Electron transfer dissociation for native peptide fragmentation facilitates enhanced identification of urinary peptides and proteins in pregnancy

Sarah R Hart¹, Jan A Fish¹, Louise C Kenny², Jenny E Myers³, Philip N Baker¹,⁴

1 Institute for Science & Technology in Medicine & School of Medicine, Keele University, Staffordshire, UK
2 The Irish Centre for Fetal and Neonatal Translational Research, 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 (200 words maximum)

Urine as a biofluid is commonly used in clinical diagnostics, including those performed during pregnancy. Urine is a rich source of polypeptides, protein degradation products, which have been filtered from blood plasma. Thus urine has strong potential as a source for novel clinical diagnostics in disease. In this study, we examine the urinary peptidome from normal healthy women during pregnancy, to demonstrate that peptides are readily observed. 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.

I have a workflow diagram which could be used for this?
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. As a clinical biofluid, taking 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 for detection of misuse of drugs in sports and recreational drug use \(^1\) \(^2\). In comparison to whole blood, urine is significantly less unstable 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 enable 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, with a significant increase in urinary protein excretion being observed in normal pregnancy \(^5\), resulting in higher concentrations of both intact proteins in the urine and proteolytic products of proteins.

Processed proteolytic products of proteins (or endogenous peptides) within urine are an interesting area as potential diagnostics for a number of reasons. Firstly, extremely low levels of peptides are robustly detectable from biofluids such as urine, with as low as attomole levels of detection being routine on modern tandem quadrupole instrumentation for medium-throughput implementation; 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 likely to yield information as to the nature of the condition as well as novel diagnostics.
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, and improved fragmentation of long polypeptides has been a notable feature of studies from a number of groups \(^{14,15,16}\). ETD further facilitates analysis of labile post-translational modifications; where CID induces bond cleavage between amino acid residue and modification, ETD preserves labile modifications upon the residue \(^{17}\).

Application to peptidome analyses is a clear avenue where ETD methods could have significant impact; many peptide products in biofluids are expected to be larger than tryptic peptides \(^{9}\), and further are anticipated to bear post-translational modifications. A number of studies have been conducted on neuropeptides of varying biological origin \(^{9,18,19}\), and have found a strong association between precursor ion charge state and preferred mode of identification (CID or ETD) \(^{9}\). ETD of plasma peptides has been performed by Shen et al. \(^{20}\), 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 \(^{21}\). An alternative approach, proposed by Savitski and Zubarev for ECD spectra \(^{22}\), combines identification data garnered from each complementary fragmentation mode (CID and ETD) to increase surety of matching to databases for endogenous peptide products \(^{23}\).

Application of electron transfer dissociation to the analysis and identification of urinary polypeptides has, to our knowledge, never previously been performed. We therefore apply this advanced technique for peptide dissociation to this class of analyte for the first time.
Experimental Section

Samples

To demonstrate feasibility of our approach, urine samples from 6 pregnant women recruited to the Screening fOr Pregnancy Endpoints (SCOPE) study were studied. Nulliparous women with singleton pregnancies were recruited to the multicentre SCOPE study. Exclusion criteria included an increased risk of major pregnancy complications. Each participant was interviewed by a research midwife at 14-16 weeks of gestation and detailed clinical information was collected, as previously described; 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. 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% trifluoroacetic 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, 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 (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’ method was employed for 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. CID was performed using an isolation width of 2amu, normalised collision energy of 35, activation Q 0.25 and activation time of 30msec. ETD was performed with an isolation width of 3amu and activation time of 150msec; supplemental activation was applied, to enable doubly-
charged precursors to provide meaningful product ion spectra via collisional warming of the activated but undissociated charge-reduced precursor ion cluster [ref Syka et al.].

**Data processing and identification of peptides**

Raw mass spectra were processed using a custom script written at UCSF, PAVA 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. Mascot data were then imported to Scaffold Q+ V 4.2.1 (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 using weblogo (weblogo.berkeley.edu) to examine emergent patterns of proteolysis.

**Results & Discussion**

**ETD provides an effective means to generating sequence data from large urinary polypeptides**

The number of peptides (and hence originating proteins) which were confidently identified by ETD outnumbered those identified by CID in our dataset (Fig 1), both in terms of unique spectra and unique peptides observed. Examination of survey spectra indicated the likely reason for this to result from both the length and charge state of the peptide species present within the urine polypeptide fraction (Fig 2). Many precursor ions were observed in survey spectra with high charge states (z≥4).

Numerous 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. Interestingly, high-confidence product ion data were generated for peptide species pertaining to the C-terminal region of the mature protein, which is cleaved from the mature protein during its secretion. Other interesting proteins which were observed in these samples included osteopontin, 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 and plasma.
In terms of the functional classes of proteins observed within this study, many different classes of protein are observed in the data (Fig 3). These include immune system process components, those involved in cellular localisation, and perhaps less surprisingly, proteins involved in developmental and reproductive processes.

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, maximising the duty cycle of the analyses performed. Whilst use of the ion trap increases sensitivity vs. using the orbitrap analyser, the reduced resolution of the product ion analyses poses some issues. 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 sufficient ion transmission efficiency that losses as a result of ion transfer between linear ion trap and orbitrap are less problematic. We anticipate that increased sensitivity high-resolution product ion analysis in conjunction with ETD would increase the number of peptides for which confident sequence identifications could be made. There remains a major issue with database searching in large search spaces; where no enzyme specificity is applied the search space for performing database searching is enormous; although increasing mass accuracy improves this, there remains a bioinformatic challenge to generate confident matches in large search space. Polypeptides were not subjected to enzymolysis as an additional goal of this experiment was to observe any emergent patterns in proteolytic action upon identified peptide species (Fig 4). With the exclusion of an over-representation of basic amino acid residues near the peptide termini, no obvious trends in peptide motif were observed. A more detailed analysis to dissect out and resolve emergent peptide sequence properties is underway.

Conclusions

This study represents the first attempt to apply electron transfer dissociation to the analysis of urinary polypeptides, particularly in the context of pregnancy.

Survey spectra were typically extremely rich in multiply-charged precursor ions (Fig 2), meaning that few peptides/proteins were identified using conventional MS/MS with CID (Fig 1). This illustrates a benefit of using ETD, as larger, more highly charged peptides gain increased surety of product ion analysis due to high mass accuracy and charge state resolution, combined with the comparatively stochastic nature of cleavage induced following ETD.

Incorporation of high-accuracy product ion data was not performed in this study as transmission efficiency of ions to the orbitrap was low enough to have a significant adverse impact upon intensity
of product ion spectra. Using a later-generation instrument with higher fidelity of transfer of ions between linear ion trap and orbitrap would improve the quality of data generated; this could increase the number of successful product ion identifications made. A further advantage of a high resolution survey/high-resolution product ion spectrum approach is the ability to include post-translational modifications and polymorphisms. These entities were 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.

For these experiments, comparisons have been performed using ‘label-free’ approaches, with Scaffold being used to examine differences between ion activation methods and samples. Further improvement in terms of comparison could be achieved by incorporation of stable isotopic tagging methods, such as dimethyl labelling. Post-labelling methods such as these, which are the only sensible choice for label incorporation where clinical samples are involved, almost invariably rely on having free amine groups at either N-terminus (i.e. non-modified N-terminal amino acid), C-terminus (as is the case in tryptic digests) or internally. Thus where N-terminally modified peptides which do not contain lysine residues are present, comprehensive coverage would not be achieved.

This methodology 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 concomitant 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. 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. Although urinary proteomic profiling of pregnancies complicated by pre-eclampsia has been attempted, 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 achieved through ETD will translate to a clinically useful predictive test.

Acknowledgements

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


Supporting Information:

Identification of peptides (Scaffold peptide reports).

Figure legends

Figure 1: Venn diagram comparing identified proteins and peptides from CID and ETD-based product ion analysis. A) Identified proteins 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: Mass spectrometry traces from LC-MS/MS analysis of urine. Typical survey A) and product ion B), C) spectra illustrating high charge states of peptides present in urinary polypeptide samples. A) shows zoomed region (inset) around a quadruply-charged precursor ion peak at m/z 448.0 B) CID and C) ETD-generated product ion spectra generated for precursor at 448.04+. This precursor gave no confident identifications by CID, but database searching of the ETD spectrum reported a peptide sequence from 14-3-3ζ, RVVSSIEQKTEGAEKK (ion score 66, expectation value 7.7x10^{-5}). Predicted product ion species from this sequence are annotated on C). Manual assignment of a small number of product ions within CID spectrum B) (annotated b and y-type products) was possible using the reported sequence information from the ETD spectrum.

Figure 3: Gene ontology analysis of identified proteins. Scaffold Q+ was used to import GO terms for all confidently identified proteins (95% peptide and protein threshold with 1 precursor/protein). These are classified according to biological function.

Figure 4: Sequence motif analysis of identified peptides following CID/ETD product ion analysis of urinary peptides. Sequence information as identified via database searching was parsed for length and redundancy, and sequences were entered into weblogo (weblogo.berkeley.edu). Height of amino acid residue indicates frequency of observation.
Figure 1:

A) Proteins

B) Peptides
ETD

Mascot match to:
RVVSSIEQKTEGAEKK
Ions Score: 66  Expect: 7.7e-005
Figure 3:

![Graph showing cellular processes and their categories]

- Growth
- Viral reproduction
- Cell killing
- Unknown
- Locomotion
- Rhythmic process
- Biological adhesion
- Multi-organism process
- Reproduction
- Reproductive process
- Immune system process
- Establishment of localization
- Developmental process
- Localization
- Metabolic process
- Response to stimulus
- Multicellular organismal process
- Biological regulation
- Cellular process

Figure 4:

<table>
<thead>
<tr>
<th>CID</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ETD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>