

# Screening for microbial protein over-expression in complex matrix, using MALDI LTQ Orbitrap™

Michiel Akeroyd<sup>1</sup>; Rob van der Hoeven<sup>1</sup>; Thomas Moehring<sup>2</sup>; Wilfried Voorhorst<sup>2</sup>; Kerstin Strupat<sup>2</sup>;

1 DSM Biotechnology Centre, Delft, The Netherlands; 2 Thermo Fisher Scientific, Bremen, Germany



## Overview

**Novel Aspect:** Protein ID of complex samples by Accurate Peptide Mass Fingerprint (PMF) using overlapping full scans enabling higher signal-to-noise spectra, and reliable isotope patterns

**Purpose:** Screening of successful over-expression in micro organisms

**Methods:** MALDI produced ions are detected in an Orbitrap™ mass analyzer; mass spectra quality of data acquired from overlapping full scan windows, such as  $m/z$  500 – 2000,  $m/z$  1500 – 3000,  $m/z$  2500 – 4000 are compared to single full scan windows  $m/z$  500 – 4000

**Results:** Overlapping full scan window acquisition allows for the better S/N ratios and better isotope pattern reflection

## Introduction

High throughput facilities for screening of successful protein over-expression in micro organisms grown in Micro Titer Plates (MTP) have been developed at DSM Biotechnology Centre. The samples generally provide a low protein expression in a complex matrix. Therefore the facilities include an automated, generic sample pre-treatment protocol, see Methods. The facilities need to handle ~8000 samples per week. For this purpose the approach is automated with robotics: MALDI plates are automatically loaded into the x-y stage of a MALDI ion trap based system using a robot arm coupled to a plate hotel ("autosampler" for MALDI plates). All actions with respect to creating and starting sequence lists and sample tracking with barcode readers are automated. Automated filtering of positives is performed with Principle Component Analysis (PCA) analysis.

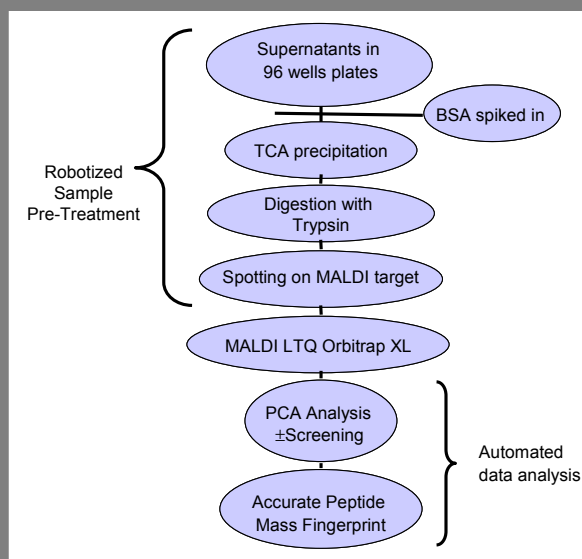
## Methods

**Samples:** Sample pre-treatment includes protein purification with TCA precipitation after spiking a known amount of Bovine Serum Albumin (BSA). BSA spiking improves precipitation of low abundant proteins. Subsequently, automated enzymatic protein digestion, dilution of the digests in MALDI matrix (alpha cyano-4-OH cinnamic acid, CHCA) solution and automated MALDI spotting are performed. The entire workflow is presented in Figure 1.

**Mass Spectrometry:** Identification of the proteins in the positive samples is performed by Accurate Peptide Mass Fingerprinting (PMF) using a MALDI source coupled to an Ion Trap – Orbitrap hybrid mass analyzer.

**Method Setup:** Typically three full scan overlapping windows, each of them 1500 mass units wide, are setup; each full scan asks for 1 e6 charges. Upon an AGC pre-scan the correct number of charges is provided for the analytical scan deriving from the identical location. The instrumentation used is described earlier (1). All spectra shown here are acquired in full profile mode for ease of reading signal to noise ratios.

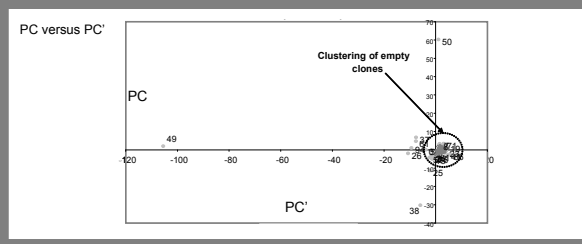
**FIGURE 1.** The complete MS-based workflow for screening of expression libraries. The robotized sample pre-treatment protocol and the automated data handling are shown. Positive clones are selected by PCA analysis and the over-expressed enzymes are identified by Peptide Mass Fingerprinting (PMF). The complete workflow is designed to handle ~8000 samples per week.



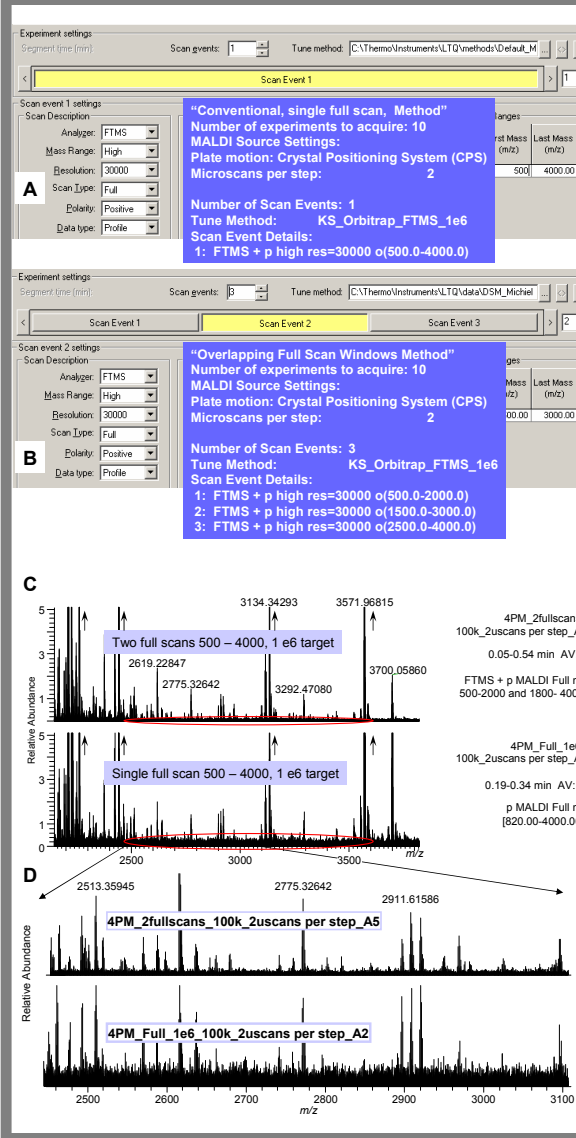
## Proof of Principle

A 96 wells microtiter plate (MTP) with the *Aspergillus Niger* empty host strain is provided. 12 Positive controls are spiked to this MTP, these positive controls are selected by nanoLC-MS/MS and range from low to very high over-expression. The robotized sample pre-treatment protocol is mimicked manually. The positive controls are filtered by PCA (Principle Component Analysis), an example is shown in Figure 2. The MS spectra are compared on 9 different principle components (PC's); the empty clones are clustered and the positive clones are dispersed from the cluster.

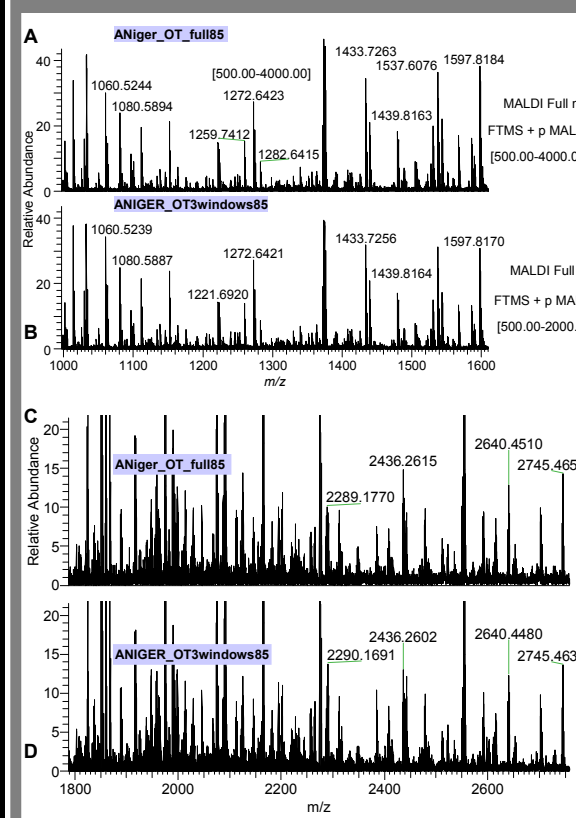
**FIGURE 2.** An example of the Principle Component Analysis PCA. The empty clones cluster and the positive clones disperse from this cluster. The PCA analysis of 9 PC's was automated.



**FIGURE 3.** Concept and Method Setup (A, B) with acquired raw files (C, D). Signal-to-Noise ratio of peaks acquired with (e.g.) two full scan windows (upper figure of C) is higher by a factor of 4 to 5 compared to a single full scan window asking for the same number of ions (1e6 charges), (lower figure of D).



**FIGURE 4.** Raw data file comparison for one of the 96 samples, sample 85 acquired, average of 10 scans each  
A) Inset from  $m/z$  1000 – 1600 of the exp.  $m/z$  500 – 4000 asking for 1e6 charges,  
B) Inset from  $m/z$  1000 – 1600 of the exp.  $m/z$  500 – 2000 (stitched together with two other windows from 1500 – 3000 and 2500 to 4000) asking for 1e6 charges.  
C) Inset from  $m/z$  1800 – 2800 of the exp.  $m/z$  500 – 4000 asking for 1e6 charges,  
D) Inset from  $m/z$  1800 – 2800 of the exp.  $m/z$  1500 – 3000 (stitched together with two other windows from 500 – 2000 and 2500 to 4000) asking for 1e6 charges.



- 1 FT full scan window (as in A, C): one cycle lasts for approx. 2 – 3.5 sec asking for 1e6 charges per full scan  $m/z$  500 – 4000 depending on the resolution chosen.
- 3 FT overlapping full scan windows (as in B, D): one cycle of 3 full scan windows lasts for 8 - 11 sec asking for 1 e6 charges per scanned / overlapping full scan window depending on the resolution chosen.
- The Signal-to-Noise Ratio for overlapping full scan windows (B, D) is already better with averaging 2 scans compared to averaging over 5 scans for the single full scan method (A, C).

**TABLE 1.** 12 Positive controls, analyzed by nanoLC-MS/MS and selected based on their relative protein abundance, were spiked to a 96 wells MT plate containing the empty host strain. The resulting MALDI MS raw files were "extracted and mono-isotopic MH<sup>+</sup> masses were used to perform PCA analysis and Accurate Peptide Mass Fingerprinting, using MASCOT™. BSA (precipitation) and Trypsin (enzymatic enzyme) proteolytic fragment masses were removed from the peaklists prior to PMF searches. Nine out of twelve positive controls were selected by PCA; these were mostly positive controls showing high relative protein abundances in nanoLC-MS/MS. All samples with relative protein abundances of >10% of the over-expressed protein, were selected as positive, using PCA. Seven of the nine PCA selected positives were identified using MALDI MS and Accurate PMF. This is possible in a significant less amount of time than using nanoLC-MS/MS.

% of total protein abundance in nLC-MS/MS	> 50%	25 - 50%	10-25%	2-10%	< 5%	Total
<b>Positive in PCA</b> (75% total hit rate, 100% hit rate for positives, > 10% expression)	3	2	2	2	3	12
<b>Identified with Accurate PMF</b> using MALDI LTQ Orbitrap XL (78% total hit rate, 86% hit rate for positives, > 10% expression)	3	2	1	1	1	7

## Discussion

In the biotechnological industry large expression libraries of microbial clones are generated to obtain biodiversity needed for the discovery of new interesting enzymes. These libraries should be tested based on the enzymatic activity, this is costly for such large numbers of samples. The number of samples for activity-based screening are significantly reduced by using a mass spectrometry based screening to select clones with successful over-expression from large expression libraries. Here, we present a generic automated workflow to select the clones showing successful extra-cellular over-expression. The results obtained in this Proof Of Principle experiment clearly show that MALDI LTQ Orbitrap instrumentation can be applied as a first filter to screen for successful over-expression. All positive controls with relative protein abundance of >10% were selected as positives with PCA. 86% (7 out of 9) of these selected positives were identified with Accurate Peptide Mass Fingerprint (Accurate PMF). Using overlapping MS scan events rather than one full MS scan increased the S/N ratios of the MS data, and by this, the reliability of peptide isotope patterns.

## Conclusions

- A generic workflow for high-throughput screening of successful protein over-expression in micro organisms based on mass spectrometry is developed
- MALDI LTQ Orbitrap XL is used to generate high resolution, high accurate MS spectra for identification by Accurate Peptide Mass Fingerprint (Accurate PMF)
- Signal-to-Noise ratios of the acquired MALDI MS spectra can be optimized by using overlapping MS scan events, rather than one full MS scan event.

## References

- (1) Strupat K., Kovtoun, V. Bui, H.; Viner, R.; Stafford, G.; Horning, S.; *J. Am. Soc. Mass Spec.* (2009) **20**, 1451 - 1463

## Acknowledgements

MASCOT is a trademark of Matrix Science Inc. All other trademarks are the property of Thermo Fisher Scientific Inc. and its subsidiaries. This information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.