Crowdsourced RNA designs outperform computer algorithms

January 27 2014

An enthusiastic group of non-experts, working through an online interface and receiving feedback from lab experiments, has produced designs for RNA molecules that are consistently more successful than those generated by the best computerized design algorithms, researchers at Carnegie Mellon University and Stanford University report.

Moreover, the researchers gathered some of the best design rules and practices generated by players of the online EteRNA design challenge and, using machine learning principles, generated their own automated design [algorithm](#), EteRNABot, which also bested prior design algorithms. Though this improved computer design tool is faster than humans, the designs it generates still don't match the quality of those of the online community, which now has more than 130,000 members.

The research will be published this week in the *Proceedings of the National Academy of Sciences* Online Early Edition.

"The quality of the designs produced by the online EteRNA community is just amazing and far beyond what any of us anticipated when we began this project three years ago," said Adrien Treiulle, an assistant professor of computer science and robotics at Carnegie Mellon, who leads the project with Rhiju Das, an assistant professor of biochemistry at Stanford, and Jeehyung Lee, a Ph.D. student in computer science at Carnegie Mellon.

"This wouldn't be possible if EteRNA members were just spitting out

designs using online simulation tools," Treuille continued. "By actually synthesizing the most promising designs in Das' lab at Stanford, we're giving our community feedback about what works and doesn't work in the physical world. And, as a result, these non-experts are providing us insight into RNA design that is significantly advancing the science."

RNA, or ribonucleic acid, is one of the three macromolecules essential for life, along with DNA and proteins. Long recognized as a messenger for genetic information, RNA also may play a much broader role as a regulator of cells. Understanding RNA design could be useful for treating or controlling diseases such as HIV, for creating RNA-based sensors or even for building computers out of RNA.

In the research being reported this week, the researchers tested the performance of the EteRNA community, EteRNABot and two state-of-the-art RNA design algorithms in generating designs that would cause RNA strands to fold themselves into certain shapes. The computers could generate designs in less than a minute, while most people would take one or two days; synthesizing the molecules to determine the success and quality took a month for each design, so the entire experiment lasted about a year.

In the end, Lee said, the designs produced by humans had a 99 percent likelihood of being superior to those of the prior computer algorithms, while EteRNABot produced designs with a 95 percent likelihood of besting the prior algorithms.

"The quality of the community's designs is so good that even if you generated thousands of designs with computer algorithms, you'd never find one as good as the community's," Lee said.

When the project began, players were asked to design RNA that folded into specific shapes selected by the Das lab. Thanks to technological

breakthroughs that now enable Das and his team to synthesize a thousand design sequences each month instead of the original 30, EteRNA has become an open research project to which researchers from labs around the world can submit design challenges.

Though EteRNA players may not be scientifically trained, they nevertheless have instincts that, when bolstered by the lab experiments, can lead to new insights. "Most players didn't have tactical insights on RNA designs," Lee said. "They would just recognize patterns—visual patterns."

"Scientifically, not all of these rules initially seemed to make sense, but people who were following them did better," he noted.

One design rule generated by the players involves "capping." RNA consists of long sequences of pairs of nucleotides and usually the easiest way to create a sequence or "stack" that won't rip itself apart when synthesized is to fill it with guanine-cytosine (GC) pairs. But too many GC pairs can produce some unexpected shapes when synthesized—"It's like doing origami with a cardboard box," as one player put it.

Lee said the players found a solution by putting the GC pairs only at the end of the stack—"capping"—and filling the rest of the stack with adenine-uracil pairs.

The project is now looking at expanding its design regimen to include three-dimensional designs. They also are developing a template that researchers in other fields can use to turn scientific projects into online challenges.

Provided by Carnegie Mellon University