



[LLNL Home](#) [About LLNL](#) [Visiting](#)

Search LLNL Web

Go

Find People

Go

[PAO Home](#)

[Public Affairs](#) : [Newsroom](#) : [News Releases](#) : [2005 News Release Archive](#) :: [NR-05-06-04](#)

[Contact Us](#)

## News Release

[Printer-Friendly](#)

[Newsroom](#)

- [News Releases](#)
- [Science Features](#)
- [Media Advisories](#)
- [Statements](#)
- [Biographies](#)
- [Photos & Images](#)
- [Video Library](#)

Contact: Charlie Osolin  
Phone: (925) 422-8367  
E-mail: [osolin1@llnl.gov](mailto:osolin1@llnl.gov)

FOR IMMEDIATE RELEASE  
June 6, 2005  
NR-05-06-04

## New technique helps identify multiple DNA regulatory sites



[Community](#)

[Employee News](#)

[Learn More](#)

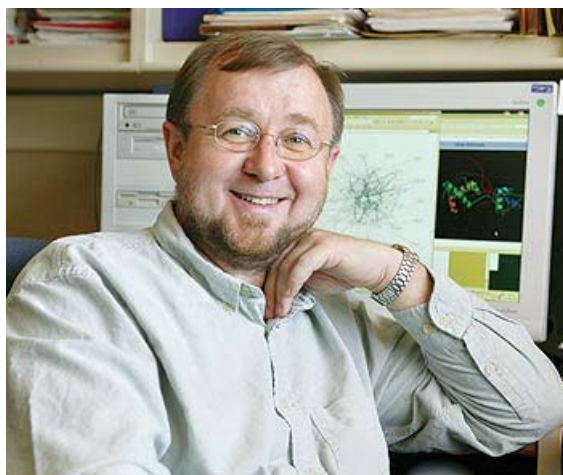
### More Information

["Discovering regulatory binding-site modules using rule-based learning"](#)  
*Genome Research*, July 2005

["Mining Genomes: Livermore computational tools are helping researchers worldwide understand how genes are regulated"](#)  
*Science & Technology Review*, June 2005

[LLNL Biosciences Directorate](#)

[The Linnaeus Centre for Bioinformatics](#)



Jacqueline McBride/LLNL

**Krzysztof Fidelis**

[\(Click here to download a high-resolution image.\)](#)

transcription factors at a time, or combinations of factors, are coming together physically and how these combinations regulate genes."

"To accomplish this," Komorowski said, "we used a machine learning technique called rough sets to mathematically model general rules that could associate known binding sites and gene expression in yeast, which is one of the most widely studied organisms." From the analysis of gene activity under a variety of environmental conditions, the teams were able to develop a set of rules for predicting the location of binding site combinations based on limited binding site and gene expression data.

"We found that the same transcription factors, in slightly different combinations, could be responsible for the regulation of different genes," said Torgeir R. Hvidsten of the LCB. "Thus we now know that binding sites can be combined to allow a large number of expression outcomes using relatively few transcription factors."

Others collaborating in the project were Jerzy Tiurny of the Faculty of Mathematics, Informatics, and Mechanics at Warsaw University in Poland; Bartosz Wilczynski of the Institute of Mathematics, Polish Academy of Sciences, and LLNL; and Andriy Kryshchak of LLNL. A report on the joint work appears in the June issue of the journal *Genome Research*.

The rough set technique was developed by Zdzislaw Pawlak in Poland in the 1980s and is particularly suitable to build models from incomplete and uncertain data. It has been used in applications ranging from medical and financial data analysis to voice recognition and image processing. Applied to gene regulation, the approach was able to predict the location of regulatory sites for about one-third of the genes in the yeast genome – a success rate as good as or better than other current techniques.

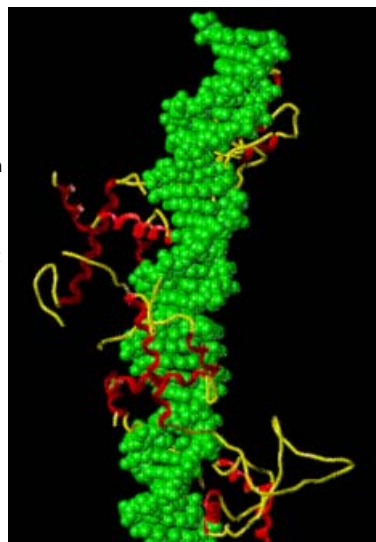
"The next step is to test this approach on different organisms, including microbes and vertebrates," Fidelis said. The growing number of organisms whose genomes have been sequenced has generated a wealth of DNA sequence information that could provide the raw material for analysis.

Primary funding for the research was provided by LLNL's Laboratory Directed Research and Development Program, the Knut and Alice

Understanding the complex regulatory mechanisms that tell genes when to switch on and off is one of the toughest challenges facing researchers attempting to discover how life works. "Binding sites," or areas of DNA that interact with the proteins that help control gene expression, can be a long distance on the DNA strand from the genes they influence. Recent research also has shown that gene expression can be controlled by several regulatory proteins working together at a combination of different binding sites.

(Regulatory proteins are known as "transcription factors"; transcription is the first step in the process by which the genetic information in DNA is decoded by the cell to manufacture proteins, the building blocks of life.)

"It's difficult to experimentally observe how transcription factors bind to DNA at a distance from a gene, or how regulation happens," said Fidelis, a computational biologist in Livermore's Biosciences Directorate. "But you can identify their binding sites in a promoter or regulatory region – there are usually a few of these for each gene. We wanted to see if we could somehow deduce how many



Wallenberg Foundation, and the Swedish Foundation for Strategic Research.

**Illustration of Rap1, an essential regulatory protein in yeast, binding to a strand of yeast DNA.**

Founded in 1952, Lawrence Livermore National Laboratory has a mission to ensure national security and to apply science and technology to the important issues of our time. Lawrence Livermore National Laboratory is managed by the University of California for the U.S. Department of Energy's National Nuclear Security Administration.

The Linnaeus Centre for Bioinformatics is a joint initiative between Uppsala University and the Swedish University of Agricultural Sciences. The LCB researchers are involved in many interdisciplinary bioinformatics projects, including epigenetics, evolutionary studies, functional genomics, and computational tool integration for life sciences.

For more information contact Krzysztof Fidelis, [fidelis@llnl.gov](mailto:fidelis@llnl.gov); or Jan Komorowski, [jan.komorowski@lcb.uu.se](mailto:jan.komorowski@lcb.uu.se).



Lawrence Livermore National Laboratory  
7000 East Avenue • Livermore, CA 94550

Operated by Lawrence Livermore National Security, LLC, for the  
Department of Energy's National Nuclear Security Administration.