Bioengineered 3:6, 311–312; November/December 2012; © 2012 Landes Bioscience

Digital biology
A new era has begun

Roy D. Sleator

Department of Biological Sciences; Cork Institute of Technology; Bishopstown, Cork, Ireland

"The biggest innovations of the 21st century will be at the intersection of biology and technology. A new era is beginning."
Steve Jobs

Just two years after the creation of Synthia (JCVI-syn1.0),[1,2] synthetic biology welcomes its second citizen; a whole-cell computational model of the uropathogenic bacterium *Mycoplasma genitalium*.[3] However, unlike JCV1-syn1.0—the first self-replicating species on the planet whose parent is a computer—Karr's creation exists only in digital form.[3] This bacterial avatar represents the first truly integrated effort to simulate the complete workings of a free-living microbe in silico.

Despite earlier attempts at creating robust cell scale computational frameworks; including approaches base on ordinary differential equations,[4] Boolean networks[5] and constraint-based models,[6] these approaches are, for the most part, limited to a subset of physiological processes within the overall metabolic context of the cell and thus fail to capture the complete picture.

The model developed by Karr et al.,[3] overcomes this limitation by employing a novel modular design which divides cellular function into modules, each representing a single biological process (e.g., transcriptomics, proteomics, metabolomics, etc.). In all, 28 modules were chosen and sub-models of each were independently built, parameterized and tested. This approach allows the flexibility to model each independent module using the most appropriate algorithm (e.g., flux-balance analysis is used to model metabolism,[7] while protein and RNA degradation are modeled as Poisson processes). The overall process is built on the premise that the sub-models exist independently in timescales of < 1 sec. Simulations are run through a loop in which sub-models interact and exchange variables (specifying the internal state of the cell) at 1 sec intervals. In other words, each sub-model runs independently at each time step but is dependent on variable values determined by sub-models at the previous step.

The overall model is based on over 900 peer reviewed publications and includes more than 1,900 experimentally observed parameters. Model training and parameter reconciliation was achieved by recreating 128 different *M. genitalium* culture simulations—each predicting both molecular and cellular properties of the in silico cell—recapitulating the key features of the training data. Model validation was achieved using data sets not used in the construction of the model and which encompass multiple biological functions (from transcriptomics to metabolomics) and scales (from single cells to microbial populations).

Despite the authors' own assertion that the model is a "first draft;" it has nonetheless revealed a number of important, and hitherto unknown, insights into the *M. genitalium* life-cycle. Of particular note is an entirely new hypothesis which identifies metabolism as an emergent controller of cell-cycle duration, independent of genetic regulation. During the simulations, the authors noticed that although there are high variances in both the time required to initiate DNA replication and the time required to copy the genome post-initiation, there is low variance in the length of the cell cycle. Before a cell can divide, it must make a complete copy of its entire DNA. Copying initiates when the proteins which constitute the replication machinery bind to the origin of replication. However, this process occurs by random diffusion; so that in some instances the proteins will attach quickly and copying will begin while the cell is young, while in other occasions the proteins will attach when the cell is relatively old. The high variance in copying time once the proteins have already bound, however, depends on the age of the cell. If the cell is young, copying periodically stalls because the cell has not had enough time to stockpile the dNTPs required to build DNA. However, if the cell is old, it already has a sufficient stockpile of dNTPs to allow copying to proceed relatively quickly. Therefore, cells that are fast initiators are slow copiers, while those that are slow initiators are fast copiers. The model thus predicts that genomic replication is rate limited by dNTP synthesis, and that cells in which the early stages of the cell cycle are prolonged are able to catch up with those that initiate replication earlier due to the accumulation of a larger dNTP pool at the onset of replication, thus reducing the variance of overall cell-cycle duration within a population—i.e., all *M. genitalium* cells take approximately the same amount of time to divide.

The model further predicts that that the chromosome is explored very rapidly, with 50% of the chromosome being bound by at least one protein within the first 6 min of the cell cycle (and 90% in the first 20 min). On average this results in the expression of 90% of the genes within the first 143 min. The model also predicts protein-protein collisions on the chromosome, with over 30,000 collisions occurring on average per cell cycle (displacing 0.93 proteins per second). Collision frequency corresponds with DNA-bound protein density across the genome with the majority of collisions being caused by RNA polymerase (84%) and DNA

Correspondence to: Roy D. Sleator; Email: roy.sleator@cit.ie
Submitted: 09/26/12; Accepted: 09/26/12
http://dx.doi.org/10.4161/bioe.22367

polymerase (8%), most commonly resulting from the displacement of structural maintenance of chromosome (SMC) proteins (70%) or single-stranded binding proteins (6%).

Finally, in an elegant exhibition of the utility of the in silico modeling approach, the authors performed multiple simulations of each of the 525 possible single-gene disruption strains (over 3,000 simulations in total). This is, in essence, the equivalent of creating a bank of 525 knockout mutants of *M. genitalium* in the wet lab and testing the physiological response of each to a number of different culture conditions. The model predicts 284 essential and 117 non-essential genes—a prediction of gene essentiality which is 79% accurate when compared with previous wet lab investigations into the minimum genome.[8] Furthermore, the model predicts 4 distinct classes of lethal gene mutation: the most debilitating disruptions involve metabolic genes and result in an inability to produce the major cell mass components; a condition which is incompatible with cell growth and division. The next most debilitating disruptions involve specific cell mass components such as RNA or protein—in this case the model predicts a near normal growth followed by a decline due to diminishing protein content. This decline in some cases may take more than one generation to manifest. The third class impairs cell-cycle processes—in this case the model predicts normal growth rates and metabolism but also projects an inability to complete the cell cycle. The fourth and final class includes strains that grow so slowly, when compared with the wild type, that they are considered physiologically nonviable.

In their accompanying editorial in *Cell*, Freddolino and Tavazoie[9] acknowledge the sheer audacity of the multiscale modeling approach taken by Karr et al.,[3] and predict two distinct paths along which these models may ultimately lead us. The first, which they call the "physicist's perspective" predicts that in silico modeling will lead to the discovery of new organizing principles that will ultimately help to frame (or in some cases refocus) our intellectual understanding of biological systems—exemplified by the current model's prediction of metabolism as an emergent cell-cycle regulator. The second, the "engineers perspective" involves the evolution of computational models to the extent that they may ultimately supplant wet lab experimentation—such as the in silico minimal genome experiment in the current study.

However, while the highly sophisticated multiscale model presented by Karr et al.,[3] is a crucial first step in the development of useful cell-scale simulations—there is still a long way to go. A significant failing of the current study is that much of the data used to build and validate the model were obtained from organisms other than *M. genitalium* (largely because it is notoriously difficult to culture—a fact which led to its substitution with the faster growing *M. mycoides* in the creation of JCVI-syn1.0[2]). Proper validation of the approach will therefore require more experimentally tractable organisms such as *Escherichia coli*—the traditional workhorse of the microbiology lab. However, when one considers that the current model takes ~9 to 10 h of compute time (about the time it takes for *M. genitalium* to divide in nature) and generates half a gigabyte of data when simulating a single cell division, then even allowing for Moore's law,[10] up-scaling to the far larger and faster-growing *E. coli* (with a genetic complement of 4,288 genes and a doubling time of just 20 min) is far from trivial. Other regulatory complexities, such as genome-wide antisense transcription,[11] spatial heterogeneity[12] and enzyme multifunctionality,[13] which are not addressed in the current model, will also have to be accounted for when designing the next generation of in silico model organisms.

References

1. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 2010; 329:52-6; PMID:20488990; http://dx.doi.org/10.1126/science.1190719.
2. Sleator RD. The story of Mycoplasma mycoides JCVI-syn1.0: The forty million dollar microbe. Bioeng Bugs 2010; 1:231-2; PMID:21327054; http://dx.doi.org/10.4161/bbug.1.4.12465.
3. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr., et al. A whole-cell computational model predicts phenotype from genotype. Cell 2012; 150:389-401; PMID:22817898; http://dx.doi.org/10.1016/j.cell.2012.05.044.
4. Castellanos M, Kushiro K, Lai SK, Shuler ML. A genomically/chemically complete module for synthesis of lipid membrane in a minimal cell. Biotechnol Bioeng 2007; 97:397-409; PMID:17149771; http://dx.doi.org/10.1002/bit.21251.
5. Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, et al. A genomic regulatory network for development. Science 2002; 295:1669-78; PMID:11872831; http://dx.doi.org/10.1126/science.1069883.
6. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? Nat Biotechnol 2010; 28:245-8; PMID:20212490; http://dx.doi.org/10.1038/nbt.1614.
7. Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, Maranas CD. A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. PLoS Comput Biol 2009; 5:e1000285; PMID:19214212; http://dx.doi.org/10.1371/journal.pcbi.1000285.
8. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, et al. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A 2006; 103:425-30; PMID:16407165; http://dx.doi.org/10.1073/pnas.0510013103.
9. Freddolino PL, Tavazoie S. The dawn of virtual cell biology. Cell 2012; 150:248-50; PMID:22817888; http://dx.doi.org/10.1016/j.cell.2012.07.001.
10. Robison RA. Moore's Law: Predictor and Driver of the Silicon Era. World Neurosurg 2012.
11. Dornenburg JE, Devita AM, Palumbo MJ, Wade JT. Widespread antisense transcription in Escherichia coli. MBio 2010; 1:e00024-10; PMID:20689751; http://dx.doi.org/10.1128/mBio.00024-10.
12. Roberts E, Magis A, Ortiz JO, Baumeister W, Luthey-Schulten Z. Noise contributions in an inducible genetic switch: a whole-cell simulation study. PLoS Comput Biol 2011; 7:e1002010; PMID:21423716; http://dx.doi.org/10.1371/journal.pcbi.1002010.
13. Sleator RD. Proteins: form and function. Bioeng Bugs 2012; 3:80-5; PMID:22095055; http://dx.doi.org/10.4161/bbug.18303.