate the standard error of the mean.

This was done by preventing the *lacI* genes from being transcribed in the model, thus eliminating the production of LacI protein. Figure 1(a) (dashed line) shows the cell cycle time when the *lacI* genes were non-expressed. The relationship between the number of non-expressed *lacI* genes and the cell cycle length is similar to that of expressed *lacI*—the cell cycle length generally increases with the addition of non-functional *lacI*. However, the magnitude of the increase is not as great as with expressed *lacI* genes.

We also examined the effect on the extra expressed *lacI* genes on the number of mRNAs produced from the native genome. We found that as the number of expressed *lacI* genes increased, the mean (per time point, averaged over the length of the cell cycle) number of mRNAs produced from the native genome decreased (Figure 1(b)). We saw a similar relationship for *araC*, *tetR*, and *gfp* (see supplementary information[24] Figure 3).

III. THE EFFECTS OF CODON USAGE ON SYNTHETIC GENE EXPRESSION

Codon optimization is a common method adopted in attempts to increase gene expression for the purposes of large-scale protein purification. From the perspective of synthetic biology, codon optimization is a potential way of tuning the dynamics of a synthetic gene network as well as the burden it imposes on the cell. We used the whole-cell model to investigate how codon optimization of four commonly used genes in synthetic biology, *lacI*, *araC*, *tetR*, and *gfp*, affects their expression. We used the JCat online codon

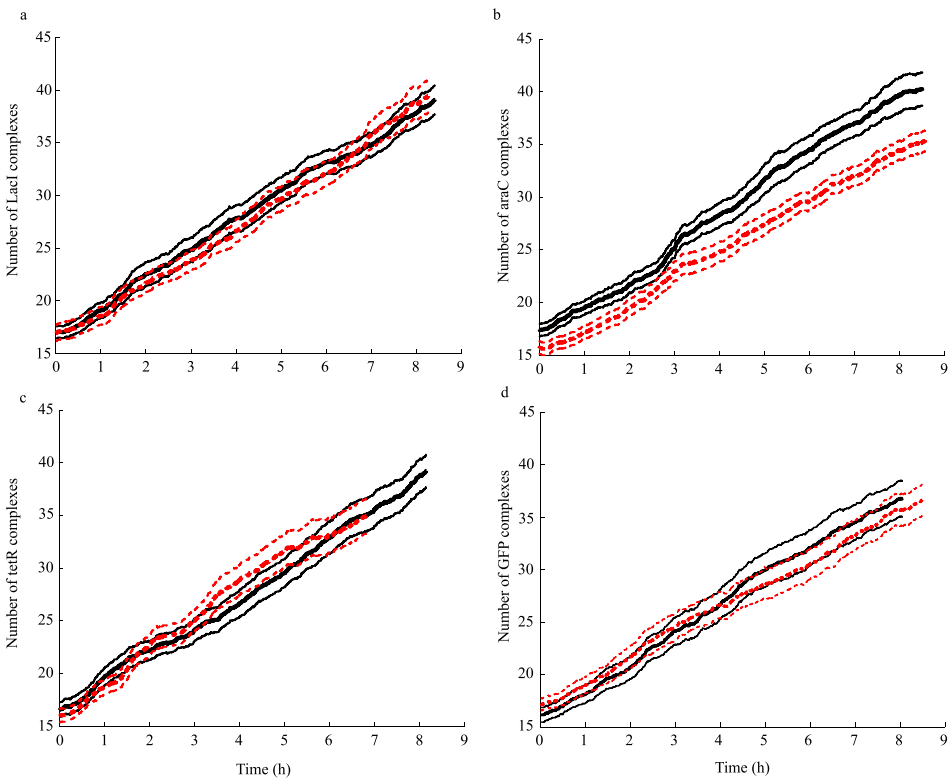

FIG. 2. *Codon usage minimally affects synthetic gene expression.* Black full lines indicate the optimized gene; red dotted lines indicate the unoptimized gene. (a) *lacI*, (b). *araC*, (c) *tetR*, (d) *gfp*. Thick lines indicate the average; thin lines represent the standard error of the mean. Results are truncated at the length of the shortest cell cycle length. In all cases, complexes denote tetramers to ease comparisons between experiments. Results are an average of 25 *in silico* simulations.

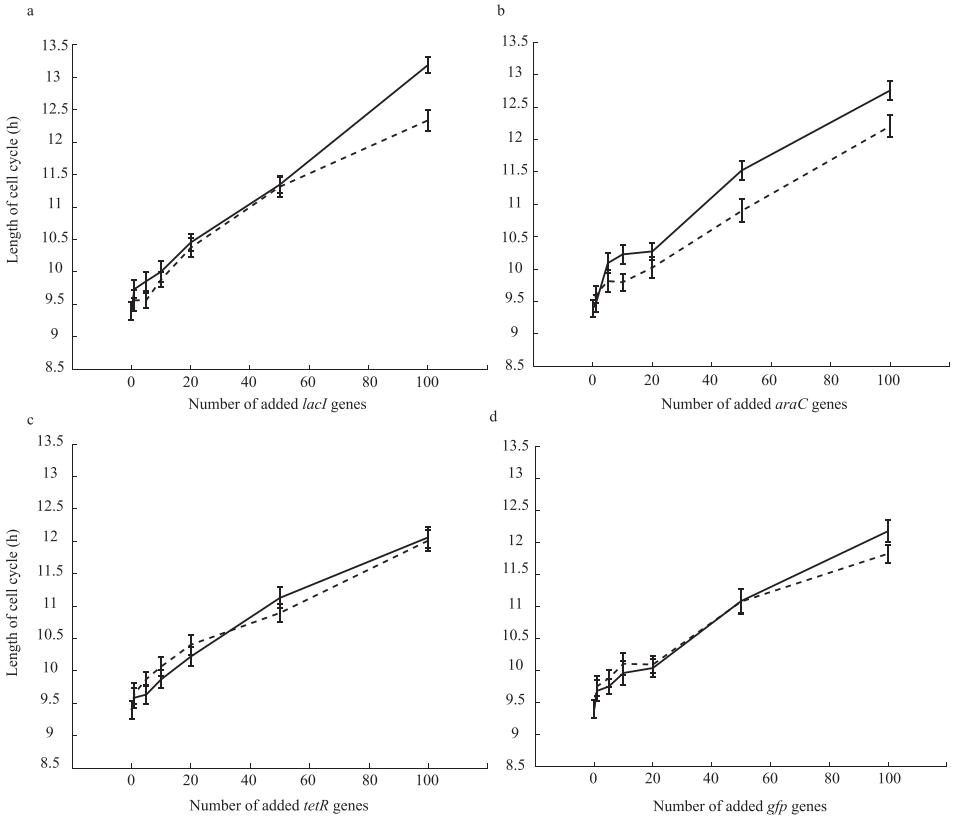

FIG. 3. *Codon optimization of synthetic genes differentially affects cell cycle length.* The mean cell cycle length is plotted for *M. genitalium* with the unoptimized gene (solid line), and the codon-optimized gene (dashed line), for 1, 5, 10, 20, 50, and 100 genes added. (a) *lacI*, (b) *araC*, (c) *tetR,* (d) *gfp*. 25 *in silico* cell simulations were performed for each circuit size. Bars indicate the standard error of the mean.

optimization tool[28] to re-encode the gene sequences for expression in *M. genitalium* (see supplementary information[24] for sequences). Briefly, JCat calculates the Codon Adaption Index (CAI) of the sequence for a particular organism; higher numbers are considered more optimal. Our re-encoded *lacI*, *araC*, *tetR*, and *gfp* genes had CAI's of 0.99, 0.96, 0.98, and 0.99, compared to 0.15, 0.18, 0.41, and 0.57 for the unoptimized sequences, respectively. Figure 2 shows the expression profiles of the optimized and unoptimized genes. In all cases, the protein expression profiles increase at approximately the same rate. This is in agreement with experimental observations in *E. coli,* where the CAI has no correlation with gene expression level.[20,29] However, experimental data have confirmed that the growth rate of *E. coli* correlates with the CAI; the higher the CAI, the faster the growth.[20] We examined the effect of the codon-optimized versions of our genes on the length of the cell cycle (Figure 3). For *lacI* and *araC* (Figures 3(a) and 3(b), respectively), which had differences in the CAI's of the original and optimized genes of 0.84 and 0.78, respectively, we found agreement with the experimental observations. Optimized *lacI* and *araC* had a lesser effect on the growth of the cell than their unoptimized counterparts (mean difference across all data points was 0.22 and 0.29 h, respectively). Conversely, for *tetR* and *gfp* (Figures 3(b) and 3(c), respectively), which had lower differences in their CAI's (0.57 and 0.42, respectively), the difference between the growth rates was approximately zero (mean difference across all data points was −0.06 and −0.01 h, respectively).

We performed a statistical analysis of the data using paired t-tests to compare the effects of the optimized and unoptimized genes on cell cycle lengths. We carried out 1-tailed paired t-tests to test the hypothesis that the mean cell cycle length with the unoptimized gene is greater than the mean cell cycle length with the optimized gene. In the case of *lacI* and *araC*, we found the cell cycle length with the unoptimized gene to be significantly greater than the cell cycle length with the optimized gene (p-values of 0.046 and 0.015, respectively). We found no significant difference in the cases of *tetR* and *gfp* (p-values of 0.8 and 0.53, respectively). Table I summarizes the results of the t-tests.

IV. GOODWIN OSCILLATOR DYNAMICS IN THE WHOLE-CELL MODEL

We implemented a Goodwin oscillator[9,30] in the context of the *M. genitalium* whole-cell model. This was done to demonstrate that a simple synthetic gene circuit could be made to function in the whole-cell model. A Goodwin oscillator is a simple genetic oscillator comprising a single gene that represses its own expression. The Goodwin oscillator was conceived 50 years ago[30] but only recently implemented as a gene network in *E. coli*.[9] We used the *lacI* sequence that is not optimized for *M. genitalium* to construct an *in silico* Goodwin oscillator within the whole-cell model (Figure 4).

TABLE I. Summary of t-test p-values.

Comparison of optimized and unoptimized genes	
gene	p-value
lacI	0.046 (1-tailed)
araC	0.015 (1-tailed)
tetR	0.8 (1-tailed)
gfp	0.53 (1-tailed)

Figure 5(a) shows the dynamics of the oscillator for two representative *in silico* cell simulations (see supplementary information[24] Figure 4 for additional *in silico* cell simulations). To ease comparisons between our simulations and experimental observations from *E. coli*[9] (Figure 5(c)), Figure 5(b) shows the simulations from Figure 5(a) with a low-pass filter applied. The simulations agree qualitatively with experimental observations in *E. coli*;[9] oscillations are irregular and noisy.

V. DISCUSSION

In this study, we used a whole-cell model of *M. genitalium* to investigate several fundamental questions in synthetic biology. Specifically, we presented a number of simple examples to demonstrate the potential uses of whole-cell models in synthetic biology and to study the effects of artificial gene circuits on their hosts.

We first investigated how the addition of genes into the genome affected the dynamics of the cell. This is a proxy for understanding how synthetic circuits, formed from externally added genes, may influence host cell behavior. We found a clear positive relationship between the number of genes added and the length of the cell cycle which was consistent for all four genes examined. Externally added genes commonly affect the growth rate of the host[20,32] although the magnitude of the effect varies between organisms.[33] The relationship between the number of *lacI* genes added and the length of the cell cycle was approximately linear. As the fraction of the total cell protein comprised of LacI was approximately linear with the number of *lacI* genes added, there is an approximately linear relationship between the fraction of total cell protein comprising LacI and the cell cycle length. This linear relationship is in agreement with results seen with the overexpression of genes in *E. coli,* where it has been observed that growth rate corresponds approximately linearly with the fraction of both β-galactosidase or an inactive form of the elongation factor TU.[34]

By performing the same analysis using non-expressed *lacI* genes, we delineated the contribution to the additional cell cycle time that was caused by the increased time required to replicate the DNA containing the *lacI* genes. Additional DNA replication time was found to account for approximately half of the increased cell cycle time. We then

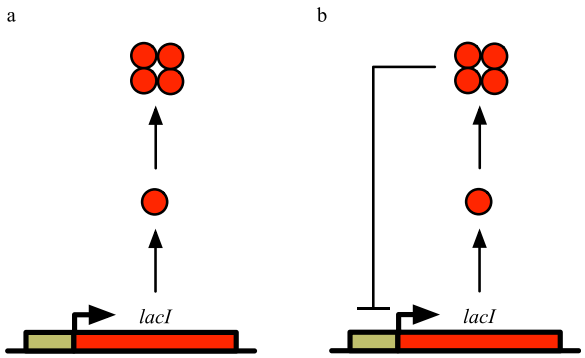

FIG. 4. *Topology of the Goodwin oscillator*. (a) LacI forms a tetramer. (b) The Goodwin oscillator is formed by LacI repressing transcription from its own promoter. Adapted from Ref. 31.

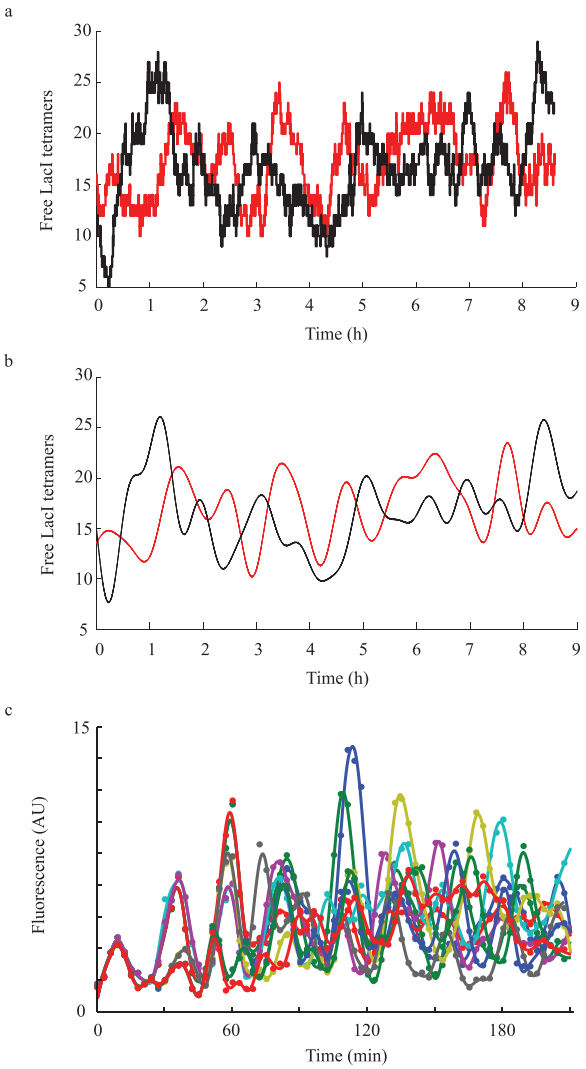

FIG. 5. *Whole-cell Goodwin oscillator displays noisy and irregular oscillations*. (a) LacI tetramer copy number dynamics in two *in silico* cells, truncated to the length of the shortest cell cycle. (b) Simulations from (a) with a low-pass filter applied (frequencies greater than 0.001 Hz were removed). (c) Experimental *in vivo* data of the Goodwin oscillator, modified from Stricker *et al.*[9] [Reproduced with permission from J. Stricker *et al.*, "A fast, robust and tunable synthetic gene oscillator," Nature **456**, 516–U39 (2008). Copyright 2008 Nature Publishing Group] Each line depicts oscillatory data from a separate *E. coli* cell.

examined the effect of the additional genes on the level of mRNAs produced from the native genome. It might be expected that the additional genes sequester cellular proteins and resources, slowing down cellular processes and therefore growth. Reduced mRNA levels, due to reduced transcription, would be a clear indication of this sequestration. We found that for all genes, there was indeed a clear effect, with the additional genes causing the number of mRNAs produced from the native genome to decrease. This behavior indicates that one of the effects of adding an increasing number of expressed genes is to reduce the overall cellular resources that can be dedicated to transcription of the native genes. Other likely contributions to the observed growth effects could be additional genes diverting limited nucleic acids and amino acids from other DNA, RNA, and proteins in the cell. This would lead to the lower production of metabolic enzymes, RNA polymerase, and ribosomes, resulting in a

lower growth rate.[35–38] It should be noted that in our studies, the expression of added genes was always on and thus, these results do not capture the fact that in many complex synthetic circuits, only a subset of the genes will be expressed at any given time.

We then examined the effect of codon usage on the expression profile of four genes commonly used in synthetic biology: *lacI*, *araC*, *tetR*, and *gfp*. In agreement with experimental results from *E. coli* (on codon usage in *gfp*),[20] we observed no difference in the rates of expression between unoptimized and optimized versions of the genes, despite substantial differences in their CAI's. In further agreement with the findings of Kudla *et al.*,[20] we found that codon optimization could have a positive effect on the growth rate. We only observed this difference for *lacI* and *araC*, where there was the greatest difference in the CAI values between the unoptimized and optimized genes. We observed no difference in the growth rate of unoptimized and optimized *tetR* or *gfp*, which had lower differences in CAI values. A gene with a higher CAI will use codons that are translated by aminoacylated-tRNAs that are more abundant within the cell. In the model, translation is affected by the availability of aminoacylated-tRNA. By using an added gene with a higher CAI, the expression of the added gene is less likely to use the less abundant aminoacylated-tRNA species, and these are therefore less likely to becoming limiting for the translation of genes in the native genome. As the expression of the genes in the native genome is directly related to the growth of the cell, an added gene with a higher CAI will have less effect on growth. This proposed mechanism is a theory, and further analysis is required to understand the relationship between the changes in the level of optimization and the corresponding growth effects. However, the relationship may be non-linear, with smaller differences in optimization causing negligible growth effects. If the observed optimization effects can be verified experimentally in *M. genitalium*, these findings would further demonstrate the predictive power of the whole-cell model. Finally, we described the successful implementation of a Goodwin oscillator within the whole-cell model and found that its behavior qualitatively matched the *in vivo* dynamics observed in *E. coli*.[9]

Integrating synthetic gene circuits within accurate whole-cell models has the potential to enable more rapid and reliable design of biological systems with less trial and error. We believe that whole-cell models will ultimately aid synthetic biologists in understanding and describing why synthetic circuits do not behave *in vivo* as expected from the predictions of isolated computational models.[10,39] This, in turn, should lead to models that can more accurately predict and thus enable the rational design of synthetic circuits. We intend this work to be taken as a preliminary investigation into the use of whole-cell models within synthetic biology. A caveat to the work presented here is that without complementary *in vivo* data, we cannot directly test the *in silico* predictions made with the whole-cell model. Nonetheless, given the previous accuracy of this whole-cell model in Karr *et al.* (e.g., its qualitative agreement with measured metabolite concentrations, reaction fluxes, and RNA expression levels)[17] and the consistency of our findings with existing experimental data, we hope that this effort may launch new studies to compare experimental results with model predictions using this approach. Such efforts will enable the further optimization of whole-cell synthetic circuit models for more rapid and accurate design.

VI. FUTURE OUTLOOK

There are many intriguing questions that a whole-cell approach to modeling synthetic gene circuits could be used to address. Ideally, any future investigations would be closely accompanied with experimental work, with the aim of comparing model predictions with experimental results and improving the accuracy of this approach. *M. genitalium* is not an ideal model organism; it is a small organism with a long doubling time, has few known genetic tools, and no known defined medium. However, *Mycoplasma* have been successfully engineered in prior work—for instance, with functioning GFP and TetR proteins[40,41] and on a much larger genome-wide scale in the recent work of Gibson *et al.*[42]

This study has focused on one side of the relationship between a synthetic circuit and the host—the effect of the synthetic circuit on the host. We have yet not examined the other side of the relationship—the effect of host processes on the behavior of the synthetic circuit. Ultimately, the effects between the host and the synthetic circuit are intertwined and feed back on each other, and whole-cell models have the potential to uncover details of these likely complex relationships.

The whole-cell model necessarily makes numerous assumptions and simplifications. A current limitation of the whole-cell model for use in synthetic biology is the absence of non-specific binding of transcription factors to DNA, although the non-specific binding of other proteins, such as RNA polymerase, is captured in the model. Though a transcription factor may have a single specific binding site, it is thought that transcription factors that are not bound specifically are likely to be bound non-specifically, elsewhere in the genome.[43,44] The remainder of the genome therefore acts a sink for transcription factors and consequently has important implications for the dynamics of gene circuits.[43,44] For synthetic biology, a useful extension to the whole-cell model would therefore be the ability to define non-specific binding for added transcription factors.

The use of whole-cell models for synthetic biology is in its infancy, and thus advantages to using such models for biological engineering will need to be demonstrated with experimental data. Nonetheless, whole-cell models are likely to form the backbone of future modeling approaches in synthetic biology and will help transform the field into a more predictive engineering discipline. Within the workflow of the synthetic biology design and construction process, we envisage whole-cell models will serve as advanced biophysical models that will exist in-between higher-level graphical design tools such as Clotho[45] and GenoCAD[46] and the physical implementation of the circuit in the cell. The software advancements made here allow for synthetic circuits to be easily integrated into the whole-cell model, allowing users to programmatically add and remove model components including genes,

transcription units, protein monomers, complexes, reactions, and regulatory interactions (see Appendix). Further improvements of the model may incorporate transcriptional,[47,48] RNA-based,[49,50] and recombinase-based circuitry.[51] We hope this will encourage the application of the whole-cell modeling approach to synthetic biology. Beyond synthetic biology, we anticipate whole-cell models will be useful for interpreting experiments and guiding biological discovery.

ACKNOWLEDGMENTS

We would like to thank Vladimir Potapov and Amy Keating for their help with running simulations on their cluster (supported by National Science Foundation under Grant No. 0821391). We thank Jacob Rubens for helpful discussion regarding our results on codon optimization. Bonny Jain was supported by the MIT Electrical Engineering and Computer Science Advanced Undergraduate Research Program, Oliver Purcell was supported by the Defence Advanced Research Projects Agency (DARPA), Jonathan Karr was supported by a NSF Graduate fellowship, Markus Covert was supported by an NIH Director's Pioneer Award (8DP1LM011510-04), and Timothy Lu was supported by an NIH New Innovator Award (1DP2OD008435) and the National Science Foundation (1124247).

APPENDIX: IMPLEMENTAION OF SYNTHETIC CIRCUIT WHOLE-CELL MODELS

1. Whole-cell models

The *Mycoplasma* whole-cell model was implemented as described previously.[17] The knowledge base Application Programming Interface (API) was improved to enable users to programmatically add and remove model components in a straightforward fashion, including genes, transcription units, protein monomers, complexes, reactions, and regulatory interactions. In addition, we created new a class called SimulationRunner to encapsulate all of the model components and parameter modifications required for specific simulations and to separate these perturbations from the submodel classes. Researchers can use the revised code to easily simulate perturbations including gene additions and deletions.

2. Synthetic circuit whole-cell models

The *lacI* knock-in model was implemented by adding a single protein-coding gene to the MATLAB *M. genitalium* whole-cell model.[17] First, we appended the *E. coli* K-12 W3110 *lacI* (b0345) DNA sequence with start and stop codons modified for *M. genitalium* as well as 100 immediate upstream random nucleotides (which includes the promoter) to the end of the *M. genitalium* chromosome at position 580077 (see supplementary information[24] for sequences). Second, we defined a new promoter with 6 nucleotide (nt) −35 and −10 boxes at the −35 and −10 positions relative to *lacI*. The *lacI* promoter RNA polymerase binding probability was set equal to the 0.001 mean RNA polymerase mRNA promoter binding probability, and the *lacI* mRNA half-life was set equal to the 4.53 min mean mRNA half-life. The half-lives of both the LacI monomers and tetrameric complexes were set to 20 h (the default half-life for proteins in the whole-cell model). The same procedure and parameters were used for the genes *araC*, *tetR* and *gfp*.

The codon-optimized *lacI* knock-in model was implemented similarly using the codon-optimized *lacI* sequence calculated by JCat[28] (see supplementary information[24] for sequence). The tandem repeat *lacI* knock-in model was also implemented similarly by repeatedly appending the *E. coli* *lacI* gene and immediate upstream region to the end of the *M. genitalium* chromosome 20, 50, and 100 times. The same procedure was used for the genes *araC*, *tetR* and *gfp*.

The Goodwin oscillator model was implemented by first adding *lacI* to the whole-cell model as described above. Next, we implemented LacI auto-repression by adding a 10 nt LacI tetramer DNA-binding site at the −40 position relative to the *lacI* gene and setting the fold-change effect of bound LacI to the *lacI* RNA polymerase recruitment rate to 10%. Finally, because the whole-cell model only simulates a single ∼9 h cell cycle, oscillations were not observed when using a 20 h half-life. We, therefore, decreased the half-lives of the LacI protein monomers and complexes from 20 h to 24 min. A short half-life may be obtainable using ssrA tags, commonly employed to increase protein degradation rates in oscillators implemented in *E. coli*[9,10] and which are also present in *Mycoplasma*.[52]

The interested reader is referred to Data S1, Sec. 1.4 of Karr *et al.*[17] for further discussion of the whole-cell model initial conditions.

3. Numerical simulations and data analysis

All simulations and data analysis were performed using MATLAB 2012b. Each whole-cell simulation typically took between 20–30 h, and 2 h, respectively. Statistical analysis was carried out using MATLAB 2012b.

4. Availability

All of the simulation code and numerical results are freely available open-source at SimTK (http://simtk.org/home/wholecell).

[1]A. A. Cheng and T. K. Lu, "Synthetic biology: An emerging engineering discipline," Annu. Rev. Biomed. Eng. **14**, 155–178 (2012).
[2]A. S. Khalil and J. J. Collins, "Synthetic biology: Applications come of age," Nat. Rev. Genet. **11**, 367–379 (2010).
[3]C. M. Agapakis and P. A. Silver, "Synthetic biology: Exploring and exploiting genetic modularity through the design of novel biological networks," Mol. Biosyst. **5**, 704–713 (2009).
[4]A. Arkin, "Setting the standard in synthetic biology," Nat. Biotechnol. **26**, 771–774 (2008).
[5]D. Endy, "Foundations for engineering biology," Nature **438**, 449–453 (2005).
[6]Y. Y. Chen, K. E. Galloway, and C. D. Smolke, "Synthetic biology: Advancing biological frontiers by building synthetic systems," Genome Biol. **13**, 240 (2012).
[7]N. J. Guido *et al.*, "A bottom-up approach to gene regulation," Nature **439**, 856–860 (2006).
[8]M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger, "A tunable synthetic mammalian oscillator," Nature **457**, 309–312 (2009).
[9]J. Stricker *et al.*, "A fast, robust and tunable synthetic gene oscillator," Nature **456**, 516–U39 (2008).
[10]M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," Nature **403**, 335–338 (2000).

[11]M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: From theories to phenotypes," Nat. Rev. Genet. **6**, 451–464 (2005).
[12]M. Yoda, T. Ushikubo, W. Inoue, and M. Sasai, "Roles of noise in single and coupled multiple genetic oscillators," J. Chem. Phys. **126**, 115101 (2007).
[13]P. Marguet, Y. Tanouchi, E. Spitz, C. Smith, and L. You, "Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology," PLoS ONE **5**, e11909 (2010).
[14]Y. Tanouchi, A. Pai, N. E. Buchler, and L. You, "Programming stress-induced altruistic death in engineered bacteria," Mol. Syst. Biol. **8**, 626 (2012).
[15]A. Pai, Y. Tanouchi, C. H. Collins, and L. You, "Engineering multicellular systems by cell-cell communication," Curr. Opin. Biotechnol. **20**, 461–470 (2009).
[16]D. Nevozhay, T. Zal, and G. Balázsi, "Transferring a synthetic gene circuit from yeast to mammalian cells," Nat. Commun. **4**, 1451 (2013).
[17]J. R. Karr *et al.*, "A whole-cell computational model predicts phenotype from genotype," Cell **150**, 389–401 (2012).
[18]C. M. Fraser *et al.*, "The minimal gene complement of *Mycoplasma genitalium*," Science (N.Y.) **270**, 397–403 (1995).
[19]J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?" Nat. Biotechnol. **28**, 245–248 (2010).
[20]G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin, "Coding-sequence determinants of gene expression in *Escherichia coli*," Science (N.Y.) **324**, 255–258 (2009).
[21]M. Lewis, "The lac repressor," C. R. Biol. **328**, 521–548 (2005).
[22]T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in *Escherichia coli*," Nature **403**, 339–342 (2000).
[23]M. R. Atkinson, M. A. Savageau, J. T. Myers, and A. J. Ninfa, "Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*," Cell **113**, 597–607 (2003).
[24]See supplementary material at http://dx.doi.org/10.1063/1.4811182 for lacI sequences used and additional Goodwin oscillator simulations.
[25]C. Berens and D. Porschke, "Recognition of operator DNA by Tet repressor," J. Phys. Chem. B **117**, 1880–1885 (2013).
[26]S. J. Remington, "Green fluorescent protein: A perspective," Protein Sci. **20**, 1509–1519 (2011).
[27]R. Schleif, "AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action," FEMS Microbiol. Rev. **34**, 779–796 (2010).
[28]A. Grote *et al.*, "JCat: A novel tool to adapt codon usage of a target gene to its potential expression host," Nucleic Acids Res. **33**, W526–W531 (2005).
[29]M. Welch *et al.*, "Design parameters to control synthetic gene expression in *Escherichia coli*," PLoS ONE **4**, e7002 (2009).
[30]B. C. Goodwin, *Temporal Organization in Cells. A Dynamic Theory of Cellular Control Processes* (Academic Press, London, 1963).
[31]O. Purcell, N. J. Savery, C. S. Grierson, and M. Di Bernardo, "A comparative analysis of synthetic genetic oscillators," J. R. Soc., Interface **7**, 1503–1524 (2010).
[32]T. Makino, G. Skretas, and G. Georgiou, "Strain engineering for improved expression of recombinant proteins in bacteria," Microb. Cell Factories **10**, 32 (2011).
[33]L. Ma, G. Zhang, and M. P. Doyle, "Green fluorescent protein labeling of listeria, salmonella, and *Escherichia coli* O157:H7 for safety-related studies," PLoS ONE **6**, e18083 (2011).
[34]H. Dong, L. Nilsson, and C. G. Kurland, "Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction," J. Bacteriol. **177**, 1497–1504 (1995).
[35]R. L. Gourse, T. Gaal, M. S. Bartlett, J. A. Appleman, and W. Ross, "rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*," Annu. Rev. Microbiol. **50**, 645–677 (1996).
[36]P. Jorgensen *et al.*, "A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size," Genes Develop. **18**, 2491–2505 (2004).
[37]S. Klumpp and T. Hwa, "Growth-rate-dependent partitioning of RNA polymerases in bacteria," Proc. Natl. Acad. Sci. U.S.A. **105**, 20245–20250 (2008).
[38]T. Moss, "At the crossroads of growth control; making ribosomal RNA," Curr. Opin. Genet. Develop. **14**, 210–217 (2004).
[39]C. Lou *et al.*, "Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch," Mol. Syst. Biol. **6**, 350 (2010).
[40]M. F. Balish, R. T. Santurri, A. M. Ricci, K. K. Lee, and D. C. Krause, "Localization of Mycoplasma pneumoniae cytadherence-associated protein HMW2 by fusion with green fluorescent protein: Implications for attachment organelle structure," Mol. Microbiol. **47**(1), 49–60 (2003).
[41]M. Breton *et al.*, "First report of a tetracycline-inducible gene expression system for mollicutes," Microbiology **156**, 198–205 (2010).
[42]D. G. Gibson *et al.*, "Creation of a bacterial cell controlled by a chemically synthesized genome," Science (N.Y.) **329**, 52–56 (2010).
[43]A. Burger, A. M. Walczak, and P. G. Wolynes, "Abduction and asylum in the lives of transcription factors," Proc. Natl. Acad. Sci. U.S.A. **107**, 4016–4021 (2010).
[44]T.-H. Lee and N. Maheshri, "A regulatory role for repeated decoy transcription factor binding sites in target gene expression," Mol. Syst. Biol. **8**, 576 (2012).
[45]B. Xia *et al.*, "Developer's and user's guide to Clotho v2.0 A software platform for the creation of synthetic biological systems," Methods Enzymol. **498**, 97–135 (2011).
[46]M. J. Czar, Y. Cai, and J. Peccoud, "Writing DNA with GenoCAD™," Nucl. Acids Res. **37**, W40–W47 (2009).
[47]Y. Benenson, "Biomolecular computing systems: Principles, progress and potential," Nat. Rev. Genet. **13**, 455–468 (2012).
[48]R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu, "Synthetic analog computation in living cells," Nature **497**(7451), 619–623 (2013).
[49]Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson, "Multi-input RNAi-based logic circuit for identification of specific cancer cells," Science (N.Y.) **333**, 1307–1311 (2011).
[50]F. Isaacs, D. Dwyer, and J. Collins, "RNA synthetic biology," Nat. Biotechnol. **24**(5), 545–554 (2006).
[51]P. Siuti, J. Yazbek, and T. K. Lu, "Synthetic circuits integrating logic and memory in living cells," Nat. Biotechnol **31**(5), 448–452 (2013).
[52]E. Gur and R. T. Sauer, "Evolution of the ssrA degradation tag in Mycoplasma: Specificity switch to a different protease," Proc. Natl. Acad. Sci. U.S.A. 105, 16113–16118 (2008).