DOI 10.1002/prca.201400163

Research Article

Comparison of higher energy collisional dissociation and collision-induced dissociation MS/MS sequencing methods for identification of naturally occurring peptides in human urine

Martin Pejchinovski[1,2], *Julie Klein*[2], *Adela Ramírez-Torres*[2], *Vasiliki Bitsika*[3], *George Mermelekas*[3], *Antonia Vlahou*[3], *William Mullen*[4], *Harald Mischak*[2,4] *and Vera Jankowski*[5]

[1] Charite-Universitätsmedizin Berlin, Berlin, Germany
[2] Mosaiques Diagnostics GmbH, Hanover, Germany
[3] Biotechnology Division, Biomedical Research Foundation, Academy of Athens, Athens, Greece
[4] Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
[5] Universitätsklinikum RWTH Aachen, Institute of Molecular Cardiovascular Research, Aachen, Germany

Received: October 21, 2014
Revised: February 27, 2015
Accepted: March 23, 2015

Purpose: The aim of this study is to determine the best fragmentation method for sequence identification of naturally occurring urinary peptides in the field of clinical proteomics.

Experimental design: We used LC-MS/MS analysis of urine samples to determine the analytical performance of higher energy collisional dissociation (HCD), CID with high and low resolution MS/MS for the identification of naturally occurring peptides in the low molecular weight urinary proteome.

Results: HCD and CID high-resolution generated a 22% error rate in peptide sequence identifications. CID low-resolution showed significantly higher error rates (37%). Excluding the error rate (i.e rejection of cysteine-containing peptides), we observed a higher degree of overlap between HCD and CID high resolution for identification of peptide sequences of rank 1 and cross-correlation ≥ 1.9 (262 peptide sequences) compared to CID low (208 peptide sequences with HCD and 192 peptide sequences with CID high). Reproducibility of detected peptides in three out of the five replicates was also higher in HCD and CID high in relation to CID low resolution.

Conclusion and clinical relevance: Our data demonstrated that HCD and CID high-resolution performed with better accuracy and reproducibility than CID low resolution in respect to the identification of naturally occurring urinary peptide sequences.

Keywords:
Biomarkers / Clinical proteomics / MS/MS sequencing / Peptide fragments

Additional supporting information may be found in the online version of this article at the publisher's web-site

Correspondence: Dr. Vera Jankowski, RWTH Aachen, Institute of Molecular Cardiovascular Research, Aachen, Germany Pauwelsstr. 30. 52074 Aachen, Germany
E-mail: vjankowski@ukaachen.de
Fax: +49-241-80-82716

Abbreviations: ETD, electron transfer dissociation; **FDR**, false discovery rate; **FTMS**, fourier transform mass spectrometry analyzers; **HCD**, higher energy collisional dissociation; **ITMS**, ion trap mass spectrometry; **LTQ**, linear trap quadrupole; **NOP**, naturally occurring peptide; **Xcorr**, cross-correlation

1 Introduction

During the last decades, clinical proteomics has evolved and is becoming one of the key components of life science research [1]. Proteomic studies are in most cases directed toward protein identification and characterization (e.g. by analyzing PTMs). However, considerable interest is now given to the quantification and identification of naturally occurring

Colour Online: See the article online to view Figs. 1–8 in colour.

Clinical Relevance

Identification of naturally occurring peptides by their amino acid sequences plays an important role in scientific research to elucidate pathophysiological processes during disease progression. In this regard, sequencing of the peptides is not only helpful to assign a particular protein involved in the disease, but also to search for differentially regulated proteases. Up to now, there have been several technical reports dealing with optimal fragmentation conditions and sequencing methods for tryptic peptides, whereas those for nontryptic peptides are still missing. Using two complementary fragmentation methods, namely higher-energy collisional dissociation (HCD) and CID, we performed a methodological comparison to determine which of them is most reliable in retrieving high quality sequences of urinary peptides.

peptides (NOPs), as potential biomarkers of diseases [2, 3]. Large-scale investigation of the peptide content of biological samples can provide notable insights into disease development and the associated pathophysiological mechanisms [4, 5]. Of note, the human urinary peptidome represents a rich source of potential biomarkers for a wide range of diseases not only affecting the kidney and the urogenital tract but also throughout the body [5–8]. Complete characterization of the native peptides, in contrast to their precursor proteins, is more representative of the (patho-) physiological state of the organism and is therefore essential for the assessment of potential biomarkers [2, 3, 9]. However, complete peptide sequencing and identification, to allow protein identification, still remains one of the main challenges in clinical proteomics. Without this it is not possible to define and understand the biological role of these molecules [10,11]. Indeed, sequencing of naturally occurring peptides display a greater problem than that the sequencing of tryptic peptides, as optimal search parameters are not restricted by the enzyme used [12]. However, certain sequence assignments can be identified in the naturally occurring urinary peptidome. A remarkable feature in NOP analysis is the ability to identify erroneous sequence assignments by using the presence of free cysteine. Peptide ssequences that include cysteine are not possible, since without reduction and alkylation cysteine forms disulphide bonds and therefore we can use this to identify incorrect sequence assignments [11]. The bond strength of the disulphide bridges are extremely stable and are used to generate the tertiary structure of the protein and are resistant to fragmentation [13, 14].

MS has emerged as the gold standard for proteome and peptidome studies due to its ability to analyze thousands of peptides/proteins in parallel. A key step for MS-based proteomics analysis for correct protein identification is accurate peptide sequencing and positive database searching. More importantly, de novo sequencing of unknown naturally occurring peptides is a major challenge in current research and peptidome analysis. MS peptide sequencing is based on the production of peptide fragments from a precursor ion that are then compared to the fragmentation patterns contained in databases. Development of MS instrumentation for protein and peptide identification utilizes a number of fragmentation methods such as CID, higher-energy collisional dissociation (HCD), and electron transfer dissociation (ETD) [10]. These different fragmentation methods can be associated with several different mass analyzers like Fourier transform mass spectrometry analyzers (FTMS) for accurate mass measurement and ion trap mass spectrometry analyzers (ITMS) for the measurements of fragment ions [15, 16]. Moreover, much attention has been given to tandem mass spectrometers, especially quadrupoles and or ion traps coupled with ion cyclotron resonance cells, such as the hybrid linear ion trap Orbitrap mass spectrometer (LTQ Orbitrap) [17, 18].

Generally, CID generates b- and y-type fragment ions and has been the method applied in the vast majority of proteomics studies. This fragmentation process is based on the collision of isolated precursor ions with noble gases (e.g argon, helium), leading to conversion of their kinetic energy to vibrational internal energy. The increase of internal energy to a certain point results in peptide bond breakage and thus fragment ions are formed in the LTQ Orbitrap. High resolution detection of the precursor ions is performed in the Orbitrap, whereas subsequent production of the fragment ions (b- and y-type ions) is typically performed in the ion trap. Low resolution detection of these fragments can occur in the linear ion trap or high resolution sequential detection can be carried out in the Orbitrap [18]. In addition, the LTQ Orbitrap has been utilized to perform HCD fragmentation. In the HCD method, fragment ions are produced in a separate collision cell and not in the ion trap and then transferred to the C-trap and injected into the orbitrap for high resolution measurements. Finally, ETD fragmentation mode induces fragmentation of cations (e.g. peptides or proteins) by electron transfer leading to cleavage of N-C_α bond and creation of c- and z-type fragment ions [19].

Each of these methods has been shown to have advantages and disadvantages and it remains unclear that is the most suitable for peptide identification. In particular, CID is fast and can be performed with high or low resolution, however it shows a low mass cutoff limitation, preventing the analysis of fragments with m/z values below approximately 33% of the parent ion [20]. On the other hand, HCD has no low-mass cutoff, employs higher energy dissociation and high-resolution ion detection, which increases the number of fragments with higher quality MS/MS spectra. However, high levels of dissociation energy in HCD mode could lead to further fragment ion breakage resulting in the production of a- type ions [21]. Finally, ETD fragmentation has been proven to be most suitable for analyzing highly charged peptides ($z > 3$) and is especially suited to identification of PTMs; however, the overall fragment ion production is lower than CID and HCD [16, 19]. Comparison of fragmentation methods has been performed recently and complementary use of different fragmentation methods has been proposed in most cases [16, 20, 22, 23]. In particular, it has been shown that complementary use of CID and ETD can increase the number of identifications and peptide coverage in tryptic digested samples [20]. On the other hand, HCD has been proven to be more effective in experiments labeling with isobaric tags for relative and absolute quantitation and tandem mass tagging labeling proteomic approaches as it does not have the low mass cut off [20]. In general, ETD is considered as the least effective fragmentation method, and therefore not frequently used in high-throughput proteomics studies and is mainly used for PTM identification [24]. In the case of naturally occurring peptides, Shen et al. suggested the complementary use of all three fragmentation methods for the identification of protein degradation products in human plasma [16]. However, a comprehensive study on the fragmentation methods for identification of NOPs in urine is still pending.

In this investigation, we focused on the evaluation of the most used peptide fragmentation modes, in order to distinguish the most appropriate method for characterization of naturally occurring peptides. For this reason, we performed LC-MS/MS analysis [25], a reproducible and highly robust technology, in a complex human urine sample, and compared the performance of HCD, CID high, and low resolution fragmentation modes, making use of the presence of cysteine in a peptide sequences as an indicator of an incorrect sequence assignment.

2 Materials and methods

2.1 Sample preparation

A pooled urine sample from healthy anonymous individual was analyzed. Consent of the individual was obtained and complied with the guidelines of the Declaration of Helsinki (www.wma.net/en/30publications/10policies/b3/index.html). Basically, pooled midstream morning urine was collected and stored at –20°C before the analysis. Collection of the urine sample was used in several recent studies [26–28]. No protease or phosphates inhibitors were added and the pH was not adjusted. Urine samples were prepared for LC-MS/MS analysis as described by Zürbig and her collaborators [29].

2.2 LC-MS/MS analysis

Aliquots of 5 μL were analyzed on a Dionex Ultimate 3000 RSLS nano flow system (Dionex, Camberly, UK). After loading (5 μL) onto a Dionex 0.1 × 20 mm 5 μm C18 nano trap column at a flowrate of 5 μL/min in 98% 0.1% formic acid and 2% acetonitrile, sample was eluted onto an Acclaim PepMap C18 nano column 75 μm × 15 cm, 2 μm 100 Å at a flow rate of 0.3 μL/min. The trap and nano flow column were maintained at 35°C. The samples were eluted with a gradient of solvent A:98% 0.1% formic acid, 2% acetonitrile verses solvent B: 80% acetonitrile, 20% 0.1% formic acid starting at 1% B for 5 min rising to 20% B after 90 min and finally to 40%B after 120 min. The column was then washed and reequilibrated prior to the next injection. The eluant was ionized using a Proxeon nano spray ESI source operating in positive ion mode into an Orbitrap Velos FTMS (Thermo Finnigan, Bremen, Germany). Ionization voltage was 2.6 kV and the capillary temperature was 250°C. The mass spectrometer was operated in MS/MS mode scanning from 380 to 2000 amu. The top 20 multiply charged ions were selected from each scan for MS/MS analysis. Peptide fragmentation in CID low-resolution MS/MS mode was performed by usage of ITMS analyzer with 60 000 resolution for MS1 and by the ITMS analyzer with nominal 1500 resolution for MS2. For HCD and CID High-resolution, FTMS analyzer was selected with 60 000 resolution for MS1 and with 7500 resolution for MS2.

2.3 Sequence data analysis

Peptides and proteins were identified using Proteome Discoverer version 1.2 (Thermo Fisher Scientific, Bremen) with SEQUEST spectral algorithm. Search parameters were set as follows: maximum threshold for MS/MS peptides 500 counts, precursor mass tolerance up to 10 ppm, fragment mass tolerance for HCD and CID high-resolution 0.05 D and for CID low-resolution 0.8 Da. No fixed modifications were selected, oxidation of methionine, and proline was set up as variable modification. No enzyme specificity was selected and the minimum precursor mass set to 790 Da, maximum precursor mass of 6000 Da with a minimum peak count of 10. The high confident peptides were defined by cross-correlation (Xcorr) ≥ 1.9 and rank 1 as most valid for our experiments.

3 Results

3.1 Generation of top scoring peptide sequences dataset

To perform accurate comparison and evaluation of the different methods, five technical replicates of one standard urine sample was analyzed with HCD, CID high, and CID low resolution (Supporting Information Table). The total number of

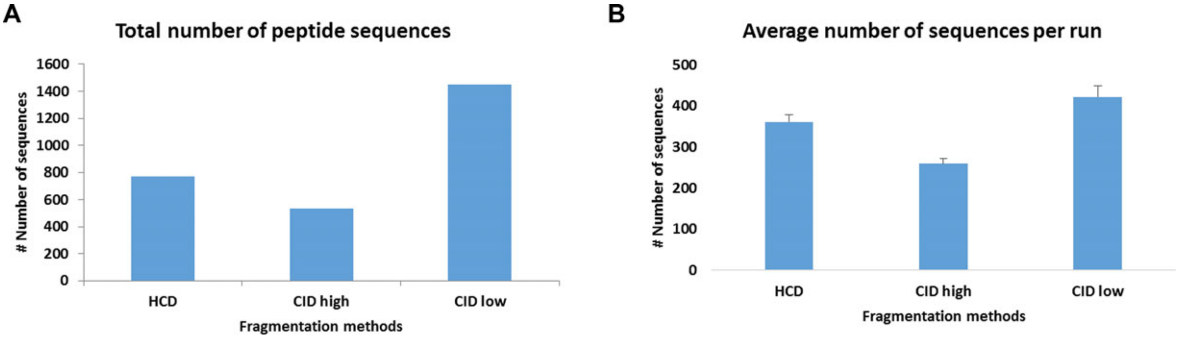

Figure 1. Analysis of standard human urine by three fragmentation methods. (A) Evaluation of total peptide sequences of rank 1 and Xcorr ≥ 1.9 obtained by HCD, CID high, and low fragmentation methods. The LC-MS/MS analysis of the urine sample was done in five replicates with (B) average number of the peptide sequences per run in each fragmentation method.

peptide sequences identified with rank 1 and Xcorr ≥ 1.9 were further evaluated. This resulted in identification of total 770 sequences using HCD, 532 using CID High and 1499 using CID Low fragmentation methods (Fig. 1A). By calculating the average number of total peptide sequences detected in each run, we identified 360 ± 17 sequences with HCD, 258 ± 13 sequences with CID high and 421 ± 27 sequences with CID low ($p = 0.008$ between all methods) (Fig. 1B).

We investigated the reproducibility of the total number of peptide sequences detected in each of the fragmentation methods. Figure 2A shows that the majority of the peptide sequences identified by CID low are detected only once out of five replicates, compared to HCD and CID high methods that have a much lower number of single sequence identification (1265 for CID low, 367 and 242 peptide sequences for HCD and CID high). The number of sequences detected repeatedly in at least three of the five replicates was higher in HCD and CID high compared with CID low (38% for HCD, 39.7% for CID high versus 10.4 % for CID low) (Fig. 2B). In CID low a very high percentage of peptide sequences appearing in less than three replicates was detected (89.6%), while lower percentage was found for CID high (60.3%) and HCD (62%) (Fig. 2B).

3.2 Overlap of peptide sequences identified by HCD, CID high-, and CID low-resolution methods with rank 1 and Xcorr ≥ 1.9

In order to examine the performance and similarity of all three fragmentation modes, we assessed how many sequences were uniquely or commonly identified by each MS/MS fragmentation approach. A total of 180 peptide sequences were common to the three fragmentation methods. The overlap between HCD and CID high consisted of 123 peptide sequences, whereas CID low demonstrated poorer overlap, with only 39 common peptide sequences with HCD and 26 common peptide sequences with CID high (Fig. 3).Our analysis also showed higher selectivity for CID low (1254 unique sequences) compared with HCD (428 unique sequences) and CID High (203 unique sequences) (Fig. 3).

3.3 Evaluation of erroneous peptide sequences (cysteine-containing peptides)

To assess which of these modes is best for the correct characterization of naturally occurring peptides in urine, we next

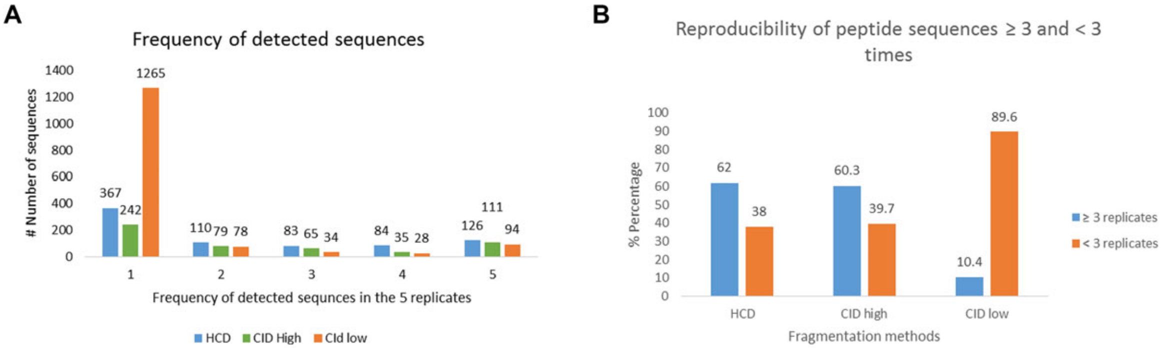

Figure 2. Frequency of detected peptide sequences in the five replicate runs. (A) Representation of reproducibility of total number of peptide sequences detected in the five replicates provided by each fragmentation method. (B) Comparison of the peptide sequences detected in more than three times (blue bar) and less than three times (orange bar) in HCD, CID high, and low resolution.

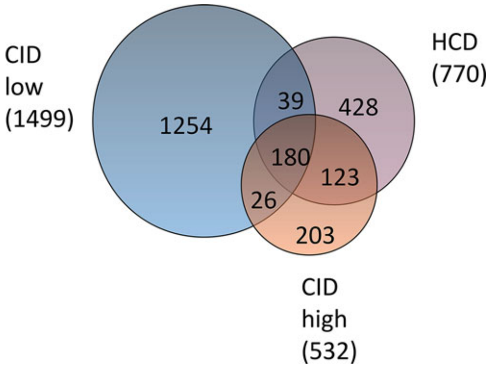

Figure 3. Overlap of the total number of peptides sequences in and between each method. Venn diagram that illustrates the overlap, common, and unique peptide sequences detected by HCD, CID high, and low resolution fragmentation methods. The analysis is with inclusion of cysteine-containing peptides as artifacts.

examined the data for possible erroneous assignments. We investigated the inaccuracies in our data, based on the number of the cysteine-containing peptides.

Cysteine is one of the least abundant amino acids making up only 3.2% of the amino acid content of proteins [30]. Cysteine in its unmodified form (i.e. Cys-SH) cannot be present in NOP, hence cysteine-containing peptides sequences can be considered as artifacts. We showed that in HCD, 23% (178/770) of the sequences identified with rank 1 and Xcorr ≥ 1.9 were cysteine-containing peptides, in CID high 21% (113/532) and for CID low 37% (561/1499) (Fig. 4A). As these numbers were very high, we then investigated the number of cysteine-containing peptides in the first 200 rank 1 peptides with the highest Xcorr values. These top-scoring sequences showed less cysteine-containing peptides, only 7% for HCD (14/200), 10% for CID high (19/200), and 17% for CID low (34/200) (Fig. 4B). We evaluated the reproducibility of the identification of the cysteine-containing peptides between all five replicates. We found that the erroneous peptide sequences were mostly identified with lower reproducibility, while sequences identified in at least three out of five replicates contained less artifacts (Fig. 4C–D). We next, analysed the distribution of the cysteine-containing peptides in the first 1500 peptides sequences ranked 1 and sorted by highest Xcorr. As before, we found that HCD and CID high resolution showed a maximum of 30% cysteine-containing peptides whereas CID low resolution showed maximum of 36% cysteine-containing peptides (Fig. 5A–C).

To assess the false discovery rate (FDR) in the standard format as described by Keller et al. [31] we generated a randomized version of the Fasta database using a Matrix Science perl script, available to download at http://www.matrixscience.com/help/decoy_help.html, under Manual Decoy Search. This script was run with ActiveState Perl 5 (v5.20.1) available from perl.org. The random FASTA database was appended to the original and the full database was searched with the sequest algorithm in proteome discoverer. The false discovery

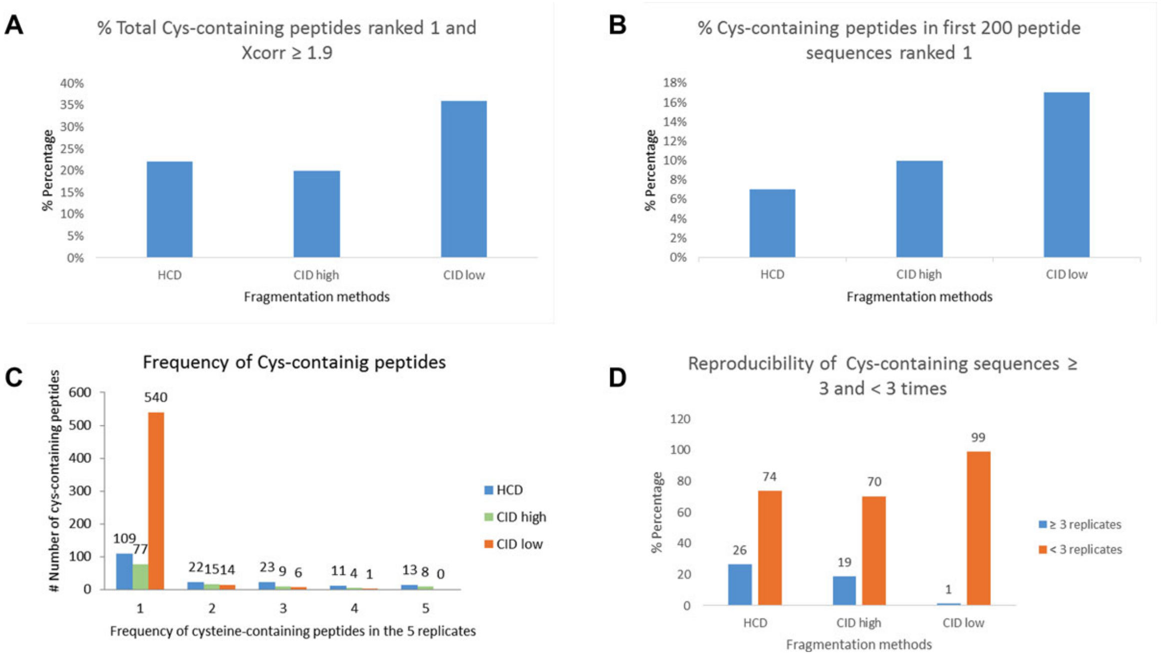

Figure 4. Evaluation of cysteine-containing peptides. (A) Percentage of total Cys-containing peptides ranked 1 with Xcorr ≥ 1.9. (B) Cys-containing peptides in the first 200 peptides filtered by highest Xcorr. (C) Evaluation of the frequency of cystein-containing peptides in the five replicates between each method and (D) Reproducibility of cysteine-containing peptides identified in three or more replicates (blue bar) and in less than three replicates (orange bar).

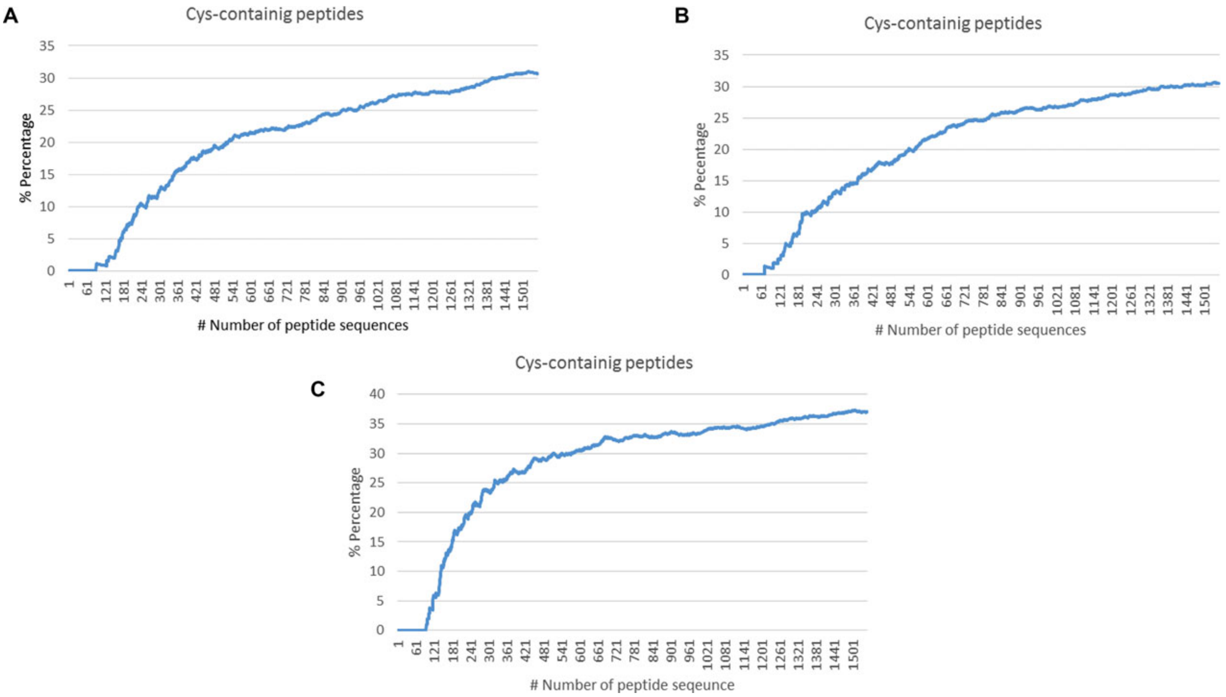

Figure 5. Interpretation of the first 1500 peptides ranked 1 and filtered by highest Xcorr verses percentage of sequences containing in each of the MS/MS fragmentation method. (A) Percentage of Cys-containing peptides by combined HCD MS/MS runs. (B) Percentage of Cys-containing peptide by combined CID high resolution MS/MS runs. (C) Percentage of Cys-containing peptides by combined CID low resolution MS/MS runs.

rate for CID low resolution was 14%, for CID high resolution it was 6.3% and for HCD it was 6.4%.

Finally, to clarify whether the assessment of cysteine-containing peptides is a systematic error or not, we analyzed data obtained from one of the HCD runs. By processing the raw data, we have compared the HCD dataset with precursor mass tolerance of 10 ppm and fragment mass tolerance of 0.05 Da versus same data set but with a reduced precursor mass tolerance of 2 ppm and fragment mass tolerance of 0.02 Da. The result showed that in both cases we obtained the high percentage of cysteine-containing peptides with maximum of 35% (Fig. 6A and B). By additional examination of the first 1000 peptide sequences ranked 1 and filtered by highest Xcorr with the same parameters, showed almost the maximum amount of 25% cysteine-containing peptides (Fig. 6C and D). The result shown here would indicate that increasing the mass accuracy of database search does not increase probability of obtaining a correct sequence.

3.4 Reproducibility and overlap of peptide sequences without cysteine containing peptides ranked 1 and Xcorr ≥ 1.9

We further examined the reproducibility of peptide sequences, this time excluding cysteine-containing peptides. Of note, CID low showed the highest number of peptide sequences detected in only one of the five replicates (725 peptide sequences) compared to HCD and CID high (258 and 165 peptide sequences, respectively) (Fig. 7A). Peptide sequences detected at least three times in all five replicates by HCD (246/592 peptide sequences) and CID high (190/419 peptide sequences) was higher compared to CID low (149/938 peptide sequences). On the other hand, CID low showed more peptide sequences that were detected less than three times out of five replicates (789/938 peptide sequences) (Fig. 7A and B). A total of 172 common peptide sequences were detected by all three fragmentation methods (Fig. 7C). Compared to the overlap when cysteine-containing peptides are included (Fig. 3), the overlap between these fragmentation methods showed a lower degree of overlap between HCD and CID high (further 90 peptide sequences), HCD with CID low (further 36 peptide sequences) and CID high with CID low (further 20 peptide sequences) (Fig. 7C). The number of unique peptide sequences for CID low (710 peptide sequences) was much higher compared HCD (294 peptide sequences) and CID high (137 peptide sequences).

In addition, we investigated the overlap between the peptide sequences that were detected in at least three replicates after excluding cysteine-containing sequences in each fragmentation method (Fig. 7D). In total the number of high quality sequences without cysteines, and detected in minimum three replicates, was 246 for HCD, 190 for CID high and only 149 for CID low, respectively. In accordance with

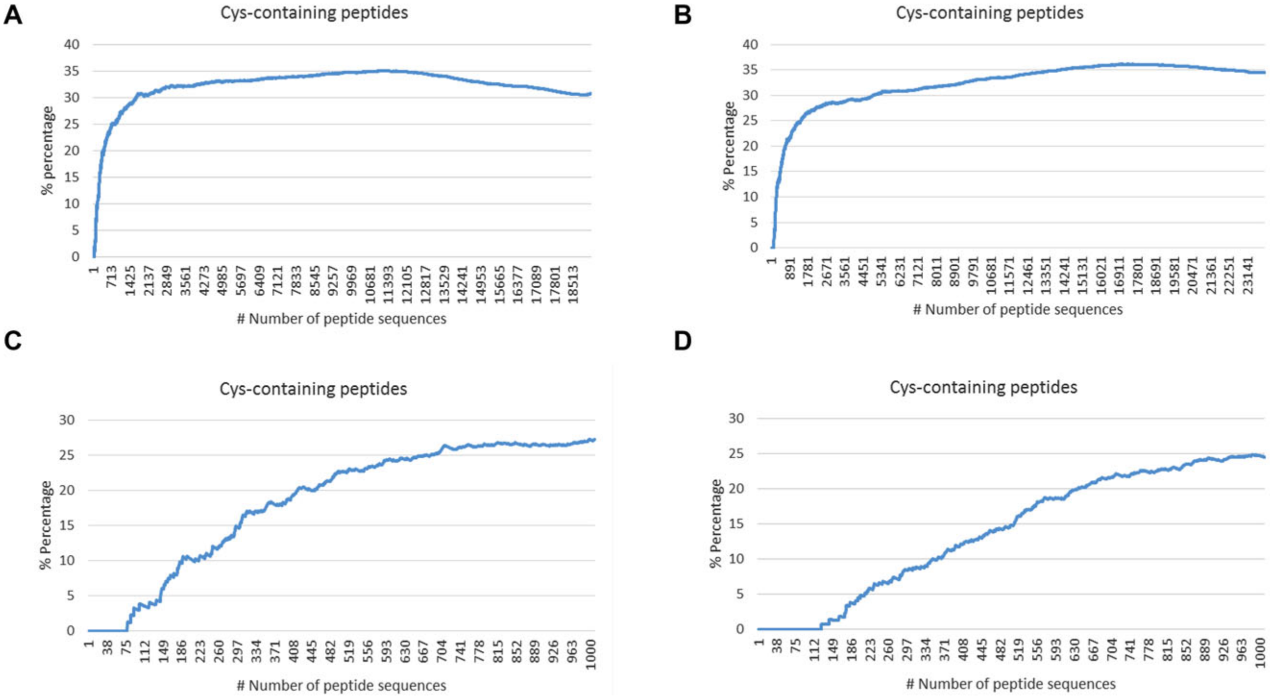

Figure 6. Incorrect peptide assignments obtained for one individual HCD run. (A) Interpretation of raw data for Cys-containing peptides in one of the HCD run with mass tolerance of 10 ppm and fragment mass tolerance of 0.05 Da and (B) with mass tolerance of two ppm and fragment mass tolerance of 0.02 Da. (C) Interpretation of raw data for the first 1000 Cys-containing peptides ranked 1 and filtered by highest Xcorr in one of the HCD run with mass tolerance of 10 ppm and fragment mass tolerance of 0.05 Da and (D) with mass tolerance of 2 ppm and fragment mass tolerance of 0.02 Da.

the above results, the overlap between HCD and CID high (further 49 peptide sequences) was higher compared to the overlap of HCD and CID low (13 peptide sequences) or CID high and CID low (further 14 peptide sequences). There were only 94 peptides sequences that were common to all three fragmentation methods. HCD showed highest number of unique peptides detected (90 peptide sequences), whereas both fragmentation methods of CID detected less unique number of peptides (33 with CID high and 28 with CID low).

3.5 Comparison of spectra associated with different sequence identification

Finally, we compared three sequences identified in HCD with high abundance and frequency that were not present in CID High, and vice versa (Table 1). We checked whether the peptide sequences were either not found at all, or found with lower rank, or whether the peptide was associated with different and/or lower quality spectra, hence interpreted with different sequence. The analysis showed that the equivalent precursor ion with same charge and retention time, HCD and CID generated different spectra but with the number of common fragment ions (Fig. 8) that were interpreted by the software as different peptide sequences.

4 Discussion

It has been reported that the accurate detection and identification of low molecular weight peptides in urine can be used for multiple disease biomarker discovery [32]. It has also been documented that NOP's from urine have advantages over other body fluids for this purpose due to their accessibility and not requiring intensive sample preparation prior to proteomic analysis [33]. The utilization and development of different fragmentation methods such as HCD and CID, as well as mass analyzers like FTMS and ITMS, can offer additional insights into sequence identification and provide an opportunity for more accurate and efficient identification of NOP's.

A number of factors can influence the identification of NOP peptides in mass spectrometry. One of these is with regard to the mass of the peptides being analyzed. According to, Zheng et al. [34], the majority of endogenous peptides with molecular weight lower than 3000 Da can be routinely analyzed using MS/MS, though these data were from the analysis of serum. Full-scan characterization of peptides with molecular weight over 3000 Da are less likely to be identified and can cause interference with other smaller peptide fragments [35], though this work was carried out using a LTQ ion trap mass spectrometer. We compared the data from HCD and CID high resolution fragmentation

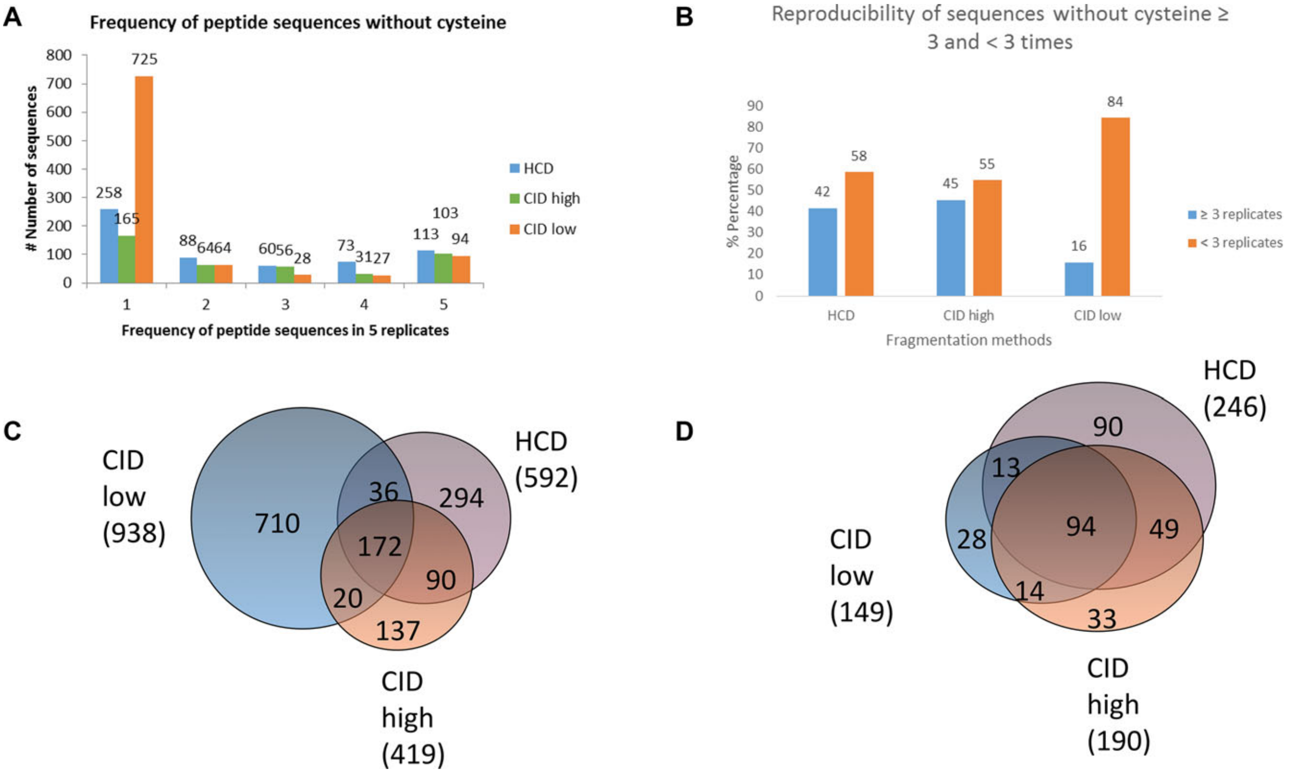

Figure 7. Reproducibility and overlap of peptide sequences without cysteine containing peptides. (A) Reproducibility of the peptides sequences without cysteines and (B) percentage of peptide sequences without cysteine identified in three or more replicates (blue bar) and less than three times (orange bar). (C) Overlap of peptide sequences without cysteine-containing peptides between each method. (D) Overlap of the peptide sequences without cysteine identified at least three times in and between each fragmentation method.

methods according to their molecular weight, from 1000 Da to 3000 Da in 400 Da increments and then between 3000 and 4000 Da and also above 4000 Da. Our results for NOP in urine indicate the majority of peptides identified by both methods (Total 832) are below 3000 Da. However, the greatest number of peptides identification in single range (364) was between 3000 and 4000 Da. HCD produced the highest number of sequence identifications (503) with CID high resolution having 329 with 198 common sequences below 3000 Da.

Comparison of all the identified peptides sequences (i.e. ranked 1, Xcorr ≥ 1.9) provided herein, displayed the complementarities of HCD and CID high versus CID low fragmentation method. Our analyses indicated that the total number of identified peptide sequences generated by CID low fragmentation method was ~twofold higher compared to HCD and CID high, showing however, less reproducible peptide sequences. It is well known that CID fragmentation performed in ion trap suffers from a low mass cut off and low resolution of the ions detected, generating spectra with lower

Table 1. Different sequence interpretation of the same highly abundant peptide peaks ranked 1 between HCD and CID high-resolution

	Area	Xcorr	MH^+	Rt.	Replicates
HCD					
GGppPGPpPPAGGGGGAGGAGGG PPPGPPGAGDRGGGGPGGGGPG	241 300 000	4.1	3406.5650	33.65	4
QPppQGpDGGGLPDGGDGpPPPQ	52 440 000	2.69	2197.9669	39.70	5
PpmDDVYApGELGPGGGGASPP	28 990 000	2.51	2085.9014	35.74	5
CID High					
PEATVEEAEAAAFGSEKPQSm FEpPQVSSPVQ	225 700 000	2.83	3406.5672	35.08	1
IGNVmGQSPSMMVGMPMPNGF	45 510 000	2.01	2197.9665	40.23	3
EPQpPPQGPDGGGLpDGGDGpP	25 180 000	2.25	2085.9036	36.46	3

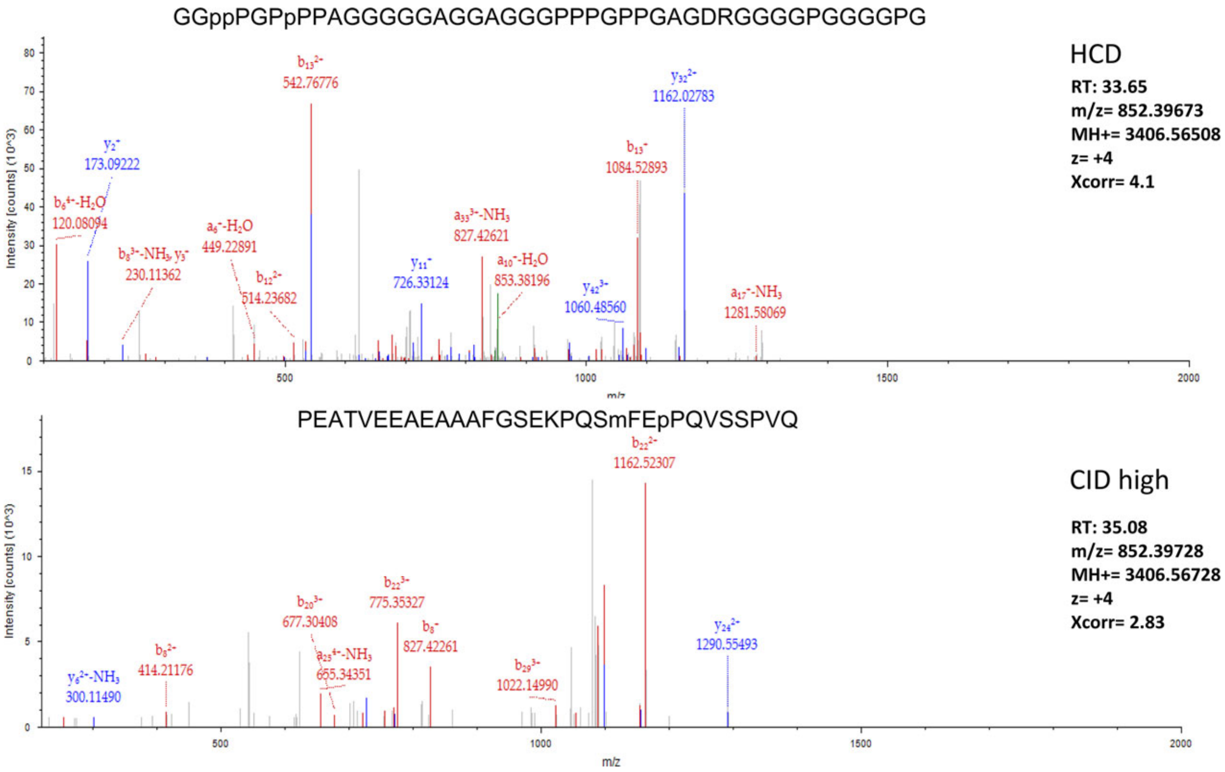

Figure 8. Spectrum of peptide sequence identified with HCD and CID high resolution fragmentation method. Representation of peptide fragmentation spectra that were found with high abundance and frequency in both HCD and CID high-resolution, but with difference sequencing results.

confident peptide identification [36]. The acquisition speed of CID fragmentation detected in the ion trap is much faster than that detected in the Orbitrap mass analyzer. This would give rise to the higher number of detected and identified peptide sequences, through with higher FDR. However, our experiments showed that most of the peptide sequences in CID low were detected just once out of five replicates (84.3%), indicating a randomness to the results. In comparison, HCD and CID high showed approximately three times more peptide sequences that were detected at least three out of five replicates. In the course of these finding, we demonstrated that the overlap between HCD and CID high was four times higher than compared to the overlap of HCD versus CID low as well as between the both CID methods.

MS/MS allows several methodological assessments of peptide FDRs, based on the construction of decoy databases or by calculating the distribution of peptide-spectrum match (PSMs) scores among all identifications. Often, a large number of low PSMs scores are problematic for the determination of true peptide and protein identifications [37]. Therefore, inadequate clustering of correct peptide identifications to correct proteins can lead to tenfold higher true FDR [38]. Thus, accurate peptide and protein false discovery estimation remains an open issue.

Our results pointed out that the classical determined false discovery rate, in NOP, for HCD and CID high were twofold lower compared to CID low. Examination of FDR calculated by cysteine method, gave a much higher amount of incorrect PSMs, almost threefold higher for CID low and fivefold higher for both HCD and CID high. These findings would indicate that the classical method of FDR is a significant underestimate of the true one for NOP. Therefore, discrimination between true and false PSMs in large-scale studies can be improved by including information about presence of specific amino acid or a sequence motif as a search methodology for statistical analysis [39]. In our study, we considered cysteine-containing peptides as an indication of incorrect assignments. We noticed approximately 15% less erroneous peptide sequences in the total number of sequences generated by HCD and CID high fragmentation methods than for CID low. When we examined the first 200 peptide sequences by their Xcorr value, HCD and CID high showed ∼twofold less cysteine-containing peptides when compared to CID low. Similar to our work, Shen et al. [16] investigated the performance of HCD, CID, and ETD for analysis of peptides using Orbitrap mass spectrometer. For evaluation of the identified peptide sequences, FDR-controlled SEQUEST, Mascot, Utags and de novo scoring algorithm were used. It was reported that CID fragmentation method using SEQUEST provided more peptide identification compared to HCD and ETD, though the accuracy of the sequences cannot be verified.

By comparison and combination of several peptide fragmentation methods, it has been well documented that different approaches could contribute to the identification of increased numbers of spectra and provide improved peptide assignments [40–42]. For this reason, we have examined the overlap between HCD, CID high, and low fragmentation methods. We demonstrated that the overlap between HCD and CID high was three times higher when compared to the overlap of HCD versus CID low as well as between the both the CID methods. However, altogether the overlap between the methods was limited, a result that should be balanced by the increased quality of the common sequences, characterized by higher frequency, and a lower number of artifacts. Part of the low overlap between the methods can be explained by the same sequences actually being present, but below the Xcorr threshold of 1.9, though we expected to see a higher overlap between HCD and CID high. Interestingly, when observing all combined runs from these three fragmentation methods, most (84%) of CID low peptide sequences were detected just once in between each of the runs. When examining the high quality peptide sequences that were detected in at least three replicates and after excluding of cysteine contained sequences, we observe the largest number of detected sequences by HCD and lowest by CID low. Also the highest number of unique best quality sequences was in the HCD method. In a similar investigation, the complementarity of fragmentations between HCD, CID, and ETD was investigated by examination of the overlap of each peptide population [20]. Frese et al. reported higher number of unique peptides for HCD (i.e doubly and triply charged peptides) compared to CID and ETD fragmentation methods. The overlap of most HCD peptides sequences showed highest Mascot score compared to those identified by either CID or ETD [20].

Although there were similarities between HCD and CID high performance, we noticed that fragmentation spectra were interpreted differently by Sequest search engine, being assigned to different sequences. Through the intrinsic differences attributed to fragmentation method, we expected that the same peptide sequences would be detected by both HCD and CID high resolution. However, we observed that this was not always the case, even in highly abundant peptides and peptides with high Xcorr scores. They consistently reported different sequences for the same peptide peak as seen in Fig. 8. At equivalent mass and retention time, HCD and CID generated different spectra with different ions, with more *y* and *a* ions in HCD, and more *b* ions in CID high. In a similar manner, the study by Nagaraj et al. [43] described the differences of identified spectra between HCD and CID. One of the most obvious discrepancies was the mass accuracy of the fragments with deviation of 0.1–0.3 Da from calculated values for ion trap measurements compared to just few ppm for the Orbitrap measurements. Peptides sequences with typical length in HCD and CID were covered by similar amino acids. CID spectra showed more *b*-ions compared to HCD because of increased fragmented in HCD [21, 44].

In conclusion, in this study we have shown that utilization of HCD and CID high-resolution MS/MS methods can successfully perform identification of NOP's in human urine. It is evident that HCD and CID high-resolution produces much superior results both in consistency of identifications and of quality of peptide sequences when compared to CID low-resolution. Classical CID carried out in the ion trap produced a higher number of single peptide identification that introduced randomness to the data that a simple FDR calculation could not account for. The use of cysteine as an indicator of a false sequence identification has shown that the classical FDR methodology greatly underestimates the true FDR in NOP sequencing. One of the most obvious facts was the number of artifacts, which was constantly higher in all experiments performed by CID low-resolution. Furthermore, we demonstrated practical advantage of HCD and CID high-resolution in improvement of the accuracy of protein/peptide characterization.

The work presented in this manuscript was supported by the grant of the EU-FP7/2007-2013, grant agreement no. 317246 (TranCYST-Translational research (training) in Polycystic Kidney Disease).

The authors have declared the following potential conflict of interest: Prof. Harald Mischak is founder and co-owner of Mosaiques Diagnostics, who developed CE-MS technology. Martin Pejchinovski and Adela Ramirez Torres are employed by Mosaiques Diagnostics.

5 References

[1] Mischak, H., Ioannidis, J. P., Argiles, A., Attwood, T. K. et al., Implementation of proteomic biomarkers: making it work. *Eur. J. Clin. Invest.* 2012, *42*, 1027–1036.

[2] Coon, J. J., Zürbig, P., Dakna, M., Dominiczak, A. F. et al., CE-MS analysis of the human urinary proteome for biomarker discovery and disease diagnostics. *Proteomics Clin. Appl.* 2008, *2*, 964–973.

[3] Good, D. M., Zürbig, P., Argiles, A., Bauer, H. W. et al., Naturally occurring human urinary peptides for use in diagnosis of chronic kidney disease. *Mol. Cell Proteomics* 2010, *9*, 2424–2437.

[4] Mischak, H., Schanstra, J. P., CE-MS in biomarker discovery, validation, and clinical application. *Proteomics Clin. Appl.* 2011, *5*, 9–23.

[5] Rodriguez-Suarez, E., Siwy, J., Zurbig, P., Mischak, H., Urine as a source for clinical proteome analysis: from discovery to clinical application. *Biochim. Biophys. Acta* 2014, *5*, 884–898.

[6] Metzger, J., Negm, A. A., Plentz, R. R., Weismuller, T. J. et al., Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders. *Gut* 2013, *1*, 122–130.

[7] Roscioni, S. S., de, Z. D., Hellemons, M. E., Mischak, H. et al., A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. *Diabetologia* 2012, *56*, 259–267.

[8] Theodorescu, D., Schiffer, E., Bauer, H. W., Douwes, F. et al., Discovery and validation of urinary biomarkers for prostate cancer. *Proteomics Clin. Appl.* 2008, *2*, 556–570.

[9] Klein, J., Eales, J., Zurbig, P., Vlahou, A. et al., Proteasix: a tool for automated and large-scale prediction of proteases involved in naturally occurring peptide generation. *Proteomics* 2013, *13*, 1077–1082.

[10] Guthals, A., Bandeira, N., Peptide identification by tandem mass spectrometry with alternate fragmentation modes. *Mol. Cell Proteomics* 2012, *11*, 550–557.

[11] Klein, J., Papadopoulos, T., Mischak, H., Mullen, W., Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification in human urine. *Electrophoresis* 2014, *35*, 1060–1064.

[12] Olsen, J. V., Ong, S. E., Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell Proteomics* 2004, *3*, 608–614.

[13] Kraj, A., Brouwer, H. J., Reinhoud, N., Chervet, J. P., A novel electrochemical method for efficient reduction of disulfide bonds in peptides and proteins prior to MS detection. *Anal. Bioanal. Chem.* 2013, *405*, 9311–9320.

[14] Mormann, M., Eble, J., Schwoppe, C., Mesters, R. M. et al., Fragmentation of intra-peptide and inter-peptide disulfide bonds of proteolytic peptides by nanoESI collision-induced dissociation. *Anal. Bioanal. Chem.* 2008, *392*, 831–838.

[15] Second, T. P., Blethrow, J. D., Schwartz, J. C., Merrihew, G. E. et al., Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures. *Anal. Chem.* 2009, *81*, 7757–7765.

[16] Shen, Y., Tolic, N., Purvine, S. O., Smith, R. D., Improving collision induced dissociation (CID), high energy collision dissociation (HCD), and electron transfer dissociation (ETD) fourier transform MS/MS degradome-peptidome identifications using high accuracy mass information. *J. Proteome. Res.* 2012, *11*, 668–677.

[17] Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E. et al., Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol. Cell Proteomics* 2011, *10*, M111.

[18] Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L. et al., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell Proteomics* 2009, *8*, 2759–2769.

[19] Wiesner, J., Premsler, T., Sickmann, A., Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* 2008, *8*, 4466–4483.

[20] Frese, C. K., Altelaar, A. F., Hennrich, M. L., Nolting, D. et al., Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome. Res.* 2011, *10*, 2377–2388.

[21] Olsen, J. V., Macek, B., Lange, O., Makarov, A. et al., Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, *4*, 709–712.

[22] Chi, H., Sun, R. X., Yang, B., Song, C. Q. et al., pNovo: de novo peptide sequencing and identification using HCD spectra. *J. Proteome Res.* 2010, *9*, 2713–2724.

[23] Kocher, T., Pichler, P., Schutzbier, M., Stingl, C. et al., High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *J. Proteome Res.* 2009, *8*, 4743–4752.

[24] Sarbu, M., Ghiulai, R. M., Zamfir, A. D., Recent developments and applications of electron transfer dissociation mass spectrometry in proteomics. *Amino Acids* 2014, *46*, 1625–1634.

[25] Klein J., Papadopoulos, T., Mischak, H., Mullen, W., Comparison of CE-MS/MS and LC-MS/MS sequencing demonstrates significant complementarity in natural peptide identification. *Electrophoresis* 2014, *35*, 1060–1064.

[26] Haubitz, M., Good, D. M., Woywodt, A., Haller H et al., Identification and validation of urinary biomarkers for differential diagnosis and dvaluation of therapeutic intervention in ANCA associated vasculitis. *Mol. Cell. Proteomics* 2009, *8*, 2296–2307.

[27] Jantos-Siwy, J., Schiffer, E., Brand, K., Schumann, G. et al., Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. *J. Proteome Res.* 2009, *8*, 268–281.

[28] Kistler, A. D., Mischak, H., Poster, D., Dakna, M. et al., Identification of a unique urinary biomarker profile in patients with autosomal dominant polycystic kidney disease. *Kidney Int*. 2009, *76*, 89–96.

[29] Zürbig, P., Schiffer, E., Mischak, H., Capillary electrophoresis coupled to mass spectrometry for proteomic profiling of human urine and biomarker discovery. *Methods Mol. Biol.* 2009, *564*, 105–121.

[30] Miseta, A., Csutora, P., Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Mol. Biol. Evol.* 2000, *17*, 1232–1239.

[31] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002, *74*, 5383–5392.

[32] Siwy, J., Mullen, W., Golovko, I., Franke, J., Zürbig, P., Human urinary peptide database for multiple disease biomarker discovery. *Proteomics Clin. Appl*. 2011, *5*, 367–374.

[33] Mischak, H., Julian, B. A., Novak, J., High-resolution proteome/peptidome analysis of peptides and low-molecular-weight proteins in urine. *Proteomics Clin. Appl.* 2007, *1*, 792–804.

[34] Zheng, X., Baker, H., Hancock, W. S., Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer. *J. Chromatogr. A* 2006, *1120*, 173–184.

[35] Yang, X., Hu, L., Ye, M., Zou, H., Analysis of the human urine endogenous peptides by nanoparticle extraction and mass spectrometry identification. *Anal. Chim. Acta* 2014, *829*, 40–47.

[36] Mann, M., Kelleher, N. L., Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. USA* 2008, *105*, 18132–18138.

[37] Li, Y. F., Radivojac, P., Computational approaches to protein inference in shotgun proteomics. *BMC Bioinformatics* 2012, *13*, S4.

[38] Jeong, K., Kim, S., Bandeira, N., False discovery rates in spectral identification. *BMC Bioinformatics* 2012, *13*, S2.

[39] Zhang, H., Yi, E. C., Li, X. J., Mallick, P. et al., High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol. Cell Proteomics* 2005, *4*, 144–155.

[40] Hart, S. R., Lau, K. W., Gaskell, S. J., Hubbard, S. J., Distributions of ion series in ETD and CID spectra: making a comparison. *Methods Mol. Biol.* 2011, *696*, 327–337.

[41] Kim, S., Mischerikow, N., Bandeira, N., Navarro, J. D. et al., The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol. Cell Proteomics* 2010, *9*, 2840–2852.

[42] Molina, H., Matthiesen, R., Kandasamy, K., Pandey, A., Comprehensive comparison of collision induced dissociation and electron transfer dissociation. *Anal. Chem.* 2008, *80*, 4825–4835.

[43] Nagaraj, N., D’Souza, R. C., Cox, J., Olsen, J. V., Mann, M., Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J. Proteome Res.* 2010, *9*, 6786–6794.

[44] Sleno, L., Volmer, D. A., Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom*. 2004, *39*, 1091–1112.