Prof. Jürgen Cox
Max Planck Institute
of Biochemistry

Customer Insights

- Unravelling high-throughput 4D-Proteomics data with dedicated software

Unravelling high-throughput 4D-Proteomics data with dedicated software

Researchers at the Max Planck Institute of Biochemistry (MPIB) are developing sophisticated software for high-throughput MS-based 4D-Proteomics

Working with Bruker
Professor Jürgen Cox, Research Group Leader of Computational Systems Biochemistry at the Max Planck Institute of Biochemistry, has developed industry-leading software for mass spectrometry (MS)-based proteomics analysis, in partnership with Bruker:

"Thanks to MS instrument developments, such as Bruker's 4D-Proteomics workflows on the timsTOF Pro, combined with our ever-developing software, the field of proteomics can work in a high-throughput way to achieve deeper insights."

Computational Systems Biochemistry at MPIB

The Computational Systems Biochemistry group is one of approximately 30 scientific departments and research groups at the Max Planck Institute of Biochemistry (MPIB) in Martinsried, Munich. The MPIB is one of the largest institutes of the Max Planck Society, with around 800 employees and expertise in the fields of biochemistry, cell and structural biology, and biomedical research.

Professor Jürgen Cox, PhD, is the research group leader of Computational Systems Biochemistry at the MPIB, which focuses its research on systems biology, proteomics, mass spectrometry (MS), and bioinformatics. Prof. Cox dedicates his work to the development of computational approaches to analyze large scale data, primarily in the field of proteomics, but also to answer questions in metabolomics, transcriptomics, and lipidomics. He joined the MPIB in 2006 as a senior scientist and became the Computational Systems Biochemistry research group leader in 2014.

Together with his team, which includes four post-doctoral researchers, between six and seven PhD students, and several masters and bachelor students, Prof. Cox develops algorithms and tools for analyzing the large volume of spectral data produced in modern proteomics experiments. The MaxQuant software platform is one of the most widely used platforms in computational proteomics, enabling the analysis of large mass spectrometric data sets. Freely available for academic and non-academic researchers, many laboratories across the globe benefit from MaxQuant's precise protein and peptide quantification algorithms. The group also developed Perseus, a software platform that supports researchers in the interpretation of protein quantification and interaction data as well as data on post-translational modifications (PTMs). MaxQuant performs quantification with labels and via the MaxLFQ algorithm on label-free data, and achieves high peptide mass accuracies thanks to its advanced nonlinear recalibration algorithms.

Proteomics research challenges

The modern high-throughput MS-based proteomics methods that are required to gain deeper insights into biological processes produce enormous amounts of data. This raw data necessitates powerful, automated computer-based methods that provide reliable identification and quantification of proteins. Prof. Cox describes the development of MaxQuant software for the proteomics field:

"When we started developing MaxQuant, it was clear the proteomics field lacked automation and standards. It was not uncommon for PhD students to manually print out and look at the data for every peptide they wanted to identify in an experiment. Quantification was also done manually. That's not scalable at all. The development of MaxQuant enabled an immediate transition to high-throughput proteomics. Now software can take the raw data output from the instruments and give you back the identification of proteins and their quantities."

MaxQuant is often used for liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) shotgun proteomics – a method of identifying proteins in complex mixtures to provide a wider dynamic range and coverage of proteins. Shotgun (or bottom-up) proteomics is the most commonly used MS-based approach to study proteins by digesting them into peptides prior to MS analysis. The software possesses a large ecosystem of algorithms for comprehensive data analysis. It incorporates the peptide search engine Andromeda and, coupled with Perseus, offers a complete solution for downstream bioinformatics analysis [1].

Over the last two decades, significant advances in technology and new methodologies have made proteomics an extremely powerful tool for protein scientists, biologists, and clinical researchers [2]. As analytical instrumentation continues to evolve, more data is produced with each technical advancement in proteomics research. That, of course, also creates new challenges for bioinformatics software development.

The addition of the ion mobility dimension in Bruker's timsTOF Pro™ delivers greater sensitivity, selectivity, and MS/MS acquisition speeds for proteomics research. The novel design allows for ions to be accumulated in the front section, while ions in the rear section are sequentially released depending on their ion mobility, and in subsequent scans selected precursors can be targeted for MS/MS. This process is called Parallel Accumulation Serial Fragmentation, or PASEF [3]. The unique trapped ion mobility spectrometry (TIMS) design allows researchers to reproducibly measure the collisional cross section (CCS) values for all detected ions, and those can be used to further increase the system's selectivity, enabling more and more reliable relative quantitation information from complex samples and short gradient analyses.

The increase in speed resulting from PASEF technology allows more samples to be analyzed in a shorter time frame but generates vast amounts of spectra data, creating challenges when dealing with large sample cohorts. The

MaxQuant shotgun proteomics workflow was adapted to extract this abundance of information from the timsTOF Pro, making it possible to manage 4D features in the space spanned by retention time, ion mobility, mass and signal intensity that benefit the identification and quantification of peptides, proteins and PTMs. Prof. Cox explains the impact of these technology advances on MaxQuant:

"TIMS adds another dimension, from ion mobility to proteomics, and that's another piece of evidence to identify.

It's challenging because it's not just one piece of new information, it's a whole dimension. So, we've adapted MaxQuant for 4D-Proteomics workflows like PASEF, data-independent acquisition (dia)-PASEF and Mobility Offset Mass Aligned (MOMA), particularly for Bruker's timsTOF Pro. Adding another dimension can lengthen algorithm processing times, creating a significant challenge for software development with 4D-Proteomics. We have optimized computation time in MaxQuant to overcome this, so users can achieve good results in a reasonable time frame."

Clinical research proteomics

One of the research areas of particular interest to Prof. Cox and his team is clinical research proteomics. However, the analysis of proteomics data from samples derived from patients requires special computational strategies. The problems that need to be addressed include: how to extract meaningful protein expression signatures from data with high individual variability, how to integrate the genomic background of the patients into the analysis of proteomics data, and how to determine biomarkers and properly estimate their predictive power. Prof. Cox describes this research:

"We make use of machine learning algorithms to classify patients and employ feature selection algorithms to extract predictive protein signatures. The question for the software developer is the extra clinical test engine. The field is working on getting mass spectrometry into clinical practice, and it may become a diagnostic tool in 5-10 years. Or it could be a translational tool. It's not clear yet if proteomics will be a guide to which molecules to look at, or if it will be a major component of clinical diagnostics. Either way is exciting.

We are working with several clinical groups to bring MS-based proteomics into clinical practice – we think clinical research proteomics will be one of the main applications of the future, and instruments like the timsTOF Pro will greatly benefit this endeavor."

Single-cell proteomics

Clinical research proteomics is only one area where Prof. Cox and his team see upcoming technical advancements for MaxQuant. While today's laboratories generate large data sets from single-cell genomic and single-cell transcriptomic research, single cell proteomics (sc-proteomics) is a nascent field. It holds the potential to enable researchers to compute the proteins in single cells, avoiding the need to infer proteins from cellular mRNA levels [4]. That also creates new challenges for computational analysis. Prof. Cox and his team are looking ahead to future needs as sc-proteomics develops, looking closely at emerging technologies to establish quantification standards. He explains the potential impact in the field:

"Single-cell proteomics is just around the corner. It's achievable in the next few years. You need sensitivity in the instrumentation, which Bruker provides with the timsTOF Pro, and you need the software to work with it."

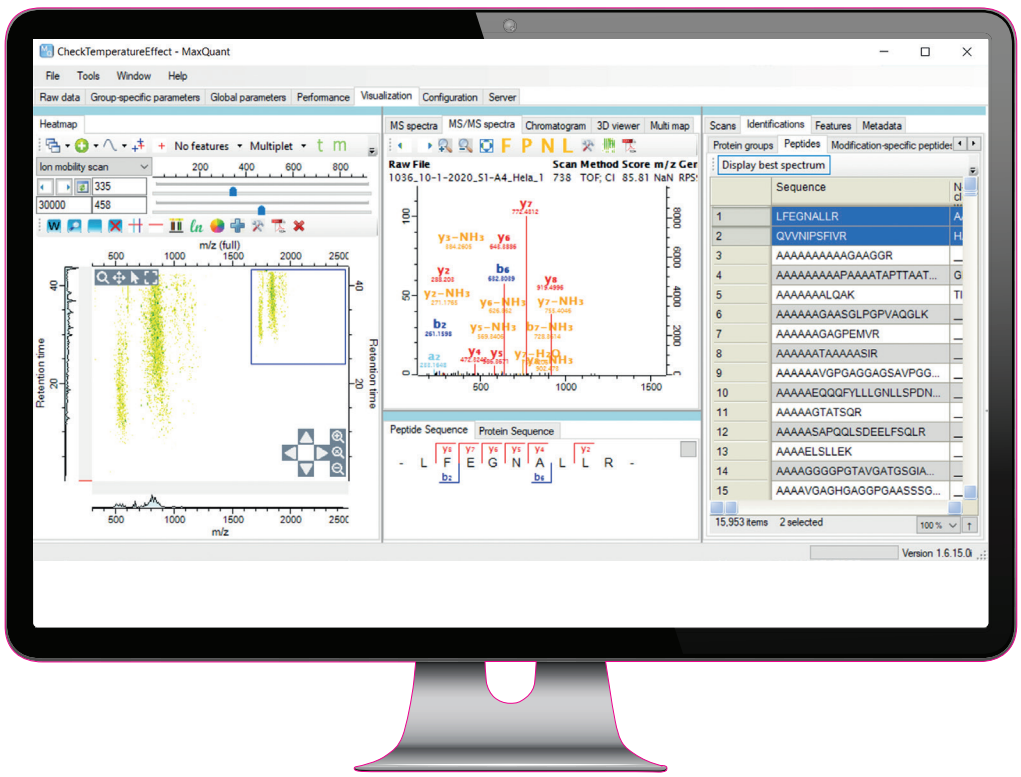
Crucial collaborations

The ongoing improvement of the MaxQuant software platform is possible due to important collaborations between the Computational Systems Biochemistry research team and other institutions and industry leaders, such as Bruker. These collaborations include projects designed to improve proteomics technologies, with MaxQuant providing the computational tools necessary to analyze data acquired on new and emerging hardware platforms. Prof. Cox describes the impact of these collaborations:

"We are all working towards the same goal, which is to achieve better algorithms and tools for proteomics. Our core competency is proteomics data in general, and we combine our expertise with that of leaders in the field. The results of these collaborations are used to develop future versions of the software to optimize 4D-Proteomics workflows."

For example, the addition of TIMS to LC-MS based shotgun proteomics, powered by PASEF, has the potential to boost proteome coverage, quantification accuracy, and dynamic range, resulting in fast ultra-sensitive analysis. However, it requires suitable software that extracts the information contained in the 4D data space spanned by mass-to-charge ratio (*m/z*), retention time, ion mobility, and signal intensity. The collaboration with Bruker on the timsTOF Pro was a natural extension of the organizations' previous work together, as Prof. Cox explains:

"I started collaborating with Bruker when I became Research Group Leader at MPIB in 2014, initially with the impact II instrument and then with the timsTOF Pro when it was launched in 2017. Based on our successful earlier work it was obvious we should work together to develop features in MaxQuant to enable 4D-Proteomics data processing. The collision cross sections obtained are proving to be very valuable in reducing false positives in the 'match between runs' analysis used in label-free quantitation" [5].



Prof. Cox comments on why he has chosen to work closely with Bruker over the years:

"Other than the company's high-quality instruments, we enjoy working with Bruker because we have a great relationship with the personnel. We like the fact that we work with scientists with a deep level of expertise."

These collaborations provide a constant feedback loop, where the MaxQuant developers receive comments and opinions from end users in the proteomics community and use that information to improve the software. These valuable contributions are just one benefit of these long-term collaborative relationships. Prof. Cox explains:

"We always try to look at MaxQuant from the user's perspective – the biologist or medical expert or mass spectrometry expert. We're really trying to get in the heads of these people. That affects how we present information in the software to the user and how the user interacts with it. It's a big issue. In many platforms, there's a divide between the mathematical component and the biological component. It can be biased towards the programmer. So, we always ask the questions, 'What does the user experience? What are the hurdles to using the software successfully?' We want MaxQuant to be biologist- or medical researcher-friendly."

Future perspectives

Prof. Cox and his team remain committed to developing MaxQuant to meet the future needs of researchers. For example, recent advances in data-independent acquisition (DIA) sensitivity have encouraged Prof. Cox and his team to integrate DIA workflows into MaxQuant using machine learning algorithms.

The success of DIA relies on key instrumental capabilities, namely: resolution, sensitivity, accuracy, and dynamic range uncompromised by a fast-spectral acquisition rate. DIA has been implemented in the timsTOF Pro in a way that takes advantage of the speed and sensitivity of TIMS and PASEF, in a method called dia-PASEF. The 4D nature of the dia-PASEF data is an advantage for DIA software developers, who can take advantage of the additional ion mobility dimension for alignment and extraction of features. Such recent technological advances, as well as developments in DIA methods, have provided new opportunities for the Computational Systems Biochemistry research group. Prof. Cox predicts MaxQuant users will be very excited about the platform's upcoming DIA capability:

"It will have quite an impact. DDA and DIA are becoming comparable because of improved sensitivity in instrumentation. The hardware is becoming simpler for users, but data has been much more challenging because the software must find which fragments belong to which molecule.

Now we have addressed this, it gives us a deeper coverage of proteomics, for example for clinical research applications."

Bruker's timsTOF Pro is an example of how powerful developments in MS technology have led to the expansion of MaxQuant and the software's ability to meet future needs in the proteomics field. These improvements are helping researchers develop new capabilities and applications by delivering more sensitivity and selectivity for the identification and quantification of peptides, proteins, and PTMs. Prof. Cox describes how instrumentation advances like the timsTOF Pro affect MaxQuant's development:

"The timsTOF Pro has definitely made a substantial contribution to the field.

As technology advances, we must deal with even larger amounts of information on the software side because instruments continue to expand their dynamic range and capabilities. We're always expanding and improving MaxQuant to meet the complexity of biological processes and novel MS instruments. It's always changing."

For more information about the Computational Systems Biochemistry group, please visit https://www.biochem.mpg.de/cox.

For more information about the timsTOF Pro, please visit https://www.bruker.com/products/mass-spectrometry-and-separations/lc-ms/o-tof/timstof-pro.html.

References

[1] Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV and Mann M (2011). *Andromeda: A peptide search engine integrated into the MaxQuant environment*, J. Proteome Res. **10**(4): 1794–1805.

[2] Cox J and Mann M (2011). *Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology*, Annu. Rev. Biochem. **80**: 273-299.

[3] Meier F, Brunner AD, Koch S, Koch H, Lubeck M, Krause M, Goedecke N, Decker J, Kosinski T, Park MA, Bache N, Hoerning O, Cox J, Räther O, Mann M (2018). *Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer*, Mol Cell Proteomics. **17**(12):2534-2545.

[4] Marx V (2019). *A dream of single-cell proteomics*, Nature Methods. **16**: 809–812

[5] Prianichnikov N, Koch H, Koch S, Lubeck M, Heilig R, Brehmer S, Fischer R, and Cox J (2020). *MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics*, Molecular & Cellular Proteomics, **19**(6):1058-1069

Customer Insights

Unravelling high-throughput 4D-Proteomics data with dedicated software

Researchers at the Max Planck Institute of Biochemistry (MPIB) are developing sophisticated software for high-throughput MS-based 4D-Proteomics

About the Max Planck Institute of Biochemistry

The Max Planck Institute of Biochemistry (MPIB) in Martinsried, Munich is one of the leading international research institutions in the fields of biochemistry, cell and structural biology and biomedical research. With about 30 scientific departments and research groups and about 800 employees, the MPIB is one of the largest institutes of the Max Planck Society.

The approximately 480 scientists, coming from 45 different nations, study the structure of proteins - on single molecules, but also on complex organisms. Their work and the support of various central service facilities make the MPIB a leading international institute in the field of protein research. The high quality of the research work is also reflected in numerous awards and prizes. Two scientists have already been awarded the Nobel Prize: Feodor Lynen in 1964 and Robert Huber in 1988.

About Bruker Corporation

Bruker is enabling scientists to make breakthrough discoveries and develop new applications that improve the quality of human life. Bruker's high-performance scientific instruments and high-value analytical and diagnostic solutions enable scientists to explore life and materials at molecular, cellular and microscopic levels. In close cooperation with our customers, Bruker is enabling innovation, improved productivity and customer success in life science molecular research, in applied and pharma applications, in microscopy and nanoanalysis, and in industrial applications, as well as in cell biology, preclinical imaging, clinical phenomics and proteomics research and clinical microbiology.

For more information, please visit: **www.bruker.com**

Bruker Daltonics GmbH & Co. KG

Bremen · Germany
Phone +49 (0)421-2205-0

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com - www.bruker.com