

JUNE 22, 2005

Bioinformatics Reveals New Gene Regulation System

By comparing 140 sequenced bacterial genomes, researchers have uncovered a system for regulating genes essential to bacterial replication - and they did it solely by computer keystrokes and mouse clicks.

Mikhail Gelfand, a Howard Hughes Medical Institute international research scholar at the Institute for Information Transmission Problems (IITP) in Moscow, and his postdoctoral fellow, Dmitry Rodionov, used comparative genomics to identify a new transcription factor system in bacteria that represses expression of genes involved in DNA replication. They scanned gene sequences and proteomes of several taxonomic groups of bacteria, identifying not only a highly conserved signal sequence, but also the regulatory transcription factor that bound it, the repressor nature of the signal, and other genes also regulated by this system.

"This is a proof of principle that you can go a very long way by comparative genomics now."

— **Mikhail S. Gelfand**

"We provided a very detailed description of a system just by doing bioinformatics alone," says Gelfand, director of the IITP's research and training center of bioinformatics. "It's a proof of principle that you can go a very long way by comparative genomics now." Their findings will be published in the July issue of *Trends in Genetics*, with early publication now online. Gelfand is presenting the work on June 24, 2005, at the annual meeting of HHMI international research scholars in Mérida, Mexico.

Gelfand and Rodionov started their search using a technique called phylogenetic footprinting to review the upstream DNA sequences of a group of genes that code for ribonucleotide reductase enzymes. These enzymes convert the ribonucleotide building blocks of RNA into the deoxyribonucleotides used to build DNA. This conversion is critical for duplicating a bacterium's entire genome before it divides to reproduce.

The search revealed a conserved palindromic sequence occurring upstream of many ribonucleotide reductase (Nrd) genes. A genetic palindrome is a sequence of nucleotides on one strand of DNA that reads the same as the sequence on the opposite strand, only backwards - a common feature of DNA

sequences that are recognized by regulatory molecules. They designated the sequence the NrdR-box.

Because the signal was found in so many diverse groups of bacteria, the researchers thought it might represent a universal regulatory mechanism. The next question was whether the signal was promoting or repressing expression of Nrd genes.

The team observed that their signal always overlapped with the promoter signal, the region of DNA required for the initiation of the conversion of gene to protein. Molecules that promote transcription recognize and bind to this sequence, which lies just outside of the gene. Repressor signals commonly work by allowing other proteins to bind on top of the promoter sequence and physically block promoters. Therefore, the duo predicted that the NrdR-box functioned as a repressor sequence.

Next, the researchers identified the transcription factor protein that binds to the NrdR-box. To do this, they used a bioinformatics approach they call phylogenetic profiling, compiling a list of genomes that clearly contained the NrdR-box and those that clearly did not have it. Then they searched the proteomes of 63 bacteria species, looking for proteins that strictly followed the same present-or-absent pattern as the NrdR-box. Only one protein cluster matched the pattern, and it represented a family of proteins that shared traits of transcription factors.

To strengthen the prediction that these proteins were the transcription factors that bind the NrdR-box, the team used another comparative genomic tool called positional clustering. Positional clustering takes advantage of the fact that functionally related gene sequences (such as the genes for Nrd and its transcription factor) frequently inhabit the same 'neighborhood' of the chromosome.

"If you are looking in one genome, many genes will be neighbors by coincidence," Gelfand noted. "But if two genes are neighbors in many diverse genomes, then they are likely to be related." Indeed, the Nrd genes and the transcription factor genes clustered together, providing additional evidence that the regulatory picture drawn by the team was correct.

Israeli researchers simultaneously showed through 'wet' biology experiments in *Streptomyces* bacteria that a transcription factor from this family represses Nrd gene expression in the living bacterial cell, confirming the Russian researchers' predictions. Confident that they had identified a new repressor of bacterial genes, Gelfand and Rodionov searched genomes for other upstream sites where the NrdR-box occurred. They found that it regulates other genes related to DNA replication, such as the enzymes that cut, paste, and untangle new DNA as it is synthesized, and enzymes that are involved in recycling nucleotide building blocks.

Although the work does not have direct application to human medicine, Gelfand pointed out that many antibiotics work by attacking the process of bacterial DNA replication. So, he said, this work has identified potential

targets for designing new antibiotic drugs. But more importantly, the work shows how molecular discoveries of whole regulatory systems can be made through careful analysis of genomes—without ever lifting a pipette, he said.

“There are 100 enzymes functioning at the core of bacterial metabolism for which the genes are still unknown,” said Gelfand. Using multiple bioinformatics tools can uncover cell systems that might have escaped experimental detection, he suggested. “By comparing hundreds of genomes, you can see patterns that are not seen when looking at just a couple of them.”