MINING URINARY PROTEOME USING AN ALTERNATIVE MS DATA ACQUISITION APPROACH AND A NOVEL DATABASE SEARCH ALGORITHM

P. Olivova[1], M. Gilar[1], S.J. Geromanos[1], G. Li[1], J. C. Gebler[1], R. Roy[2], and M.A. Moses[2]
[1] Waters Corp., Milford, MA, USA.
[2] Children's Hospital, Harvard Medical School, Boston, MA, USA.

OVERVIEW

- The human urine proteome was investigated by liquid chromatography (LC) with mass spectrometry (MS) detection.
- Female **breast cancer** patient urine samples have been examined and compared in order **to define differences in urinary proteome for potential biomarker discovery**.
- **A mixed-mode cation-exchange reversed-phase sorbent (MCX)** was used to **cleanup** the samples prior to LC-MS.
- **A novel mass spectrometry method employing alternate scanning (MS^E)** was used for peptide/protein identification.
- Three different database search algorithms were employed for peptide identification, including MASCOT, ProteinLynx Global Server (PLGS), and **novel Ion Accounting (IA) software**.

INTRODUCTION

Human urine contains a mixture of proteins produced from the filtration of blood within the glomerules, proteins secreted from the kidneys, and those derived from the urogenital tract. Thus far, in clinical laboratories, the identification and the measurement of known disease-associated proteins or hormonal markers have relied highly on immunological analyses, such as RIA or ELISA, in a one marker at a time manner.

In this work we have investigated the human urine proteome by LC-MS^E analysis. The urine samples from breast cancer female patient have been examined and compared to control samples. The mixed-mode cation-exchange reversed-phase sorbent (MCX) was applied for sample cleanup and was found to be critical for successful LC-MS^E analysis.

In order to eliminate the run to run peptide/protein identification variability usually observed for tandem mass spectrometry (MS/MS), an alternative low and elevated energy scanning approach was used for parallel monitoring of precursors and resulting fragment ions (Figure 1). The precursors and their corresponding high-energy ions were grouped via IA software, using accurate mass and retention time parameters.

Three different database search algorithms were employed for peptide identification, including MASCOT, ProteinLynx Global Server (PLGS) and the Ion Accounting (IA) software. IA is a novel database search approach that matches the theoretical tryptic peptide b/y ions with acquired multiple product ion spectra. The rules for confident identification include the parameters such as the number of matched ions, corresponding precursor/fragment ion intensities, and robustness of the matches in the replicate experiments. Only peptides detected in at least in two out of three LC-MS replicates were considered as confident hits, and were counted towards protein identifications. The novel MS scanning approach, in conjunction with IA database search algorithm, improves the repeatability of LC-MS^E experiments, and provides better protein coverage. A number of confidently identified peptides/proteins in normal and cancer patient urine samples is shown in Table 1.

MCX cleanup of urine samples

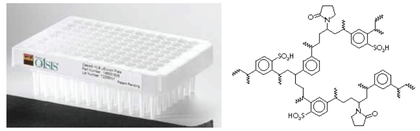
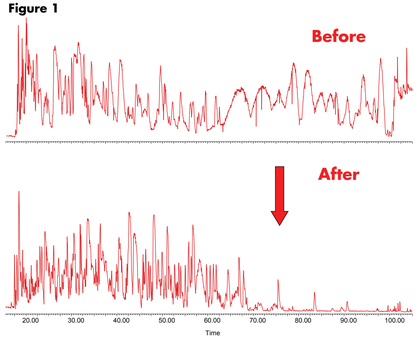


Figure 1



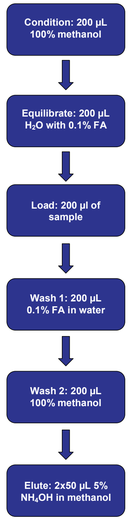


Figure 1: LC/MS^E analysis of urine sample.
Gradient 0-40 %B in 96 minutes. Flow rate 0.3 µl/min. Injection ~0.4 µg of human urine.
Column temperature 40 °C.
Sample cleanup protocol is listed in the flow chart.
Sample loading condition for MCX: 30% ACN (v/v) in aqueous 0.1% FA.
Sample was reconstituted in 0.1% FA with 1% ACN after evaporation to dryness .

METHODS

Sample Preparation

Urine samples were obtained from a collaborator (Professor M.A. Moses, Children's Hospital in Boston). These samples were desalted using ultrafiltration with 5 kDa MWCO Vivaspin (devices). Precipitates formed during the process were removed by a centrifugation at 10 000 rpm for 10 min. The protein concentration was measured using a Bradford assay (Pierce). Samples were reduced, alkylated and digested with modified porcine Trypsin (Promega) according to a standard protocol. The mixed-mode cation-exchange reversed-phase sorbent (MCX) was applied to sample cleanup prior to LC-MS^E analysis.

LC-MS Conditions

Tryptic peptides were analyzed by LC-MS^E on Q-Tof Premier (Waters). Data analysis was performed using PLGSV2.2 software, MASCOT, and IA software.

RP column: BEH dC18, 1.7µm 75x100mm (Waters)
RP flow: 0.3 µl/min
Solvent A: 0.1% Formic Acid in water
Solvent B: 0.1% Formic Acid in acetonitrile
Temperature: 40°C
Gradient: 1-97 min 0-40% B; 0.4% ACN/min.

Protein Expression Analysis

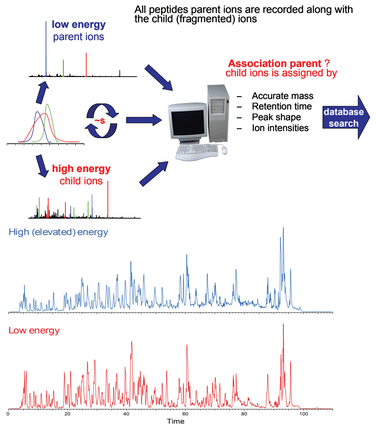


Figure 3. LC-MS^E low and elevated energy chromatograms.
The bottom picture represents low energy MS chromatogram (precursor ions), while the data in upper one were acquired at elevated collision cell energy, producing mostly peptide fragment ions. Both chromatograms have similar profile.

Figure 2. MS^E data acquisition.
In contrast to DDA, the alternate scanning methodology does not rely on peptide precursor pre-selection prior to the fragmentation. The alternate scanning mode data acquisition is configured to acquire in parallel for both MS and MS/MS data (the data collection mode is termed MS^E). Both the low energy (precursors) and elevated energy scans (fragments) are collected throughout the entire LC-MS experiment as outlined in this figure.
The MS^E data acquisition method provides an accurate mass LC-MS data for both precursors and associated fragment ions (<10 ppm). All charge states for a given peptide are fragmented in the collision cell, thereby affording an increased sensitivity for those peptides which exist in multiple charge states. The scanning frequency for both low and elevated energy is configured such that the chromatographic peaks of detectable precursors and associated fragment ions are well defined (collecting multiple data points across the peak).

The precursors and high-energy ions assignment

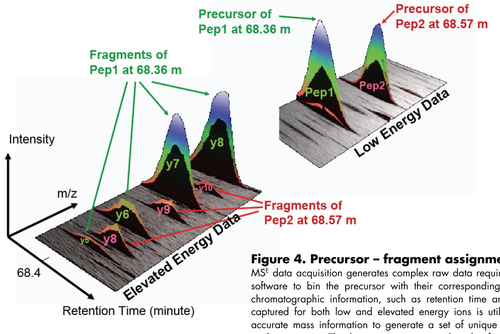


Figure 4. Precursor – fragment assignment.
MS^E data acquisition generates complex raw data requiring a dedicated software to bin the precursor with their corresponding fragments. The chromatographic information, such as retention time and peak shape captured for both low and elevated energy ions is utilized along with accurate mass information to generate a set of unique MS^E spectra for each precursor. The basic assumption is that the fragment ions can originate only from precursor eluting exactly in the same retention time. Generated MS^E spectra are searched against databases via conventional algorithms, such as MASCOT.

Ion Accounting software

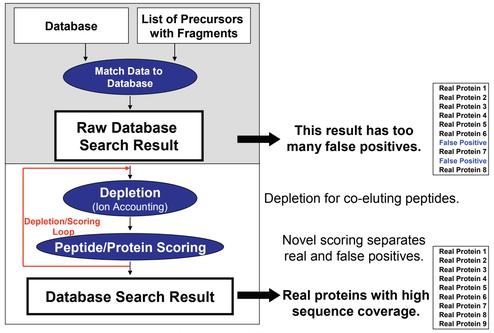


Figure 5. Ion Accounting software workflow.

Table 1: Database search results

	No. of Proteins after filtering		No. of Peptides after filtering		False positive rate
	control	cancer	control	cancer	
IA	58	85	334	556	0%
PLGS	55	80	219	320	4%
MASCOT	46	80	98	130	3.5%

Table summarizes the number of peptides and proteins identified from LC-MS^E experiment of non-depleted human urine samples (breast cancer versus control were compared).

Table 1. A number of confidently identified peptides/proteins in normal and cancer patient urine samples.
The experiment was carried out in triplicate analyses; the MS^E spectra were searched against a Swiss Prot Human Database using three different search engines (MASCOT, PLGS or IA software). The database also contains the randomized proteins sequences; the number of randomized proteins is 3 times greater than the number of human proteins in database. The randomized proteins were used to assess the degree of false positive identifications. Only the peptides replicating in at least in two out of three analyses were accepted as valid hits. The valid peptide hits were later collapsed into protein identifications.
The database was indexed using strict trypsin cleavage rules, and one missed cleavage was allowed. Fixed modifications of carboxyamidomethylation on cysteine was selected. Tolerance used for peptides identification was ± 50 ppm, tolerance for MS/MS fragments was ± 0.03 Da.

CONCLUSIONS

- MCX cleanup method significantly improved LC-MS^E analysis results.
- A novel LC-MS^E method was successfully applied for the analysis of a highly complex proteomic sample.
- The approach provided reliable protein ID`s with high sequence coverage.
- Novel Ion Accounting (IA) software rendered more protein/peptide identification compared to conventional algorithms for database search.
- Proteins unique for breast cancer samples will be further examine by orthogonal methods (e.g. ELISA) as potential biomarker discovery.

720001940EN