

# New massive dataset of bacterial proteins

December 7 2015

---

Scientists from Switzerland and the Netherlands have conducted a quantitative and qualitative analysis of the proteins that the bacterium *Escherichia coli* expresses in 22 different growth conditions. More than 2,300 proteins were identified, some at average levels of one copy per cell. The resulting dataset describes most (>90 percent) of the protein mass in the cells and will be a treasure trove for cell biologists. A first description is published on 7 December in *Nature Biotechnology*.

In order to understand the connection between the genomic information present in cells and their physiology, it is important to assess which genes are active in producing proteins under different conditions. The most direct way to gather this information is with a quantitative measurement of proteins present in the cell.

Technical advances have only recently made the large-scale measurement of absolute [protein](#) levels possible. Scientists from the Universities of Basel and Zürich (Switzerland) and the University of Groningen (the Netherlands) joined forces to measure proteins in the *E. coli* bacteria, grown under 22 different conditions. Using a mass-spectrometry based proteomic approach, they not only identified which proteins were present, but also at how many copies per cell.

## Massive dataset

The results from a massive dataset that will inspire much new research, says Groningen Professor of Systems Biology Matthias Heinemann, who coordinated the experiment together with Alexander Schmidt (Basel).

'We managed to analyse 90 percent of the protein mass of these cells', he explains. 'We found over 2,300 different proteins, representing over half of the 4,300 bacterial genes.' This doubled the number of proteins that have been absolutely quantified in *E. coli*. For some of these proteins, no function has as yet been established. 'But looking at the pattern of expression over the 22 different growth conditions, we now have a clue about what they are doing.'

The proteins have very different expression levels, ranging from over 100,000 copies per cell to two, one or even less on average. 'First, this shows how sensitive our methods are, but it also makes you wonder what the function is of proteins that are expressed at very low levels.'

Although some genes can be active through purely stochastic effects (and thus produce proteins by chance), Heinemann doesn't rule out a proper function for single copies of a protein in a cell. 'After all, other biological entities which appear as single copies, like genes, do have a function.' The study has also uncovered new post-translational adaptations that are made to [bacterial proteins](#).

## **New questions**

The dataset described in the *Nature Biotechnology* paper is already being used by other scientists and triggering new exciting investigations. 'Our data will act as reference data for new research and has already led to a number of studies, which are in the process of being published. This dataset allows scientists to ask and answer new questions.'

For this study, bacteria were grown under different conditions at the University of Groningen. Samples were taken and shipped to Basel, where the protein content (including membrane-bound proteins) was isolated and analysed using mass spectrometry. Finally, the results were analysed by the entire team.

**More information:** The quantitative and condition-dependent Escherichia coli proteome, [dx.doi.org/10.1038/nbt.3418](https://doi.org/10.1038/nbt.3418)

Provided by University of Groningen

Citation: New massive dataset of bacterial proteins (2015, December 7) retrieved 7 May 2024 from <https://phys.org/news/2015-12-massive-dataset-bacterial-proteins.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.