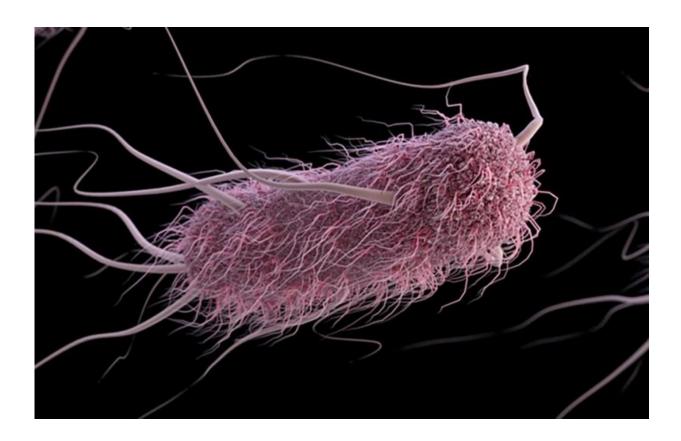# Computer model is 'crystal ball' for E. coli bacteria

October 28 2016, by Andy Fell

Escherichia coli is one of the most-studied of all bacteria, but there's still much we do not know about it. A new computer model developed at UC Davis will help scientists design experiments to study its biochemistry and metabolism. This could be useful in medicine and biochemistry. Credit: Centers for Disease Control

It's difficult to make predictions, especially about the future, and even more so when they involve the reactions of living cells—huge numbers of genes, proteins and enzymes, embedded in complex pathways and feedback loops. Yet researchers at the University of California, Davis, Genome Center and Department of Computer Science are attempting just that, building a computer model that predicts the behavior of a single cell of the bacterium Escherichia coli.

The results of their work were published Oct. 7 in the journal *Nature Communications*.

The new simulation is the largest of its kind yet, said Ilias Tagkopoulos, professor of computer science, who led the team.

"The number of layers, and the amount of data involved are unprecedented," he said. The dataset on which the model is based includes, for example, over 4,389 profiles of the expression of different genes and proteins across 649 different conditions. Both the dataset, named "Ecomics" and the integrated model, MOMA (Multi-Omics Model and Analytics) are available to other researchers to use and test.

The model could be useful to researchers as a fast and inexpensive way to predict how an organism might behave in a specific experiment, Tagkopoulos said. Although no prediction can be as accurate as actually performing the experiment, this would help scientists design their hypotheses and experiments. Applications range from finding the best growth conditions in biotechnology to identifying key pathways for antibiotic and stress resistance.

The "Blue Waters" supercomputer at the National Center for Supercomputer Applications was used to build the model. Credit: UC Davis

A week to download, 2 years to build

Collecting and downloading the data took a week, but processing the data into a single dataset took two years of the three-year project, Tagkopoulos said. The team built models for four layers, starting with gene expression and working up to the activity at the whole-cell level. Then they integrated the layers together. They used techniques in machine learning to train the models to predict the behavior of each layer, and ultimately of the cell itself, under different conditions.

The model was built on computer clusters at UC Davis, and on supercomputers available through a national network. The researchers received a National Science Foundation grant of computing time on "Blue Waters," one of the world's most powerful supercomputers, at the National Center for Supercomputer Applications.

Although E. coli is a well-known organism, we are far from knowing everything about its biochemistry and metabolism, Tagkopoulos said.

"We are exploring a vast space here," he said. "Our aim is to create a crystal ball for the bacteria, which can help us decide what is the next experiment we should do to explore this space better."

With collaborators at Mars Inc., Tagkopoulos hopes to begin building similar databases and models for bacteria involved in foodborne illness, such as Salmonella enterica and Bacillus subtilis. He expects other researchers to draw on the Ecomics database, and hopes to make the MOMA model interface more accessible for biologists to use.

"We're living in an amazing era at the intersection of computer science, engineering and biology," he said. "It's a very interesting time."

More information: Minseung Kim et al. Multi-omics integration accurately predicts cellular state in unexplored conditions for Escherichia coli, *Nature Communications* (2016). DOI: 10.1038/ncomms13090

Provided by UC Davis