SYNTHETIC BIOLOGY

Life, simulated

A computer model of a bacterium accounts for gene function and predicts unexpected emergent properties.

If you can build something from scratch, it follows that you understand how its parts can combine to produce the whole. This is the idea that underlies attempts to simulate cells with a computer. In recently published work, Markus Covert and colleagues at Stanford University have generated a whole-cell model of the human pathogen *Mycoplasma genitalium* in which they simulate the function of all of its genes.

Mycoplasma has only 525 genes, most of which have homologs that have been studied in other organisms even if some *Mycoplasma* genes have not been studied directly, Covert says, explaining his choice of organism. “We were guided very much by the Human Genome Project in this, which is to say to start with the simplest organism.” Also, *Mycoplasma* can be cultured in the laboratory, which Covert considered important as he wished to combine theoretical and experimental work, and *Mycoplasma* is the first organism to have had its chromosome synthesized *de novo*.

Using data gleaned from hundreds to thousands of published papers, on *Mycoplasma* as well as on *Escherichia coli* and other bacteria, the researchers set out to put the parts back together in the computer. They broke the cell’s overall functions down into 28 separate functional modules describing different subprocesses—RNA decay, metabolism or DNA replication, for instance—and modeled each of these separately based on the information available about that subprocess in the literature.

This turned out to be a very good way to attack the problem, Covert explains, because it allowed them to model each subprocess at the appropriate level of detail. If they expected the parameters describing a certain process not to be very accurate—for instance, if the only available data came from distantly related organisms or if the process had simply not been studied in much detail—then they could choose not to constrain that submodel too strictly. Finally,

they devised a way to integrate the submodels into a whole.

“As molecular biologists,” says Covert, we’ve been working under the hypothesis for the last 50+ years of reductionism: that is, if we identify all the parts and we figure out what they do and how they relate to one another then we should be able to put that all together and predict how a cell behaves. So in some ways by building this model we’re testing that hypothesis.”

But does the model look anything like the real thing? Indeed, Covert and colleagues report that their simulated bacterium has many expected properties, in its metabolic fluxes, global distribution of protein and RNA, metabolite amounts and gene function (they predicted gene essentiality with 79% accuracy).

The simulated bacterium also led to new hypotheses. For instance, by simulating the division cycle of hundreds of cells, the researchers saw much more variability in the time taken to initiate DNA replication than in the cell-cycle time overall. Initially perplexed at this discrepancy, they eventually identified a possible compensating effect in their model. Slow initiation of replication, the model suggests, leads to a buildup of nucleotide pools that then speeds up the subsequent replication process proper, making up for lost time. The researchers are actively testing this idea experimentally. “To discover [...] an emergent control that not only had we not included [in the model], but that we had never thought about, that was very exciting,” says Covert. “It’s fun when your creation takes on a life of its own.”

In a separate example, the model predicts dynamic genome occupancy by 30 DNA-binding proteins and even predicts collisions

that may occur between them. Experimentally obtaining such data would be very difficult with current techniques. Genome-wide footprint data of DNA-binding proteins is typically static, whereas dynamic information on the speed and direction of protein movement along DNA comes largely from single-molecule or other biophysical studies. In the model, these two types of data are brought together to generate ideas that can be tested experimentally in a more directed fashion.

The simulations will not stop here. Covert hopes to extend this approach to model *Escherichia coli* and then, in what will no doubt be a difficult leap, to simulate yeast and

Modeling a bug. The subprocesses of simulated *Mycoplasma*. Reprinted with permission from Elsevier.

then a mammalian cell. In the meantime, the model bacterium is freely available to other researchers. Covert says it has already been downloaded hundreds of times since the paper was published. He hopes that other scientists will “tear the model apart” and will use it to test alternate hypotheses about the biological processes that interest them. “This model certainly isn’t the last word on *Mycoplasma*,” he says. “It’s more like a first draft.”

Natalie de Souza

RESEARCH PAPERS

Karr, J.R. *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401 (2012).