



Computational Nanogami

RNA Structure Search Stretches Across Georgia Tech Disciplines

Back in 2009, Josh Anderson didn't know much about biology. But he knew that a summer undergraduate research assistantship working on something called "RNA folding" had to be better than the job his mother had lined up for him. Prashant Gaurav recalls that in 2009 he was thinking about applying to graduate school in computer science—not about the base pairings of a nasty RNA virus like Hepatitis C.

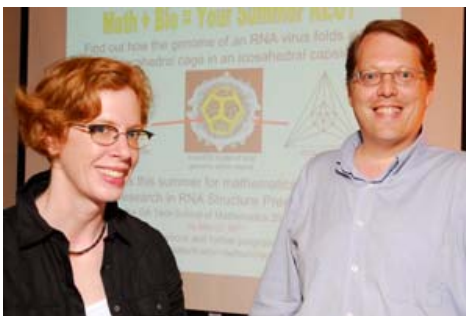
Yet in 2011 both Anderson (double-major in discrete mathematics and computer science) and Gaurav (M.S. student in computational science and engineering) are hard at work on problems in computational molecular biology as part of an interdisciplinary research team led by Christine Heitsch, associate professor in the School of Mathematics, with professors Steve Harvey from Biology and David Bader from CSE. The computational side of this math/bio/CSE research collaboration is rounded out by Emily Rogers, a CSE Ph.D. student, and CS major Andrew Ash. (Ash and Anderson both graduated in May 2011.) Together with other students and postdocs from math and biology, the team is trying to reverse engineer *in silico* some of the mysteries of "molecular origami," as Heitsch describes RNA folding.

"This is one of Nature's versions of origami," explains Heitsch. "Instead of a flat piece of paper which is folded along creases to form a complicated 3D shape, an RNA sequence 'folds' along complementary Watson-Crick pairings between nucleotides in different parts of the sequence. Think of the base pairing of an RNA sequence like a string made up of bits of Velcro that can and do stick together." Those Watson-Crick base pairings are also what make up the structure of RNA's more famous relative, the double-stranded DNA helix.

Knowing the structure of DNA is fundamental to understanding how genetic information is stored and passed on. But the genomes of many, if not most, viruses are actually RNA, not DNA, and there is experimental evidence suggesting that the folded structure an RNA viral genome affects the functioning of the virus. So advancing understanding of the base pairings in RNA viruses should help with treating the diseases they cause.

Since RNA base pairings are both strong and stable, predicting those pairings can be approached as a thermodynamic optimization problem. This and related approaches work reasonably well for short sequences, but definitely do not scale up to lengthy RNA viral genomes. Hence, the ongoing scientific challenge is to develop quantitative methods to find the proverbial needle (that is, native base pairings) in the very large haystack of possible structures for an RNA virus like Hepatitis C.

Enter Anderson, Ash, Gaurav and Rogers, who are helping to design new computational approaches to understanding where base pairs can and do occur, and how they influence RNA structure.



Professors Christine Heitsch and David Bader, along with

"We're essentially building complicated machines that people are running right on top of us," says Anderson, who started on the project in 2009 as a summer Research Experience for Undergraduates (REU) student. "I've been a bit of a jack of all trades: I've done code optimization, algorithm design. What I brought into the project was a lack of fear of coding."

"I didn't have any biology background," echoes Gaurav, "but I've been able to achieve some significant speed improvements by bringing more efficiency to the program's parallelism. We leveraged multicore computing techniques that enable the biologists to analyze and process large RNA sequences—like the HIV virus—much faster."

What a lovely mess

This research team traces its origins to a conference that Heitsch attended shortly before starting

Steve Harvey (not pictured), lead the multidisciplinary project examining base pairings in RNA viruses. Pictured above are (standing, l to r) Bader, Prashant Gaurav, Svetlana Poznanovik, Dragos Ilas, Shel Swenson, Zsuzsanna Sukosd and Andrew Ash, along with (sitting, l to r) Emily Rogers, Josh Anderson and Heitsch.

her faculty position at Georgia Tech. As it happened, a lunch conversation about potential computational collaborators pointed her to Bader as a computer scientist with interests in biological applications. Another serendipitous connection was made shortly after Heitsch arrived on campus and met Harvey, a biophysicist whose research interests include 3D structures of RNA viral genomes.

"Biology is inherently complicated," Heitsch says. "Mathematical theories are very powerful tools for gaining new insights into why the current folding methods don't provide sufficiently accurate results.

But translating these theoretical insights into practical applications requires dealing with all the biological complexity that you can. For me, that means collaborating with a computational scientist who is excited by the algorithmic challenges this presents."

Bader says he's thrilled to participate in a research project integrating mathematics and computer science to tackle a fundamental problem in biology. Too often in computer science, he says, researchers fall into the groove of only doing research that promises tidy outcomes. It's part of the computer scientist's neural wiring, he says, to seek out problems that tease with objectively "correct" solutions.

"While we are focusing on computing innovations to solve real-world problems like those in biology and physics, it's a relatively new phenomenon among computer scientists," Bader says. "It's just not a traditional mindset among computing experts, but it's in our genes at Georgia Tech."

"In too many places, mathematics is completely separate from computer science, and both are separate from biology," Heitsch says. "One of the things that's so great about Georgia Tech is the overlap between disciplines, so projects like this can happen."

Using CS and math to solve biological problems

The junior members of this interdisciplinary team each contribute a unique mix of mathematical, computational and biological expertise. With undergraduate degrees in computer science and in biology from the University of California at Berkeley, and a master's in bioinformatics from Georgia Tech, Rogers is a great fit for this project—and the CSE Ph.D. program. "I really like this area, where computation and biology meet," says Rogers. "I enjoy my CS classes but I'm not so interested in pure CS research; I like to use CS tools to solve biological problems."

"It's a very creative project," says Zsuzsanna Sukosd, a visiting graduate student on leave from her nanoscience Ph.D. program at Denmark's Aarhus University. "When people do bioinformatics work, they're often running programs that other people wrote. In this project we're working together to write the program ourselves."

Postdoctoral researcher Shel Swenson acts as the project "lynchpin," Heitsch says, with a background that positions her at the center of the three disciplines involved: Swenson earned her Ph.D. in mathematics from the University of Texas at Austin, advised by computer scientist Tandy Warnow, one of the world's leading computational phylogeneticists.

Swenson also has a successful track record of incorporating students into research projects, including three summers of leading REU programs at Austin. "The best part about mentoring undergraduates is when they start to actively question information they're given, instead of just passively absorbing it," Swenson says. This summer, she and another mathematics postdoctoral researcher, Svetlana Poznanovik, are introducing new REU students to the challenges of RNA folding.

For the undergraduates, the project has been a full-on immersion into the world of academic research. Mathematics undergrad Dragos Ilas, who was mentored by Poznanovik and also graduated this spring, says that he "could see the energy with which people worked on their projects, and the sense of productivity and achievement becomes contagious after a while."

"It's hard but rewarding," Anderson says. "Half the time it feels like you are banging your head against the wall, but every now and then you get something to work, and it's such a rush."

"I feel like our contributions were highly valued, and we were able to work on critical portions of the project," says Ash. "Computing accelerates the pace of all other disciplines, and it's in demand everywhere."

"It's hard but rewarding, half the time it feels like you are banging your head against the wall, but every now and then you get something to work, and it's such a rush."

[ShareThis](#)