# N news

Kirk L. Kroeker

The Rise of Molecular Machines

The field of molecular computing is achieving new levels of control over biochemical processes and fostering sophisticated connections between computer science and the biological sciences.

Taking cues from both speculative fiction and hard science, today's most prolific futurists have envisioned a point in the future when developments in genetics, nanotechnology, and robotics make it possible to sidestep the constraints of human durability and intelligence. Controversial assumptions notwithstanding, even the most optimistic speculation about the future symbiotic convergence of humans and technology is deriving at least some measure of credibility from emerging work in molecular computing. Researchers in this field are achieving new levels of control over biological processes and fostering sophisticated crossovers between computer science and the biological sciences.

In one recent development, scientists in the department of molecular computing at the California Institute of Technology (Caltech) have built what they are calling the most complex biochemical circuit ever created from scratch. These circuits, the Caltech researchers say, will allow scientists to explore the principles of information processing in biological systems and design biochemical pathways with
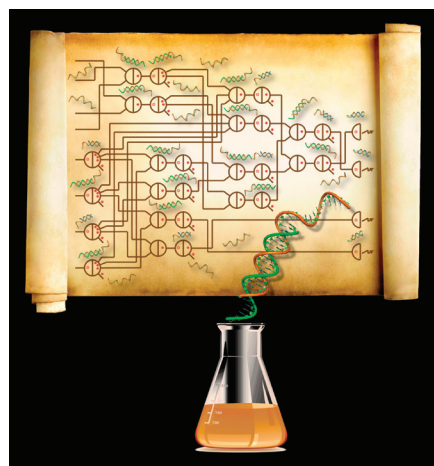
Wiring diagram specifying a biochemical circuit that consists of 74 different DNA molecules. The circuit, developed at Caltech, demonstrates an approach for implementing arbitrary digital logic in biochemical systems. The lines correspond to single-stranded oligonucleotides, while the nodes correspond to partially double-stranded molecules.

decision-making capabilities. Such circuits, they say, will give biochemists unprecedented control over chemical reactions for biological and chemical engineering and may even lead to the proliferation of molecular-scale biological machines.

Lulu Qian, a senior postdoctoral scholar in bioengineering at Caltech, and Erik Winfree, a Caltech professor of bioengineering and computer science, computation, and neural systems, used DNA-based components to build the circuit. Instead of depending on electron flows through transistors, the DNA logic gates receive and produce molecules as their signals. The molecular signals travel from one gate to another, connecting the circuit as if the molecules were wires.

With his colleagues Georg Seelig, David Soloveichik, and David Zhang, Winfree first built a biochemical circuit in 2006. In that work, DNA signal molecules connected several DNA logic gates to each other, forming a multilayered circuit consisting of 12 molecules. In the new design, Qian and Winfree made the logic gates from pieces of single- and double-stranded DNA. The two researchers have made several circuits with this approach; the largest, containing 74 different DNA molecules, can compute the square root of any number up to 15 and round the answer down to the nearest integer.

During the calculation process, the custom-built molecules float around the solution and bump into each other, prompting strands with a certain DNA sequence to zip themselves to compatible strands while simultaneously unzipping other strands. The unzipped strands are released back into the solution to continue the cycle until the calculation process is complete. The researchers simply monitor the concentrations of output mole-

Heading to the instrument room at Caltech. In the box are the pipettes and test tubes containing molecules for a biochemical circuit: the 74 tubes each contain one type of DNA logic gate or fuel, and the eight tubes contain input signal strands. To get the DNA circuit running, the 74 gate and fuel molecules must be mixed to form the functional circuit, and a combination of the eight input signal strands must be added. Fluorescence must then be monitored during the next 10 hours to determine the circuit's output.

cules to determine the answer, which takes some 10 hours to compute.

While the logic gates have identical structures and can therefore be standardized, Winfree says the research still faces several significant challenges, such as automating the design and analysis process and finding ways to control the inevitable faulty behavior of the floating molecules. In addition, Winfree says it is not clear how well such molecular systems can be scaled to take on tasks more complex than relatively basic math. According to Winfree, the millions of logic gates that are common in silicon-based computing will not be possible for DNA circuits, at least in the near term.

Despite the evident limitations of DNA circuits, researchers working in this area are already achieving more control over biochemical processes than has ever before been possible. "We're trying to design DNA systems that self-assemble, implement biochemical circuitry, and act as molecular robots," says Winfree, who is modest in his assessment of the science and engineering involved in creating the molecular circuit with Qian. "It took previously demonstrated proofs of principle, simplified them, auto-

mated them, and got them to work at a much larger scale," he says. "Doing so, of course, involved some conceptual advances that might be considered a breakthrough."

To Winfree, the potential applications for such programmable biochemical systems are limitless. Eventually being able to use such systems for diagnosing disease or delivering custom-designed cellular therapies, for example, is a common refrain in discussions about the utility of molecular computing. "You wouldn't say that the purpose of electronic computers is to play video games and keep track of banks' financial records," he says. "The same goes for circuitry at the molecular scale; there are a lot of things you can do with molecules and chemistry."

DNA-based Data Storage

In another molecular-computing project, researchers at the Chinese University of Hong Kong (CUHK) are using DNA for data storage. The project, which recently won the gold medal at the International Genetically Engineered Machine (iGEM) competition at the Massachusetts Institute of Technology, is said to be the first to use the DNA of *E. coli* to store and en-

crypt text and images. The researchers working on the project, headed by Ting Fung Chan, estimate that one gram of bacteria can store up to 650,000 gigabytes, roughly equivalent to 325 hard drives, each with two terabytes of capacity.

Chan, a professor in the school of life sciences and the deputy director of the center for microbial genomics and proteomics at CUHK, explains that the idea for the project initially came from a conversation about *Assassin's Creed*, a video game in which the protagonist has inherited his ancestor's memory through DNA. In the original iGEM competition, Chan and his team encoded a short text message in a single piece of DNA and inserted it into bacterial cells, where the DNA was reshuffled, then retrieved and decoded.

Storing small bits of data, according to Chan, is not the project's ultimate goal. Currently, he and his team are designing a parallel storage system in which the DNA sequence would be cut into segments and stored in multiple cells. "We are pursuing a true massively parallel storage system," Chan says. "Storing a large piece of information, such as a photograph or a dictionary, is impossible to do within a single piece of DNA because of the limits associated with current DNA synthesis technology."

Simply fragmenting the information and inserting it into multiple cells without having an effective method for rebuilding the original file would destroy the data because the order of fragments would be unknown, explains Chan. So he and his team devised a biological system that uses standard storage mechanisms—such as headers and checksums—so each piece of information can be mapped and located for retrieval. The method involves removing DNA from the cells, altering it with enzymes, and returning it to new cells where the DNA sequence will be shuffled. "The idea is to demonstrate that we can, in principle, store any digital information," says Chan.

Chan notes that while he and his team have made progress at encoding pictures in several DNA sequences and storing them in parallel in multiple cells, research in this area has essentially just begun. He says it will

take at least another five to 10 years for the core ideas to develop. "Synthetic biology is an emerging field, and there is special research funding only in some parts of the world," says Chan. "This is multidisciplinary research, and training of a new generation of young researchers is very much needed."

While most research in molecular computing has focused on constructing small modules in which individual genetic units carry out a particular task, Chan says developing a complete machine will require genome-scale engineering that depends not on the individual building blocks, but on a deep understanding of how these individual building blocks interact with each other. To this end, he says, a key challenge to overcome in this area is to gain a much more complete understanding of individual cellular components so they can be accurately modeled as an integrated system.

Despite such challenges, the tools and techniques used by scientists working in molecular computing are beginning to mature, suggesting the possibility that some of these projects may soon form the groundwork for commercial applications. Caltech's Winfree, for his part, predicts the field will gain significant momentum during the next 10 years. "Molecular com-

The largest circuit, which contains 74 different DNA molecules, can compute the square root of any number up to 15 and round the answer down to the nearest integer.

puting is beginning to be a real contender for some simple applications," he says, noting that every few years there is a breakthrough that changes what people inside and outside the field think is possible.

By way of example, he cites Paul Rothemund's DNA origami project, the goal of which was to fold DNA into tiny shapes and images. "Before DNA origami, most folks, including me, would not have thought it possible, and afterwards it was routine," says Winfree. "What's next is impossible to predict."

Still, as the pace of innovation in this area of research continues to accelerate, more crossovers between molecular computing and traditional computing will emerge. Many principles from electrical engineering and computer science are already being used to design molecular systems, as was done in the Caltech project, which used the principles of digital abstraction and signal restoration to make the circuits capable of withstanding imperfections, and in the CUHK project, which used traditional storage and cryptography paradigms to encode data in bacteria.

Winfree points out that if the pace of innovation in molecular computing continues, in 30 years scientists will be able to make systems built with 20,000,000 nucleotides, which is larger than *E. coli*'s complete genome. "It's hard to imagine what such designed-from-scratch molecular systems will be capable of doing," says Winfree. While it remains to be seen whether such systems will eventually be able to augment human capabilities in a manner consistent with the visions of today's futurists, Chan, like Winfree, remains optimistic about the potential of molecular computing. "I do believe that this is no longer science fiction, as I would have thought when I was a kid," he says. **C**
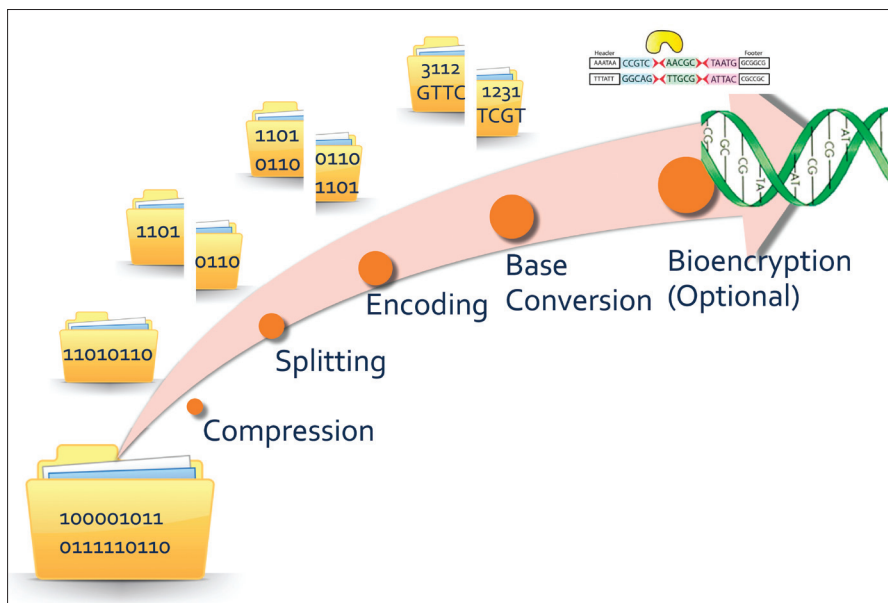
An overview of the biological data storage system developed at the Chinese University of Hong Kong. In the bacteria-based storage system, binary files are compressed and split into data packets. Each packet contains a payload, an address, error-correction code, and an optional encryption marker. A binary-to-quaternary base conversion is performed on the encoded data, followed by substituting the quaternary numbers for the four DNA bases.

Further Reading

Qian, L. and Winfree, E.
Scaling up digital circuit computation with DNA strand displacement cascades, *Science 332*, 6034, June 3, 2011.

Ran, T., Kaplan, S., and Shapiro, E.
Molecular implementation of simple logic programs, *Nature Nanotechnology 4*, 10, October 2009.

Rothemund, P.
Folding DNA to create nanoscale shapes and patterns, *Nature 440*, 7083, March 16, 2006.

Storm, D.
Unhackable data in a box of bacteria: Future of InfoSec? *Computerworld*, January 18, 2011.

Zyga, L.
Biomolecular computer can autonomously sense multiple signs of disease, PhysOrg.com, July 6, 2011.

Kirk L. Kroeker works in communications and has written extensively about the impact of emerging technologies.