

Proteomics

Optimized one-pot single-cell proteomics workflow

Maximize a combination of high-confidence coverage and throughput

Authors

Karl Mechtler^{1,2,3,4}, Manuel Matzinger^{1,2,3,4},
David Hartmayr⁵, Maciej Bromirski⁶

¹Research Institute of Molecular Pathology (IMP), Vienna BioCenter, Vienna, Austria

²Gregor Mendel Institute of Molecular Plant Biology (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Vienna, Austria

³Institute of Molecular Biotechnology (IMBA), Austrian Academy of Sciences, Vienna BioCenter (VBC), Vienna, Austria

⁴Vienna BioCenter Core Facilities (VBCF), Vienna, Austria

⁵Cellenion, Lyon, France;

⁶Thermo Fisher Scientific, Warsaw, Poland

Keywords

Single-cell proteomics, wide window acquisition, proteoCHIP, FAIMS Pro interface, Vanquish Neo UHPLC, μ PAC column, Orbitrap Exploris 480 mass spectrometer, Proteome Discoverer software, CHIMERYs, DIA, WWA, LFQ

Introduction

Mass spectrometry-based proteomics focuses on bulk-cell analysis. Such approaches can fail to uncover the differences between individual cells, which is crucial for understanding diseases. Single-cell proteomics is essential to understanding biological processes and disease states, revealing the full complexity of cellular diversity. In the past, single cell analyses relied on measurements of nucleic acids such as mRNA, where polymerase chain reaction (PCR) technology can be used to amplify low-level nucleic acids to enhance sensitivity. While these studies provide useful biological information, protein measurements offer a more direct depiction of cellular states. Cellular mRNA does not necessarily correlate with protein abundances, nor does it provide information about protein post-translational modifications (PTMs).¹

A single mammalian cell contains about 100 to 300 pg of protein,²⁻⁴ which is at least 1,000 times less than the amounts measured in traditional bulk proteomics analyses. Without PCR-like amplification to enhance detection of low-level proteins and peptides, obtaining a comprehensive picture of the proteins in a cell has been difficult. Another challenge is that in individual cells, concentrations of various proteins can span a range of over ten orders of magnitude. At the same time, research groups are looking for ways to increase speed of analysis, measure more samples per day, analyze large sample cohorts and accelerate completion of impactful large-scale studies.

When performing liquid chromatography-mass spectrometry (LC-MS)-based single-cell proteomics, every workflow step and technology used, from single cell isolation and preparation to LC-MS separation and data acquisition and analysis, must be optimized to obtain the best results. With advances in the technologies used in LC-MS-based proteomics workflows, sensitive, reproducible, and efficient single-cell proteomics is now possible. This technical note presents these advances and how they work in concert to maximize the combination of protein and peptide identifications and sample throughput. Additionally, we also introduce a novel acquisition strategy called wide window acquisition (WWA).^{5,6} WWA uses a broad isolation window ($\geq 4 m/z$) for data-dependent precursor selection in concert with CHIMERYs™ intelligent search algorithm by MSAID™ improving peptide identification for single-cell proteomics experiments.

While an isolation width of 4 m/z was found ideal to maximize protein IDs for standard injection amounts (200–400 ng), the optimal width shifted towards even wider windows with lowered input reaching an optimum around 8–12 m/z for single cell level inputs (see Figure 5, results section). Indeed, it seems intuitive that wider isolation windows are beneficial to reach similar levels of complexity for single cell samples as compared to bulk samples. This intentional co-isolation of precursors furthermore gives an advantage in resulting quantitative accuracy since more datapoints and more peptides are assigned per peptide on average.⁵

Overview: advanced workflow setup

The following components combine to create an optimized workflow platform that addresses the challenges of performing sensitive, robust, and rapid single-cell proteomics (Figure 1):

- cellenONE® and proteoCHIPs from Cellenion
- Thermo Scientific™ Vanquish NEO™ UHPLC system
- Thermo Scientific™ μ PAC™ Gen2 HPLC columns
- Thermo Scientific™ FAIMS Pro interface
- Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer
- Thermo Scientific™ Proteome Discoverer™ software featuring CHIMERYs™ intelligent search algorithm from MSAID™

Advanced technologies enable optimized single-cell proteomics workflow

The cellenONE performs image-based isolation of cells and particles of a wide range of sizes and shapes, at a rate of 40 cells/min. The user defines the isolation parameters, including size (diameter), morphology (degree of elongation), and imaging method (brightfield or fluorescence). Using brightfield imaging, cells between 3 and 80 μm can be isolated. In the fluorescence imaging mode, cells and particles as small 0.5 μm can be selected for isolation based on four fluorescence channels. Though scientists typically choose to isolate round cells, it is possible to reliably isolate very elongated cells such as cardiomyocytes. By adjusting the fluorescence (optional), elongation and diameter parameters for cell selection, individual cells can be accurately isolated, avoiding undesired selection

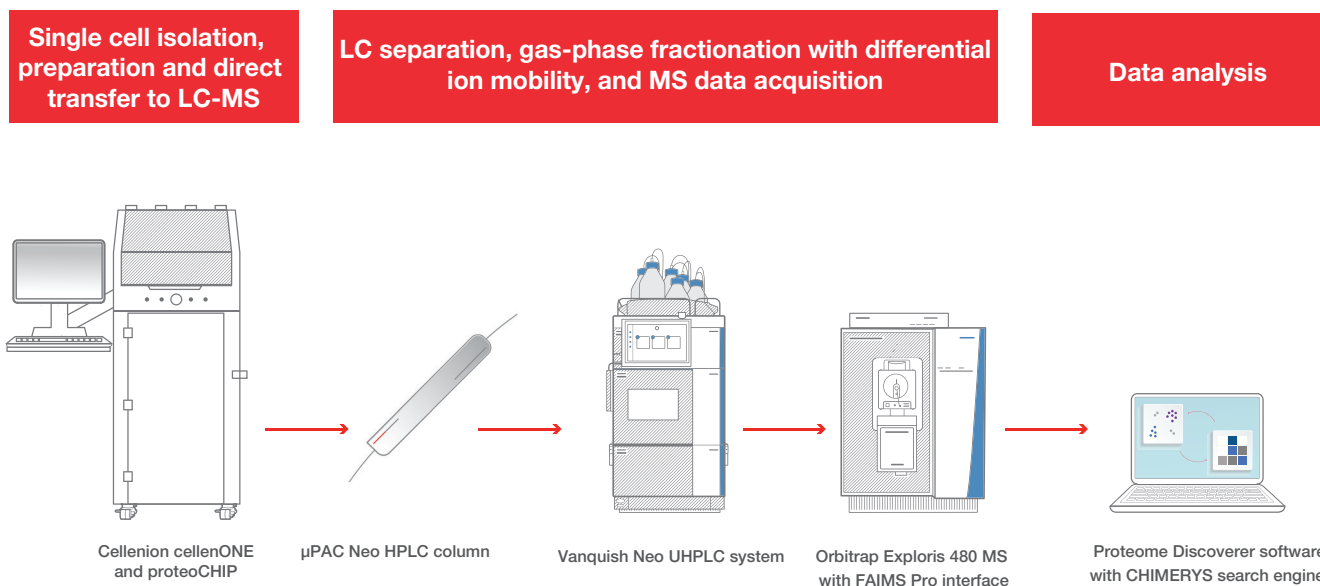


Figure 1. Advanced single-cell proteomics workflow technology

of cell agglomerates, multiple cells or apoptotic cells, or cellular debris. The optimal isolation parameters can be mapped and defined before each experiment. Compared to other cell sorting technologies, only about 1–50 μL of cell suspension is required at comparatively low concentration (100–200 cells/ μL), which makes the cellenONE ideal for sorting rare cell species. In combination with the instrument's climate controls and pL-volume dispensing capabilities, the cellenONE allows for automatization of single-cell proteomics experiments in sub- μL volumes, where evaporation effects otherwise would become problematic.

High precision acoustic-based dispensing is applied for liquid handling. Droplets are produced by a glass capillary surrounded by a piezoelectric ceramic that induces acoustic waves when a voltage is applied. This process generates droplets of a defined size, usually about 300 pL (range of 100–500 pL) per drop, depending on the voltage, pulse, and nozzle size. The drop volume is determined by the parameters chosen and is very stable with a CV of <0.2%. Organic as well as aqueous solutions can be dispensed.

Figure 2 illustrates how the cellenONE device dispenses single cells. To begin, 1–50 μL of cell suspension is aspirated into its glass capillary and the capillary is placed in front of the camera. The capillary tip is segmented into two virtual zones: the ejection zone (Figure 2, green zone) that contains exactly one cell and corresponds to the volume of the next generated droplet, and the sedimentation zone (Figure 2, pink zone) used as a safety margin towards other cells. If a single cell fitting the user-defined parameters is detected, the cell is dispensed into a target well on the proteoCHIP. If not, the contents are dispensed into a recovery vial directly below the camera. The images are saved and can be retrieved for re-inspection.

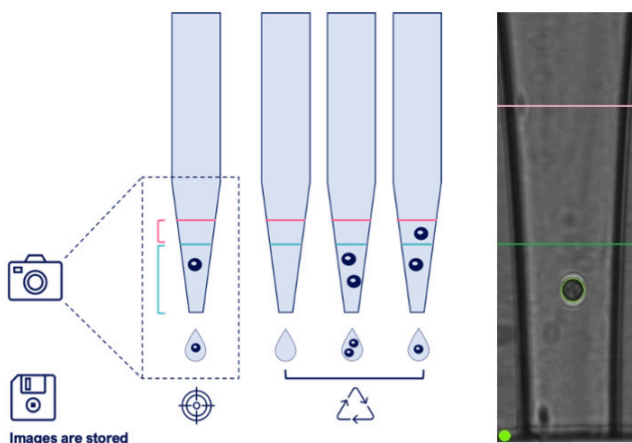


Figure 2. Precision single cell isolation using cellenONE. The ejection zone is shown in green and the sedimentation zone is shown in pink (Figure and photo courtesy of Cellenion).

Label-free, automated, high-recovery one-pot sample preparation using standard 384-well plates

Sample lysis and digestion is performed within a clean 384-well plate inside the cellenONE, as described earlier.⁷ In brief, 1 μL of a master mix containing 0.2% DDM, 100 mM TEAB, 3 ng/ μL trypsin, 0.01 % enhancer (ProteaseMAX™) is predispensed into the wells. To limit evaporation, humidity is set to 85%. Individual cells at 18–25 μm diameter and a max elongation of 1.5 are sorted by the cellenONE into the respective wells. This is followed by incubation for 2 h at 50 °C at 85% relative humidity inside the instrument. Samples were kept hydrated every 15 min by automated addition of 500 nL water to each well. After 30 min of incubation, an additional 500 nL of 3 ng/ μL trypsin is added, which replaces one hydration step. After digestion, samples are acidified for quenching and supplemented with 5% DMSO to improve the solubility of hydrophobic peptides, thereby improving recovery. Samples can be directly injected from the 384-well plate.

To generate single cell stocks, 20x HeLa cells are isolated per well and are processed exactly as explained for single cells above. They are supplemented with 8 μL of 5% DMSO/0.1% TFA, and the equivalent of one cell (0.4 μL) is used per injection. Since technical replicates can now be injected from the same well, this strategy enables methodological performance benchmarking of single cell like samples without any biological variance and allows visualization of acquisition dependent differences using a lowered total replicate number.

Figure 3A illustrates the importance of each improved workflow parameter. In the initial setting, adding trypsin only once, having the sample dried, and omitting DMSO, allowed for ~1,000 proteins to be identified from single cell stocks mimicking real single cell samples. Each of the following: sequential trypsin addition, sample hydration, and DMSO supplementation improves proteome coverage, yielding more than 1,300 proteins and 4,500 peptides identified from the equivalent of a single cell. These improvements led to the identification of 527 additional proteins compared to the initial settings, while re-identifying more than 90% of the proteins seen with the initial settings (Figure 3B) and suggesting the superiority of the improved workflow.

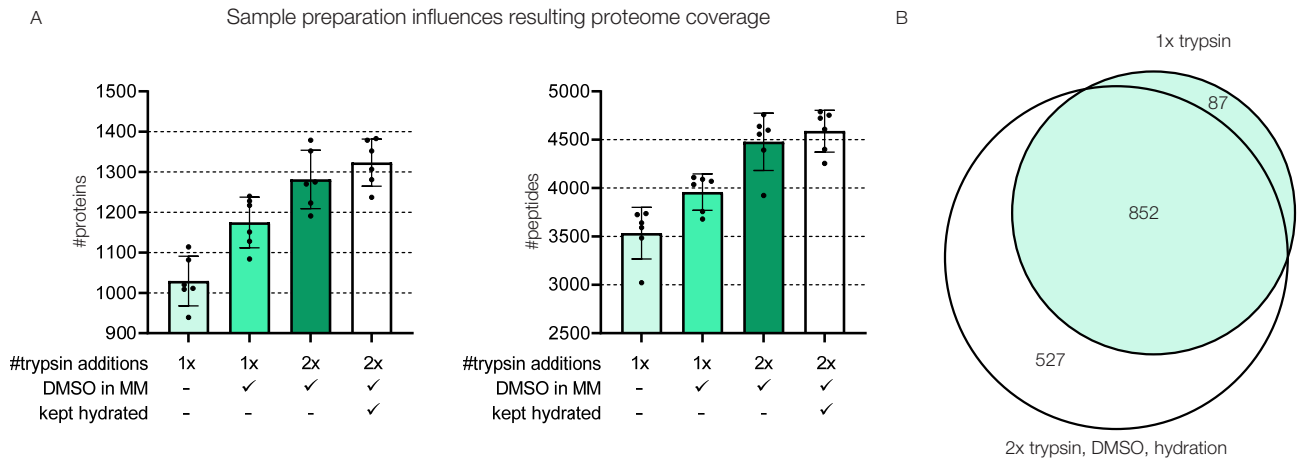


Figure 3. (A) Single cell stocks, LF-SCP workflow as published previously⁷; WWA(iso12) acquisition as published previously⁵, data analysis in Proteome Discoverer 3.0 software using CHIMERYS, 1%FDR, no MBR; **(B)** Venn overlap of identified proteins of representative replicates from (A)

Next-generation separating power: Vanquish Neo UHPLC system with μ PAC Neo HPLC columns

The Vanquish Neo UHPLC system is designed to provide excellent LC system performance and unmatched retention-time precision for shallow gradients to high-throughput applications. The Vanquish Neo UHPLC system is the first UHPLC system built to excel across the entire low-flow range that is ideal for proteomics applications including ultralow flow rates down to 100 nL/min required for single cell analysis.

Thermo Scientific™ μ PAC™ HPLC columns deliver the ultra-high resolution needed to extract the maximum amount of information from highly complex samples in low-flow methods. These silicon-based micropillar array HPLC columns are comprised of highly ordered, superficially porous brick-shaped pillars that enable excellent peptide separation and fast loading and equilibration due to exceedingly low backpressure. Each column is manufactured using the same lithographic mask, providing exceptional column-to-column reproducibility, low-flow rate flexibility from 100–1,000 nL/min at moderate LC pump pressures up to 350 bar, and excellent retention-time stability. Perfect order in the column eliminates peak dispersion in low-flow applications, providing substantially sharper and thus higher intensity peaks, increasing chromatographic resolution and sensitivity for proteomics experiments. With a factor-of-two reduction in pillar diameter and interpillar distance, the μ PAC Neo HPLC columns provide a substantial increase in separating power, expanding the boundaries of single-cell proteomics coverage (Figure 4).

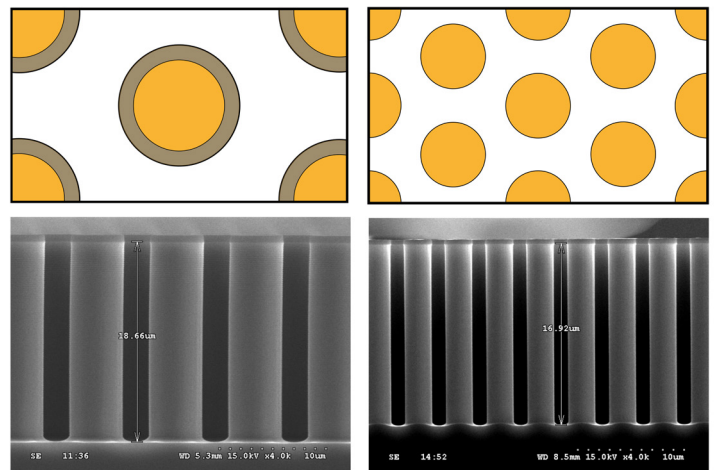


Figure 4. Comparison of μ PAC (left) and μ PAC Neo (right) HPLC column

Orbitrap Exploris 480 mass spectrometer with FAIMS Pro interface: a powerful combination

The Orbitrap Exploris 480 mass spectrometer empowers life-science researchers with single cell level sensitivity, quantitative confidence and rigor, and sample throughput that together led to impactful results. From label-free quantitation (LFQ) and quantitative multiplexing of proteomes for discovery proteomics studies, to targeted quantitative proteomics, the instrument enables measurement of sample proteins with a high standard of sensitivity, accuracy, and precision that facilitates investigation of global protein kinetics, even with low-input and limited sample.

Identifying and characterizing proteins PTMs by bottom-up LC-MS relies on the acquisition of high-quality MS and MS² data. Combining the FAIMS Pro interface with an Orbitrap Exploris 480 mass spectrometer increases proteome profiling efficiency. The

FAIMS Pro interface is a next-generation, differential ion mobility device that uses gas-phase fractionation to achieve selective enhancement of peptides and improve the signal-to-noise ratios (S/N) by removing the singly charged ions, which is beneficial for overall sensitivity. Moreover, this improves sample coverage and sample throughput for single-cell proteomics experiments.

Choice of MS data acquisition mode increases IDs

Depending on the experimental goals, single-cell proteomics experiments can be carried out on the Orbitrap Exploris 480 mass spectrometer in either data-dependent acquisition (DDA), WWA, or high-resolution data-independent acquisition (HR-DIA) mode. DDA identifies as many peptides as possible in selected m/z ranges. The highest peaks in the full scan MS spectrum in each narrow m/z isolation window are selected for isolation and then fragmented to produce a MS/MS spectrum. Either the extracted and integrated peptide peak areas, or the intensity at the highest point of the chromatographic peak, are compared across sample sets (i.e., control versus experimental) for relative quantitation. The disadvantage of DDA is that not all peaks in the full scan spectrum trigger collection of a MS/MS spectrum, resulting in missing values (missing peptides and thus proteins). Using a DDA approach, it can be challenging to quantify relatively low-abundance peaks. Recent innovations in software algorithms can reduce missing values by improving the extraction and mapping of LC-MS peaks to identified spectra and with enhanced feature detection across data files using retention-time alignment and feature linking. On the other hand, DDA is ideal for Tandem Mass Tags™ (TMT™) experiments because the cycle time is usually lower in DDA vs. DIA, yielding more data points per peak and resulting in better quantitation accuracy.

The HR-DIA approach collects fragment ions for all eluting peptides. A precursor mass range is selected and then divided into a series of relatively wide isolation windows, for example, increments of 25 m/z each. The MS² spectrum is acquired from all detected precursor ions in the first isolation window and then the process is repeated for each consecutive adjacent isolation window until the entire precursor mass range is covered. Like DDA, the full scan MS intensities of peptides are collected and used for quantitation. The advantage of the HR-DIA method is that it increases proteome coverage and data completeness by acquiring MS/MS data from all detected precursor ions, reducing missing values for quantitation. Thus, HR-DIA methods can increase confidence that a protein of biological significance will be less likely to be missed in complex samples.

WWA combines the advantages of DDA and HR-DIA. WWA uses a broad isolation window ($\geq 4 m/z$) for data-dependent precursor selection. Like DIA, precursor ions within the isolated window are fragmented together, producing chimeric spectra that can be used by advanced software algorithms to increase peptide spectrum matches (PSMs) and protein IDs, even for low-abundance peptides that would be otherwise be missed.

Since single cell samples are not as complex as standard proteomics samples, the isolation window can be set relatively wide. The ideal isolation window can be determined experimentally. For example, Figure 5 shows the PSMs and proteins identified in a diluted 250 pg HeLa bulk digest at various Orbitrap Exploris 480 MS isolation width settings. At an isolation width of 8 m/z , protein identification essentially reaches saturation at 600 proteins identified. However, the PSMs obtained continue to increase as the isolation window width is increased. In this example, the number of PSMs increased to more than 80%, which is impressive for a low-input sample.

CHIMERYs intelligent search algorithm deciphers chimeric spectra, doubling proteomic depth

The effects of WWA on increasing IDs and thus proteomic depth are enhanced by using the CHIMERYs search algorithm, which allows confident identification of up to eleven peptides from a single chimeric spectrum. The algorithm makes use of accurate fragment-spectrum predictions based on millions of spectra and additional spectral properties, which are neglected by traditional search engines, such as relative signal intensities, to significantly improve the identification rates, sensitivity, and accuracy. This approach leads to deeper data mining and substantially increases the number of PSMs in DDA data. PSM-level false discovery rate (FDR) control can be performed using the Percolator algorithm.

Figures 6A and B demonstrate the difference in peptides and protein IDs provided by CHIMERYs in a 30-minute gradient compared to MS Amanda in an equivalent 30-minute gradient and SEQUEST™ in a 10-minute gradient. MS Amanda can perform searches using chimeric spectra but does not make artificial intelligence-based spectral predictions. In this example, CHIMERYs's artificial intelligence routinely doubled the number of peptides, and protein IDs compared to MS Amanda and SEQUEST regardless of the gradient length.

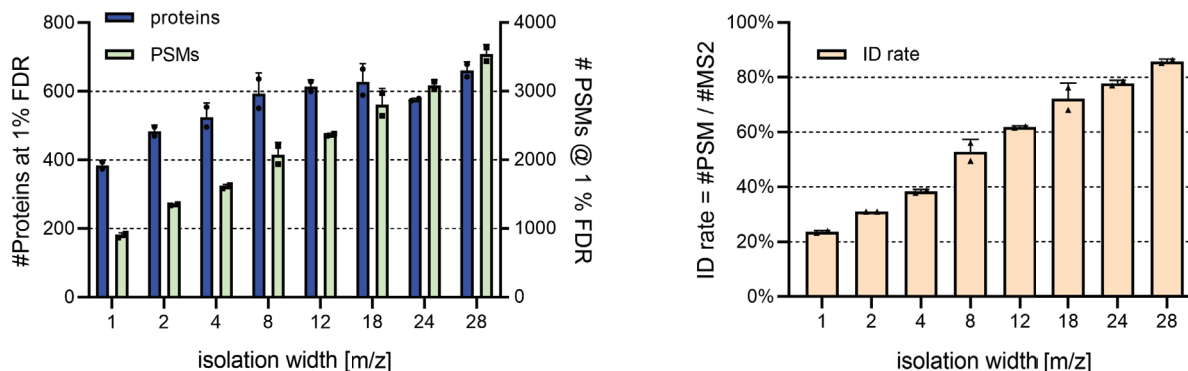


Figure 5. Experimental determination of the optimal isolation window setting for WWA on the Orbitrap Exploris 480 mass spectrometer. 250 pg HeLa diluted bulk digest was analyzed.

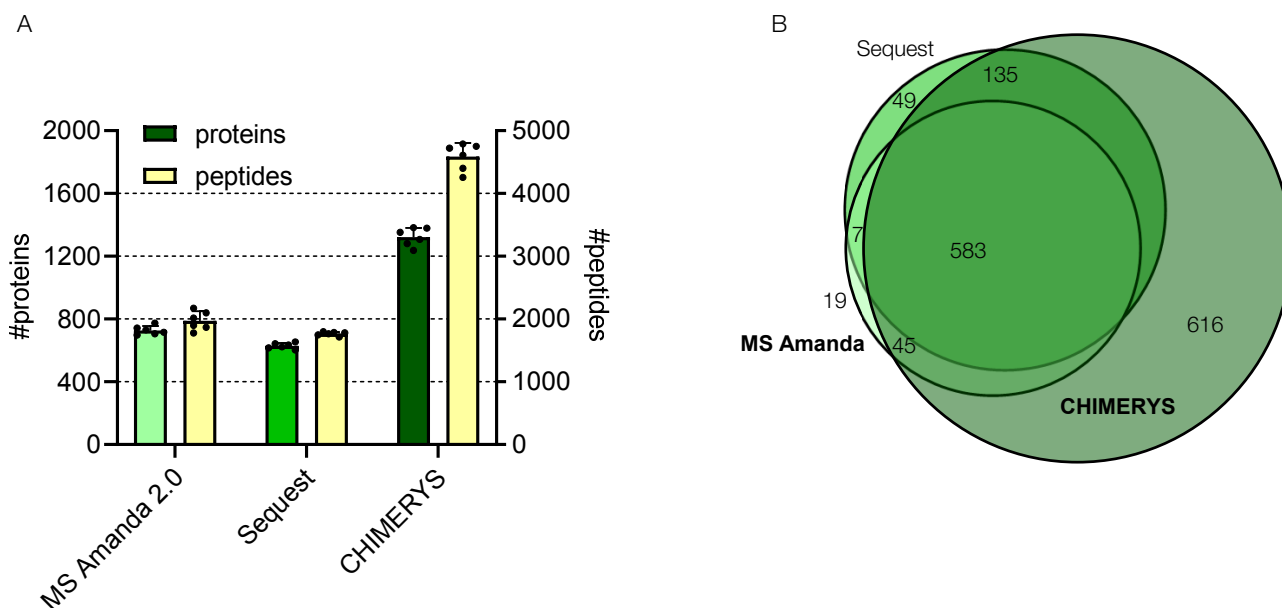


Figure 6. (A) Single cell stocks, LF-SCP workflow as published before⁷; WWA(ISO12) acquisition as published before⁵; data analysis in Proteome Discoverer 3.0 software using MS Amanda,⁸ SEQUEST,⁹ or CHIMERYs, 1%FDR, no MBR; (B) Venn overlap of identified proteins of representative replicate from (A)

CHIMERYs is a node embedded in Proteome Discoverer 3.0 software, a platform that provides raw data preprocessing and a large variety of data and statistical analysis capabilities. Its node-based workflow editor enables data analysis customization and automation to achieve maximum coverage and confidence. Results from multiple search engines such as SEQUEST™ HT, Mascot, MS Amanda, Byonic™ from Protein Metrics, and Thermo Scientific™ ProSightPD™ software can be integrated into data processing. Use of third-party nodes, and customization using the scripting node, maintains full integration in the Proteome Discoverer software framework to leverage all its capabilities

including visualization and statistical tools. The Proteome Discoverer software data processing workflow used in the experiments presented here is illustrated in Figure 7.

Optimized one-pot single-cell proteomics workflow performance

Dynamic range

The optimized one-pot single-cell workflow enables PSM identification and subsequent protein identification and quantification from single cells over an impressive dynamic range of four to five orders of magnitude (Figure 8).

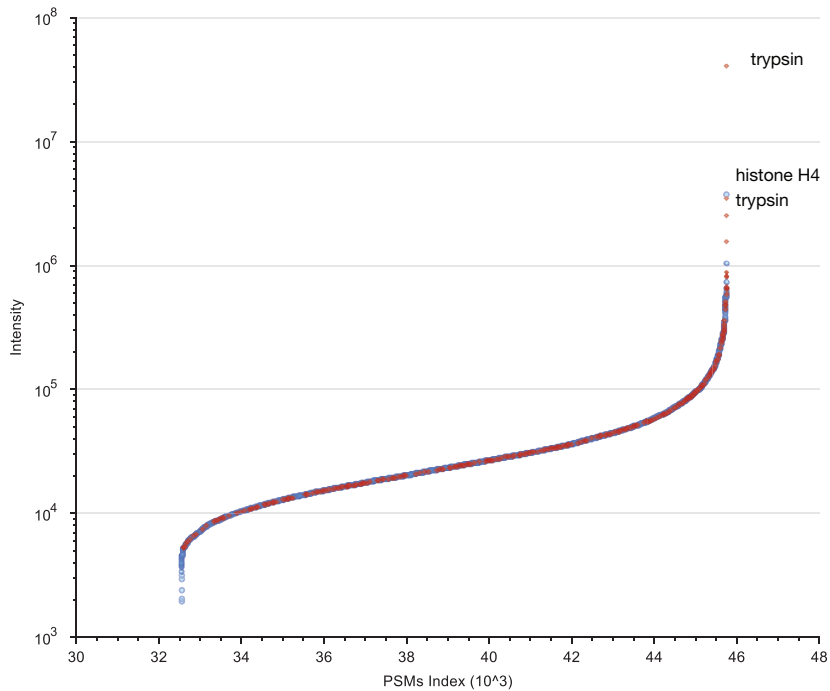


Figure 8. Dynamic range of precursor intensities of identified PSMs: Single cell stocks, LF-SCP workflow as published previously⁷; WWA (iso12) acquisition as published previously⁵, Data analysis in Proteome Discoverer 3.0 software using CHIMERYS, 1%FDR, no MBR. Contaminants shown in red, human peptides in blue.

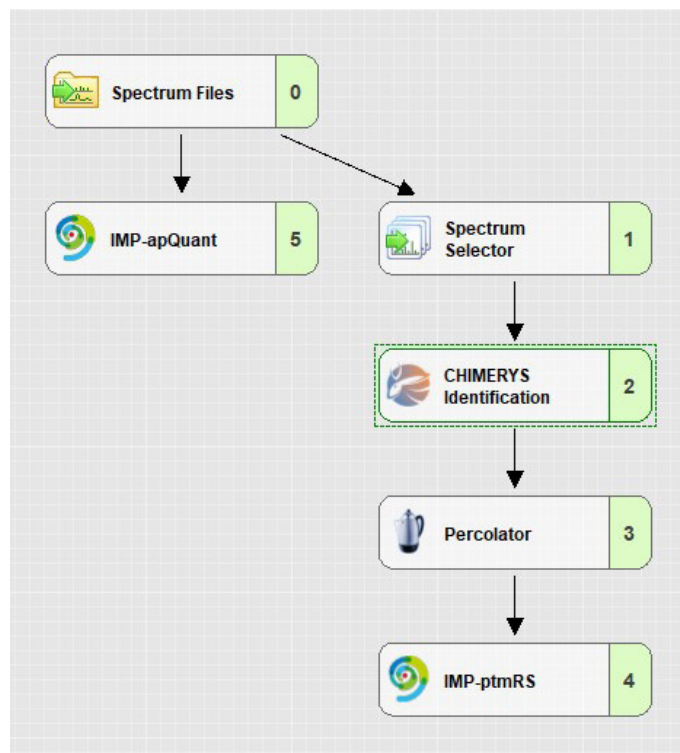


Figure 7. Proteome Discoverer software data processing workflow using the CHIMERYS intelligent search algorithm node. Percolator improves the discrimination between correct and incorrect spectrum identifications. Accurate and precise LFQ was performed using IMP-apQuant. IMP-ptmRS enables automated and confident localization of modification sites in validated peptide sequences and can be applied to commonly used fragmentation techniques (CID, ETD and HCD). Both IMP-apQuant¹⁰ and MP-ptmRS¹¹ were developed by the Protein Chemistry Facility IMP/IMBA/GMI, (Dr. Bohrgasse 3, 1030 Vienna, Austria) and are available without charge at <https://pd-nodes.org>.

Conclusion

Scientists are interested in single-cell proteomics measurements because analyzing bulk cells can result in difficult-to-interpret results and hide important biological differences. To facilitate this workflow with comprehensive proteome coverage and high throughput, single-cell proteomics studies require the use of next-generation technologies and optimized workflows.

In combination with cellenONE and proteoCHIP-based cell isolation and sample preparation, Thermo Fisher Scientific offers an advanced workflow platform composed of the Vanquish Neo UHPLC system with the μ PAC Neo HPLC columns, FAIMS Pro interface with the Orbitrap Exploris 480 mass spectrometer, and Proteome Discoverer software with the CHIMERYS intelligent search algorithm. In concert with optimized workflow methodologies, the platform offers demonstrated sensitive, reproducible, and efficient single-cell measurements that maximize the combination of protein and peptide identifications and sample throughput, enabling differentiation of cell lines in low input samples.

This leading-edge performance is due to several factors. Notably, the cellenONE robot provides reproducible loss-minimized sample preparation and improved recovery in ultra-low volumes with automated sample handling that includes water supplementation and addition of 5% DMSO to sample wells. Sensitivity is further enhanced by the separation power of the μ PAC Neo HPLC columns and the Orbitrap Exploris 480 mass spectrometer with the FAIMS Pro interface, allowing for deep label-free analyses of single cells. The CHIMERYS algorithm increases identifications with intelligent analysis of the substantial chimeric spectra produced in WWA experiments. Outstanding workflow sensitivity is achieved with a dynamic range of four to five orders of magnitude for identification and LFQ of more than 1,000 proteins in a single cell without using a carrier or matching between runs. If desired, sample throughput increased by using shorter chromatographic gradients and by omitting MS fragmentation.

Without a doubt, single-cell proteomic technologies have evolved to a stage where first applications are in sight and now including a variety of biological and biomedical questions that cannot be addressed by conventional assays instead. The technology is promising to profile individual cells in tumor tissues to study how cancer cells initiate, progress, and metastasize. Another exciting application is the investigation of brain tissues at different anatomical regions to learn about brain activity or the analysis of brain organoids to understand brain development on a molecular level. Furthermore, single-cell proteomics will be of the highest value in finding new biomarkers for early diagnosis and treatment monitoring in a wide variety of diseases, which has the potential to fundamentally change the practice of medicine.¹²

References

- Liu, Y.; Beyer, A.; Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 2016, 165, 535–550.
- Macromolecular Components of E. coli and HeLa Cells - AT. <https://www.thermofisher.com/uk/en/home/references/ambion-tech-support/rna-tools-and-calculators/macromolecular-components-of-e.html>.
- Volpe, P.; Eremenko-Volpe, T. Quantitative Studies on Cell Proteins in Suspension Cultures. *Eur. J. Biochem.* 1970, 12, 195–200.
- Kulak, N. A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* 2014, 11, 319–324.
- Mayer, R. L. et al. Wide Window Acquisition and AI-based data analysis to reach deep proteome coverage for a wide sample range, including single cell proteomic inputs. 2022.09.01.506203 Preprint at <https://doi.org/10.1101/2022.09.01.506203> (2022).
- Truong, T. et al. Data-Dependent Acquisition with Precursor Coisolation Improves Proteome Coverage and Measurement Throughput for Label-Free Single-Cell Proteomics. 2022.10.18.512791 Preprint at <https://doi.org/10.1101/2022.10.18.512791> (2022).
- Matzinger, M.; Müller, E.; Dürnberger, G.; Pichler, P.; Mechtler, K. Robust and easy-to-use one pot workflow for label-free single-cell proteomics. *Anal. Chem.* (2023) <https://doi.org/10.1021/acs.analchem.2c05022>.
- Dorfer, V. et al. MS Amanda, a Universal Identification Algorithm Optimized for High Accuracy Tandem Mass Spectra. *J. Proteome Res.* 2014, 13, 3679–3684.
- Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 1994, 5, 976–989.
- Doblmann, J. et al. apQuant: Accurate Label-Free Quantification by Quality Filtering. *J. Proteome Res.* 2019, 18, 535–541.
- Taus, T. et al. Universal and Confident Phosphorylation Site Localization Using phosphoRS. *J. Proteome Res.* 2011, 10, 5354–5362.
- Pham, T.; Tyagi, A.; Wang, Y.-S.; Guo, J. Single-cell proteomic analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2021, 13, e1503 (2021).

Learn more at thermofisher.com/singlecellproteomics