NewsBytes

Automating Scientific Discovery

Robots already have a place in many labs, automating tedious tasks such as pipetting samples. But a new system designed at Aberystwyth University in the United Kingdom has taken laboratory automation a step further.

"The idea of using a robot is not news, but what's different about ours is the robot was also involved in developing hypotheses and experiments on its own," says **Ross King, PhD**, head of computational biology at Aberystwyth University's computer science department. The work was published in the April 2009 issue of *Science*.

The robot, named ADAM, was programmed to find the genes that encode "orphan" enzymes in yeast. Armed with information from bioinformatics databases such as KEGG (the Kyoto Encyclopedia of Genes and Genomes), ADAM hypothesized, from sequence similarities, which genes could encode the enzymes.

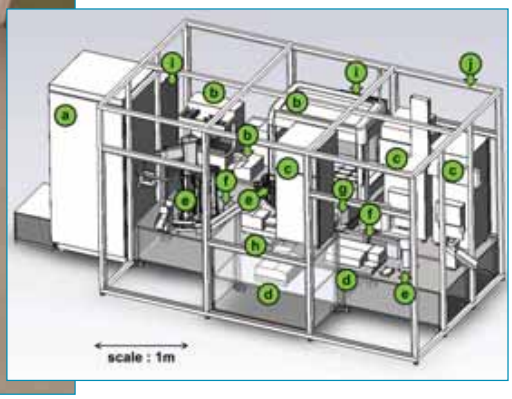
ADAM owes its brainpower in part to databases of formalized knowledge. One component is a detailed model of yeast metabolism written in the logic language Prolog; another is an ontology describing laboratory experiments, based on the open-source project EXPO. The robot also recorded its own experimental information as it worked. "One of the advantages of a robot scientist is that you get all that metadata for free," says King. "We can understand far more about the structure of the experiment than we would if only humans had been involved."

ADAM's four computers directed the experiments, with robot arms moving yeast mutants from freezer to incubators to plate readers. Ultimately, it found 12 gene-enzyme pairings that the authors were able to confirm. In some cases, the link between gene and enzyme was found to be supported by literature even though it was missing from ADAM's starting data. For others, the authors double-checked ADAM's results by purifying and testing the protein themselves.

> "The idea of using a robot is not news, but what's different about ours is the robot was also involved in developing hypotheses and experiments on its own," says Ross King.

The successful matches "are mostly to do with odd pieces of biochemistry that hadn't been sorted out yet," King says, which explains why the enzymes remained orphans for so long. Some were isozymes, with more than one gene encoding the same function, and others were promiscuous enzymes that catalyze more than one reaction.

King and his collaborators have started work on the next generation of robot scientists, beginning with a robot called EVE that will work to discover new drugs for tropical diseases.



ADAM is a 5-meter-long robot whose equipment includes cameras, sensors, and computers in addition to (a) an automated -20°C freezer, (b) three liquid handlers, (c) three automated +30°C incubators, (d) two automated plate readers, (e) three robot arms, (f) two automated plate slides, (g) an automated plate centrifuge, (h) an automated plate washer, (i) two air filters, and (j) a plastic enclosure. Diagram reprinted with permission from King, RD,* et al.*, The Automation of Science,* Science*, 324:85 (2009). Photo: Courtesy of Aberystwyth University.

As King describes it, "ADAM and EVE are special purpose, but our goal for the future is to make more general purpose automation."

"People ask if this is going to put scientists out of business, but the answer is no," says **David Waltz**, **PhD**, director of the Center for Computational Learning Systems at Columbia University. Instead, he says, "this will make scientists more productive," but they would also have to learn new skills. "Scientists would have to learn to be proficient in Artificial Intelligence and to create formal representations of knowledge."

—By Beth Skwarecki

The Function of DNA Form

According to a new computational analysis of DNA structure, variations in DNA shape—along the grooves of the double helix—may play an important role in defining how the genome works. The analysis revealed that six percent of the DNA ladder's shape is conserved across a range of different mammals—even though the sequences that produce those conserved shapes could vary.

"We've found a new way that evolutionary selection is working in the human genome, beyond just preserving the strict sequence of nucleotides," says **Tom Tullius**, **PhD**, chemistry professor at Boston University and one of the authors of the report, published April 17 in the journal *Science*. "I hope that this finding will open up some new ways of thinking about how the genome works. It's more than just a collection of letters."

A 2007 study by the ENCODE (Encyclopedia Of DNA Elements) research consortium hinted that something other than nucleotide sequence was at play in determining genome function. Looking at one percent of the human genome, the researchers found that only about half of the known functional regions (for example, sections of DNA where proteins bind) showed sequence conservation across a range of mammals (from mouse to human). "We were struck by the fact that you may not be looking at the complete story if you only look at sequence conservation to define function," Tullius says.

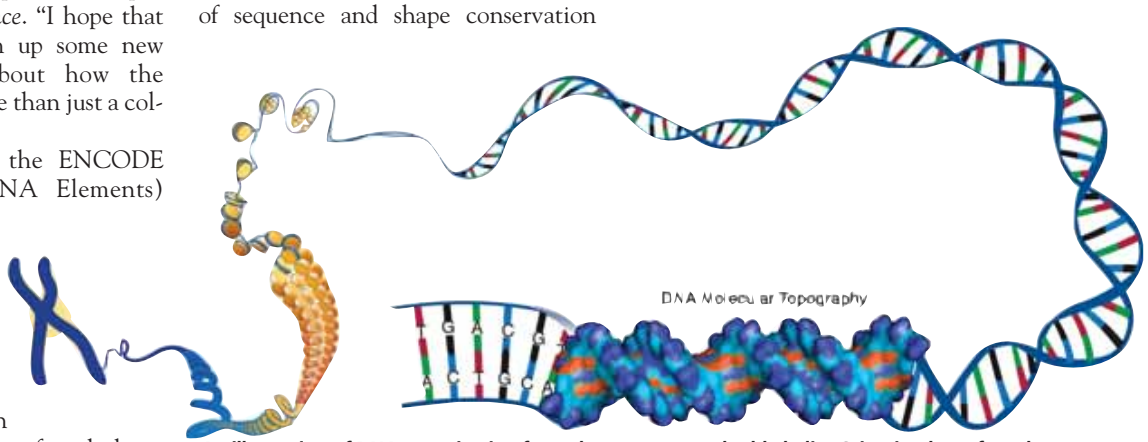
Tullius and his colleagues wondered if shape might be a factor. They had previously discovered, experimentally, that different DNA sequences can have similar structures. Using the reactive hydroxyl radical molecule, they had probed for subtle differences in DNA shape. Small variations in the radical's accessibility to the DNA yield a detailed structural map. These variations are often in the DNA's minor groove width, which can range from four angstroms at the narrowest to 11 at the widest, Tullius says. This finding led them to wonder whether sequences could diverge through evolution while form remained the same.

To answer that question, Tullius, **Elliott Margulies**, **PhD** of the National Institutes of Health, **Steve Parker** of Boston University and their colleagues created a computer program called Chai. The program compares computational predictions of DNA shapes from the same one percent of the human genome studied by ENCODE, and other mammalian genomes. They found that certain parts of the genome are conserved solely by structure, not sequence. Moreover, the combination of sequence and shape conservation almost entirely covers the functional sites identified by the ENCODE study. Tullius and his colleagues also found that polymorphisms associated with disease are more likely to cause structural changes in DNA than neutral polymorphisms—meaning that these shape changes could be disrupting the binding of some essential protein.

"I hope that this finding will open up some new ways of thinking about how the genome works," Tom Tullius says. "It's more than just a collection of letters."

In a 2007 report in the journal *Cell*, **Barry Honig**, **PhD** of Columbia University, had concluded that DNA shape influenced the binding of a homeodomain protein to developmental genes. "The combination of these two studies makes it clear that DNA shape is important in function," Honig says. "This gives us a new avenue to study how DNA functions that we didn't have before."

—By Rachel Tompa, PhD



An illustration of DNA organization from chromosome to double helix. Scientists have found subtle structural differences at the molecular level between different regions of DNA, often in the width of the helix's minor groove. Surprisingly, different sequences can yield the same shapes in DNA. Tullius, Margulies and Parker found that these subtle shapes are conserved between humans and other mammals, meaning evolution is acting not only on our DNA sequence, but its form. Courtesy of Darryl Leja, NHGRI, NIH.