Title: Electron transfer dissociation of native peptides facilitates enhanced identification of urinary peptides

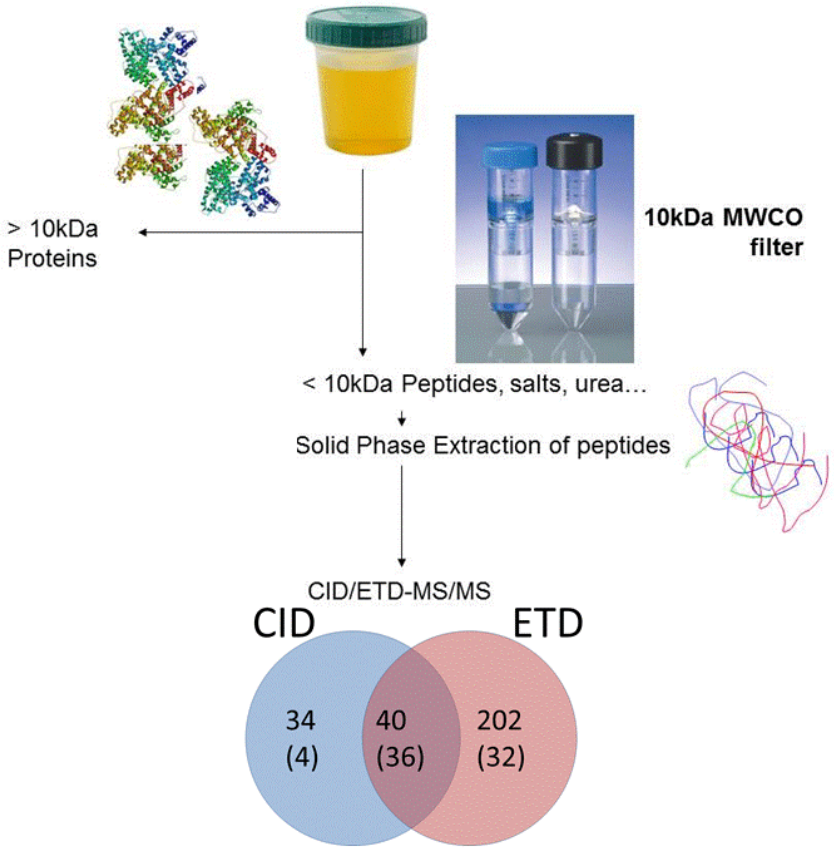
Author: Sarah R. Hart Louise C. Kenny Jenny E. Myers Philip N. Baker

PII: S1387-3806(15)00285-7
DOI: http://dx.doi.org/doi:10.1016/j.ijms.2015.08.025
Reference: MASPEC 15495

To appear in: *International Journal of Mass Spectrometry*

Received date: 19-6-2015
Revised date: 19-8-2015
Accepted date: 25-8-2015

Please cite this article as: S.R. Hart, L.C. Kenny, J.E. Myers, P.N. Baker, Electron transfer dissociation of native peptides facilitates enhanced identification of urinary peptides, *International Journal of Mass Spectrometry* (2015), http://dx.doi.org/10.1016/j.ijms.2015.08.025



Highlights:

- Native urinary polypeptides identified using CID/ETD product ion analysis
- ETD product ion spectra show improved properties for identification of native polypeptides
- More peptides identified using ETD than CID

Electron transfer dissociation of native peptides facilitates enhanced identification of urinary peptides

Sarah R Hart[1], Louise C Kenny[2], Jenny E Myers[3], Philip N Baker[1,4]

1 Institute for Science & Technology in Medicine & School of Medicine, Keele University, Staffordshire, UK
2 The Irish Centre for Fetal and Neonatal Translational Research (INFANT), University College Cork, Cork, Ireland
3 Maternal & Fetal Health Research Centre, St Mary's Hospital, Manchester, UK
4 The Liggins Institute, University of Auckland, Auckland, New Zealand

Contact: Sarah R Hart, Institute for Science & Technology in Medicine & School of Medicine, Keele University, Keele, Staffordshire, ST5 5BG, UK.
email: s.r.hart@keele.ac.uk tel: +441782 734639

Abstract

Urine as a biofluid is commonly used in clinical diagnostics, including those performed during pregnancy. Urine is a rich source of polypeptides and polypeptidic protein degradation products, which have been filtered from blood plasma, thus urine has potential as a source for novel clinical diagnostics in disease. In this study, we examine the urinary peptidome from normal healthy women during pregnancy, and demonstrate ready observation of large polypeptide. We utilise the dissociation method, electron transfer dissociation (ETD) to increase the identification rate of the peptides present within these samples, as the polypeptide species observed in these samples are large and highly charged. An increase in the number of peptides whose identities could be ascribed using routine database searching methods was enabled via the use of ETD.

Keywords:

Peptidome, urine, electron transfer dissociation, pregnancy, non-tryptic peptides.

Introduction

The use of urine as a readily-available biofluid for the development of novel diagnostics is a clear area of significant potential for clinical studies. Obtaining urine samples is trivial, non-invasive and highly acceptable to patients. Urine samples are widely used in diagnostic assays for renal function, drug metabolism and toxicology screens to detect recreational drug use and misuse of drugs in sports and [1] [2]. In comparison to whole blood, urine is stable and has both lower complexity and a reduced dynamic range [3]. If we can reliably observe markers of disease in urine samples, development of urine tests therefore forms a highly-desirable development area for clinical diagnostics.

Urine is formed via filtration of blood plasma in the kidney, and is principally comprised of water, sugars, amino acids and inorganic salts[4]. In general, intact proteins are not passed into urine due to their high molecular weights, and selective reabsorption of major serum proteins (albumin, immunoglobulins) via specific receptor molecules removes the bulk of this protein from the urine. The protein content of urine in normal healthy subjects is therefore typically low (30-150 mg/day) [5]. In pregnancy, kidney size and functional capacity are altered in response to rising progesterone levels, to facilitate an increased blood volume, encompassing the additional excretory requirements of the fetoplacental unit. This growth in renal capacity enables increased glomerular filtration requirements to be managed [6]. Raising the glomerular filtration rate influences the protein excretion via urine, resulting in a significant increase in urinary protein excretion in normal pregnancy [5], and higher concentrations of both intact proteins and proteolytic products of proteins in the urine.

Processed proteolytic products of proteins (or endogenous peptides) within urine are of interest as potential diagnostics for a number of reasons. Firstly, extremely low levels of peptides are robustly detectable in biofluids such as urine, with as low as attomole levels of detection being routine on modern tandem quadrupole instrumentation for targeted analysis at medium- to high-throughput; moreover, high levels of accuracy and low false positive rates are readily achieved [7]. Analysis of urine samples represents a rich source of polypeptide molecules as biomarkers for disease conditions, with approximately a ten-fold molar excess of peptides over proteins being excreted each day [8]. Proteolysis is altered in a number of disease conditions [9] [10], therefore observing the altered actions of proteases upon the peptidome is ultimately likely to yield information as to the pathogenesis of the condition as well as generating novel diagnostic strategies.

A significant body of work exists in the field of urinary proteomics, with between 1500 and 2300 proteins having been identified as components of urine [11] [12], following combined approaches incorporating GeLC-MS and shotgun proteomics of proteins following centrifugal concentration. For example, using a capillary electrophoresis-time-of-flight mass spectrometry approach, a panel of marker peptides were selected from urine polypeptides (nominal molecular weight <20kDa) which delineated prostate cancer patients from normal controls [13]; a limited number of these markers were amenable to routine identification via collision induced dissociation (CID).

In recent years, electron-mediated dissociation techniques, initially electron capture dissociation (ECD) and latterly electron transfer dissociation (ETD) have been applied to peptides [14]. In comparison with CID, the cleavage observed has been shown to provide more complete sequence coverage, with fewer second-and third generation product ions being observed than in CID-generated product ion spectra; this is particularly valuable for longer polypeptides and intact proteins [14] [15] [16] [17]. Electron-mediated dissociation techniques further facilitate analysis of labile post-translational modifications; where CID typically induces cleavage between amino acid residue and modification, ETD preserves labile modifications upon the sidechain [18].

Application to peptidome analyses is a clear avenue where the impact of ETDis now becoming clear; many peptide products in biofluids are expected to be larger than tryptic peptides [8], and further are anticipated to bear post-translational modifications. A number of studies have been conducted on polypeptides of varying biological origin[9] [19] [20] [21], and have found varying association between precursor ion charge state and preferred mode of identification (CID or ETD) [9]. ETD analysis of plasma peptides has been performed by Shen *et al.* [22], who have demonstrated that high mass accuracy (achieved by use of Orbitrap mass spectrometers) is instrumental in improving peptide matching to database entries in their unbiased Unique Sequence Tags (UStags) approach [23]. An alternative approach, proposed by Savitski and Zubarev for ECD spectra [24], combines identification data garnered from each complementary fragmentation mode (CID and ETD) to increase surety of matching to databases for endogenous peptide products [25].

Application of electron transfer dissociation to the analysis and identification of native urinary polypeptides has, to our knowledge, never previously been performed. We therefore apply this advanced technique for peptide dissociation to this class of analyte to demonstrate the feasibility of the approach for normal human urine samples, taken from pregnant subjects. We compare the data generated from the two typical fragmentation techniques used within conventional switching experiments (CID and ETD). Within this dataset we find a large number of peptides with physicochemical characteristics precluding ready identification by collisional dissociation, many of

which are amenable to identification using ETD.

Material and methods

Samples

Urine samples were obtained from low risk nulliparous women with singleton pregnancies recruited to the Screening fOr Pregnancy Endpoints (SCOPE) study were studied [26]. The exclusion criteria and inclusion criteria have been described in detail elsewhere[27]. Each participant was interviewed by a research midwife at 14-16 weeks of gestation and detailed clinical information was collected, as previously described [28]; Mid-steam urine specimens were collected at this antenatal attendance, processed within 4 hours, and stored at -80°C. Women were tracked prospectively and information about pregnancy outcomes was obtained; any subject who subsequently developed a pregnancy complication was excluded from the study.

Preparation of urine sample

Samples were shipped on dry ice and subsequently prepared using a modified version of a previously-published protocol [3]. Urine samples were thawed and spun to pellet urinary tract cellular debris (3,000 rpm, 4°C, 20 min). Peptides were separated from proteins (arbitrary NMWCO of 10kDa) using centrifugal concentrators (Vivaspin 20, Sartorius AG), with the spin through peptide fraction being subjected to solid phase extraction (SPE) to concentrate and clean up the peptides. SPE was performed using HLB cartridges (Aldrich), which were used according to manufacturers' instructions, eluting bound materials using 60% acetonitrile, 0.1% trifluoracetic acid. SPE-eluted components were subjected to strong cation exchange treatment to segregate bile components from peptides. This was performed as per the protocol of Cutillas [3], using bulk SCX media (polysulphoethyl A, PolyLC, Columbia MD). Peptides were eluted in 500 mM ammonium acetate in 20 % ACN, volatile buffer components were then removed by vacuum centrifugation prior to peptides being subjected to LC-MS/MS analysis.

LC-MS/MS analysis of urinary peptides

LC-MS/MS was performed on an orbitrap instrument (LTQ Orbitrap XL MS, ThermoFisher Scientific, San Jose, CA) equipped with ETD source, coupled to a nanoAcquity HPLC (Waters Corp, Milford, MA); separation was performed over 100min gradients from 0-80% acetonitrile on a BEH column (75um x

200mm, 1.7um particle size). A 'Top 3' data dependent acquisition, whereby the three most abundant multiply-charged precursors from a given survey spectrum, collected in the orbitrap mass analyser (resolution 30,000), were selected for independent fragmentation by both CID and ETD within the linear ion trap (AGC 30,000). CID was performed using an isolation width of 2amu, normalised collision energy of 35, activation Q 0.25 and activation time of 30 msec. ETD was performed with an isolation width of 3amu and activation time of 200 msec and anion AGC of 100,000; supplemental activation was applied [29]. Dynamic exclusion parameters were: repeat count 1, duration 30 sec, exclusion duration 60 sec, exclusion mass width -0.1- +1.1 amu, exclusion list 500 entries.

Data processing and identification of peptides

Raw mass spectra were processed using a custom script written at UCSF, PAVA [30] to generate mgf-formatted text files, and subjected to Mascot searching (Matrix Science, London). Search parameters were: SwissProt database (download date Feb 19 2014) with taxonomy restricted to *homo sapiens* (20,271 sequences), no enzyme filter applied, precursor ion tolerance 5ppm, product ion tolerance 0.6Da, search type CID+ETD, decoy search applied. Methionine oxidation was allowed as a variable modification. Mascot identification data were imported to Scaffold Q+ V 4.3.3 (Proteome Software, Portland, OR), applying peptide and protein thresholds of 95% confidence, with a minimum of 1 peptide for identification. Gene ontology information was imported from NCBI. Identified peptide sequences were parsed as gapped text files. These files were then submitted to weblogo (weblogo.berkeley.edu) to examine emergent patterns of proteolysis. This analysis provides a visual readout of sequence properties, with relative frequency of amino acid presence vs. expected abundance being shown proportional to one-letter code size [31].

The mass spectrometry and identification data have been deposited to the ProteomeXchange Consortium [32] via the PRIDE partner repository, with the dataset identifiers: PXD002312 & 10.6019/PXD002312, PXD002346 & 10.6019/PXD002346 and PXD002372 &10.6019/PXD002372.

Results & Discussion

ETD provides an effective means to generating sequence data from large urinary polypeptides vs. collisional dissociation

The number of polypeptides (cleaved from originating precursor proteins) which were confidently identified by ETD substantially outnumbered those identified by CID in our dataset (Fig 1), both in

terms of unique spectra and unique peptides observed. Manual examination of a subset of survey and product ion spectra indicated several reasons for this. One proposed explanation for the disparity in identification frequency from the two ion activation methods is attributed to intrinsic properties of the peptides. An abundance of large precursor ions bearing high charge states (z≥4) was observed in typical survey spectra (exemplified in Fig 2a). Previous studies have proposed that larger highly-charged precursors show poor fragmentation properties by CID due to vibrational energy dissipation [33]; ETD effectively fragments such species [9,34]. Qualitative assessment of example CID product ion spectra (Fig 2b) in cases where successful identification was made by ETD (Fig 2c), but not using collisional dissociation, indicated a strong difference in the quality of the product ion spectra. In the example case, a number of sequence-related product ions could be manually assigned to the product ion spectrum using sequence information derived from the ETD-based identification data (Fig 2b). The lack of CID-based identification is likely to derive from both the mass spectrometric methodology utilised in our analyses and the resultant database searching parameters. Spectra were acquired in low resolution mode using the ion trap, providing product ion spectra with low mass accuracy (0.6Da product ion tolerance being used in our database search). This reduces the likelihood of making above-threshold identifications. Increased mass accuracy, achieved via using the high-resolution mass spectrometer, would remove this problem; poor efficiency of ion transfer however meant that this was not performed in this particular analysis. A further contribution to the poor identification of polypeptides in this dataset using collisional dissociation is that, in contrast to tryptic peptides, a number of polypeptides lack C-terminal basic residues (Fig 3, Supplementary Data). The precise location of basic residues within peptide sequence is known to strongly influence peptide fragmentation by CID, with a potential source of difficulties in non-tryptic peptide identification resulting from the formation of non-canonical (scrambled) product ions [35].

The majority of peptides observed in our dataset have molecular weights of less than 5kDa (Fig 2a, Supplementary Data: Scaffold Reports). Whilst the natural distribution of polypeptidic species in the urine samples is reasonably likely to be biased towards species of lower molecular weight, the use of centrifugal filters in sample preparation and the tuning parameters of the mass spectrometer will have made additional contributions to this observation. Tuning parameters favouring larger polypeptides, and the use of larger poresize filters for sample preparation, could see an increase in the net molecular weight of polypeptide species observed.

Numerous polypeptides from originating proteins identified, with differing proteolysis patterns

A number of different functional classes of protein were observed within this study, as defined by Gene Ontology, including immune system process components, those involved in cellular localisation and proteins involved in developmental and reproductive processes. Multiple distinct peptides originating from abundant proteins were identified within urine samples, including uromodulin, a 64kDa glycoprotein expressed in the kidney and secreted into urine, whose functions are believed to be in regulating colloid pressure of the urine and regulating inflammation processes in the urinary tract. High-confidence product ion data were generated for polypeptide species pertaining to the C-terminal region of the mature protein sequence, which is removed from the mature protein by cleavage during its secretion, representing products which derive from the non-secreted domain [36]. A number of closely-related 'ragged' peptides with overlapping sequences, differing in precise site of cleavage were observed both for this parent protein and for other precursors identified. This reflects heterogeneity in the cleavage mechanisms involved during proteolytic processing of these species (see Supplementary Table 1).

Protein-derived polypeptides with sequences derived from osteopontin were also observed in these samples (see Supplementary Table 2). Osteopontin is a protein which has been shown by some groups to have a positive correlation with endothelial damage in the pregnancy complication pre-eclampsia; altered levels of osteopontin have been observed in pregnancies complicated by pre-eclampsia, as compared to uncomplicated pregnancies, in both trophoblast [37] and plasma [38]. The method as developed is therefore capable of assigning polypeptide sequences to biologically-relevant species.

Polypeptides analysed within this study were not subjected to targeted *in vitro* enzymolysis, as a goal of this experiment was to observe any emergent patterns in proteolytic action upon identified peptide species. (Fig 3) No clear patterns were observed in this analysis, bar an abundance of uncharged glycine residues throughout the peptides, and a small over-representation of basic amino acid residues, particularly lysines, near peptide termini. Detailed analysis of this phenomenon is beyond the scope of the present manuscript, and overinterpretation of these trends should be avoided, as these observations could result from a combination of the cleavage properties of endoproteinases active at some stage in the generation or preparation of these biological sample, or from inherent properties of peptides whose fragmentation properties are favoured by ETD and which were therefore identified, and are thus predominant in this dataset. The relative contributions of these factors remain unknown.

Future directions

Comparisons herein employ label-free approaches, with Scaffold spectral counting being used to examine differences between ion activation methods and samples. This method has clear limitations, particularly given the clear differences in the ability of the two fragmentation techniques to effectively dissociate and hence enable identification of the polypeptide ions under study. challenge. Similar, although not identical, results were obtained for three biological replicates (samples obtained from different patients with normal pregnancy outcomes, see Supplementary Data). Comparison to pregnancies complicated by pre-eclampsia or other common conditions is an important next step.

Low resolution (linear ion trap) data were generated for all product ion spectra; this was largely to retain sensitivity, speed and parallel nature of analyses, thus maximising the duty cycle of the analyses performed and number of peptides confidently identified. Transmission efficiency of ions to the orbitrap using the older-generation hybrid LTQ Orbitrap XL instrument was sufficiently low to have a significant adverse impact upon intensity of product ion spectra, and concomitant detrimental effect upon sequence assignment in preliminary experiments (data not shown). Use of the ion trap increased sensitivity for product ion analysis, but the low resolution of the product ion analyses has implications both for false discovery rate of peptide identification data, and reliability of observed products' charge states [9]. Increased mass accuracy for product ion analyses has the potential to significantly increase the certainty of matched sequence candidates, and more recent generations of Orbitrap instrumentation have improved ion transmission efficiency, meaning that losses as a result of ion transfer between linear ion trap and orbitrap are less problematic; this could increase the number of successful product ion identifications made for both fragmentation methods [22 23]. A further advantage of a high resolution survey/high-resolution product ion spectrum approach is the ability to include post-translational modifications and polymorphisms as variable search modifications. These important biological entities could not be not included in our study due to the massive search space of "no-enzyme" searches, but form an part of normal human variation, with polymorphisms being largely unrepresented in standard database searching methods [23]. Future studies incorporating high-resolution product ion data could begin to encompass this information.

Conclusions

This study applies ETD to the analysis of urinary polypeptides in the urine of pregnant women with uncomplicated outcomes. Survey spectra were typically extremely rich in multiply-charged

precursor ions, meaning that relatively few peptides/proteins were identified using conventional MS/MS with CID.

The use of peptidomics methodologies has the potential to enhance the accuracy of prediction or diagnosis of pregnancy complications. For example, almost 1 in 20 first pregnancies are complicated by pre-eclampsia, a disease characterized by the co-occurrence of hypertension and proteinuria. The condition is associated with serious maternal and perinatal morbidity and mortality, accounting for 70,000 maternal and 500,000 infant deaths annually [39] [40]. Identification of risk of pre-eclampsia is the first step to effective intervention and prevention, however, the overwhelming majority of first time mothers have no identifiable clinical risk factors in early pregnancy [28]. Although urinary proteomic profiling of pregnancies complicated by pre-eclampsia has been attempted [41] [42] [43], the methodologies used have not been able to discriminate between normal and pre-eclampsia pregnancies at gestations early enough to enable preventative strategies. It remains to be determined whether our approach, employing enhanced polypeptide identification of native urinary proteolytic fragments, achieved with the assistance of ETD, will eventually lead to a clinically useful predictive test.

Acknowledgements

The SCOPE study was funded by the Health Research Board Ireland (CSA/2007/2) and is housed within INFANT supported by Science Foundation Ireland (12/RC/2272).

EPSRC grants EP/E043143/1 & EP/E043143/2 supported SRH. LTQ-orbitrap hybrid data were generated at the Bio-Organic Biomedical Mass Spectrometry Resource at UCSF (Director A.L. Burlingame), supported by the Biomedical Research Technology Program of the NIH National Center for Research Resources, NIH NCRR P41RR001614 and NIH NCRR RR019934. The contributions of MSF staff, particularly David Maltby, Shenheng Guan and Aenoch Lynn in support of mass spectrometric and informatic analyses are gratefully acknowledged. Kate Murray (St Mary's Hospital) is acknowledged for technical assistance with sample selection.

References

(1) Zürbig, P.; Dihazi, H.; Metzger, J.; Thongboonkerd, V.; Vlahou, A. *Proteomics. Clin. Appl.* 2011, *5*, 256–268.

(2) Ryan, D.; Robards, K.; Prenzler, P. D.; Kendall, M. Recent and potential developments in the analysis of urine: A review. *Analytica Chimica Acta*, 2011, *684*, 8–20.

(3) Cutillas, P. R. *Methods Mol. Biol.* 2010, *658*, 311–321.

(4) Hall, J. E. *Guyton and Hall Textbook of Medical Physiology*; 2010.

(5) Kumar, A.; Kapoor, S.; Gupta, R. C. *J Clin Diagn Res* 2013, *7*, 622–626.

(6) Carlin, A.; Alfirevic, Z. *Best Pract. Res. Clin. Obstet. Gynaecol.* 2008, *22*, 801–823.

(7) Percy, A. J.; Chambers, A. G.; Yang, J.; Hardie, D. B.; Borchers, C. H. Advances in multiplexed MRM-based protein biomarker quantitation toward clinical utility. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 2013.

(8) Ling, X. B.; Mellins, E. D.; Sylvester, K. G.; Cohen, H. J. *Urine Peptidomics for Clinical Biomarker Discovery*; 2010; Vol. 51.

(9) Sasaki, K.; Osaki, T.; Minamino, N. *Mol. Cell. Proteomics* 2013, *12*, 700–709.

(10) Huesgen, P. F.; Lange, P. F.; Overall, C. M. *PROTEOMICS – Clin. Appl.* 2014, n/a – n/a.

(11) Adachi, J.; Kumar, C.; Zhang, Y.; Olsen, J. V; Mann, M. *Genome Biol.* 2006, *7*, R80.

(12) Kentsis, A.; Monigatti, F.; Dorff, K.; Campagne, F.; Bachur, R.; Steen, H. *Proteomics. Clin. Appl.* 2009, *3*, 1052–1061.

(13) Theodorescu, D.; Schiffer, E.; Bauer, H. W.; Douwes, F.; Eichhorn, F.; Polley, R.; Schmidt, T.; Schöfer, W.; Zürbig, P.; Good, D. M.; Coon, J. J.; Mischak, H. *Proteomics Clin Appl* 2008, *2*, 556–570.

(14) Good, D. M.; Wirtala, M.; McAlister, G. C.; Coon, J. J. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Molecular & Cellular Proteomics*, 2007, *6*, 1942–1951.

(15) Hart, S. R.; Lau, K. W.; Hao, Z.; Broadhead, R.; Portman, N.; Hühmer, A.; Gull, K.; McKean, P. G.; Hubbard, S. J.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* 2009, *20*, 167–175.

(16) Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. *Nat. Methods* 2008, *5*, 405–407.

(17) Bunger, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L. *Anal. Chem.* 2008, *80*, 1459–1467.

(18) Wiesner, J.; Premsler, T.; Sickmann, A. *Proteomics* 2008, *8*, 4466–4483.

(19) Hui, L.; Cunningham, R.; Zhang, Z.; Cao, W.; Jia, C.; Li, L. *J. Proteome Res.* 2011, *10*, 4219–4229.

(20) Altelaar, A. F. M.; Mohammed, S.; Brans, M. A. D.; Adan, R. A. H.; Heck, A. J. R. *J. Proteome Res.* 2009, *8*, 870–876.

(21) Azkargorta, M.; Soria, J.; Ojeda, C.; Guzmán, F.; Acera, A.; Iloro, I.; Suárez, T.; Elortza, F. *J. Proteome Res.* 2015, *14*, 2649–2658.

(22) Shen, Y.; Tolić, N.; Xie, F.; Zhao, R.; Purvine, S. O.; Schepmoes, A. a; Moore, R. J.; Anderson, G. a; Smith, R. D. *J. Proteome Res.* 2011, *10*, 3929–3943.

(23) Shen, Y.; Tolić, N.; Hixson, K. K.; Purvine, S. O.; Pasa-Tolić, L.; Qian, W.-J.; Adkins, J. N.; Moore, R. J.; Smith, R. D. *Anal. Chem.* 2008, *80*, 1871–1882.

(24) Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A. *Angew. Chem. Int. Ed. Engl.* 2006, *45*, 5301–5303.

(25) Hayakawa, E.; Menschaert, G.; De Bock, P.-J.; Luyten, W.; Gevaert, K.; Baggerman, G.; Schoofs, L. *J. Proteome Res.* 2013, *12*, 5410–5421.

(26) McCowan, L.; North, R. A.; Taylor, R. S. Screening for pregnancy endpoints: preeclampsia, growth restricted baby and spontaneous preterm birth. Australia New Zealand Clinical Trial Registry: ACTRN12607000551493, 2007.

(27) McCowan, L.; North, R. A.; Taylor, R. S. Screening for pregnancy endpoints: preeclampsia, growth restricted baby and spontaneous preterm birth. Australia New Zealand Clinical Trial Registry: ACTRN12607000551493, 2007.

(28) North, R. A.; Mccowan, L. M.; Dekker, G. A.; Poston, L.; Chan, E. H.; Stewart, A. W.; Black, M. A.; Taylor, R. S.; Walker, J. J.; Baker, P. N.; Kenny, L. C. *BMJ* 2011, *342*, d1875.

(29) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. *Anal. Chem.* 2007, *79*, 477–485.

(30) Lynn, A. J.; Chalkley, R. J.; Baker, P. R.; Medzhiradszky, K. F.; Guan, S.; Burlingame, A. L. In *American Society for Mass Spectrometry*; Santa Fe, NM, 2008.

(31) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* 2004, *14*, 1188–1190.

(32) Vizcaíno, J.; Deutsch, E.; Wang, R. *Nat. …* 2014, *32*, 223–226.

(33) Lifshitz, C. In *Principles of Mass Spectrometry Applied to Biomolecules*; John Wiley & Sons, Inc., 2006; pp. 239–275.

(34) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 2006, *1764*, 1811–1822.

(35) Chawner, R.; Eyers, C. E.; Gaskell, S. J. *Int. J. Mass Spectrom.* 2012, *316-318*, 284–291.

(36) Santambrogio, S.; Cattaneo, A.; Bernascone, I.; Schwend, T.; Jovine, L.; Bachi, A.; Rampoldi, L. *Biochem. Biophys. Res. Commun.* 2008, *370*, 410–413.

(37) Gabinskaya, T.; Salafia, C. M.; Gulle, V. E.; Holzman, I. R.; Weintraub, A. S. *Am. J. Reprod. Immunol.* 1998, *40*, 339–346.

(38) Stenczer, B.; Rigó, J.; Prohászka, Z.; Derzsy, Z.; Lázár, L.; Makó, V.; Cervenak, L.; Balogh, K.; Mézes, M.; Karádi, I.; Molvarec, A. *Clin. Chem. Lab. Med.* 2010, *48*, 181–187.

(39) Sibai, B.; Dekker, G.; Kupferminc, M. *Lancet* 2005, *365*, 785–799.

(40) Wildman, K.; Bouvier-Colle, M.-H. *BJOG* 2004, *111*, 164–169.

(41) Buhimschi, I. A.; Zhao, G.; Funai, E. F.; Harris, N.; Sasson, I. E.; Bernstein, I. M.; Saade, G. R.; Buhimschi, C. S. *Am. J. Obstet. Gynecol.* 2008, *199*.

(42) Carty, D. M.; Siwy, J.; Brennand, J. E.; Zürbig, P.; Mullen, W.; Franke, J.; McCulloch, J. W.; Roberts, C. T.; North, R. A.; Chappell, L. C.; Mischak, H.; Poston, L.; Dominiczak, A. F.; Delles, C. *Hypertension* 2011, *57*, 561–569.

(43) Lee, S. M.; Park, J. S.; Norwitz, E. R.; Kim, S. M.; Kim, B. J.; Park, C.-W.; Jun, J. K.; Syn, H. C. *J. Perinat. Med.* 2011, *39*, 391–396.

Figure legends

Figure 1: Venn diagram comparing urinary proteins and polypeptides represented by significant Mascot matches to typical product ion spectra from CID and ETD-based product ion analysis. A) Identified 'proteins' (multiple urinary polypeptides matching to a single SwissProt entry) from dataset at 95% protein threshold (1.4% reported FDR). B) Identified peptides from dataset at 95% peptide threshold (0.26% decoy FDR).

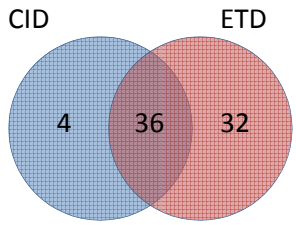
Figure 2: Example of typical mass spectra from LC-MS/MS analysis of urine. Full scan survey A) and product ion B), C) spectra illustrating typical data obtained from high charge state peptides present in urinary polypeptide samples. A) shows positive ion mode survey spectrum (m/z 350-1500), zoomed region (inset) around a quadruply-charged precursor ion signal at m/z 448.0 B) CID and C) ETD-generated product ion spectra generated for precursor, m/z 448.0^{4+}. CID analysis of this precursor gave no confident Mascot identification, database searching of the ETD spectrum reported a peptide sequence derived from the protein 14-3-3ζ, RVVSSIEQKTEGAEKK (ion score 66, expectation value $7.7x10^{-5}$). Predicted product ion species from this sequence are annotated on both B) and C), with manual assignment of unambiguous product ions within 0.6Da mass tolerance to CID spectrum B) being derived from reported sequence identification data obtained from ETD.

Figure 3: Sequence motif analysis of identified peptides following CID/ETD product ion analysis of urinary peptides. Sequence information as identified via database searching was exported as .csv format, extracted and parsed for length and redundancy, with identical sequences being removed, and sequences were entered into weblogo (weblogo.berkeley.edu). Height of amino acid residue is proportional to frequency of observation within the dataset.

Figure 1:

A)

Protein precursors represented by proteolytic fragments



B)

Peptides observed

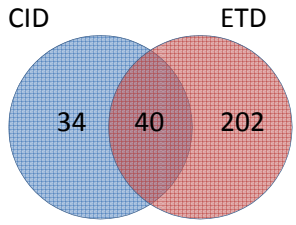
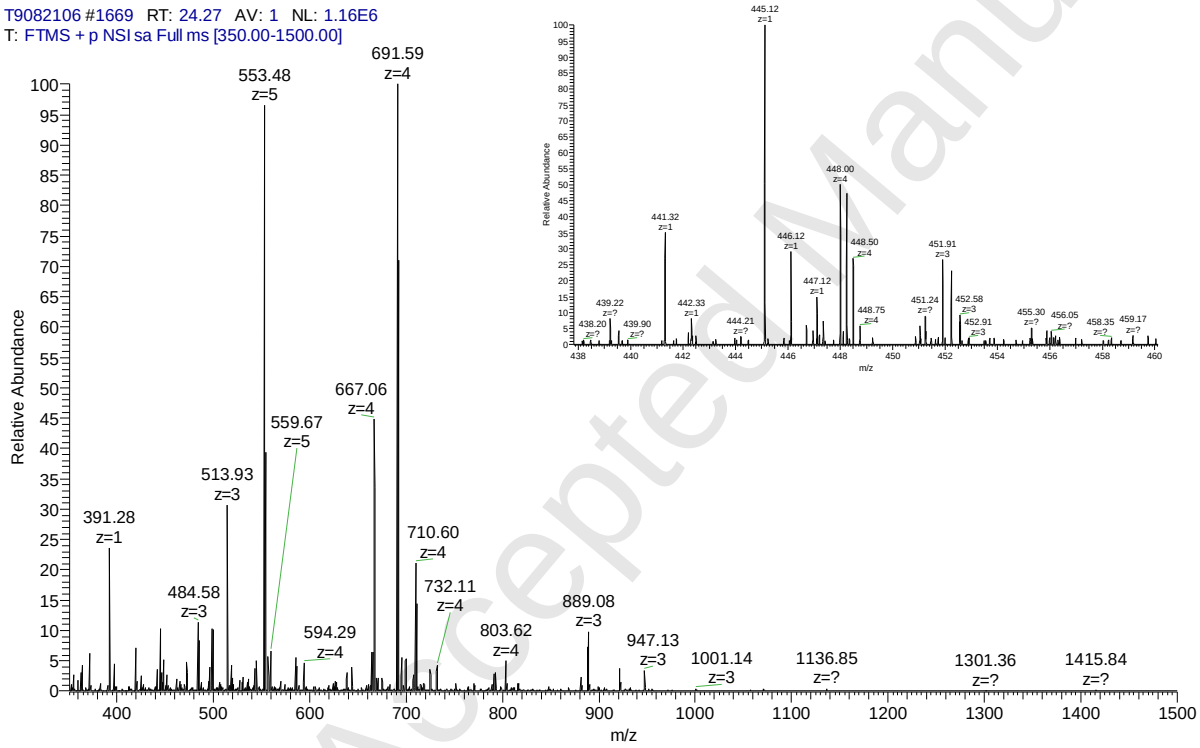
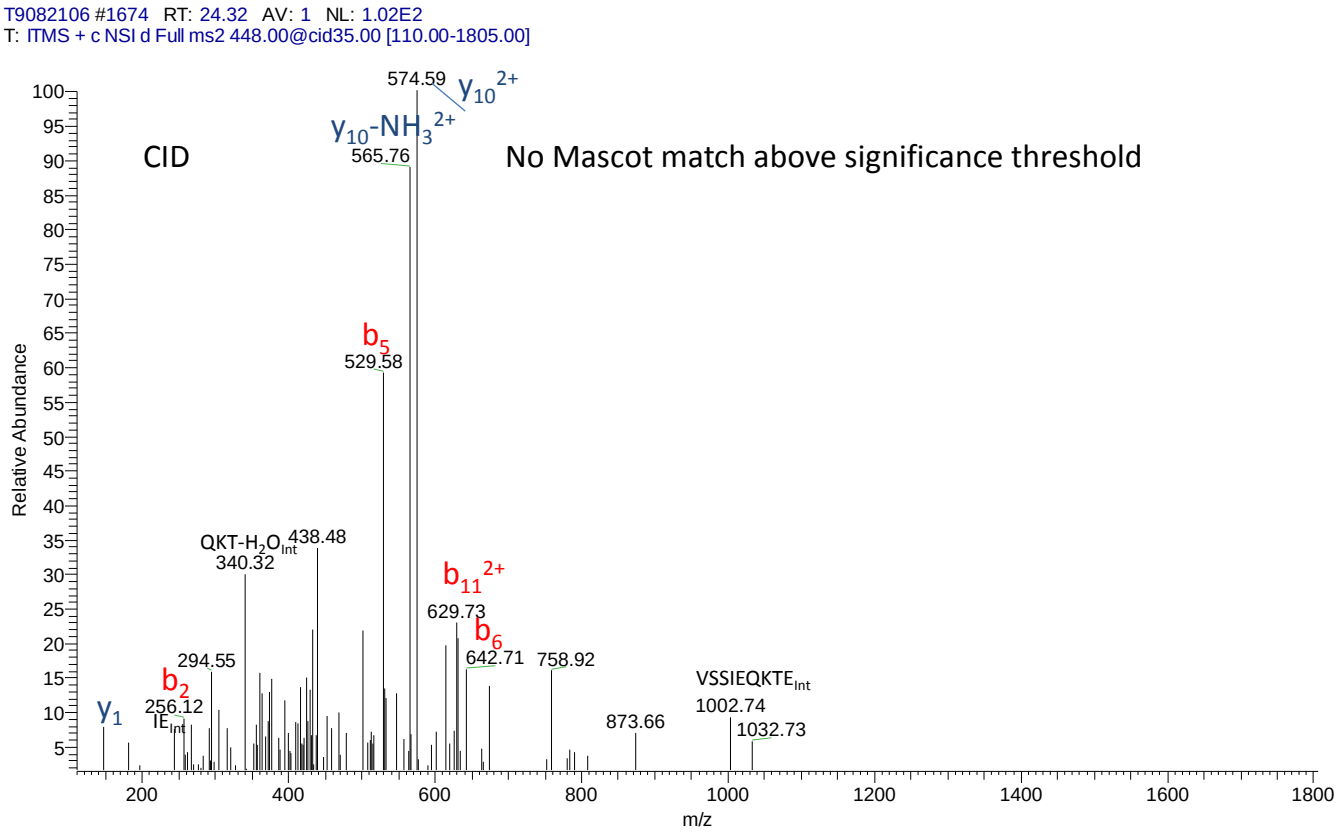


Fig 2

A)



B)



C)

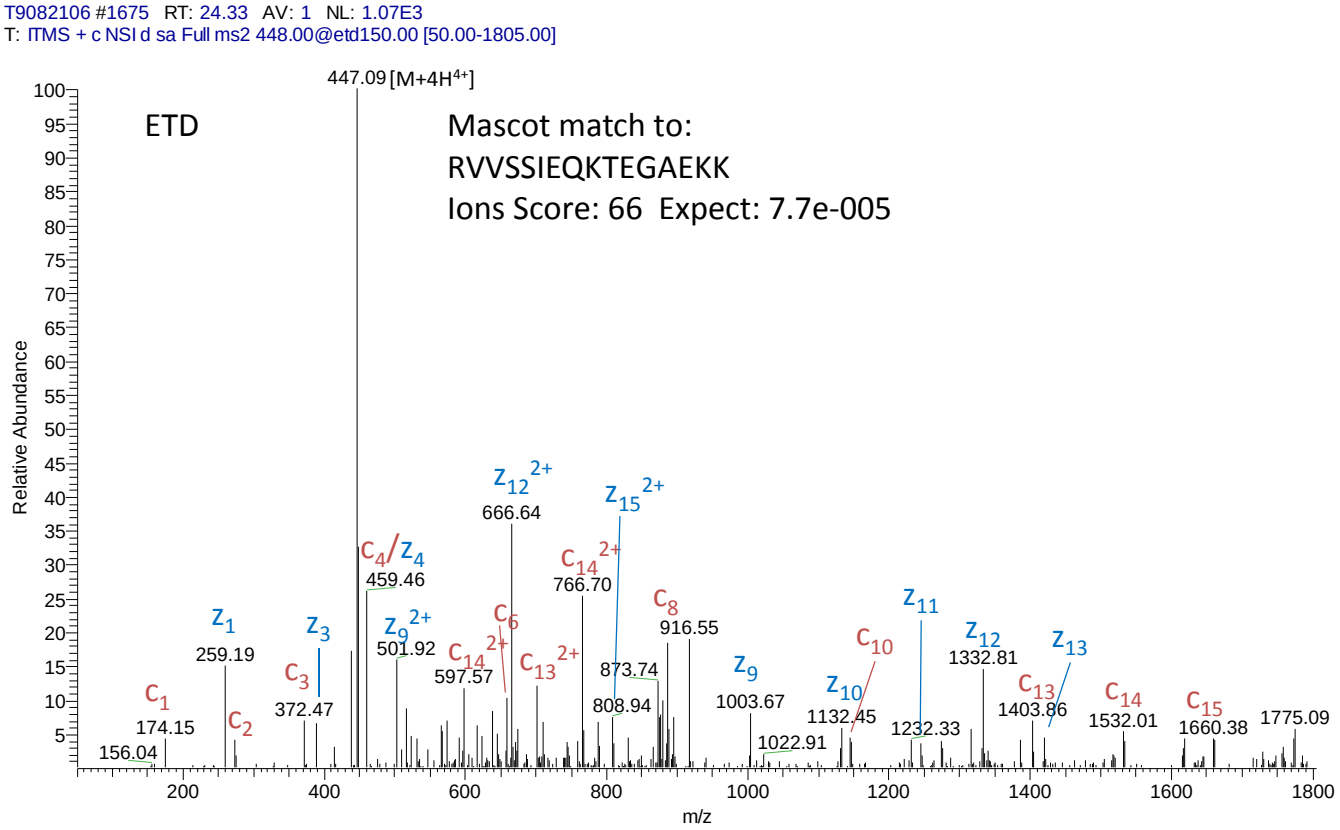


Figure 3: