double-stranded DNA breaks in any sequence for genome targeting or editing.

Several studies have shown that ZFNs and TALENs (**Fig. 1c**,**d**) can be engineered to induce mutations at specific locations in genome sequences[8,9] in plants (*Zea mays*), bacteria (*Escherichia coli*), model animals (*Drosophila melanogaster*), human embryonic stem cell lines and induced pluripotent stem cell lines. However, these proteins have to be redesigned for every different DNA sequence targeted. The 3-nt recognition motif of ZFNs can be promiscuous and generate off-target cleavage, and target frequency has been estimated at ~1 in every 500 bp[8]. The ability to modulate TALEN specificity by modification of the repeat variable di-residues within their amino acid sequence has been an advantage of TALENS compared with ZFNs[9]. However, the nonspecific nuclease domain of FokI in TALENs can generate nonspecific and non-targeted cleavage. Also, the sole product is strictly dsDNA cleavage, and target frequencies have been estimated at ~1 in every 35 bp[8].

Chimeric RNA-Cas9 systems could provide several advantages over ZFNs and TALENs, including sequence-recognition specificity (by customized crRNA sequences), PAM dinucleotides that can be frequent (GG in the case of Jinek *et al.*[1]), the convenience of redesigning nucleotides in nucleic acid sequences as opposed to modifying amino acids in protein sequences, and the ability to nick either or both DNA strands. The latter provides a molecular basis for precise genome editing by a programmable DNA nicking enzyme[10]. Because Cas9 has two distinct domains (RuvC and HNH), which each nick a specific DNA strand, wild-type Cas9 and functional domain mutants can generate either dsDNA breaks or single-stranded DNA nicks, as desired (**Fig. 1b**). Single-stranded breaks are repaired by error-free homologous recombination at the nicking site, producing a precise mutation, as opposed to error-prone, nonhomologous, end-joining repair for dsDNA cleavage, which frequently creates insertions and deletions after ZFN or TALEN cleavage.

Molecular analysis of Cas9 cleavage has shown that both linearized and supercoiled DNA can be cleaved, suggesting that multiple distinct chimeric RNAs can be concurrently or sequentially used to process DNA at multiple target sites using a single enzyme, opening avenues for genome stacking and shuffling. Moreover, specific cleavage sites can be engineered to generate DNA molecules that have custom extremities for iterative genome build-up. Accordingly, crRNA-Cas–directed nicking and cleavage set the stage for more precise DNA surgery and genome editing.

Although immediate applications of this new tool include customized DNA nicking and/or cleavage in bacteria, there are intriguing possibilities for genome editing and genome engineering of eukaryotes. This will require testing whether crRNA-Cas systems can efficiently cleave chromatin DNA *in vivo* and be readily transferred into organisms of interest, notably yeast and fungi, but also plants, for crop and agricultural applications, and human cells, for medical purposes. Only the future will tell whether this programmable molecular scalpel can outcompete ZFN and TALEN DNA scissors for precise genomic surgery.

COMPETING FINANCIAL INTERESTS
The author declares competing financial interests: details are available in the online version of the paper.

1. Jinek, M. *et al. Science* advance online publication, doi:10.1126/science.1225829 (28 June 2012).
2. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. *Nature* **482**, 331–338 (2012).
3. Deltcheva, E. *et al. Nature* **471**, 602–607 (2011).
4. Garneau, J.E. *et al. Nature* **468**, 67–71 (2010).
5. Makarova, K.S. *et al. Nat. Rev. Microbiol.* **9**, 467–477 (2011).
6. Andersson, A.F. & Banfield, J.F. *Science* **320**, 1047–1050 (2008).
7. Sapranauskas, R. *et al. Nucleic Acids Res.* **39**, 9275–9282 (2011).
8. Cermak, T. *et al. Nucleic Acids Res.* **39**, 7879–7889 (2011).
9. Miller, J.C. *et al. Nat. Biotechnol.* **29**, 143–148 (2011).
10. Kim, E. *et al. Genome Res.* **22**, 1327–1333 (2012).

Silicon dreams of cells into symbols

Jeremy Gunawardena

Diverse mathematical models combine to create a comprehensive whole-cell computational model of a human pathogen.

Writing in *Cell*, a team from Stanford University and the J. Craig Venter Institute has reported a whole-cell simulation of *Mycoplasma genitalium*, a urogenital parasite adored by synthetic biologists for its reduced genome, a mere snip at 525 genes. This audacious accomplishment by Karr *et al.*[1] of what has been called "a grand challenge of the 21st century"[2] goes beyond previous attempts[3] in its comprehensiveness and especially in capturing the elusive and subtle chicken-and-egg recursion through which a cell creates itself.

Models of living processes have a long history in science and fiction. From Jacques de Vaucason's automata to The Terminator, our creations are rooted in a yearning to control the uncontrollable. In the afterglow of cybernetics, Arthur Guyton built a model of the systemic blood circulation[4] that still inspires awe, if not actual use. Physiologists and bioengineers have long been assembling a Virtual Heart[5] and have even been aspiring to a Virtual Rat (http://www.systemscenters.org/centers/virtual-physiological-rat/). Industrial applications include the computational platform PhysioLab, developed by the biotech company Entelos, which models dynamic interactions among organ systems in diseases like diabetes[6].

The origins of the *M. genitalium* simulation lie in the constraint-based models of whole-cell metabolism developed by Bernhard Palsson[7]. Such a constraint-based model implements one of the 28 processes into which Karr *et al.*[1] have partitioned the cell's operations (**Fig. 1**). These include processes that track exchanges with the extracellular medium and all the metabolic fluxes, the state of the supercoiled chromosome, transcription of all active genes, processing of all mRNAs, translation of all proteins, formation of all macromolecular complexes including RNA polymerases and ribosomes, and progress of cytokinesis and FtsZ polymerization. It would be a nightmare to tie all of these processes together without a beautiful idea: between timesteps of one second, the processes operate independently of each other. What links them are 16 variables that record the overall cellular state (e.g., counts of all mRNAs, protein monomers, protein complexes). The variables are read by the processes at the start of each timestep and written back at the end, in a loop that repeats until the cell divides. This takes slightly longer than *M. genitalium* itself does, 10 h on a 128-node cluster that runs the publicly available, object-oriented Matlab code.

The one-second time-scale separation enables two vital features. First, it allows painless implementation of the subtle recursion

Jeremy Gunawardena is in the Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA.
e-mail: jeremy@hms.harvard.edu

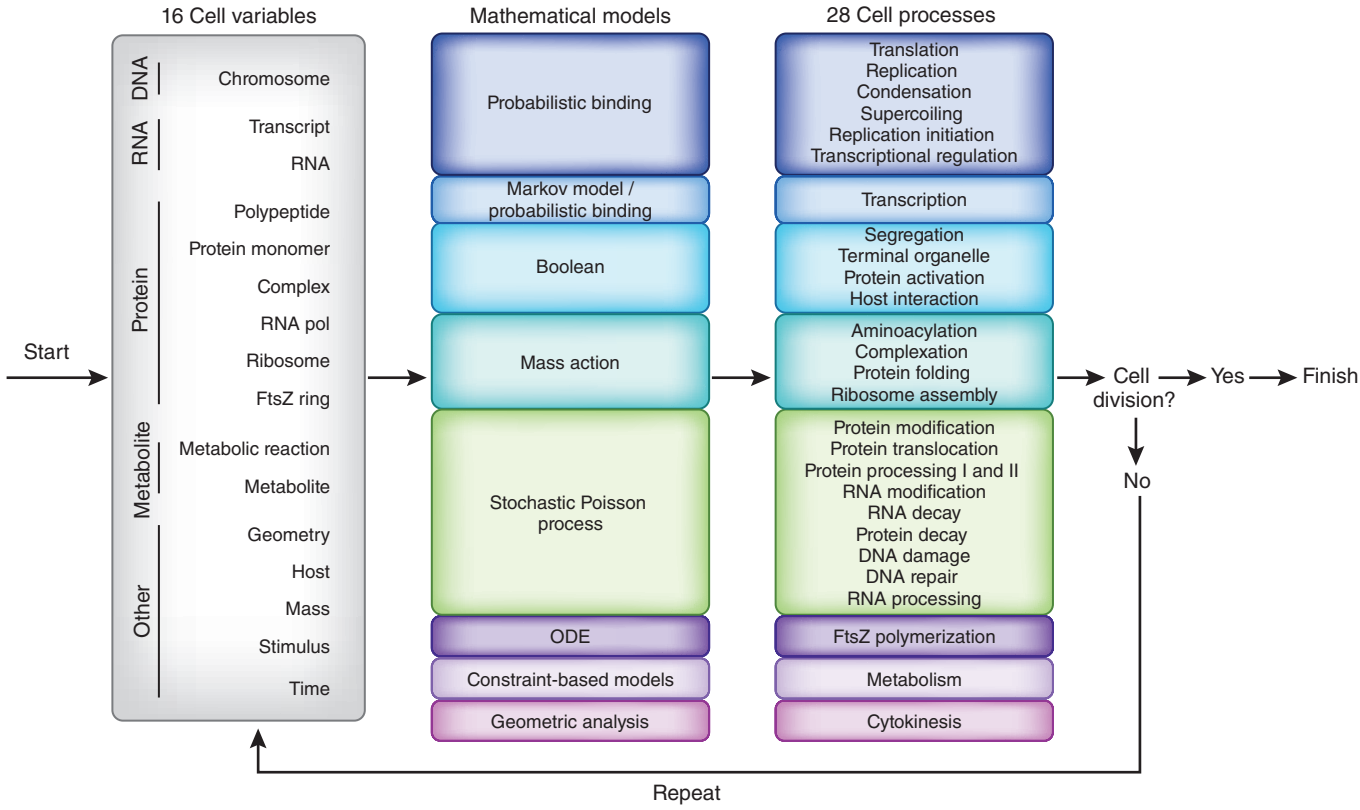

Figure 1 Whole-cell simulation of *M. genitalium*. The functionality of the cell was divided into 28 processes (right column). Each process was independently modeled using the most appropriate mathematical representation (middle column). The background color of each process matches the background color of the mathematical representation used to model it. The simulation is broken down into one-second timesteps. At the start of each timestep, the processes read the values of the 16 cell variables (left column), carry out their computations independently of each other and then update the cell variables at the end of the timestep. This procedure integrates the different processes, which influence each other only through the cell variables, and is carried out repeatedly until the simulation terminates at cell division, as measured by the progress of FtsZ polymerization. RNA pol, RNA polymerase; ODE, ordinary differential equation.

mentioned above, by which proteins make the ingredients out of which they themselves are made. Robert Rosen struggled with the profound challenges of such "metabolic closure"[8,9], and successful implementation of it distinguishes this model from any other. Second, it allows each of the 28 processes to be independently implemented by the most appropriate method (e.g., constraint-based models, stochastic processes, discrete models, computer code, ordinary differential equations) (**Fig. 1**). The right horse can be chosen for the right course. The 1,900 resulting parameters were estimated using experimental data garnered from a variety of organisms, from *Mycoplasma pneumoniae* to *Escherichia coli*. Karr *et al.*[1] had to wade through over 900 publications in the process. The model is neither simple nor elegant but it is effective. Knockout of individual genes *in silico* identifies essential genes with 80% accuracy compared to data from *M. genitalium*.

Although such accuracy is encouraging, what we always demand of a model is emergent novelty. What we often forget is that a model is not a description of reality; it is a description of our assumptions about reality[10]. Despite the complexity, all that can be deduced from a model is implicit in its assumptions. Implicit does not mean obvious: the assumptions underlying the whole numbers, 1, 2, 3,…, have kept mathematicians on their toes since Diophantus's *Arithmetica*. The assumptions of the *M. genitalium* model are not so profound. However, they have not been brought together in this way before. The novelty arises from looking inwards, from learning whether we really understand what we think we know about how metabolism, transcription, translation and replication recursively create a new cell. Indeed, the model yields a very interesting prediction of this kind. Simulations show that replication initiation, and replication itself, vary substantially in duration between stochastically different cells, whereas cell cycle durations are tightly clustered around nine hours. According to the model, cells that take longer to initiate replication accumulate a greater pool of the dNTPs needed to make DNA so they replicate faster, making up for lost time. Metabolism coordinates the cell cycle, independently of genetic regulation. It is a beguiling prediction, but whether artifact or real biology remains to be tested.

Because models are only as good as their assumptions, they can fall short of expectations. Constraint-based models are sometimes touted as requiring no assumptions beyond reaction stoichiometries. Such models do well at predicting exchange fluxes with the environment but, as for the internal fluxes, they may not get even the directions correct. If thermodynamics is not put in, it does not magically emerge. Similarly, the *M. genitalium* model cannot provide insights into lysine acetylation[11] because that is not part of its program. Models cannot supplant the kind of experiments undertaken by Luis Serrano and colleagues, whose comprehensive unraveling of the metabolome[12], transcriptome[13] and proteome[14] of *M. pneumoniae* revealed how noncoding RNAs and multifunctional enzymes can compensate for a reduced genome. The expectation that, with enough details, a model will miraculously spring to life and make such experiments unnecessary is the stuff of fiction.

This is not to say that models cannot tell us about things we do not yet know exist. Physicists have spent the gross domestic product of a small country confirming the physical existence of something conjured up conceptually

in Peter Higgs' 1964 theoretical paper. Much the same conjuring has happened in biology. Michaelis and Menten conjured up enzyme-substrate complexes more than 30 years before Britton Chance showed they existed, using a simple model now familiar to all biochemists[10]. Mendel conjured up the discrete particles later called genes using high school–level algebra. These kinds of models trade detail for abstraction; they focus on a particular question to the exclusion of all else. They require more inspiration, less perspiration.

Karr *et al.*[1] already have their sights set on *E. coli*, a more experimentally tractable organism of interest to a broad range of biologists. With 4,288 genes, a division time of 30 min and far more copious data, computational power may become the limiting resource. One wonders whether specialized hardware, like that which David Shaw and colleagues have exploited so spectacularly in molecular dynamics[15], will eventually be required, if whole-cell simulation is to aspire to mammalian complexity.

Having pulled off a tour-de-force of computational biology, it is a shame that Karr *et al.*[1] succumb to genocentrism in their title. The model no more "predicts phenotype from genotype" than DNA makes RNA makes protein. DNA does not make anything. For that you need a cell. How strange it is that cell theory, the first true theory in biology, predating both Darwin's evolution and Mendel's genetics, is so readily ignored in our current fixation with the genome, only one of the cell's many components. The irony is that the paper itself is profoundly antigenocentric; metabolic closure could hardly be otherwise[8]. Moreover, the real paper—the supplementary information, all 120 pages of it—is modern cell biology compiled into instructions, symbols and formulas. The authors have published a new kind of molecular and cell biology textbook—one that is executable. Our students will want to download the app.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Karr, J.R. *et al. Cell* **150**, 389–401 (2012).
2. Tomita, M. *Trends Biotechnol.* **19**, 205–210 (2001).
3. Ishii, N., Robert, M., Nakayama, Y., Kanai, A. & Tomita, M. *J. Biotechnol.* **113**, 281–294 (2004).
4. Guyton, A.C., Coleman, T.G. & Granger, H.J. *Annu. Rev. Physiol.* **34**, 13–44 (1972).
5. Noble, D. *Physiology (Bethesda)* **19**, 191–197 (2004).
6. Shoda, L. *et al. Clin. Exp. Immunol.* **161**, 250–267 (2010).
7. Edwards, J.S., Covert, M. & Palsson, B.O. *Environ. Microbiol.* **4**, 133–140 (2002).
8. Rosen, R. *Life Itself: a Comprehensive Enquiry into the Nature, Origin and Fabrication of Life. Complexity in Ecological Systems* (Columbia University Press, 2005).
9. Letelier, J.-C. & Soto-Andrade, J. *J. Theor. Biol.* **238**, 949–961 (2006).
10. Gunawardena, J. *Mol. Biol. Cell* **23**, 517–519 (2012).
11. van Noort, V. *et al. Mol. Syst. Biol.* **8**, 571 (2012).
12. Yus, E. *et al. Science* **326**, 1263–1268 (2009).
13. Güell, M. *et al. Science* **326**, 1268–1271 (2009).
14. Kühner, S. *et al. Science* **326**, 1235–1240 (2009).
15. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. & Shaw, D.E. *Annu. Rev. Biophys.* **41**, 429–452 (2012).