STRATEGIES FOR STRINGENT CATALOGING OF *METHYLOPHAGA THIOOXIDANS* AND *METHYLOCELLA SILVESTRIS* USING AN ALTERNATIVE SCANNING LC-MS APPROACH

Joanne B. Connolly [1]; Nick Tomczyk [1]; Susan E. Slade [2]; Vibhuti Patel [2]; Rich Boden[2]; Konstantinos Thalassinos [2]; Hendrik Shaefer[2]; James Scrivens [2]
[1]Waters, Manchester, UK; [2]University of Warwick, Coventry, UK

INTRODUCTION

Identity^E, an alternative scanning LC-MS^E approach was used to conclusively and stringently identify *Methylophaga thiooxidans* and *Methylocella Silvestris* proteins from tryptic digests of whole cell lysates. The system used comprised of a NanoACQUITY™ coupled to a SYNAPT™ mass spectrometer operating in LC-MS^E acquisition mode. Data was processed using a novel database search algorithm 'Ion Accounting', which uses the physiochemical characteristics of fragmented tryptic peptides to aid stringent protein identification (Figure 1).

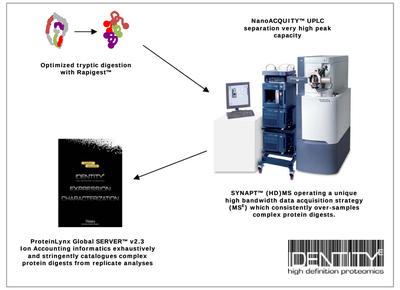

Figure 1. Identity^E system solution workflow.

The system solution was developed in response to concern from the proteomics community as to the reproducibility and security of protein identification in published data. The following quotes highlight a few instances of issues with other current methodologies;

"…the risk of a false-positive protein assignment is greater when only a single peptide is used to identify a protein…" [1]

"…Too many papers that I receive and/or read in scientific journals have no evidence of any repeatability, much less reproducibility…Thus, I suspect that >50% of papers now being published in the proteomics field have data that was performed but once (n = 1), which makes the entire work unacceptable for publication…"[2]

"… the degree of stringency required in proteomic data generation and analysis appears to have been underestimated…As a result, there are likely to be numerous published findings that are of questionable quality…" [3]

Triplicate analyses for all Identity^E experiments, and stringency of data processing and database searching addresses all these perceived problems and allows the publication of comprehensive, confident, and repeatable protein identifications.

METHODS

Sample preparation

100 µg of whole cell lysates of *Methylophaga Thiooxidans* and *Methylocella Silvestris* were washed twice with 500 µL of 0.1% Rapigest using a MW 5000 cut-off spin column. The complex protein solution was then concentrated to 100 µL using a MW 5000 cut off spin column. The solution was the heated at 80°C for 15 minutes. Reduction and alkylation was performed with 5 µL 100 mM DTT and 5 µL 200 mM IAA, respectively. The proteins were digested with 1:50 (w/w) sequencing grade trypsin for 16 hours. RapiGest was removed by the addition of 2 µL 15 M HCl, followed by centrifugation. The supernatant was collected and diluted to 1 µg/µL with 0.1% formic acid. The dilution of the samples 1:1 with a glycogen phosphorylase B tryptic digest of a known concentration gave sample concentrations of 0.5µg/µL. The inclusion of the internal standard protein allowed the PLGS v2.3 software not only to identify the components of the complex mixture but also to calculate the absolute amounts of identified proteins[4].

LC-MS Conditions

NanoACQUITY SYNAPT MS experiments were performed using a 1.5 hr reversed phase gradient (5 to 40% acetonitrile over 90 minutes) utilizing a 75 µm x 20 cm C18 column operated at 300 nL/min. Samples were run in triplicate.

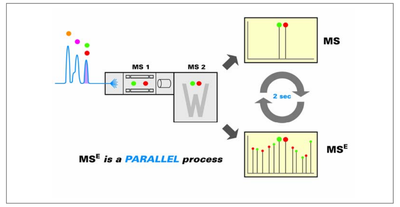

Figure 2. Schematic of MS^E mode of analysis.

The data-independent analysis mode (MS^E) , shown in figure 2, was employed on the SYNAPT MS mass spectrometer enabling precursor and fragment ions from the tryptic digest to be analyzed simultaneously. The mass spectrometer was programmed to step between low energy (4 eV) and elevated (15-35 eV) collision energies on the gas cell, using a scan time of 1 s per function over 50-1990 *m/z*. MS^E data independent analysis provides accurate mass measurements of all detectable precursor and product ions. Chromatographic alignment of precursor and product ion data reduces miss-assignment of product ions to parent ions of similar mass or retention time. Protein identifications and absolute quantification information were extracted using the dedicated algorithms employed in PLGS v2.3 by searching species specific databases.

RESULTS

Figure 3 shows overlayed triplicate low energy MS^E chromatograms of Methylophaga Thiooxidans. Each analysis consumed 500ng tryptic digest according to supplied sample concentration information. This figure highlights the reproducibility of the chromatography. Similar chromatograms were obtained for Methylocella Silvestris (not shown).

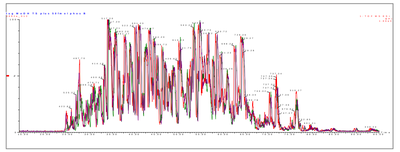

Figure 3. Overlayed triplicate MS^E chromatograms of Methylophaga Thiooxidans

The quality of the MS^E data obtained can be assessed by clustering the data acquired across replicate injections according to mass and retention time. Figure 4 shows the mass accuracy (ppm) of common data points found across 6 replicate injections of Methylophaga thiooxidans. The mean mass precision is shown to be 2.13 ppm. The mean retention time reproducibility of replicating EMRTs is ~ 0.2% CV. Similar data was obtained for Methylocella Silvestris.

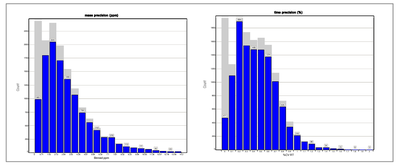

Figure 4. Mass accuracy (ppm) and retention time reproducibility of EMRT data points over 6 replicate analyses of Methylophaga Thiooxidans

Quantitative measurement of low energy precursor ions is facilitated by the data independent analysis and an increase in dynamic range is observed as the limitations posed by conventional Data Directed Analysis MS/MS duty cycle are negated. The addition of a protein digest of known concentration allowed the calculation of the absolute amounts of the identified proteins.[4] In the example of *Methylophaga thiooxidans* tryptic digest, over 3 orders of magnitude of protein concentration can be detected, ranging from 0.01 ng to 74 ng on column (Figure 5). The total column loading can also be calculated using this methodology – in this case a total column loading of 367 µg allowed the confident identification of 309 proteins under stringent conditions. A false positive rate of 2.27% was calculated for this experiment.

In the analysis of the tryptic digest of cell lysate from *Methylocella Silvestris*, similar results were obtained - 3 orders of magnitude of protein concentration was detected from 0.01 ng to 74 ng.

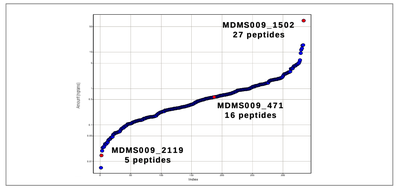

Figure 5. Methylophaga Thiooxidans Dynamic Range of Protein Identification

The total column loading in the case of *Methylocella Silvestris* was found to be 367 µg, this allowed the confident identification of 309 proteins under stringent conditions. A false positive rate of 2.27% was calculated for this experiment.

The Identity^E analysis of these two bacterial strains was compared to previous analyses using traditional methods. *Methylocella Silvestris* had been analyzed using a GeLC approach whilst the *Methylophaga Thiooxidans* was analysed using a shotgun SCX-RP approach. The instrumental set-up was very similar in all cases. For traditional analyses the mass spectrometric mode of analysis used was data directed (DDA) whereas in Identity^E the analyses, MS^E is data independent. As can be seen in Table 1, the Identity^E analyses were the only experiments performed on the whole cell lysate without prior fractionation. This gave advantages in total analysis time required—even considering that the analysis was performed in triplicate. The total sample consumption was also lowest for the Identity^E experiments.

	Identity^E *Methylophaga Thiooxidans and Methylocella Silvestris*	SCX-RP / DDA Approach *Methylophaga Thiooxidans*	1D gel approach *Methylocella Silvestris*
Columns	RP analytical column nanoACQUITY™ BEH C18 1.7 µm, 75 µm x 200 mm trap column nanoACQUITY™ Symmetry C18 5µm, 180 µm x 20 mm	SCX PolyLC PolySULFOETHYL A 200x2.1 mm 5 µm 300A RP analytical column nanoACQUITY™ BEH C18 1.7 µm, 75 µm x 100 mm trap column nanoACQUITY™ Symmetry C18 5µm, 180 µm x 20 mm	RP trap MCA-30-05-C18 analytical column Acclaim PepMap 100 C18 5 µm 100A 15 cm
LC system	NanoACQUITY	NanoACQUITY	CapLC
MS system	SYNAPT™ MS	SYNAPT™ HDMS	Q-ToF Ultima Global
Mobile phases	H2O and ACN with 0.1%FA	H2O and ACN with 0.1%FA	H2O and ACN with 0.1%FA
Replicates	3	1	1
Fractionation	None - whole cell soluble protein lysate in single shot	43 SCX	39 gel slices
RP gradient	90 mins	43 x 20 min/fraction	40 min/fraction
Analysis Mode	MSE	DDA no lockmass correction	DDA no lockmass correction
Total time	6 hours	30.5 hours	26 hours
Sample consumption	1.5 µg	280 µg (iTRAQ labeling)	14 µg

Table 1 System Set-up Comparison

Table 2 shows the similarities and differences of the data processing and database searching between the traditional and the Identity^E experiments. The advantages in using PLGS v2.3 software are highlighted, in that the user can dictate the minimum number of fragment ions a peptide must contain to be a correct identification. It also allows the selection of the minimum number of peptides a protein identification must have to be reported. In most cases for stringency this would be set to two — but for ease of comparison with the provided traditional data sets this was set to one.

	Identity^E	Conventional Approaches
Processing and database search software	PLGS v2.3 with Ion Accounting	PLGS v2.2.5 and Mascot 2.2
Database	Species specific MT or MS	Species specific MT or MS
Minimum fragment ions per protein	5	-
Minimum peptides per protein	1 (usually 2 - specific requirement)	1
Minimum number of times an ID must replicate	2 out of 3 analyses	-
False positive rate	Calculated as MT 2.14 and MS 0.28%	2% by Mascot
Peptide and fragment ion tolerances	10 and 20ppm	50 ppm (no lock mass correction)

Table 2. Processing and Database Search Comparison

Table 3 highlights the numbers of proteins identified and the ability to determine the absolute quantity of proteins present using Identity^E and compares this to traditional methods. Notably the number of proteins is reported for each individual replicate analysed by Identity^E and each provides approximately the same number of protein identifications. For stringency of identification, however, we combine the results of all those replicate analyses and only count those proteins which replicate in 2 or more of the replicate injections. After completing this filtering of the data it is shown that Identity^E has many more protein identifications than the traditional methods. For Methylophaga thiooxidans we identify 374 proteins compared to 67, whereas for the Methylocella Silvestris we can identify 355 proteins compared to 183.

	IdentityE	SCX-RP/ DDA Approach *Methylophaga Thiooxidans*	1D gel Approach *Methylocella Silvestris*
Number of identified proteins	*MT* 451, 483 and 432 *MS* 468, 467 and 474	67	183
Number of proteins replicating 2/3 times	*MT* 374 *MS* 355	None	None
One Hit Wonders	*MT* 6 *MS* 4 (all replicate 2/3 times)	38	92
Average number of identified peptides per protein	*MT* 8 *MS* 12	2	2.4
Absolute quant possible	Yes	Not Possible	Not Possible
Calculated average column loadings	MT 398ng - MS 584ng	Not Possible	Not Possible

Table 3. Protein Identifications and Absolute Quantification Results Comparison

Approximately half of the identifications made using conventional analysis methods were made with only 1 peptide identified per protein. In contrast to this, on average 9 peptides per protein were identified using the data independent methodology combined with Ion Accounting. Triplicate injections were undertaken to allow filtering of protein identification on replication and to give statistical information on the absolute quantification calculations. The result of this type of analysis in practical terms is the detection of higher numbers of proteins with additional peptides per protein identification, giving increased confidence in the protein assignment. The improvement in protein identification becomes more apparent with increasing complexity of analytes.

CONCLUSION

Alternative scanning LC-MS approach (Identity^E) allows stringent cataloging and simultaneous absolute quantification of *Methylophaga Thiooxidans* and *Methylocella Silvestris* proteomes using novel 'Ion Accounting' software.

The Identity^E High Definition system solution provides:

— A higher number of correct protein identifications compared to traditional methodologies

— A very stringent protein identification strategy using replicate analyses

— A highly reproducible protein identification strategy

—Protein identification and absolute quantification of identified components in a single step

—Highest band-width analysis (UPLC/MS^E)

—Comprehensive data sets

—Expression^E additional software can also provide label-free relative quantification using the same data sets

References

1. Steven Carr *et al.* (2004) The Need for Guidelines in Publication of Peptide and Protein Identification Data. *Mol. Cell. Proteomics, 3.6,* 531-533
2. Ira Krull and William Hancock. (May 2004) Editorial Journal of Proteome research
3. M.R. Wilkins et al. (2006) Guidelines for the Next 10 years of Proteomics. Guidelines for the Next 10 years of Proteomics. Proteomics, 6, 4-8
4. Silva et al, Mol. Cell Proteomics 5 (2006) 144

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS

720002649EN