

The Proteome Landscape of the Kingdoms of Life

PharmaFluidics spoke with Philipp Geyer and Johannes Mueller of Prof. Matthias Mann's Group, Max Planck Institute of Biochemistry, Martinsried, Germany, on the occasion of their *NATURE* publication.

Proteins perform the vast majority of functions in biology, but still large-scale proteome investigation has lagged behind compared to the genomics initiatives; what challenges do you see and what is missing?

On the project side the genome research got a lot of visibility worldwide due to the human genome project. On the technology side, the main hurdle for proteomics so far has been the lack of standardized, reliable workflows with high throughput, and which are truly reproducible across time and amongst laboratories. This relates to sample preparation, mass spectrometry, bioinformatics, but chromatography has been an especially weak link which we have addressed here in particular with PharmaFluidics μ PAC™ technology.

Where do you think the next steps for the project will go, how would you envision the proteomics community could build further on this and what opportunities will this open?

Next steps will obviously be to gather and analyze even more organisms and also cellular subsets to complement the current database, and to progress the understanding about the relative importance of proteins and about the way they interact. Such knowledge can also be used to identify model organisms with adequate protein targets for drug and toxicity testing.

The existence of this huge dataset will allow researchers to develop workflow strategies based on accurate retention time prediction.

A key insight from the study was on machine learning and big data, which allows to generate de novo proteome data based on high quality spectra and subsequently validate and reverse-annotate genomic data.

From the biology side it was interesting to observe that key household proteins which control and regulate the production of proteins - such as chaperone proteins and ribosomal proteins - are highly abundant and expressed with great constancy over the entire tree of life.

Could you tell something about the impact of the publication for your careers as researchers?

"We started working with the μ PAC™ columns about 2 years ago, and progressively started to take on more organisms to feed our big data algorithms. By the time we had collected about 40 organisms, we did some serious data crunching and saw insights emerging across 2 dimensions: on the peptide Retention Time stability, for the technology, and on organism meta-proteomes, for the biology." Says Johannes. "We then started to realize that we were on track to make a novel, substantial step ahead in machine learning based proteomics".

Both Philipp and Johannes are very excited and it is a blessing to have a first "first-author" paper in *NATURE*, one of the journals with the highest impact factor.



How would you qualify the impact of high quality HPLC/MS data on the performance of the machine learning algorithm?

This is obviously a key parameter. The variance on LC retention time is a potential limitation of our machine learning model.

We observed the Retention Time stability of the μ PAC™ columns to substantially outperform packed bed columns at similar separation performance. This is obviously a key advantage, as more and better-quality data are essential for machine learning.

How do you foresee the importance of Retention Time stability for fast spectral database searches with low false discovery rates?

When using match-between-runs algorithms, this would definitely have a positive effect on the number of correctly annotated features and therefore increase the value of the data generated.

How would you qualify the importance of inter-laboratory reproducibility for data pooling and community effort?

The microstructured and extremely reproducible μ PAC™ chromatographic system is conducive to generation of standardized data sets across time and site; this is also of high importance for future clinical and diagnostics purposes.

How will the “proteomesoflife.org” website, database work? Is it consultative or accretive? Who can contribute and in what conditions?

For the moment, the database is a consultative tool providing data about identified proteins and their abundance. It would be our intention to open up the database for adding newly annotated proteins; obviously retention time data could be added for further standardization, which would then require to use the same gradients and columns.



Philipp Geyer has done his PhD in Matthias Mann’s laboratory ‘Proteomics and Signal Transduction’ at the Max Planck Institute of Biochemistry (MPI) in Martinsried, focusing on new technology developments to unlock the potential of the plasma proteome and its clinical applications. After finishing his PhD, he continued his research at the MPI and the Center for Protein Research in Copenhagen, where he headed the plasma proteomics efforts of the Mann laboratories. Philipp has the vision that mass spectrometry-based plasma proteomics can change diagnostics and clinical decision making in a way that will substantially increase our life expectations and the quality of how we live. To fulfill this vision, he has founded together with proteomics and AI experts the clinical proteomics company OmicEra Diagnostics GmbH.



Johannes Mueller is a PhD student at the Max Planck Institute of Biochemistry in Martinsried at the Department of Proteomics and Signal Transduction lead by Matthias Mann. He joined the group in 2018 after receiving a Masters degree in Biochemistry from the Technical University of Munich and is focused on application and development of mass spectrometry (MS)-based proteomics techniques, especially concerning liquid chromatography (LC) separation. With the publication of ‘The proteome landscape of the kingdoms of life’, Johannes’ dream to employ the highly specialized LC-MS based method to increase the knowledge about organisms from all domains of life comes true. He aims to follow up the topic and explore more proteomes of all known life forms in the future.



Prof. Dr. Matthias Mann is a pioneer, eminent researcher and global authority in the field of proteomics. With his research teams at the Max Planck Institute in Martinsried, Germany and at Novo Nordisk Foundation Center for Protein Research at the University of Copenhagen, Denmark, Prof. Mann progressively turned LC/MS into a highly effective tool for characterizing the totality of proteins in an organism or even a single cell. With more than 700 publications in proteomics and bioinformatics, Prof. Mann achieved a h-index ranking of 231 and Google Scholar lists him with more than 247,000 scientific citations. From his research group in Munich originated in 2016 PreOmics GmbH – a company commercializing sample prep sets and recently OmicEra Diagnostics GmbH – a clinical proteomics company. Current interests include machine learning with large proteome data sets; and the standardization of routine LC/MS workflows to bring proteomics into clinical and diagnostic labs.