

Analysis of 89 models of metabolic processes finds flaws in 44 of them—but suggests corrections

October 8 2014, by Larry Hardesty



Credit: Jose-Luis Olivares/MIT

Metabolic networks are mathematical models of every possible sequence of chemical reactions available to an organ or organism, and they're used to design microbes for manufacturing processes or to study disease. Based on both genetic analysis and empirical study, they can take years to assemble.

Unfortunately, a new analytic tool developed at MIT suggests that many of those models may be wrong. Fortunately, the same tool may make it fairly straightforward to repair them.

"They have all these models in this database at [the University of California at] San Diego," says Bonnie Berger, a professor of applied mathematics and computer science at MIT and one of the tool's developers, "and it turns out that many of them were computed with floating-point arithmetic"—an approximate numerical representation that most computer systems use to increase efficiency. "We were able to prove that you need to compute them in exact arithmetic," Berger says. "When we computed them in exact arithmetic, we found that many of the models that were believed to be realistic don't produce any growth under any circumstances."

Berger and colleagues describe their new tool, and the analyses they performed with it, in the latest issue of *Nature Communications*. First author on the paper is Leonid Chindelevitch, who was a graduate student in Berger's group when the work was done and is now a postdoc at the Harvard School of *Public Health*. He and Berger are joined by Aviv Regev, an associate professor of biology at MIT, and Jason Trigg, another of Berger's former students.

Floating-point arithmetic is kind of like scientific notation for computers. It represents numbers as a decimal multiplied by a base—like 2 or 10—raised to a particular power. Though it sacrifices some accuracy relative to exact arithmetic, it generally makes up for it with gains in computational efficiency.

Indeed, in order to perform an exact-arithmetic analysis of a data structure as huge and complex as a metabolic network, Berger and Chindelevitch had to find a way to simplify the problem—without sacrificing any precision.

Pruning the network

Metabolic networks, Chindelevitch says, "describe the set of all reactions that are available to a particular organism that we might be interested in. So if we're interested in yeast or E. coli or the tuberculosis bacterium, this is a way to put together everything we know about what this organism can do to transform some substances into some other substances. Usually it will get nutrients from the environment, and then it will transform them by its own internal mechanisms to produce whatever it is that it wants to produce—ethanol, different cellular components for itself, and so on."

The network thus represents every sequence of [chemical reactions](#) catalyzed by enzymes encoded in an organism's DNA that could lead from particular nutrients to particular chemical products. Every node of the network represents an intermediary stage in some chain of reactions.

To simplify such networks enough to enable exact arithmetical analysis, Chindelevitch and Berger developed an algorithm that first identifies all the sequences of reactions that, for one reason or another, can't occur within the context of the model; it then deletes these. Next, it identifies clusters of reactions that always work in concert: Whatever their intermediate products may be, they effectively perform a single reaction. The algorithm then collapses those clusters into a single reaction.

Most crucially, Chindelevitch and Berger were able to mathematically prove that these modifications wouldn't affect the outcome of the analysis.

"What the exact-arithmetic approach allows you to do is respect the key assumption of the model, which is that at steady state, every metabolite is neither produced in excess nor depleted in excess," Chindelevitch says. "The production balances the consumption for every substance."

When Chindelevitch and Berger applied their analysis to 89 metabolic-network models in the San Diego database, they found that 44 of them contained errors or omissions: If the products of all the reactions in the networks were in equilibrium, the organisms modeled would be unable to grow.

Patching it up

By adapting algorithms used in the field of compressed sensing, however, Chindelevitch and Berger are also able to identify likely locations of network errors.

Compressed sensing exploits the observation that some complex signals—such as audio recordings or digital images—that are computationally intensive to acquire can, upon acquisition, be compressed. That's because they can be converted into a different mathematical representation that makes them appear much simpler than they did originally. It might be possible, for example, to represent an audio signal that initially consists of 44,000 samples per second of its duration as the weighted sum of a much smaller number of its constituent frequencies.

Compressed sensing performs the initial sampling in a clever way that allows it to build up the simpler representation from scratch, without having to pass through the more complex representation first. In the same way that compressed sensing can decompose an audio signal into the constituent frequencies with the heaviest weights, Chindelevitch and Berger's algorithm can isolate just those links in a metabolic network that contribute most to its chemical imbalance.

"We're hoping that this work will provide an impetus to reanalyze a lot of the existing [metabolic-network](#) model reconstructions and hopefully spur some collaborations where we actually perform this analysis and

suggest corrections to the model before it is published," Chindelevitch says.

"This is not an area where one would expect there to be a problem," says Desmond Lun, chair of the Department of Computer Science at Rutgers University, who studies computational biology. "I think [the MIT researchers' work] will change people's attitudes in the sense that it raises an issue that most people would have thought was not an issue, and I think it will make us a lot more careful."

"Computers operate with limited precision because there are only so many digits that you can store—even though, I must say, they store a lot of digits," Lun explains. "Through software, you can be more or less careful about how much precision you lose in that way. There are very, very good packages out there that try to minimize that problem. And mostly, I would have thought, and I think most people would have thought, that that would be sufficient for these metabolic models."

Errors in the models may have gone unnoticed because analyses performed on them often comported well with empirical evidence. But "those floating-point errors vary from package to package," Lun says. "Certainly, it would be very concerning to find that because somebody used this software package, they got these great results, and then if I used a different software package, I would not."

More information: "An exact arithmetic toolbox for a consistent and reproducible structural analysis of metabolic network models." Leonid Chindelevitch, Jason Trigg, Aviv Regev & Bonnie Berger. *Nature Communications* 5, Article number: 4893 [DOI: 10.1038/ncomms5893](https://doi.org/10.1038/ncomms5893)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT

research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Analysis of 89 models of metabolic processes finds flaws in 44 of them—but suggests corrections (2014, October 8) retrieved 12 May 2024 from <https://phys.org/news/2014-10-analysis-metabolic-flaws.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.