# Exploring the potential of the last generation UHR-Q-TOF for rapid generation of accurate information on proteoforms distribution and relative abundancy.

**Nicolai Bache[1],** Wolfgang Jabs[2], Stuart Pengelley[2], Christian Albers[2], Klaus Meyer[2] Pierre-Olivier Schmit[3]

[1]Bruker Daltonics, Thriges Plads 6, DK- 5000 Odense C, Denmark
[2]Bruker Daltonik, Farenheitstrasse 4, D-28359 Bremen, Germany
[3]Bruker Daltonique, 34, rue de l'industrie, 67166 Wissembourg, France

## Introduction

All along their life cycle, proteins undergo various transformations that can alter their functions while keeping a good part of their primary sequence intact. These multiplication of PTM patterns, alternative splicing forms or products of proteolitic processing cannot be simply resolved with a bottom-up approach, as very few peptides are specific from the given proteoform. However, the information relative to the distribution of different proteoforms is encoded in their intact masse. Being able to catch this information efficiently will drive the biologist into a new analytical dimension, far beyond the traditional gel-based approaches.

## Objectives

In this study we were willing to evaluate how the latest generation UHRQ-Tof capabilities of delivering resolution, mass accuracy and isotopic fidelity even in highly complex mixtures separated with Fast LC could help to quickly resolve proteoforms in complex mixtures, in order to map their distribution and relative abundancies.

## Material and Methods

Undigested protein mixtures of E.Coli (Bruker Daltonics), Yeast (Promega) , and non-depleted plasma (Sigma), have been separated on a Phenomenex Aeris Widepore, 3.6 µ, C4, 150x2.1 mm column, using 20 min or 45 minutes chromatographic methods. Separation were performed on a Dionex RSLC system coupled to an impact II benchtop UHR-Q-Tof (Bruker)

or a floor-standing maXis II UHR-Q-Tof (Bruker) , operating in MS, Auto MS/MS or targeted MS/MS acquisitions.
Mixtures of yeast intact protein (Promega) spiked with various ratios of UPS2 Proteomics Dynamic Range Standard Set (Sigma) have been separated on a 250 mmX100 µm monolithic pepswift PS-DVB column after preconcentration on a monolithic pepswift 200µm x 5mm trap (ThermoFischer Scientific). Separation has been performed on an Ultimate nano-RSLC system (Thermo Fischer Scientific) coupled to an impact II benchtop UHR-Q-Tof (Bruker Daltonics) via a CaptiveSpray nanoBooster ion source (Bruker Daltonics), and operated in MS and auto MS/MS modes.
All data have been automatically processed (calibration, protein signal extraction with Dissect™, deconvolution and obtention of monoisotopic masses with SNAP™, export of deconvoluted monoisotopic masses with corresponding retention time and intensities) in Data Analysis 4.2 (Bruker Daltonics). Statistical analyses have been performed on a slightly modified version of the Profile Analysis 2.1 Software. Identifications have been performed using the Top-Down Sequencing search functionality of BioTools 3.2 (Bruker Daltonics) and Mascot 2.4(Matrix Science).

Results

Using High-Quality threshold protein detection, we could easily detect more than 800 proteoforms (from doubly charges peptides up to 35Kda proteins) out of the E.Coli mixture, and more than 1500 proteoforms out of 50µg of the Yeast mixture, using a 15 minute gradient (20 minutes method). The separation of yeast over a 35 minutes gradient (45 minutes method) enabled to distinguish more than 1000 proteoforms from a 1µl injection of non-depleted plasma.
The spectral quality observed for single compounds was preserved while measuring these highly complex mixtures : the average mass error for the monoisotopic peak of the RS 19 protein (10,2 Kda) over 4 injections was 0,07 ± 0,2 ppm. The isotopic fidelity (express as shift of the theoretical abundance, relative to the most abundant isotope) was always better that 2%, and the resolution exceeded 50 000. An Auto-LC MS/MS run enabled the identification of RS 19 using CID from several charge states. In the absence of on-the flight deconvolution, automated LC-MS/MS enabled to fragment the high abundance low MW (<20KDa) proteins. Lower intensity proteins could be fragmented and identified using a targeted MS/MS Strategy.

A proteoform distribution profile could be obtained for glycoproteins spiked in the sample, which was similar to the one observed for the isolated glycoprotein, therefore suggesting that the approach enables to give a relative quantitation information for the compounds present in the initial mixture. Using unsupervised statistics approaches, we have been able to separate the yeast extracts spiked with various amount of the UPS II mixture.

## Conclusions

The last generation UHRQ-TOF, by combining a large spectral dynamic range to the capacity of preserving a high spectral quality over a large mass range in complex mixtures, are now capable of delivering rapidly a high-quality proteoform distribution information that has the potential to complement the information delivered by the bottom-up approaches.
In this first approach we observed a discrimination of the higher masses in favor of the lower MW proteins. As this is probably the results of the absence of sample pre-fractionation combined to the use of very short gradient, we will continue to explore the system's potential using size-filtered samples and adapted chromatographic conditions.