Non-Target Chemical Analysis Using Liquid Chromatography, Differential Ion Mobility and Tandem Mass Spectrometry

by

Daniel Beach

A Thesis
Presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Chemistry

Guelph, Ontario, Canada
© Daniel Beach, April, 2013
ABSTRACT

NON-TARGET CHEMICAL ANALYSIS USING LIQUID CHROMATOGRAPHY, DIFFERENTIAL ION MOBILITY AND TANDEM MASS SPECTROMETRY

Daniel Beach
University of Guelph, 1992
Advisor: Professor W. Gabryelski

Identification of trace unknown analytes in complex samples remains a significant challenge for analytical chemistry. Mass spectrometry (MS) and analytical separations techniques can now be used to develop and support a new analytical strategy called non-target analysis which aims to provide comprehensive identification and quantification of all detectable chemical species in a complex sample. This thesis addresses challenges currently limiting the utility of this non-target approach by developing analytical methods for acquiring MS data suitable for identification of trace unknowns and investigating current tools available for unknown identification from MS spectral data.

Liquid chromatography (LC) - MS, a widely used technique in trace analysis, was used to develop an analytical method capable of simultaneously acquiring high resolution MS and tandem mass spectrometry (MS/MS) data for hundreds of metabolites in urine. An emerging separation technique called high field asymmetric waveform ion mobility spectrometry (FAIMS) was also investigated, as an alternative to LC, for the identification of non-target analytes in urine. Modifications were carried out to the FAIMS-MS source interface allowing for transmission of small metabolite ions from FAIMS to MS. The challenge of direct electrospray (ESI) in urine analysis using ESI-FAIMS-MS was addressed by using sample dilution and
extending MS data acquisition time using FAIMS. This allowed for higher quality MS data to be acquired for low abundance urinary metabolites than was possible by LC-MS and the complete elimination of ionization suppression in dilute urine samples. Insight gained into ESI suppression in complex samples allowed for two methods of semi-quantification to be proposed for non-target analytes in complex samples without using unavailable chemical standards.

To address the challenge of unknown identification, faced throughout this thesis, an integrated approach was implemented to identify metabolites based only on spectral data without the usual requirement of availability of chemical standards. This approach combined spectral libraries, literature reports on ion chemistry and *de novo* identification based on gas phase ion chemistry with a detailed fragmentation study on nucleic acid bases, notably protonated uracil. Together, the instrumental methods and approaches to data analysis described allowed for the identification of 110 abundant chemical species detected in urine.
Acknowledgements

First I would like to thank my supervisor, Professor Wojciech Gabryelski for all of his support of my work throughout my time at Guelph. I am grateful for the opportunity to come here for my PhD, the time he has spent teaching me his craft and the resources he has given me to do my work. I am also very thankful for the opportunities he has given me to travel and present my work at world class conferences and for the hard work he has put in to nominating me for awards and scholarships, these have had a significant impact on the advancement of my career. I would also like to thank Professors Paul Rowntree, Tadeusz Gorecki and Richard Manderville for serving on my advisory committee as well as Professors Philip Britz-Mckibbin and Janusz Pawliszyn for agreeing to review my thesis and for serving on my defense committee. Thanks to the members of the Gabryelski research group, as well as all those I have had the pleasure of working with throughout my time at Guelph, for all their help and support. I would also like to acknowledge some of my mentors from previous points in my life, Dr. Michael Quilliam, Dr. Jocelyne Hellou, Dr. Stephen Duffy and Dr. Stephen Westcott. They have each played a large role in my development as a scientist and a person and have continued to support me throughout my time at Guelph.

Finally and most importantly I am grateful to all my family and friends for their support as I have worked through graduate school. To my wife Margaret and my daughter Clara, thank you for moving to Guelph and supporting me as I pursued my studies. I love you both very much.
Table of Contents

Abstract............................................................................................................................................ii

Acknowledgements ........................................................................................................................ iv

Table of Contents ............................................................................................................................ v

List of Tables ................................................................................................................................vii

List of Figures ............................................................................................................................... ix

List of Reaction Schemes ................................................................................................................. xii

Chapter 1: Introduction ................................................................................................................... 1

1.1. Objectives and Scope of Chemical Analysis .................................................................... 1

1.2. Mass Spectrometry Techniques for Non-target Analysis ................................................. 7

1.2.1. Ionization Techniques ............................................................................................. 8

1.2.2. Mass Analyzers ..................................................................................................... 10

1.2.3. Tandem Mass Spectrometry and Hybrid MS Instruments ................................... 13

1.3. Sample Preparation and Analytical Separations in Non-target Analysis ....................... 20

1.3.1. Sample Preparation for Non-target Analysis ........................................................ 21

1.3.2. Liquid Column Separations .................................................................................. 25

1.3.3. High-Field Asymmetric Waveform Ion Mobility Spectrometry ......................... 31

(a) Separation of Ions by Differential Mobility ................................................................. 32

(b) A Brief History of FAIMS-MS .............................................................................. 36

(c) FAIMS as a Primary Separation Tool ........................................................................ 39

1.4. Identification of Unknowns from Spectral Data ............................................................. 42

1.4.1. Known Unknowns: Spectral Matching and CID Reproducibility ........................ 43

1.4.2. Unknown Unknowns: De novo Structural Elucidation ......................................... 48

(a) Gas Phase Ion Chemistry For Unknown Identification ............................................. 51
Chapter 2: Identification of Abundant Species Detected in Urine by Non-Target Liquid Chromatography - Electrospray Ionization - Tandem Mass Spectrometry

Abstract

2.1. Introduction

2.2. Experimental

2.3. Results and Discussion
   2.3.1. Detection of Urinary Metabolites Using Data Dependent Acquisition
   2.3.2. Identification of Acylcarnitines in Urine
   2.3.3. Identification of Prolyl Diketopiperazines
   2.3.4. Other Small Endogenous and Dietary Metabolites
   2.3.5. Identification of Nucleic Acid Bases and Nucleosides

Chapter 3: Non-Target Analysis of Urine by Electrospray Ionization - High Field Asymmetric Waveform Ion Mobility - Tandem Mass Spectrometry

Abstract

3.1. Introduction

3.2. Experimental

3.3. Results and Discussion

3.4. Conclusions

Chapter 4: Gas Phase Ion Chemistry of Uracil During Collision Induced Dissociation

Abstract

4.1. Introduction

4.2. Experimental

4.3. Results and Discussion
   4.3.1. Ionization, Tautomerization and Collisional Activation of Uracil
   4.3.2. Ammonia Elimination from Protonated Uracil
4.3.3. Elimination of Water from Protonated Uracil .................................................... 155

4.3.4. Elimination of Isocyanic Acid from Protonated Uracil ................................. 159

4.4. Conclusion ............................................................................................................. 161

Chapter 5: Linear and Nonlinear Regimes of Electrospray Signal Response in Analysis of Urine by ESI-FAIMS-MS ............................................................................................. 164

Abstract ......................................................................................................................... 164

5.1. Introduction ............................................................................................................. 165

5.2. Experimental .......................................................................................................... 168

5.3. Results and Discussion ........................................................................................ 170

5.3.1. Signal to Noise in ESI-FAIMS-MS Using Extended Data Acquisition Time in MS................................................................................................................................. 170

5.3.2 Ionization Efficiency and Ionization Suppression in ESI................................. 173

5.3.4. Quantitative Non-target Analysis in the Linear and Suppression Regimes of ESI................................................................................................................................. 180

5.4. Conclusions .......................................................................................................... 186

Chapter 6: Final Conclusions ......................................................................................... 188

References ...................................................................................................................... 195
List of Tables

Table 1.1: Typical figures of merit of common mass analyzers........................................13

Table 1.2: Summarized analytical workflow used in the metabolomics of human tears.........17

Table 2.1: Acylcarnitine identified in urine by data dependent LC-ESI-MS/MS..............74 - 75

Table 2.2: Cyclic dipeptides identified in urine by non-target LC-ESI-MS/MS.................84

Table 2.3: Small endogenous and dietary metabolites detected in urine by non-target LC-ESI-MS/MS.........................................................................................................................90 - 91

Table 2.4: Nucleosides and nucleic acid bases identified in urine by non-target LC-ESI-MS/MS.................................................................................................................................102 - 103

Table 4.1: Interpretation of data from the CID of stable isotope labeled uracil..............147

Table 5.1: Relative sensitivities of urinary metabolite standards.....................................181
List of Figures

Figure 1.1. Targeted detection of products of nucleic acid oxidation, 8-oxoguanosine. ........... 4

Figure 1.2. The metabolomic approach to chemical analysis. ..................................................6

Figure 1.3. High resolution precursor ion isolation of individual components of fulvic acid by ESI-FTICR-MS.................................................................................................................. 20

Figure 1.4. Structure of polymeric OASIS stationary phases from Waters corp. ......................24

Figure 1.5. Base peak chromatogram from two dimensional liquid chromatography analysis of urine using reverse phase and normal phase chromatography with column switching.....................................................................................................................28

Figure 1.6. The workflow and representative data used in non-target identification of guanosine using capillary electrophoresis migration time modeling ......................................................31

Figure 1.7. Ion mobility for different ion types at high and low electric field strength.............33

Figure 1.8. Trajectories of an ion passing between two plates during application of an asymmetric waveform...............................................................................................................................35

Figure 1.9. Non-target analysis of drinking water by ESI-FAIMS-MS.....................................40

Figure 1.10. Schematic representation of single stage quadrupole collision induced dissociation (CID) of glutathione VS sequential CID..............................................................................................................46

Figure 1.11. Competing pathways of CID reactivity of acylglycines.......................................49

Figure 1.12. Collision induced dissociation (CID) spectra for distinct protomers of 4-aminobezoic acid separated by FAIMS prior to CID..........................................................................................52

Figure 1.13. Reactivity of N-dimethylnitrosoamine during collisional excitation prior to dissociation.....................................................................................................................................................53

Figure 1.14. Competing charge remote (A) and charge directed (B) processes in the dissociation of a deprotonated non-8-enoic acid.................................................................................................54

Figure 1.15. Dissociation of protonated aliphatic ethers by ion neutral complex.......................55

Figure 1.16. Electrospray ionization and tautomerization of uracil during collisional activation by CID ......................................................................................................................................................57

Figure 2.1. The most abundant components in urine observed by LC-ESI-MS..........................70

Figure 2.2. CID spectra of selected acylcarnitines from data dependent LC-ESI-MS/MS analysis of urine extracts............................................................................................................................................73
Figure 2.3. Structures of urinary cyclic dipeptides identified by data dependent LC-ESI-MS/MS analysis........................................................................................................................86

Figure 2.4. Product ion spectra of two urinary cyclic dipeptides identified by data dependant LC-ESI-MS/MS analysis...................................................................................................87

Figure 2.5. Collision induced dissociation spectra from non-target LC-ESI-MS/MS analysis of urine.................................................................................................................................93

Figure 2.6. Comparison of MS/MS spectra from data dependent acquisition (A, C) and targeted data acquisition (B, D) LC-ESI-MS/MS analysis methods........................................106

Figure 3.1. Additive effects of sequential optimization of MS signal intensity in ESI-FAIMS-MS of nucleoside standards..............................................................................................................119

Figure 3.2. Cross section of the ESI-FAIMS-MS interface.................................................................................................................121

Figure 3.3. Extracted ion CV spectra showing the separation of the standard mixture of nucleosides and modified bases at signal intensity optimized conditions................122

Figure 3.4. Urine sample analyzed by ESI-FAIMS-MS..................................................................................................................124

Figure 3.5. Identification of modified nucleosides and bases in urine by ESI-FAIMS-MS/MS.................................................................................................................................126

Figure 3.6. Comparison between spectral quality in LC-ESI-MS/MS and ESI-FAIMS-MS/MS.................................................................................................................................130

Figure 4.1. MS analysis of protonated uracil.................................................................................................................................145

Figure 4.2. Product-ion spectra of protonated uracil at m/z 113 obtained by quadrupole ion trap MS and by quadrupole-time-of-flight.................................................................146

Figure 4.3. Sequential dissociation of major products of protonated uracil in MS⁴ using QIT-MS.................................................................................................................................155

Figure 5.1. Linear relationship between signal and scan time using ESIFAIMS-MS.........................................................................170
Figure 5.2. Standard calibration of equimolar mixture of palmitoyl-L-carnitine, 2′-deoxyinosine, N2,N2-dimethylguanine, and alanine.................................................................174

Figure 5.3. Sequential dilution of 2′-deoxyinosine (dI) spiked urine by ESI-FAIMS-MS.......176

Figure 5.4. Matrix matched standard calibration curves of 2′-deoxyinosine.........................180

Figure 5.5. Quantification of urianryl-methyleytidine detected in the suppression regime.......184
List of Reaction Schemes

Scheme 2.1. Gas phase chemistry of protonated acylcarnitines during collision induced
dissociation...............................................................................................................78

Scheme 2.2. The impact of specific acyl chain modifications of the gas phase chemistry of
protonated acylcarnitines during collision induced dissociation.........................81

Scheme 2.3. The gas phase reactivity of protonated prolyl-diketopiperizines..............85

Scheme 2.4. The gas phase chemistry of protonated N1-methyl-2-pyridone-5-carboxamide
during collision induced dissociation........................................................................95

Scheme 2.5. Gas phase chemistry of protonated N, N’-diacetylputricine during collision induced
dissociation..................................................................................................................97

Scheme 2.6. Gas phase chemistry of protonated prolamide during collision induced
dissociation..................................................................................................................98

Scheme 3.1. The elimination of water from protonated 3-methylcytosine...............132

Scheme 3.2. Proposal for formation of product ions detected for 3-methylcytosine......134

Scheme 4.1. The protonated of uracil during positive electrospray and the tautomerization during
collision induced dissociation of protonated uracil................................................149

Scheme 4.2. Gas phase ion chemistry processes leading to the formation of m/z 96 product ion
by elimination of ammonia from protonated uracil....................................................152

Scheme 4.3. Dashed bond shorthand used to demonstrate the contributions of multiple resonance
structure to overall charge stabilization.........................................................................153

Scheme 4.4. Hierarchy of product ions of collision induced dissociation of protonated uracil
considering tautomerization prior to dissociation.......................................................156

Scheme 4.5. The elimination of water from protonated uracil.................................157

Scheme 4.6. The elimination of isocyanic acid and cyanic acid from protonated uracil........159
Chapter 1

Introduction

1.1. Objectives and Scope of Chemical Analysis

The objectives and capabilities of qualitative and quantitative chemical analysis are defined differently depending upon the type of chemical species and sample matrix being analyzed. Classical methods of chemical analysis typically deal with sample components in the percent concentration range. Modern instrumental analysis is routinely able to carry out detection of chemicals at the ultra-trace (below parts per million) level in complex sample matrixes. Instrumental analytical methods consist of several steps including: sample preparation, separation, detection and data processing. In trace analysis, each of these steps is designed to be as selective as possible but only specific for the pre-determined analyte(s). Analyte specificity is defined as the ability of a method to produce a signal for an analyte. Analyte selectivity is the ability of a method to resolve signals from different chemical species.

The detection and quantification of analytes present in complex samples at trace levels or lower is carried out using highly analyte specific techniques of chemical analysis which have been developed using authentic chemical standards and which rely heavily upon their availability. In order for a highly selective and analyte specific analytical method to be developed, prior knowledge of the analyte is always required. Whether this is a known or postulated chemical compounds, a chemical standard is first acquired which enables development of the analytical method. This method development consists of finding analytical techniques for each of the steps of the chemical analysis which are most selective and specific
for the analyte(s) of interest. Each of these steps is then optimized to further improve sensitivity, selectivity and specificity. The result of such an analytical method development, is a method capable of selective detection of the targeted analytes in a sample which also contains many, often much more abundant, species that are not detected by the method.

The development of such highly sensitive and selective analytical methods of trace analysis which are specific to pre-determined analytes has been enabled by the combination of liquid chromatography (LC) with tandem mass spectrometry (MS/MS) detection using electrospray ionization (ESI). These techniques are reviewed in detail in section 1.2 and 1.3 of this introduction chapter. An illustration of the power and limitations of such an approach can be found in the analysis of several products of nucleic acid damage in urine by LC-ESI-MS/MS. These analytes, 8-oxoguanine (8-oxoGua), 2,6-diamino-4-hydroxy-5-formamidopyrimidine (Fapy-Gua), 8-oxoguanosine (8-oxoGuo) and 8-oxo-2’-deoxyguanosine (8-oxodG) have been known as products of oxidative damage of DNA for many years, and serve as biomarkers of oxidative damage of DNA. The availability of powerful analytical methodology has allowed for their detection and quantification in urine, where they exist at low to sub nano-molar concentrations. In the development of such a method, chemical standards labeled with $^{15}$N and $^{13}$C stable isotopes were required for each analyte, allowing for its differentiation from naturally occurring isotopomers detected at different mass to charge ratio ($m/z$) by MS. A sample preparation method involving ultra-centrifugation and spiking with the labeled standards was then developed, along with an LC method allowing for separation of the analytes from many of the other chemical species in urine. The MS detection method used an MS/MS scan mode called multiple reaction monitoring (MRM) or single reaction monitoring. This highly analyte specific scan mode is widely used in trace chemical analysis, and represents the bulk of the selectivity of
the LC-ESI-MS/MS method. For an MRM method to be developed the retention time in LC, the 
m/z of the analyte, the MS conditions optimized for its detection, the m/z of the fragment ions 
which are produced upon its dissociation along with MS parameters for its dissociation must all 
be known. In order for these parameters to be establish and the method developed, analysis of an 
authentic standard of each analyte is required.

The data that can be acquired from a method like this consist of ion chromatograms 
where the analytical signal represents the detection of product ions with a known m/z which 
originate from fragmentation of precursor ions with a known m/z. Examples of such 
chromatograms are shown in Figure 1.1 for the detection of the four products of NA oxidation.¹ 
For analyte identification and quantification, these would be compared to the analogous ion 
chromatograms for the isotopically labeled standards spiked into the sample. This analytical 
method represents the highest sensitivity and selectivity that can be achieved in most cases, but 
no additional structural information about the analytes or the other components of the sample can 
be obtained by this method.

The above approach is termed targeted analysis and in a very general sense answers the 
questions “Are these pre-selected substances in this sample?” for qualitative analysis and “What 
are the concentrations of these substances in this sample?” for quantitative analysis. For all but a 
handful of specific analytical challenges yet to be addressed, targeted analysis is able to detect 
and quantify any analyte present at the ultra-trace level in a complex sample provided an 
authentic standard can be acquired.
Figure 1.1. Targeted detection of products of nucleic acid oxidation, 8-oxoguanosine (A), 8-oxo-2’-deoxyguanosine (B), 8-oxoguanine (C), 2,6-diamino-4-hydroxy-5-formamidopyrimidine (D), using LC-ESI-MS/MS with multiple reaction monitoring. Adapted from Malayappan 2007.¹

Much more recently, an alternative approach to chemical analysis with a different set of objectives and requirements has been developed which aims to selectively acquire analytical signals for a very large number of chemical species in a complex sample. This approach has been enabled by the development of high resolution MS instrumentation which have very high selectivity but are capable of being operated with specificity for a large number of analytes without prior knowledge of their presence in a sample. However, the targeted approach outlined above is not compatible with the analysis of very large numbers of analytes in a sample of unknown composition. In the targeted approach, analytes must be chosen prior to method
development so new or unexpected chemical species are not detected. Further, the requirement for chemical standards for each analyte is not compatible with very large numbers of analytes as availability of standards is limited. These limitations to targeted analysis have led a very different approach to chemical analysis to be developed for the analysis of metabolites in a biological samples called metabolomics. The field of metabolomics has created a demand for powerful analytical methods capable of selectively producing a signal for a large number of analytes from a complex biological sample, usually across multiple classes of analytes. Multivariate statistical analysis is then used to determine significant differences in these detected signals between sampling groups in order to detect biologically significant metabolites to a certain application.

The untargeted metabolomic analysis of human urine which was carried out to identify possible biomarkers of breast cancer demonstrates the differences between this approach and the targeted approach discussed above. In this study, urine samples are prepared and analyzed in a similar fashion, but without spiking of a labeled standard, and a similar LC separation is used. The most significant difference in analytical method lies in the MS data acquisition, which uses high resolution MS detection to acquire signals for all detected ions as they elute from LC. Figure 1.2 shows the total ion chromatogram (sum of all detected ion intensities) for a urine sample. The significance of each analytical signal was evaluated using multivariate statistical analysis and a small number of species, including one shown in Figure 1.2 at 2.9 min, were determined to be statistically significant and were targeted for identification using high resolution MS and MS/MS data.
The success of metabolomics experiments such as this relies heavily on the capability to identify unknown substances, but metabolomics methods themselves are not usually suitable for identification. In this respect metabolomics poses the question “Which of these many signals is likely to be important for this application?” as a way of more efficiently using resources on challenging unknown identification.

The chief limitation of analytical methods used in metabolomics lies in the identification of the unknown species detected. Modern analytical and MS instrumentation is beginning to enable another analytical approach called Non-target analysis which aims to incorporate the comprehensive detection of metabolomics methods with the best possible qualitative MS data for identification of unknowns. In the simplest terms, non-target analysis, which is the topic of this thesis, poses the questions “Which chemical species are present in this sample, and at what conditions?
concentration?”, which has traditionally not been a reasonable question to ask of instrumental chemical analysis, particularly in complex biological or environmental analysis. Structural elucidation of purified substances based on a combination of nuclear magnetic resonance spectroscopy, x-ray crystallography, mass spectrometry and other spectroscopic techniques is well established but in no way extends to substances present below trace (mg/kg) levels within complex samples.

There has been a trend in the recent chemical literature towards the goal of comprehensive non-target analysis using tandem mass spectrometry (MS/MS), although this is rarely discussed as a central goal in the field. This introduction reviews work in several sub-disciplines of analytical chemistry that are central to the advancement of non-target analysis. These studies come from diverse research areas including research on MS and separations instrumentation, gas phase ion chemistry, spectral databases and spectral prediction. The ideal analytical method for non-target analysis it would (1) allow detection of as broad a range of analytes as possible, (2) be highly selective, (3) provide high quality accurate mass MS and MS/MS data for structural elucidation, (4) allow for quantification of detected species. Additionally, approaches to structural elucidation from the MS/MS data acquired by the above method must be improved to increase the possibility, reliability and ease of unknown identification.

1.2. Mass Spectrometry Techniques for Non-target Analysis

All MS instrumentation requires a number of important elements including an ionization source to produce gas phase ions from sample components, one or more stages of mass analysis and some form of ion detection. These must be carefully chosen to provide instrumental figures of merit suitable for the goals of the analysis. Non-target analysis of complex samples has been
enabled by the development of sophisticated MS instrumentation without which identification of low abundance unknown analytes in complex samples would not be possible. Requirements for non-target MS analysis of complex biological samples include effective ionization of non-volatile, thermally labile bio-molecules, high or ultra-high resolution mass analysis to provide elemental composition of detected species and the capability for tandem MS experiments which provide information on structure.

1.2.1. Ionization Techniques

Possibly the most significant advancement since the inception of MS as an analytical tool has been the development of techniques capable of producing intact gas phase ions from non-volatile and thermally labile species in solution. In particular, electrospray ionization (ESI) has enabled the field of biological mass spectrometry by greatly expanding the size and type of molecules which can be analyzed by MS and allowing for online combination with liquid phase analytical separations. Prior to its development, trace MS analysis was largely limited to analytes which were volatile and thermally stable since electron ionization (EI) and chemical ionization (CI) require analytes to be in the gas phase prior to ionization.

Electrospray Ionization uses a high DC voltage applied to a needle tip through which a liquid sample is flowing to produce gas phase protonated or deprotonated ions from dissolved sample components. This occurs by a complex process of charge separation, ejection of charged droplets from the needle tip, droplet evaporation and disintegration, desolvation and de-clustering which ultimately yield bare gas phase ions which can be effectively sampled by mass spectrometry.\(^3\) Compared with EI, ESI is considered a soft ionization technique that results in little fragmentation of the ionized species enabling detection of intact ions for molecular weight determination. Unlike EI, ESI is highly suitable for the ionization of polar and nonvolatile
species which would degrade rather than enter the gas phase as neutral species upon heating, for example in a GC injector. However ESI has the limitation that analytes must be capable of bearing charge, usually through protonation (positive ESI) or deprotonation (negative ESI). With the exception of cases where only volatile analytes are of interest and GC-EI-MS is still the technique of choice, the relatively broad selectivity of ESI, and its ability to produce intact ions that can be further analyzed by collision induced dissociation (CID) make ESI the primary ionization technique of interest for trace analysis of biological and environmental samples.

Conventional ESI is operated at flow rates of hundreds of microliters per minute or greater when the spray plume is heated and high gas flows are used to aid in desolvation. Another effective approach to allow efficient conversion of analytes to gas phase ions is to significantly reduce the analytical flow rate directed to ESI along with the diameter of the ESI emitter, a technique referred to as nanospray. Nanospray usually operates at flow rates in the range of 300 – 800 nL min\(^{-1}\) driven either by specialized low-flow pumping, splitting of higher flow rates or by a combination of electrical and capillary forces without the need for additional pumping. Because the efficiency of ionization by ESI is related to the initial size of the charged droplets produced after ejection from the needle tip, the smaller droplets produced in nanospray lead to an overall increase in efficiency of gas phase ion production.\(^4\) This has the principal advantage of allowing for sensitive analysis of very small biological samples of limited availability but also shows reduction in ionization suppression and the formation of salt adducts of analytes when compared with conventional flow ESI.\(^5,6\)

Electrospray-MS is considered a poor technique for absolute quantitative analysis and is often characterized by nonlinear signal response as a result of ionization suppression or detector saturation. Ionization suppression arises from a limited ionization capacity of ESI when sample
matrix and buffer components compete with analytes for charge or surface area on the ESI droplet during ionization.\textsuperscript{3,7} Suppression can be observed, differentially depending on the properties of the analyte and the sample matrix, any time the limited ionization capacity of ESI is exceeded.\textsuperscript{8,9} Below this capacity, typically cited as $\sim 10^{-5} \text{ M}$ total concentration of ionizable species, all analytes exhibit linear response.\textsuperscript{7–10} The only method of true quantitative analysis by ESI-MS uses an isotopically labeled standard for each analyte. By this approach, analyte and standard show identical sensitivity regardless of matrix effects and linearity of the absolute signal is not required. For an analyte just identified by non-target analysis, authentic standards are usually unavailable and analytical sensitivity remains unknown until one can be acquired. Thus ionization suppression observed in complex sample analysis by ESI represents a limitation of the technique and a significant challenge in non-target analysis.

\subsection{1.2.2. Mass Analyzers}

The most obvious way to introduce selectivity into an MS analysis is to increase the mass resolving power of the mass analyzer. Typical quadrupole and ion trap mass analyzers exhibit unit resolution; they are able to resolve nominal $m/z$ from one another but not to differentiate ions with different elemental composition but the same nominal mass. More sophisticated instruments provide much higher mass resolutions, which can resolve ions differing in mass by less than the mass of an electron and can provide accurate masses suitable for determining the elemental composition of relatively large ions. These high and ultra-high resolution mass analyzers are critical tools enabling non-target analysis.

Two figures of merit are used to characterize the suitability of a MS technique for high resolution, accurate mass work. The first, MS resolving power is defined as the $m/z$ of the measured ion divided by the MS peak width in equation 1.
\[
R = (m/z) \times \Delta m/z^{-1}
\]  
(1)

Thus the higher the resolving power of an instrument, the higher the \( m/z \) at which it is able to nominally resolve two ions differing by a single \( m/z \) and the greater the separation between two ions of different composition at the same nominal \( m/z \). The second important figure of merit is the mass accuracy of the analysis which is calculated by comparing the measured accurate \( m/z \) of an ion with its theoretical calculated \( m/z \) and is presented as parts per million in equation 2.

\[
\text{mass accuracy} = (m_{\text{mes}} - m_{\text{theo}}) \times m_{\text{theo}}^{-1} \times 10^6
\]  
(2)

A highly accurate measurement of \( m/z \) is of high utility because of the defects in mass from integer values of all elements other than \(^{12}\text{C} \), which is defined as having a mass of exactly 12 Da, while \(^1\text{H} = 1.007826 \text{ Da}, \(^{16}\text{O} = 15.994910 \text{ Da} \) and the mass of an electron is 0.0005486 Da. This means that with sufficiently high mass accuracy the elemental composition of ions of low enough \( m/z \) can be unequivocally determined solely based on mass measurement. For larger ions the number of possible elemental compositions for an unknown can be greatly reduced.

The mass analyzer capable of the highest resolution and most accurate mass measurement available is the Fourier Transform - Ion Cyclotron Resonance (FTICR) Mass Spectrometer. This type of mass analyzer makes use of high magnetic fields to cause ions to orbit within an ion trap at a frequency related to their \( m/z \). Ion detection is carried out by passing packets of charged ions close to the outer electrodes of the ICR trap thus inducing an AC signal called the image current. Fourier transform is carried out on the signal to de-convolve the frequencies of oscillations which are converted to an \( m/z \) scale. FTICR instruments offer resolutions of up to 2 000 000 with mass accuracies of less than 0.1 ppm.\(^{11}\) One of the primary drawbacks limiting the widespread use of FT-ICR-MS is the high cost of the instruments and their upkeep.
Much more recently a new type of Fourier transform mass analyzer has been developed which has greatly increased access to ultra-high resolution MS. This technology, called Orbitrap, is based on orbital trapping of ions using electrical fields rather than magnetic fields used in FTICR, but uses a similar approach of obtaining $m/z$ from the image current of oscillating packets of ions. The use of electrical rather than magnetic fields significantly reduces the cost and footprint of the instrument compared to FT-ICR in exchange for a moderate decrease in MS power. The most recent generation of commercial Orbitrap instruments report mass accuracies of 100 000 at $m/z$ 400 when operating around a 1 sec detection time suitable for LC-MS detection with much higher resolution possible at longer acquisition times. With internal mass calibration, mass accuracies of 1-2 ppm are routinely reported for Orbitrap MS.

The chief drawback of both types of Fourier transform mass analyzers in trace analysis is their relatively slow data acquisition time. The sensitivity and resolution of these instruments is proportional to the length of time the image current is detected. While this may only be a few seconds, the peak widths in modern high resolution analytical separations reviewed in section 1.3.2 can also be on the order of a few seconds, limiting their compatibility with ultra-high resolution mass analysis.

The mass analyzer which possesses the highest spectral acquisition rate is time of flight (TOF) MS which collects spectra on the order of 20 kHz. Time of flight operates at a lower resolving power of 5000 to 20 000 and mass accuracy of 5 - 10 ppm. However, the high spectral acquisition rate of the TOF allow for its effective coupling with even the narrowest peak widths in chromatographic or electrophoretic separations. The figures of merit of these three high and ultra-high resolution mass analyzers along with those of common lower resolution instruments
(quadrupole and ion trap), and sophisticated hybrid instruments capable of tandem MS experiments (described in detail in section 1.2.3) are shown in table 1.1.

Table 1.1: Typical figures of merit of common mass analyzers.

<table>
<thead>
<tr>
<th>Mass Analyzer</th>
<th>Typical Resolution $^a$</th>
<th>Mass Accuracy $^b$</th>
<th>MS Cycle Time</th>
<th>MS/MS configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>quadrupole (Q) unit</td>
<td>nominal</td>
<td>ms</td>
<td>QQQ</td>
<td></td>
</tr>
<tr>
<td>quadrupole - ion trap (QIT)</td>
<td>nominal</td>
<td>ms</td>
<td>capable of MS$^n$</td>
<td></td>
</tr>
<tr>
<td>orthogonal time of flight</td>
<td>5 000 – 20 000</td>
<td>10 ppm</td>
<td>µs</td>
<td>QQTOF</td>
</tr>
<tr>
<td>Orbitrap</td>
<td>100 000</td>
<td>2 ppm</td>
<td>sec.</td>
<td>LIT - Orbitrap</td>
</tr>
<tr>
<td>ion cyclotron resonance</td>
<td>2 000 000</td>
<td>0.5 ppm</td>
<td>sec.</td>
<td>capable of MS/MS but usually hybrid</td>
</tr>
</tbody>
</table>

$^a$ as defined by equation 1
$^b$ as defined by equation 2

1.2.3. *Tandem Mass Spectrometry and Hybrid MS Instruments*

One of the central goals of non-target analysis is the identification of unknown species. Information on the elemental composition of detected ions is of critical importance to identification, but further manipulation of the intact ions is required in order to obtain structural information beyond their mass to charge ratio ($m/z$). This is done by applying additional energy to ions once they are in the gas phase which leads to their fragmentation, with the $m/z$ of the product ions being correlated to the structure of the precursor. The most common form of excitation energy applied in MS/MS of small ions is excitation by collisions with neutral gas molecules called collision induced dissociation (CID). In order to obtain this additional structural information without losing the $m/z$ information of the intact species, tandem mass spectrometry (MS/MS) is carried when an additional stage of mass selection is used prior to fragmentation. In an MS/MS experiment, an intact precursor ion formed by ESI is selected based on its $m/z$ by the
first stage of mass analysis and all ions of other $m/z$ are excluded. Energy is then applied to fragment the selected ions. After dissociation, a subsequent stage of mass analysis is used to measure the $m/z$ of each product ion formed from the selected precursor.

The MS/MS capabilities of several common mass analyzers are summarized in Table 1.1. Precursor ion selection, CID and product ion detection can be carried out ‘in-time’, by trapping a packet of ions using an ion trapping instrument and applying combinations of radio frequency (RF) and DC voltage to selectively trap, eliminate, excite and detect ions during each step of MS/MS. This is most often accomplished using a quadrupole ion trap or linear ion trap instrument, but it is also possible in the ICR cell of the FTICR. For mass analyzers not capable of multiple stages of MS/MS such as quadrupoles and TOF mass analyzers, hybrid configurations that use ion transmission optics to couple multiple mass analyzers allow for MS/MS experiments to be conducted ‘in space’. In the configuration commonly used for quadrupole and TOF instruments, a mass resolving quadrupole is used to select a precursor ion of interest, which is then transmitted to a quadrupole collision cell with an increased pressure of inert collision gas followed by either a quadrupole or a TOF mass analyzer to detect product ions. The robustness and high sensitivity of the triple quadrupole has made it one of the most prevalent tools in environmental and biological trace analysis. Combining a TOF mass analyzer with a mass selecting quadrupole and a collision cell yields a hybrid MS/MS instrument called the quadrupole time of flight (QTOF), which is highly suited to coupling with fast chromatographic separations. For FT-MS mass analyzers which are characterized by a relatively slow MS cycle time, hybrid configurations are almost exclusively used because of their added flexibility and robustness. Usually a linear ion trap is used to collect ions prior to high resolution MS analysis in the Orbitrap or ICR cell. Typical all MS/MS experiments are carried out in the LIT, including
fast unit resolution detection of precursor and product ions. Only when the ultra-high resolution capabilities are required are ions transmitted to the ICR or Orbitrap cell.

One of the principal challenges of complex sample analysis using EIS-MS(/MS) is interference between isobaric ions of different structure but which share a common nominal $m/z$. Whether these represent other analytes, matrix components or system contaminants, in order for an unknown to be identified or quantified it must first be detected selectively. In trace analysis one of the most common and powerful ways of increasing selectivity using MS/MS is called multiple reaction monitoring (MRM) or single reaction monitoring. In this MS/MS scan mode, carried out using a triple quadrupole (QqQ) type hybrid instrument, the analytical signal detected results only from detections of ions with a pre-determined product ion $m/z$ resulting from a pre-determined precursor ion $m/z$. In this way, MRM also introduces a high level of analyte specificity which can be used to detect ultra-trace level analytes in complex samples as in the example described in section 1.1. This high degree of specificity along with the fact that a knowledge of the dissociation of the analyte during CID is required in order to carry out the experiment makes MRM fundamentally unsuited for use in non-target analysis.

Using an MS/MS instrument with suitably fast detection, usually a QTOF, it is possible to simultaneously acquire both MS and MS/MS data for analytes in a sample in a single experiment. This scan mode, called data dependent acquisition (DDA) or information dependent acquisition, provides $m/z$ of all ions produced by ESI and MS/MS spectra for many of the most abundant species, including accurate mass data for all precursor and product ions. Of all MS scan modes, this represent the most compatible with the objectives of non-target analysis described above. Only recently has DDA begun to be used in the analysis of complex biological samples in order to allow for comprehensive detection of sample components with the qualitative data
required for identification of detected species. The work on the non-target analysis of urine presented in Chapter 2 of this thesis is one such example. Another example, published as the with the current work was being carried out, describes the use of DDA with QTOF MS to carry out human tear metabolomics using a very similar non-target instrumental method.\textsuperscript{14} The authors combined the statistical data handling of a global metabolomics experiment with non-target MS/MS data acquisition using DDA. This was carried out using a new generation of QTOF instruments capable of resolution of over 35 000 at $m/z$ 400 and mass accuracy of around 2 ppm. The DDA method consisted of an MS survey scan and 20 MS/MS scans with a total duty cycle of 1.2 seconds per LC chromatographic data point. Identification of unknowns was carried out primarily using freely available spectral databases (reviewed in section 1.4.1) with manual interpretation of spectra used only to verify identification and for complex cases of unclear ion origin such as isobaric species, adduct ions and false positives.\textsuperscript{14} Figure 1.2 shows the workflow used, which employed many of the exclusion criteria used to maximize the amount of MS scan time used for detection of new analytes of interest rather than additional spectral features of the metabolites already sampled such as isotopes, adducts and background ions. The authors also point out some of the challenges of non-target identification using LC-ESI-MS/MS and DDA. In particular, the identification of low abundance signals was obscured by interfering background signals and automatic exclusion of metabolite ions which did not match isotopic pattern criteria because of low abundance of M + 1 peaks. This raises an important point that automatic or even online data handling can often inadvertently exclude compounds that would otherwise be identifiable by manual identification.\textsuperscript{15} The principle limitation to the utility of the overall approach used in this previous work to comprehensive non-target analysis is that only previously known analytes could be identified. Analytes for which no spectral match could be found in
databases were detected but the workflow in Table 1.2 does not include their identification based on MS spectral data acquired. This is a significant analytical challenge discussed in section 1.4.2.

Table 1.2: Summarized analytical workflow used in the metabolomics of human tears.\textsuperscript{14}

<table>
<thead>
<tr>
<th>Step in Analytical Workflow</th>
<th>Method Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquid Chromatography</td>
<td>Reverse and normal phase</td>
</tr>
<tr>
<td>Electrospray Ionization (ESI)</td>
<td>Positive and negative ESI</td>
</tr>
<tr>
<td>Data Dependent Acquisition</td>
<td>Background subtraction, dynamic exclusion, charge state and isotope exclusion</td>
</tr>
<tr>
<td>Automatic Data Processing (Peak Mining)</td>
<td>Elemental formula calculation</td>
</tr>
<tr>
<td>Identification of ‘known unknowns’\textsuperscript{a}</td>
<td>Matching with database searching</td>
</tr>
</tbody>
</table>

\textsuperscript{a} as defined by Little et al.\textsuperscript{16} and reviewed in section 1.4 of this introduction.

Another instrumental parameter which is particularly important to non-target analysis but which is rarely discussed in the MS literature is the resolution of the mass selection window used to isolate precursor ions prior to dissociation. As discussed above, modern MS instruments are capable of impressive resolving powers which make unequivocally determining of the elemental composition of even fairly large species a reality. However, it is not common place to isolate precursor ions in MS/MS experiments with a similar high resolution to that used in MS detection. In the case of the QTOF, mass selection is carried out using a quadrupole which is usually operated at unit resolution or slightly better (up to 0.2 Da) selection window. In Orbitrap and most FT-ICR instruments mass selection is carried out by the linear ion trap or quadrupole preceding the ultra-high resolution mass analyzer. The use of more sophisticated ion exclusion waveforms in ion traps to provide higher resolution mass selection is possible and has been used
to isolate single isotopologues of multiply charged species. However, operating at this selection resolution requires very long analysis times of several seconds not compatible with chromatographic separation and is rarely utilized. Carrying out mass selection and CID within an ICR cell offers the highest resolution ion selection possible, approaching the detection resolution of the instrument. This approach also suffers from the limitation of long precursor isolation times, but has shown impressive capabilities when used for non-target analysis of complex samples.

The resolution of the precursor ion selection window is important because of the relatively high abundance of background species produced and detected by ESI, especially in complex samples. These background species, discussed in detail in Chapter 5 of this thesis, originate from sample and buffer components, system contaminants and the electrochemical products of all of these formed during ESI and result in detection of some signal at every single nominal m/z. These background species can interfere with mass measurements in MS mode, which can be almost completely mitigated by the use of ultra-high resolution MS, but they also lead to detection of product ion interference in MS/MS when background species of the same nominal m/z as a precursor of interest are introduced into the collision cell for dissociation. The presence of product ions not originating for a selected precursor of interest is rarely a problem in targeted analysis because they do not interfere with the detection of known product ions used for identification or quantification. However, product ion interference can significantly hinder the identification of unknown species based on MS/MS data regardless of whether database searching, spectral matching or de novo spectral interpretation are used for identification.

An excellent recent example of the concept of non-target analysis demonstrated the utility of ultra-high resolution precursor ion selection within an FTICR cell. The authors used
ESI-MS/MS to conduct non-target analysis on an extremely challenging and complex analyte mixture, a solution of fulvic acid extracted from surface water. Fulvic acid is a complex mixture of thousands of specific organic compounds which make up the most polar fraction of humic substances, an important class of soil and dissolved organic matter. Figure 1.3A shows the complexity of the fulvic acid mixture analyzed by ESI-MS and Figure 1.3B shows large number of ions differing in composition which were detected at the same nominal mass, $m/z$ 365. By carrying out ion selection in the ICR cell, the authors were able to selectively isolate ions with each accurate mass prior to collision induced dissociation as shown in Figures 1.3B and 1C. By this method, precursors differing in ionic formula by only 36 mDa (CH$_4$ vs O) could be selected for dissociation and unique insight into the structure of isolated fulvic acid ions was gained.
Figure 1.3. High resolution precursor ion isolation of individual components of fulvic acid (FA) by ESI-FTICR-MS. (A) full scan ESI-MS spectrum showing all ions detected from analysis of a FA solution. (B) All ions detected at the nominal $m/z$ 365 but differing in composition, (C) the selective isolation of each component using single shot ejection prior to CID dissociation. Adapted from Witt 2009

1.3. Sample Preparation and Analytical Separations in Non-target Analysis

Regardless of the resolution of an MS instrument, MS alone is never able to resolve structural isomers with identical elemental composition. For this purpose analytical separations such as chromatography or ion mobility spectrometry are required.
1.3.1. Sample Preparation for Non-target Analysis

The goal of sample preparation in an analytical chemistry workflow is to produce a sample suitable for analysis by a given method and to eliminate substances that could interfere with the analytical measurement. In complex sample analysis, analyte specific sample preparation is required if the analysis itself is not suitably selective to resolve all components of the mixture. However, modern analytical separations coupled to high resolution MS detection are highly selective techniques capable of resolving thousands of analytes. This greatly reduces the requirements for sample preparation techniques intended to increase selectivity when using such techniques. However, sample preparation is still required to make complex samples suitable for analysis, particularly when ESI is being used.

The biggest challenge of using ESI for the analysis of complex samples, with or without a liquid separation prior to ESI, is ionization suppression prevalent in ESI. Ionization suppression arises from a limited ionization capacity of ESI when sample matrix and buffer components compete with the analyte for charge or surface area on the ESI droplet during ionization. Suppression can be observed, differentially depending on the structure of the analyte and the sample matrix, any time the limited ionization capacity of ESI is exceeded. Ionization suppression affects the quality of quantitative measurements made using ESI by any method save isotope dilution with stable isotope labeled standards which is applicable only in targeted analysis. At its worst, ionization suppression leads to detection of only a small number of abundant suppressant ions and complete suppression of analyte signal, impacting even qualitative work. Thus, the chief aim of sample preparation in ESI-MS based non-target analysis is to reduce the concentration of matrix compounds which might hinder ESI ionization without eliminating components of the mixture which might be of interest. In metabolomics, where
analytes of interest span a broad range of polarities, sizes and functionalities, sample preparation is practically limited to removing particulate matter, desalting a sample and removing large proteins which, if of interest, would be treated separately. The requirements for a suitable sample preparation technique in the field of global metabolomics have recently been outlined succinctly and apply almost verbatim to the goal of non-target analysis.

“An ideal sample-preparation method for global metabolomics should (i) be as non-selective as possible to ensure adequate depth of metabolite coverage; (ii) be simple and fast to prevent metabolite loss and/or degradation during the preparation …; (iii) be reproducible…” – Vukovic 2012

In non-target analysis, where the goal is to obtain the most realistic view of the composition of a sample as possible using a given technique, sample dilution prior to analysis, when possible, is the most suitable technique to reduce the concentration of suppressant ions. This approach, often referred to as “dilute and shoot” is intended to reduce the total concentration of ESI ionizable species and to minimize ESI suppression. The chief advantage of this method is that it has little impact on the sample composition, and the dilution buffer is easily analyzed for background ion composition. This is widely used in analysis of complex samples using LC-ESI-MS\textsuperscript{26–28} and also in non-target studies using ESI-FAIMS-MS (described in section 1.3)\textsuperscript{29–31}. The principal disadvantage of dilution as a method of sample preparation is that it decreases the concentration of analytes in the sample, which for low abundance non-target analytes can result in dilution below the method detection limit. As discussed in Chapter 5 of this thesis, this limitation can be largely overcome by the use of FAIMS rather than liquid column separations in non-target analysis.

When some sample cleanup is required to enable effective ionization or separation of a sample, one of the most suitable techniques of sample preparation in non-target analysis is solid
phase extraction. For experiments aiming to cover as broad a range of species as possible in samples needing some sample preparation, reversed phase SPE chemistries with some polar functionality such as the Oasis phases offered by Waters corp. These phases are based on a copolymer between divinylbenzene and N-vinylpyrrolidone with the structure as in Figure 1.4. As a stationary phase, these polymers show reverse phase retention with additional retention of polar species due to the added polar functionality and have gained widespread popularity in fields of metabolomics and environmental analysis where retention of species of a broad range of polarities is desirable.\textsuperscript{32–34}

When non-target investigation is driven by knowledge of a class of compounds of interest, such as acylcarnitines or acylglycines examined in recent work, packing chemistries that limit specificity to that class such as mixed mode ion exchange columns are used to obtain a sample highly suitable for identification of new analogues.\textsuperscript{35,36} Selective extraction of urinary acylglycines was carried out using the strong anion exchange phase shown in Figure 1.4 which incorporates a tetraalkylammonium functionality.\textsuperscript{35} This allowed for the non-target identification of 65 urinary acylglycines when only 18 had been reported by targeted methods.\textsuperscript{35} A similar study of urinary acylcarnitines used the strong cation exchange phase which incorporates sulfonate functionality.\textsuperscript{36} Selective enrichment of cationic species by this method led to the non-target identification of over 350 acylcarnitines in urine.\textsuperscript{36}
Figure 1.4. Structure of polymeric OASIS stationary phases from Waters corp.

An alternative to SPE, which is an exhaustive extraction technique, is solid phase microextraction (SPME), an equilibrium extraction which has been shown to have high utility in metabolomics and environmental analysis. The technique uses a fiber coated with an extraction phase that can be used to sample analytes in solution or volatiles from headspace. Because only a relatively small proportion of analyte is removed from the sample by SPME, equilibria in complex samples are maintained and short lived intermediates can be effectively sampled, even in vivo. Recently, a large number of SPME extraction phases were evaluated for their suitability for global metabolomics and three coatings, (two proprietary chemistries and a phenylboronic acid phase) were found to be highly suitable for simultaneous extraction of hydrophobic and hydrophilic metabolites from solution. These phases were then used to sample...
metabolites from human plasma where up to 3000 spectral features could be detected using LC-ESI-Orbitrap-MS. In this case though, identification of non-target species was not possible due to the lack of MS/MS capability of the MS instrument used. One of the principal advantages of using SPME over other methods of sample preparation in non-target analysis is the fact that only a relatively small proportion of the total sample concentration is sampled. This was found to have the desirable effect of almost completely eliminating ionization suppression when compared to other methods of sample preparation in plasma metabolomics experiments. This has a very similar effect to sample dilution of allowing for absolute linearity of signal response using ESI for complex sample analysis which, as discussed in detail in Chapter 5 of this thesis is essential for quantitative measurements in non-target analysis.

1.3.2. Liquid Column Separations: Liquid Chromatography and Capillary Electrophoresis

Some form analytical separation is required to introduce the required selectivity for trace analysis in complex samples by MS, particularly in resolving isobars and isomers prior to MS/(MS) detection. For biological samples and many environmental samples where detection of analytes of a wide range of polarities is required, reverse phase liquid chromatography has remained the analytical separation of choice. In recent years, trends in reverse phase LC are towards the use of smaller (< 2 µm) stationary phase particles and narrow column diameter in order to provide an overall increase in resolution. These characteristics also lead to narrower peak widths and shorter retention times which are highly suitable for the high throughput analysis of complex samples.

In non-target analysis the total efficiency of the separation, or its plate count, is the most important parameter for enabling identification of detected unknowns while sample throughput is less of a concern since data analysis is currently the limiting step in reducing analysis time. Both
reducing particle size, as in UPLC, and extending column length increase the efficiency in LC. However, both also lead to an increase in pressure drop across the LC column, which is the limiting factor in many LC pumping systems and also leads to band broadening from an increase in parabolic flow profile. Since additional column length can be added at the cost of a linear increase in pressure drop but decrease in particle size results in an exponential increase in pressure, the most effective way of increasing separation efficiency when analysis time is not a factor is increasing column length.\textsuperscript{39,40} Practically, this high efficiency has been achieved by coupling multiple LC columns with moderate particle size to achieve very high efficiencies in HPLC.\textsuperscript{39,40} Other parameters which affect pressure drop can also be controlled to increase efficiency. In particular, increasing column temperature from 30°C to 80°C allowed for column length to be quadrupled and resolution to be doubled with the same pressure drop being maintained.\textsuperscript{40}

A large proportion of the unknown species that can be detected in non-target analysis of biological and environmental samples are small polar analytes which are poorly retained in reverse phase separations. By changing the selectivity of the mobile phase and the column stationary phase it is possible to achieve a better separation of highly polar analytes. This can be done by using hydrophobic interaction liquid chromatography (HILIC), a normal phase mode of separation employing polar functionally modified stationary phases and mobile phases with a high percentage of organic solvents. Groups of cellular metabolites which have been challenging to separate and detect by reverse phase LC but where HILIC has been used include sulfur containing endogenous metabolites from cellular extracts, 69 polar metabolites from \textit{Escherichia coli} extracts, metabolites of the raffinose family of oligosaccharides from plant cells and in the analysis of many other targeted metabolites.\textsuperscript{41} Hydrophilic interaction liquid chromatography has
also been used in urinary metabolomics, usually in combination with another complimentary separation technique such GC-MS or reverse phase LC-MS. These studies report HILIC is able to give better coverage of polar metabolites in urine but little detail is available on the structure or class of these additional compounds. In a recent report, a two dimensional liquid chromatography method with column switching combining HILIC and reverse phase columns has been developed and applied to the metabolomics analysis of urine from tumorous rats. In this work, samples were first injected onto a HILIC column where polar analytes were separated in a normal phase gradient between 90 % and 50 % acetonitrile before column switching to reverse phase gradient between 25 % and 95 % acetonitrile. Figure 1.5 shows a typical chromatogram for such an analysis and it is evident that most abundant analytes are being separated in normal phase. The result of the metabolomics comparison in this study found 17 potential biomarkers, 14 of which were identified by comparison with authentic standards and three of which remain unidentified along with all other detected species not targeted for analysis. This example further highlights the current limitations relating to metabolite identification in the field of metabolomics.
Another analytical separation which is suitable for coupling to ESI-MS/MS detection and for the analysis of complex samples is capillary electrophoresis (CE). This separation technique is based on the differential mobility of a charged species in an electric field as they travel through a conductive medium. Capillary electrophoresis has been used in numerous metabolomics studies and in the targeted analysis of chemical species in complex environmental samples. In these applications the chief advantage to using CE over reverse phase LC, like HILIC separations, is the enhanced ability to perform high resolution separations on highly polar species in aqueous solution. The chief drawback, particularly in non-target analysis and the identification of unknowns in complex samples, is the sensitivity of the technique, which is low due to the small sample size injected onto capillaries and sample dilution by the makeup buffer that is often needed for coupling CE to ESI for MS analysis. This can be overcome by the use of the CE separation to carry out online sample preparation and pre-concentration prior to analysis. Also important to consider from a non-target perspective is the fact that peak widths
eluting from CE are on the order of a few seconds, requiring fast MS detection, particularly when using DDA.

In urinary metabolomics using CE, two different sets of separation conditions are used to detect different known classes of analytes as either cations or anions. Amino acids, amines and nucleosides are effectively separated at low pH under one set of analytical conditions and have been identified using authentic standards. Under high pH conditions carboxylic acids, phosphorylated species and nucleotides can be analyzed using a different set of buffer conditions.

The unique advantage to CE as a separation tool in non-target analysis is that migration time in CE can be effectively modeled from the physicochemical properties of an ion such as molecular volume, pKₐ and intrinsic valence charge. These can be determined theoretically from chemical structure. This approach has been effectively used as part of an integrated approach for in the identification of unknown urinary metabolites. One group has developed a method of identifying unknown metabolites detected by CE-MS based on their predicted migration time and accurate mass alone. Using 375 standard metabolites a model was created that gave good correlation between predicted and calculated migrations times. This was then extrapolated to almost 3000 metabolites described in online databases including HMDB (reviewed in section 1.4.1), however all but 80 of these were peptides, representing a significant limitation of the study for non-peptide species. The authors report an increase in the number of metabolites identified in a urine sample from 179 when accurate mass matching from metabolite databases was used alone to 240 when an additional filter of predicted migration time was used. However, no additional criteria for structural confirmation was applied. This study is another good
example of the bias in species identified in metabolomic studies towards those metabolites already known and present in chemical and spectral databases.

Another study used a more rigorous, integrated approach to unknown identification which included a similar approach to CE migration time modeling of database metabolites. The steps to their integrated approach to unknown identification are shown in Figure 1.6, and include (A) MS\textsuperscript{n} characterization to determine the analyte class of an unknown, (B) online hydrogen/deuterium exchange to determine the number of exchangeable protons in the structure, (C) metabolite database search using all available data from (A) and (B) to narrow down possible candidates. These candidates, often isomeric species, were then positively identified by comparing their observed electrophoretic mobility to the theoretical electrophoretic mobility calculated based on their chemical structure (D in Figure 1.6). The ability to carry out hydrogen deuterium exchange experiments online proved useful in narrowing down the possible elemental compositions at the nominal m/z of the detected unknowns and partially compensated for the relatively low mass accuracy of the ITMS instrument used.
1.3.3. High-Field Asymmetric Waveform Ion Mobility Spectrometry

High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) (also known as differential mobility spectrometry or differential ion mobility spectrometry), is an atmospheric pressure gas phase ion separation technique that acts as an ion filter between an ionization source and a detector. As a separation technique in MS analysis, FAIMS serves as an effective means of separating ions produced by ESI prior to MS detection, either as a primary separation (ESI-FAIMS-MS) or in a multidimensional LC-ESI-FAIMS-MS configuration. While the separation mechanism in CE, described above, is based on ion mobility in solution under an applied electric field, the FAIMS separation is based on ion mobility in the gas phase.

Figure 1.6. The workflow and representative data used in non-target identification of guanosine using capillary electrophoresis migration time modeling. Adapted from Lee et al. 2007
(a) **Separation of Ions by their Differential Mobility**

The separating power of FAIMS lies in the differential mobility of gas phase ions subjected to high and low electric fields. In linear or drift tube ion mobility spectrometry (IMS), ions are passed through a region of uniform and constant electric field (E), usually on the order of 200 V m\(^{-1}\), at pressure close to 1 atm. Ions are gated into a drift tube and are separated based on differences in their drift velocity. For comparison, this experiment can be considered analogous to the time of flight (TOF) separation in mass spectrometry with the important differences of high pressure and constant electric field in IMS rather than vacuum and field free drift region in TOF. The mobility \(K\) of an ion is defined as its velocity per field strength in units of m\(^2\) (V \times s\(^{-1}\)). In this low field regime, shown in Figure 1.7, \(K\) of an ion is constant with respect to E. That is, increasing E increases velocity of an ion linearly in this low field region. This allows for the calculation of the collisional cross sections of ions and detailed study of the conformation of large biomolecules.\(^{49}\)

Differential mobility separations exploit the significant change in behavior of ions in an electrical field that is observed at high field strengths during ion mobility separations. While \(K\) at low field strength is independent of E, \(K\) at high field strength \((K_h)\) becomes highly dependent on E with the magnitude and even the sign of \(K\) varying depending on the differential mobility behavior of the analyzed ions, as shown in Figure 1.7. This difference between \(K\) at low and high field strength or differential mobility \((K_h/K)\) is the parameter by which ions are separated in FAIMS. The differential mobility of an ion is dependent on its size and shape as in IMS, but it also depends on changes induced by high field strength. These are a combination of not yet completely understood effects including changes in polarizability of ions, differential clustering with solvent molecules and interactions with neutral buffer gas.\(^{50}\) These factors depend heavily
on the structure of ions and lead to differences in differential mobility. Figure 1.8 shows three common types of high field mobility behaviors used to classify ion behaviours in FAIMS.\textsuperscript{51} Type A ions show an increased mobility at high fields, type C ions show a decrease in mobility at increased electric field and the behavior of type B ions changes depending on the magnitude of the electrical field. These designations become significant when determining the set of FAIMS conditions used to selectively transmit certain types of ions.

![Diagram of ion mobility and electric field strength](image)

**Figure 1.7.** Ion mobility for different ion types at high and low electric field strength. Adapted from Purves et. al. \textsuperscript{51}

Practically, the differential mobility of ions can be exploited for analytical separation of gas phase ions using two electrodes as an ion filter which transmits only one mobility type of ion at a time. In this sense, where IMS was analogous to TOF-MS, FAIMS can be thought of as being analogous to a quadrupole mass filter which filters ions by $m/z$, transmitting only one $m/z$ ion at a time. A simplified cross section of a FAIMS device is shown in Figure 1.7 as two...
parallel plate electrodes with a stream of neutral buffer gas running between them and gas phase ions being carried along in the gas flow. A high field strength asymmetric waveform, with $V$ around 4000 V and $t$ around 2 $\mu$sec is applied to one of the plates which induces ion motion in the plane perpendicular to the flow of gas. The two components of the asymmetric waveform are characterized as having an equal $V \times t$ product, with one component composed of a relatively short time portion of high field strength and the other a comparatively long time portion of low electric field strength. During each of these stages, ions experience a displacement towards one or the other of the electrodes depending on their charge. Then during the subsequent stage of opposite polarity ions are drawn back in the opposite direction. In the hypothetical case where an ion exhibits identical mobility at high and low field strength, shown as trajectory D in Figure 1.8, there will be no net displacement of an ion after each period of the waveform. However, as shown in Figure 1.7, the high and low field mobility of ions differs significantly and hence a net displacement will be observed after each period which can ultimately result in an ion striking an electrode and being eliminated. Figure 1.8 shows a simplified square version of a waveform with a positive high field portion which results in attraction of the positive ion shown during the low field negative portion and repulsion during the positive high field portion. By the ion naming conventions used in Figure 1.7, ions which experience a net displacement away from the upper electrode under this set of ion and waveform polarities are called type A ions (shown as trajectory A in Figure 1.8) and those that experience a net mobility towards the upper electrode are of type C (shown as trajectory C in Figure 1.8). Thus, a mixture of ions introduced between the FAIMS electrodes are separated in space in the plane perpendicular to their flow in the stream of buffer gas based on their differential mobility behaviors with only ions of a selected differential mobility behavior passing through the filter.
The net displacement of ions depending on their differential mobility can be exploited in an analytical separation by using the FAIMS electrodes as a filter to selectively transmit one mobility type of ion at a time. Practically, this is accomplished by applying a relatively small (<100 V) direct current (DC) offset to an electrode in addition to the asymmetric waveform described above. This voltage, called compensation voltage (CV), has the net effect of adjusting the trajectories of all ions between the electrodes so that ions of a selected differential mobility are selectively transmitted while others strike the electrodes. For example, in Figure 1.8, ions initially traveling along trajectory A could be adjusted so that they travelled along trajectory D and strike the detector by applying a negative CV to the top electrode. Similarly, ions of
trajectory C could be detected by applying a positive CV to the top electrode. Thus, at a fixed CV, a continuous stream of selected ion can be transmitted through the device. Alternatively, to detect ions of a range of differential mobilities in sequence, CV can be scanned over a set of voltages that span a range of differential motilities of interest. This produces a CV spectrum which shows the ion current transmitted at each CV. It is also possible to hold CV at a fixed value to transmit a continuous flow of an analyte of interest. This represents a significant difference from the separations in elution chromatography where retention time represents the separation parameter.

(b) A Brief History of FAIMS-MS

On the time scale of most modern analytical chemistry techniques, differential ion mobility spectrometry is still a relatively young technique. Early on during the commercialization of the FAIMS device as a separation tool in analytical mass spectrometry, great promise was shown in the fields of environmental and biological analysis, but even after its commercialization, adoption by analytical chemists working in these areas has been relatively slow. This thesis describes work carried out during a period when FAIMS research appeared to have stagnated. However, its publication comes at a time of renewed interest and research activity in the field making it worthwhile to consider the context of some of the technical and historical events over the history of this technique.

Compared to linear ion mobility spectrometry which has been known since the 1960’s, FAIMS as an analytical technique is comparatively young with the first devices emerging from the former USSR in the 1980s. This early FAIMS device was designed in order to offer portable explosives detection for military and security applications and many important advances were made in the development of the technique, including the coupling of FAIMS to MS. However,
due to the secrecy surrounding military research and development during the cold war, this work did not appear in the scientific literature until later.

The first English-language paper to discuss FAIMS was by the Mine Safety Appliances Company which used the technology to construct a portable air quality analyzer. While the earliest FAIMS devices employed a flat plate design similar to Figure 1.8, this first commercial device employed two concentric cylinders as electrodes. This portable device consisted of either a $^{63}\text{Ni}$ or ultraviolet ionization source and ion detection was carried out using an electrometer. While these are ideal for a portable device in question and offered suitable selectivity for the desired application, the true power of FAIMS as an analytical separation resulted from its coupling to MS for ion detection.

It was not until the Mine Safety Appliances device came to the National Research Council of Canada (NRC) in the mid-1990s that the development of FAIMS as a separation tool in analytical MS began. Here the device was first coupled to ESI, enabling the analysis of biomolecule and a new prototype using hemispherical domed electrodes was developed. This design showed enhanced sensitivity due to ion focusing in the domed portion of the analyzer and was adopted for commercialization by an NRC spinoff company called Ionalytics. Their device, called the Selectra, was first developed as a prototype and then as a first generation of commercial instrument sold with different source interface kits depending on the manufacturer of the mass spectrometer to which it would be coupled.

The utility of the FAIMS device as a method of chemical analysis in challenging fields such as the analysis of complex environmental and biological samples was successfully demonstrated by the NRC team and their collaborators using prototypes of the FAIMS instrument. Unique applications of FAIMS explored included separations of difficult
to resolve isomers\textsuperscript{59} or enantiomers\textsuperscript{31} as well as analysis of complex environmental samples\textsuperscript{30,31} and mixtures of biolomolecules\textsuperscript{57,60-62}. One of the most important contributions of the NRC group was the improved engineering of the high field waveform generator, prototypes of which could be considered dangerous and prone to blown operational amplifiers and discharge. The Selectra prototype waveform generator was a comparatively robust device that only needed to be calibrated upon installation and was heavily shielded for safety and electronically protected from discharge.

Despite the potential of this technology that was demonstrated with the prototypes, there was a relative lack of success with this first generation of commercial instruments. A detailed review of the literature carried out for this thesis yielded only a few publications where the first generation commercial device was successfully applied.\textsuperscript{52,61,63} Also present in the literature are examples of work which use the name Selectra but experimental details of the described method reveal it was actually the prototype that was used to collect the experimental data.\textsuperscript{64,65}

Since 2005, the FAIMS device has been re-designed by Thermo-Fisher for use with their ion trap and now Orbitrap mass spectrometers. The availability of these second generation commercial instruments has resulted in a resurgence of FAIMS research over the last couple of years.\textsuperscript{66,67} Even more recently, differential mobility devices have been commercialized or licensed by a number of other major mass spectrometry manufacturers including AB Sciex (Selexion),\textsuperscript{68} Agilent (the Owlstone Nanotech \textmu FAIMS device)\textsuperscript{69,70} and Bruker (prototype developed by Gary Glish at UNC, but not yet available)\textsuperscript{71}. This new wave of FAIMS research has focused largely on exploiting the added selectivity of a FAIMS for improving detection limits by reducing chemical interference in LC-ESI-FAIMS-MS experiments, particularly in highly targeted MRM methods (described in sections 1.1 and 1.2.3). This approach, using low
resolution MS detection, aims to sensitively and selectively detect a relatively small number of target analytes in complex mixtures. However, when high resolution MS is used, FAIMS can serve as the primary separation method, eliminating the need for liquid separations and simplifying sample preparation.\textsuperscript{22,30,31,72}

\textbf{(c) FAIMS as a Primary Separation Tool}

When high resolution MS is used for detection, FAIMS can be used as a primary separation tool in combination with continuous sample infusion into ESI. In targeted LC-ESI-FAIMS-MS experiments it can be beneficial to focus analytes to a narrow CV, separate from interfering species. In the direct ESI-FAIMS-MS approach analytes are separated over as wide a range of CV values as possible with the goal of using FAIMS as a primary separation tool. In this approach, the high peak capacity of high- or ultra-resolution MS is essential because the overall peak capacity of FAIMS is relatively low (20-200) compared to other analytical separations (many hundreds).

The application of ESI-FAIMS-MS/MS to the non-target analysis of disinfection byproducts in chlorinated drinking water is an excellent example of the capabilities of this new analytical separation and of the non-target analytical approach in general. Drinking water represents an extremely complex mixture of over 50,000 chemical species that result from natural dissolved organic matter, anthropogenic contaminants and the products of chemical reactions of these with water disinfectants such as chlorination agents and UV light.\textsuperscript{73,74} Of this large number of species, many halogenated disinfection byproducts are known to be harmful and a few are regulated as known human carcinogens. Challenges in drinking water analysis are the large number of species detected, many of which have never been identified, the wide range of
analyze concentrations and polarities as well as the complex matrix, often high in ionic strength. These challenges are largely the same for analysis of complex biological samples.

![Figure 1.9.](image)

**Figure 1.9.** Non-target analysis of drinking water by ESI-FAIMS-MS showing (A) the total ion compensation voltage spectrum showing the sum of all ion intensities transmitted at each CV and (B) all ions transmitted through FAIMS at compensation voltage of 11 V. Adapted from Sultan and Gabryelski 2006.31

The huge numbers of unknown species detected by ESI-MS analysis of drinking water was evident from data acquired during targeted studies of DBPs by ESI-FAIMS-MS.76 This shows multiple species present at each nominal mass in the low mass range but no established analytical method had the capabilities to identify these small polar analytes present at a wide range of concentrations. However, since these unknown analytes represented relatively small
species, they represented good candidates for identification based on MS/MS spectral data. This was accomplished by using ESI-FAIMS in combination with high resolution MS/MS analysis and de novo interpretation of MS/MS data. Of particular significance was that the only sample preparation required for this analytical method was sample dilution using methanol. This allowed for all species in the sample to be detected without the need for analyte specific sample preparation. By detecting ions over the entire range of CVs of a FAIMS separation as shown in Figure 1.9A, the authors showed the wide range of chemical species detectable in drinking water, from dissolved natural organic matter to inorganic ions. Also evident from Figure 1.9A is the wide range of ion intensities detectable in drinking water including abundant ions such as glycolic acid, which had not been previously detected as a contaminant in drinking water but was identified at part-per-million concentrations. At much lower abundance was a huge diversity of small organic species like those detected at CV = 11V in the MS spectrum in Figure 1.9B.

The spectrum taken at CV = 11V in Figure 1.9B represents a mixture of small highly polar carboxylic acid containing species which were detected at ppt concentrations. By keeping CV fixed at a single value of 11 V for extended periods of time, a continuous flow of these species isolated from the other abundant components detected in Figure 1.9A was detected by QTOF MS/MS. Identification was carried out by manual interpretation of MS/MS spectral data which gave accurate mass measurements for all precursor and product ions. From this data it was possible to propose dissociation pathways based on principles of gas phase ion chemistry that are described in section 1.4.2. This de novo structural elucidation showed that in addition to the known DBP of concern trichloroacetic acid, drinking water contains a large number of other highly polar carboxylic acids, many of them also halogenated. Possibly the most significant finding of this non-target analysis of disinfection by-products was the identification of the most
abundant species detected, glycolic acid shown at CV = 52 V in Figure 1.9. Several decades of previous research and thousands of reports on drinking water using target analysis have failed to identify this abundant species which was readily detected and identified by ESI-FAIMS-MS/MS.

1.4. Identification of Unknowns from MS Spectral Data

The task of identifying an unknown chemical species is invariably challenging. When that compound has been enriched and purified, a combination of nuclear magnetic resonance spectroscopy, X-ray crystallography and MS can be used for unequivocal structural identification. This is rarely possible or practical when an unidentified compound is an ultra-trace component of a complex mixture. Here the powerful MS and separation tools described in sections 1.2 and 1.3 of this introduction must be relied upon as primary identification tools.

A novel parallel between the identification of species in chemical analysis and anti-terror intelligence in the United States has recently been made with the help of a quote from former United States Secretary of Defense Donald Rumsfeld:

“There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say, we know there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know.” – United States Secretary of Defense, Donald Rumsfeld

This speech, which was initially panned for being an example of poor use of plain language by politicians, has also been praised for being a succinct description of a complex concept relating to uncertainty. Surprisingly, the distinction made by this quote has relevant implication when applied to non-target analysis and unknown identification where different types of unknown analytes each have a distinct set of challenges and tools for identification. Known knowns can be considered as those species for which a targeted analytical method has or
could be developed based on pre-existing knowledge of their presence in a sample. Known unknowns were defined as compounds which are known in the chemical literature but which have not yet been identified in a sample. Unknown unknowns were defined as compounds which are not known in the chemical literature. The cited work used large chemical databases to match unknown accurate mass data as the first step to identification. Defining known unknowns as being “compounds which are known in the chemical literature” was appropriate in this context of the original article. In the context of this thesis and other work where MS/MS data is being used for primary identification it is more appropriate to define known unknowns as being “compounds for which MS/MS spectral data is available.” Thus, known unknowns are those compounds which can be identified by spectral database searching, spectral matching and from data published in the literature while unknown unknowns require de novo identification based on interpretation of MS spectral data and a knowledge of gas phase ion chemistry. Each of these tasks poses a unique challenge and both are equally important in non-target analysis.

1.4.1. Known Unknowns: Spectral Matching and CID Reproducibility

The original work to propose the distinction between known unknowns and unknown unknowns aimed to demonstrate the capabilities of non-target analysis by LC-MS first using accurate mass data to conduct database search in large chemical databases that do not contain tandem mass spectrometry data, namely the Chemical Abstract Service Registry and Scifinder. In the context of an industrial chemical manufacturer who published this work, it was found that often the results of accurate mass searching could be narrowed down by the number of references for that compound and a “knowledge of the sample history” to give identification of an unknown species. It was also shown how candidate structures determined in this manner could then be confirmed or excluded based on MS/MS data. This was not demonstrated for
complex biological samples, but represents an important part the integrated strategy required for identification of unknowns from complex biological and environmental samples.

Another important tool for identification of known unknowns is searching of databases of CID spectra of standard compounds. When spectral data for a detected compound can be found in such a database, this usually represents the fastest, simple solution to unknown identification. It is now routine for instrument manufacturers to provide some sort of spectral databases with ESI-MS/MS instruments. There are also free spectral databases which have begun to appear online in recent years. Massbank, which hosts both an online database and a free standalone program has grown significantly since its inception in 2008 and now contains nearly 20,000 MS/MS spectra for many hundreds of different compounds. One of the principal strengths of this database is that in addition to presenting MS/MS data from a number of different instruments at a variety of collision energies, it also provides merged spectra from a range of collision energies. This enables database searching using spectra collected at different collisional energies than reference spectra, as discussed below. Also freely available online are large databases of information compiled about known metabolites, the Human Metabolome Database and Metlin. These databases compile information about the origin, concentrations, significance and identification of known metabolites as well as MS/MS spectral data for a small proportion of their entries. While not as easily searchable for matches to experimental MS/MS data as other stand-alone databases, the MS/MS entries in the HMDB are highly relevant for identification of non-target analytes in biological samples as they represent well characterized component for which chemical standards are readily available. The proprietary NIST spectral database (Scientific Instrument Services, Ringoes, NJ), which previously only contained EI spectra from
GC-MS experiments now also contains CID spectra for nearly 8000 chemicals collected using a variety of ESI-MS techniques and instrumentation.

Although the utility of these spectral databases in unknown identification is constantly increasing, there remain significant challenges which have thus far prevented database searching of CID spectra from becoming the routine solution to unknown identification that database searching of EI spectra has become for identification of species by CG-MS. The primary reason that EI spectra are so amenable to database searching is the constant energy of 70 eV at which compounds are ionized in EI, regardless of instrument design or manufacturer. This choice of ionization energy came about naturally as the wavelength of electrons traveling at 70 eV is roughly equivalent to the length of a chemical bond in an organic molecule, and energy is very effectively transferred to the molecule for ionization. In CID no such universal choice of collision energy exists, with collision energies required to induce dissociation differing widely based on the stability of an ion. Further, large differences can be observed between ion trapping instruments capable of sequential, multi-stage MS/MS (MS\textsuperscript{n}) experiments and radio frequency (rf) only quadrupole collision cells in hybrid instrument configurations such as triple quadrupole (QqQ) and quadrupole time of flight (QqTOF) instruments. Simply, fragmentation of a precursor in MS\textsuperscript{n} is carried out over multiple sequential stages of CID, as demonstrated graphically in Figure 1.10A while in single stage quadrupole MS/MS all product ions are collected concurrently as demonstrated in Figure 1.10B. Further, these two distinct types of instruments offer different chemical environments in which to carry out CID with ions in ion trap MS/MS undergoing a relatively large number of relatively low energy collisions during excitation compared with ions in quadrupole collision cells. First generation MS/MS spectra collected with ion trap instruments capable of MS\textsuperscript{n} usually show significant differences to
MS/MS spectra acquired for the same compounds by single stage CID. However, combining data from all stages of sequential MS^n dissociation of a compound can produce composite spectra that are often similar to those from single stage quadrupole CID as shown in Figure 1.10 for glutathione. The hierarchy of product and precursor ions from the multi-stage experiment, represented in Figure 1.10B contains significantly more information about reactivity than does the single stage experiment, but is only conducted at low resolution. Data represented in such fragmentation trees is being used with new search algorithms to improve database searching of spectra acquired using any type of CID instrument as well as to determine elemental composition in the absence of accurate mass data. Since the relative intensities of product ions in MS/MS spectra from different instruments can be highly variable, database search algorithms that do not rely on relative intensity are required unlike dot-product algorithms used for EI-MS spectral searching which weigh relative abundance heavily.
Another approach to obtaining MS/MS spectra suitable for database searching of CID spectra acquired with any type of MS instrument uses powerful new hybrid MS instrument which combines a linear ion trap and a quadrupole collision cell with an ultra-high resolution Orbitrap mass analyzer. In addition to proteomics applications for which this sophisticated instrument was developed, it has also been suggested that it has the capability to produce hybrid MS/MS spectra that might be most suitable for building a universal spectral database. This could be done by accumulating ions from quadrupole CID and sequential ion trap CID prior to accurate mass measurement of all product ions by Orbitrap to produce a comprehensive, searchable CID spectrum.
1.4.2. Unknown Unknowns: De novo Structural Elucidation

*De novo* identification of unknown compounds detected by MS/MS aims to correlate observed spectral features in MS/MS to the structures of investigated species based on established reactivity for small organic ions during CID. Interpretation of the reactivity of radical cation dissociation in EI has been developed to the point where spectral interpretation based on simple reaction mechanism is now sophomore lecture material.\(^{86}\) In general, the same is not yet true for even electron ions dissociating during CID, although a great deal of effort is expended on interpretation of such spectra across the wide range of disciplines which make use of MS/MS for identification of unknown species.

One of the most feasible approaches to non-target identification is to focus on the identification of new members of a class of compounds by first analyzing those standard compounds of that class which are available in order to establish their reactivity during CID. Such a study of reactivity can be carried out to different depths with varying utility for its ultimate goal of identification of new species. The simplest way of doing this is by comparing fragmentation patterns of analogous species. Modified analogues are expected to have analogous fragmentation patterns with some neutral losses and some product ions showing \(m/z\) shifted by the mass of the modifying moiety, depending on its location within the structure. Unfortunately for the approach of simple spectral matching, structural analogues with only minor differences in structure can show very different reactivity.\(^{35,87}\) A much more rigorous approach is to determine the impact of structural modification on reactivity in terms of chemical principles of resonance and inductive stabilization or steric effects such as the availability of protons for abstraction.

Recently, a fragmentation study of an important class of fatty acid metabolites called acylglycines was carried out which enabled the identification of many more structural analogues
of this class of metabolites than had previously been detected by targeted methods.\textsuperscript{35} This work effectively demonstrates the concept of structural modification altering CID reactivity. The main purpose of this work was to detect as many acylglycines as possible from urine samples, so sample preparation that selectively extracted acylglycines from the complex urine matrix using solid phase extraction with a mixed mode ion exchange phase (described in section 1.3.1) was developed. For this class of compounds, 17 authentic standards could be acquired for use in analytical method development and a fragmentation study. In this type of work which targets an entire class of compounds with the goal of identifying new species, a fragmentation study serves a dual purpose of providing the means to identify newly detected species and also allowing for the optimization of MS and MS/MS parameters for detection of this class of compounds. Of particular importance to the reactivity of acylglycines were the competing pathways for formation of protonated glycine at \( m/z \, 76 \), considered a diagnostic product ion for acylglycines, and the loss of neutral glycine (- 75 Da) shown in Figure 1.11. Most acylglycines have a proton available for abstraction on the acyl chain at the \( \alpha \) position (as shown in Figure 1.11.). Transfer of this proton to the glycine moiety leads to elimination of the ketene of the acyl group and formation of an abundant protonated glycine ion as in Figure 1.11A. This pathway is almost completely suppressed for ions without such a proton available for abstraction, with the alternative reaction of neutral glycine elimination as in Figure 1.11B dominating. The authors point out correctly that targeted methods developed for detection of aliphatic acylglycines using predictive MRM method with \([M+H]^+ \rightarrow m/z \, 76\) would not be suitable for the detection of several structurally modified acylglycines detected by their non-target method.\textsuperscript{35} In total, their non-target experiments developed based on a fragmentation study of available standards were
able to detect 65 acylglycines in real urine samples. Of these, only 18 had been previously reported in urine by highly targeted methods.\textsuperscript{35}

![Diagram of acylglycines]

**Figure 1.11.** Competing pathways of CID reactivity of acylglycines. Adapted from Lewis-Stanislaus and Li, 2009\textsuperscript{35}

Recently a broader approach has been applied to the generalization of reactivity of small ions during CID by examining over 1000 database CID spectra belonging to 30 functional classes of compound.\textsuperscript{88} For each functional group (such as amides, thiocarbamates, ketones and phosphates) typical cleavages around that functional center for ions bearing that functional group were identified and examples of compounds for which that pathway represents a major dissociation pathway are given. However, the authors do not offer a link between the structure and reactivity within these classes, something that would greatly increase the utility of this type of work. Upon evaluation of the structures of the example compounds for each cleavage in each class it can be seen that a chemical rationale exists for each reaction being favored in a given compound. These include the availability of nucleophilic sites and protons for transfer (as in the
case of acylglycines in Figure 1.10), stabilization of charge by resonance in a conjugated system of $\pi$ electron density and the potential for structural rearrangement or tautomerization during collisional excitation. An understanding of these concepts is critical to the manual interpretation of CID data and they should be incorporated into any generalization of reactivity.

(a) **Gas Phase Ion Chemistry for Unknown Identification**

When examining the spectrum of a possible analogue to a known compound in non-target analysis it is important to have criteria with which to evaluate the feasibility of a detected product ion which does not fit the fragmentation pattern of any known analytes. Just as simple generalizations exist and are used to interpret EI-MS spectra, a similar approach is useful for interpretation of CID spectra. Unlike for EI spectra, there is not however one definitive resource for CID reactivity, and divergent rationalizations of CID reactivity for the same types of processes are routinely observed in the literature. What follows is a description, including specific examples from the literature, of those concepts which were most generally applicable to the interpretation of CID reactivity of small even electron ions in this thesis. These generalizations can be viewed as a general ‘tool box’ for explaining observed reactivity, and an understanding of these principles is essential for *de novo* identification of an unknown compounds based on MS/MS data.

The site of protonation or deprotonation during ESI is often considered as the starting point for interpretation of CID reactivity. It is important to consider that the site of protonation or deprotonation in solution is not necessarily the lowest energy position for ionization in the gas phase. This is because in solution, charge is stabilized by participation of solvent molecules but in the gas phase only intramolecular effects such as resonance, inductive effects or intramolecular hydrogen bonding can stabilize the charge site. For this reason, determining the
functional group with the greatest proton affinity (in positive ESI) is not sufficient to determine the most stable protonation site. Furthermore ion dissociation during CID occurs at vibrationally excited energy levels which often induce charge transfer from the lowest energy structure. In fact it is possible for multiple protomers of the same ion with differing CID reactivity to be produced by ESI. This has recently been unequivocally demonstrated for 4-aminobenzoic acid which, after ESI produces ions which can be separated at two distinct sets of conditions using FAIMS, a differential ion mobility technique described in section 1.3.3 of this introduction. Figure 1.12 shows the structures of the two distinct protomers of aminobenzoic acid and their distinguishable CID spectra. However, most protomers are never resolved prior to CID, even when using FAIMS. Thus, pathways originating from multiple protomers of a precursor must be considered in the interpretation of CID reactivity.
For gas phase ions where charge is stabilized by resonance throughout a conjugated system of double bonds, charge can be considered to be delocalized throughout the system of π orbitals. This concept is illustrated in Figure 1.13A for dimethylnitrosoamine where partial positive charge is located on both the amino nitrogen and the nitroso oxygen. During ESI, the o-protonated species is preferentially formed because unlike N-protonation, charge is resonance stabilized as shown in Figure 1.13A. During collisional excitation prior to dissociation, changes in the conformation of an ion lead to charge being localized at the nitroso oxygen which induces proton transfer to the amino nitrogen as in Figure 1.13B. This higher energy species is also a precursor to nitrosoamine dissociation, and each tautomer in Figure 1.13A dissociates by hemolytic cleavage of the single bond to yield different products. Thus, any site where partial charge can be delocalized has the potential for inducing a reaction upon collisional excitation.
Each reaction step in Figure 1.13 involves participation of the charge, either through charge redistribution as in Figure 1.13A or proton transfer as in Figure 1.13B. Charge directed processes such as these occur during CID of small even electron ions whenever there is potential for conjugation and charge stabilization. The alternative to charge directed processes are charge remote processes. These occur much less favorably than do charge directed processes in small even electron species with conjugated systems of $\pi$ electron density. Figure 1.14 shows competition between charge remote and charged directed processes which occur for a deprotonated aliphatic carboxylic acid with an unsaturation at the $\omega$ position. Figure 1.14A shows a charge remote process which results in elimination of butadiene and molecular hydrogen. The charge directed alternative also observed during CID of this species, where nucleophilic attack by the charge bearing $\alpha$ carbon after proton transfer to the carboxyl group, results in elimination of propene. The process in Figure 1.14A is equivalent to the thermolysis of the neutral analogue. Such charge remote processes are only expected as higher energy dissociation processes in ions with poor charge stabilization and little opportunity for conjugation.
Figure 1.14. Competing charge remote (A) and charge directed (B) processes in the dissociation of a deprotonated non-8-enoic acid. Modified from Cheng and Gross.

A key feature of charge directed dissociation is that it occurs by formation of an ion neutral complex in the gas phase followed by separation of the ion and neutral moiety. This is potentially important because chemical reactions, usually proton abstraction from the neutral by the ion are known to occur in the short amount of time spent as an ion-neutral complex, and these can complicate spectral interpretation. During ion dissociation, even if enough kinetic energy is present for an ion and neutral to separate, when that energy is not deposited into the translation of one fragment relative to another, they briefly exist as an ion-neutral complex. For example, in the dissociation of protonated aliphatic ethers, as shown in Figure 1.15, scrambling of O bound protons and those bound to the β carbon, as well as detection of both fragments as protonated product ions confirms the presence of an ion-neutral complex. Though ion-neutral complex formation during dissociation is understood, the prediction of whether ion-neutral reactions occur in these complexes remains poorly understood. It is often difficult to
distinguish between products formed as the result of reactions in an ion-neutral complex and intra-ionic rearrangement or tautomeration reactions which occur prior to dissociation in a single step. It is however also important to note that ion-neutral reactions between neutral and ionic products, such as those presented in Figure 1.15 all represent simple charge directed processes, usually proton transfer reactions, that once characterized for a class of ions or functionality are as easily understood and explained as any other fragmentation mechanism.

![Diagram](image)

**Figure 1.15.** Dissociation of protonated aliphatic ethers by ion neutral complex. Adapted from Morton et al. 2002

When establishing the potential for stabilization of charge at a particular site of an ion by resonance stabilization, considering a single lowest energy conformation of that species is not always adequate. After ionization and desolvation, tautomerization reactions that are not favored in solution can often be easily accessed by gas phase ions, particularly during collisional excitation. These reactions can significantly alter the conjugated system of double bonds of
an ion and affect the stabilization of charge sites in precursor and product ions. Tautomerization reactions that have been reported in recent work include keto-enol and imine-eneamine tautomerization which can occur in a precursor ion before dissociation.\textsuperscript{93–95} A good example of tautomerization reaction which have been carefully studied by ion-spectroscopy, isotopic labeling CID and theoretical calculation is the nucleic acid base uracil, whose reactivity during CID is the topic of research in Chapter 4 of this thesis. Figure 1.16 shows the charge directed tautomerization reaction that occurs in the gas phase after ionization of uracil during ESI. This combination of charge redistribution and proton transfer reaction leads to tautomer 2 which is the most stable tautomer in the gas phase. The presence of multiple tautomers of uracil at energies consistent with dissociation has been confirmed by infra-red ion spectroscopy,\textsuperscript{96} and data from isotope labeling CID study can be used to correlate these different tautomers to specific pathways of dissociation and unique product ions.\textsuperscript{95}

\textbf{Figure 1.16.} Electrospray ionization and tautomerization of uracil during collisional activation by CID.\textsuperscript{95}
A solid fundamental understanding of the gas phase ion chemistry of small organic ions is currently the best tool available for *de novo* identification of unknowns based on MS/MS spectral data. Generalizing the principles which govern this reactivity so that those analyzing spectral data for *de novo* identification are using the same approach to interpret spectra is one of the first steps in expanding the utility of non-target analysis. However, for non-target analysis to be widely adopted by MS users across the wide range of disciplines that currently employ MS, methods of *de novo* identification that do not rely solely on the expertise of the user will need to be developed. This represents possibly the biggest challenge in non-target analysis, but some unique approaches are being investigated towards this goal including *in silico* prediction of CID spectra from chemical structure.

Two distinct approaches exist for the computational prediction of CID spectra from structure, each then allowing experimental MS/MS spectral data to be searched against databases of known compounds for which CID spectra from standards are not currently available. One approach is based on computationally establishing a set of fragmentation rules from literature reports of fragmentation of known species. This approach adopted by two proprietary software packages, Mass Frontier (High-Chem, Ltd) and ACD Fragmenter (ACD/Labs), is reported to offer highly specific identification of detected ions of very high structural diversity based on interpretation of thousands of known reactions. To date, the capabilities of these proprietary tools for the identification of unknowns from CID data has not been openly evaluated in the refereed scientific literature. Evaluation of similar approaches for the identification of species in EI-MS have recommended caution when interpreting match values using only this approach. The alternative approach for computational spectral prediction does not rely on learning from reported reactivity but rather generate a list of all possible masses of fragments based on
combinatorial breaking of chemical bonds in a compound along with a computation of the internal energy of each bond (Fragment Identifier\textsuperscript{98} and MetFrag\textsuperscript{99}). The breaking of bonds with lower internal energy is considered more favorable and weighted at higher ion abundance for spectral prediction. This approach is limited to small ions because of computational power required and structure of the product ions is not taken into account. It is also likely that this approach would be unable to predict product ions resulting from internal rearrangement or tautomerization prior to dissociation.

1.5. Objectives and Scope of This Thesis

This introduction provides a review of current research in the areas of mass spectrometry and analytical separations instrumentation as well as MS spectral databases and \textit{de novo} spectral interpretation for unknown identification. These reviewed studies, while important in their own sub-disciplines, all contribute to a central goal of enabling non-target analysis. Some of the stated goals of these studies include (1) to improve separation and MS detection for comprehensive analysis of complex samples,\textsuperscript{42} (2) to improve the quality of MS/MS spectral data collected from complex samples,\textsuperscript{14} (3) to improve the reproducibility of CID spectra across different MS platforms,\textsuperscript{85} (4) to compile and effectively search databases of CID spectra of known compounds,\textsuperscript{48,77} and (5) to develop a generalized understanding of the reactivity of small even electron ions is CID.\textsuperscript{88} If all of these goal were to be realized it is easy to envision a widely applicable non-target analytical strategy bringing together analytical methods capable of acquiring comprehensive high resolution MS and MS/MS data from complex samples, automated identification of previously known species by database search, and a strategy for \textit{de novo} structural elucidation of new compounds based on their reactivity in CID. The objective of this thesis is to bring these elements together towards improving the utility of MS as a tool for
non-target analysis. Although the instrumental techniques and approaches to data analysis reviewed above have been under development for many years, most of the examples of their use in studies approaching non-target instrumental analysis and spectral interpretation are extremely recent and were published since the work described in this thesis was carried out. In this respect, the field of non-target analysis is only beginning to be recognized as a central goal of MS research. As the work described in this thesis was being planned and carried out, data dependent acquisition methods were not widely used for the analysis of complex biological samples aside from in the field of proteomics. Also, MS/MS spectral databases of small organic ions were only beginning to be populated and mostly consisted of a few hundred spectra.

The first question addressed in this thesis relates to the utility of LC-ESI-MS/MS for non-target analysis of a complex biological sample, urine. Little information was available in the literature about the most abundant species detected in urine by this widely used analytical method. Further, a relatively new MS/MS scan mode called data dependent acquisition, the most promising scan mode for non-target analysis, had not been investigated for the comprehensive analysis of urinary metabolites. The objective of this work was to (1) identify the most abundant species detected by LC-ESI-MS, (2) evaluate the capabilities and limitations of LC-ESI-MS/MS using DDA in non-target analysis and (3) develop the best approach currently available for the identification of detected species based solely on MS data. This work is unique in urine analysis in that neither the analytical method nor the identification approach targets a specific class of analytes and analysis was carried out without prior knowledge of which species would be detected. This allows for the most comprehensive description of those abundant species detected in LC-MS analysis of urine and provides an opportunity to examine some of the limitations of LC-MS for identification of low abundance non-target analytes.
In this thesis, limitations to the identification capabilities of LC-MS for low abundance analytes are investigated and a relatively new technique of gas phase ion separation, FAIMS, is proposed as an alternative separation method. The Gabryelski research group has had previous success using FAIMS in the non-target analysis of environmental samples such as drinking water and mixtures of naphthenic acids.\textsuperscript{30,31} However, this technique had not been applied in the direct analysis of complex biological samples. In order to accomplish this, technical limitations of the FAIMS-MS source interface which prevented sensitive analysis of small analyte ions first needed to be overcome. The direct electrospray ionization of a complex urine sample was also seen as a significant limitation for the use of ESI-FAIMS-MS to the analysis of complex biological samples. In Chapter 3 the use of sample dilution as a method of sample preparation for urine samples in ESI-FAIMS-MS is also investigated.

During the \textit{de novo} identification of unknown species detected by LC-ESI-MS/MS and ESI-FAIMS-MS/MS in this thesis, the gas phase ion reactivity of many classes of metabolites was studied. The most detailed of these investigations, relating to uracil, was initiated when it was found that previous studies of uracil fragmentation were inadequate for understanding its gas phase ion chemistry during CID. While an abundance of data was published from the comprehensive isotope labeling CID study of uracil, it was practically un-interpreted with respect to gas phase ion chemistry as described above in section 1.4.2. In Chapter 4, these previous results are combined with CID experiments using modern accurate mass and sequential ion trap dissociation to describe in detail the dissociation chemistry of protonated uracil during CID. This work has also allowed for a set of generalizations for the interpretation of the gas phase reactivity of small even electron ions that have greatly enabled the identification of unknown metabolites throughout this thesis.
Without authentic standards, quantification of non-target analytes identified in a sample is usually not possible using ESI-MS. In the analysis of urine by ESI-FAIMS-MS in Chapter 3, dilution was implemented as a method of sample preparation. This thesis investigates in detail the quantitative implications of sample dilution and uncovers some unique insight into the nature of ionization suppression in ESI of complex samples. The unique properties of FAIMS also allow for the first proposal of a method of quantification of non-target analytes in complex samples which do not rely on the availability of authentic chemical standards.

Together, this thesis represents a significant contribution to the advancement of non-target MS analysis. Contributions are made to many of the areas of research reviewed above and cover the breadth of those areas of analytical chemistry research where advancement is needed in order for non-target analysis to become an established analytical strategy and directions towards this goal are outlined throughout the thesis.
Chapter 2

Identification of Abundant Species Detected in Urine by Non-Target Liquid Chromatography - Electrospray Ionization - Tandem Mass Spectrometry

Abstract

Liquid chromatography - electrospray ionization - tandem mass spectrometry (LC-ESI-MS/MS) is a widely used technique for target analysis of urinary metabolites and untargeted metabolomic screening. Analysis of urine by LC-ESI-MS provides detection of thousands of chemical compounds but little is known about the identity of even the most abundant species because such experiments usually focus on a small number of known compounds specifically targeted for quantification. The goal of the present work is to characterize these abundant metabolites detected during LC-ESI-MS analysis of urine subjected to routine sample preparation by solid phase extraction. Nano-flow capillary LC coupled to quadrupole time of flight MS operated in data dependent acquisition mode was used to acquire accurate and tandem MS data for hundreds of species in a single chromatographic run. By using an integrated approach for identification of non-target analytes, involving (1) matching experimental data to MS/MS spectra in spectral databases and literature reports, (2) using literature reports of gas phase ion chemistry to understand the structure-reactivity relationship of various classes of metabolites in order to identify new structural analogues or (3) de novo identification based exclusively on gas phase ion chemistry, many of the most abundant ions detected could be identified. In total 110 distinct chemical species from four broad analyte classes were identified including 24 nucleosides and nucleic acid bases, 13 prolyl diketopiperazines (DKPs, cyclic
dipeptides), 49 acylcarnitines and 24 other small urinary metabolites of other classes. These include species not previously reported in urine: a glucuronic acid conjugate of an abundant acylcarnitine, several DKPs, proline amide, dehydroadenosine and an isomer of dimethyladenosine methylated at both the sugar and base moiety. Thus, LC-ESI-MS is highly suitable for the non-target identification of the most abundant ions detected in such a complex biological sample. Some limitations of this technique in non-target identification of low abundance ions are also described: the limit to MS analysis time for a species imposed by the width of the eluting chromatographic peak and interference from chemical background of ESI. This work will be of broad interest to those carrying out analysis of urinary metabolites by LC-ESI-MS as well as those facing the challenge of identifying unknown compounds detected by MS in complex samples.

2.1. Introduction

Liquid chromatography coupled to tandem mass spectrometry detection (LC-MS/MS) is a widely employed technique for the analysis of metabolites and other chemical species excreted in urine. Analytical studies using this technique either involve analysis of targeted compounds or a “global” metabolomic approach. In the former approach, methods are designed to be highly specific to a relatively small number of known, targeted analytes. In the latter approach, a large number of ions are detected but only a small fraction of detectable species are selected for chemical identification based mainly on statistical comparison of spectral data between sample sets. This is usually done using a separate instrumental method targeting a specific analyte for identification. In either approach, analytes not targeted remain unidentified but could be better
characterized by developing better methods of comprehensive chemical analysis which integrate quantitative and qualitative data.

Urine is a mixture of many thousands of compounds and hence an extremely complex matrix in which to carry out chemical analysis. The chemical complexity of urine can hinder chemical analysis for particular species through chemical interference, false positive signals, ionization suppression or poor sample stability. The actual composition of urine is important for the understanding of all of these aspects as well as for the study of the biological processes leading to excretion of urinary metabolites. Studies on the composition of urine have been carried out for decades and include a characterization of the major constituents of urine around or above the trace (mg/L) levels as well as many specific targeted species at much lower concentrations. The most comprehensive of such studies was contracted by the National Aeronautics and Space Association and published in a report in 1971 as part of an initiative to recycle human urine into potable water in space. This report identified and quantified 67 of the most abundant organic and inorganic constituents of humane urine. These analytes range from inorganic salts, to amino acids and small nitrogenous species at concentrations ranging from 1 mg L\(^{-1}\) to 46 g L\(^{-1}\), which were determined by fractionation, selective enrichment and spectroscopy. More recently, data on metabolites known to occur in urine at a wide range of concentrations has begun to be compiled in large chemical databases for use in metabolomics experiments (including the Human Metabolome Database (HMDB) and Metlin). Recently, information in the HMDB on human blood serum has been compiled into a comprehensive publication with tabulated data on over 4000 known serum metabolites. To date no such resource existed for urine analysis at the time of submission of this thesis, though it is listed as “in preparation”.
In targeted analysis, metabolites can only be definitively identified by comparing their properties with those of authentic chemical standards. However, these standards can only be acquired for a very small proportion of species present in biological samples. Modern methods coupling analytical separations to high resolution MS are now capable of providing MS and MS/MS spectral data for species in a complex sample. The comprehensive structural characterization of all species detected in a complex sample is a new approach in chemical analysis termed non-target analysis.\textsuperscript{29,102,103} Towards this goal, work is being carried out to (a) develop better instrumental methods\textsuperscript{29,84,85}, (b) collate spectral data for known compounds in databases,\textsuperscript{48,77,78} (c) develop algorithms capable of effectively searching MS/MS spectral databases,\textsuperscript{83,84} and (d) generalize the reactivity of small organic ions during CID for the purpose of structural elucidation of new metabolites\textsuperscript{88,95,104}. Unlike traditional instrumental analysis, non-target analysis does not rely on the availability of chemical standards for identification. Instead, an integrated approach must be used which combines all information available in databases,\textsuperscript{48,77,78} literature reports on ion reactivity studies\textsuperscript{93,105–112} and supplementary fragmentation studies,\textsuperscript{89,94,95,113} to establish structures of detected species based solely on spectral data.

Because of the huge variety of compounds present in urine and the inherent inter-instrumental variability of ESI and CID, spectral database searching cannot currently be exclusively relied upon for the identification of unknown compounds to the extent that it was possible with electron ionization mass spectra from gas chromatography. However, spectral databases including some that are freely available on the internet, are extremely useful as part of an integrated approach to unknown identification by LC-ESI-MS/MS. One such database, MassBank, contains a few thousand CID spectra, but is currently biased towards drug
metabolites which limited its usefulness in the current research.\textsuperscript{77} One useful feature of this database is that spectra from multiple instruments at multiple experimental conditions are available for each included compound as well as merged spectra less prone to instrumental and experimental variability which impacts searching capabilities. Large chemical databases of metabolites, HMDB and Metlin, include entries for previously identified human metabolites, hundreds of which currently have CID spectra appended.\textsuperscript{48,78} For these resources to be useful in unknown identification, accurate mass or molecular formula (calculated from accurate mass measurement) of detected unknowns is used as the initial searching parameter and then spectral MS/MS features are manually compared when a spectrum is available.

The focus of the current work is to investigate the capabilities of LC-ESI-MS/MS for comprehensive analysis of non-target analytes in complex biological samples with an emphasis on a few critical but rarely discussed aspects including: (1) the characterization of the most abundant species detected in the LC-ESI-MS analysis of urine, (2) the various approaches which must be combined to identify such species from MS spectral data and (3) the relationship between ion abundance, spectral quality and capability of structural elucidation in LC-ESI-MS/MS experiments. Sample preparation and separation parameters were chosen to be representative of current practices in urine analysis by LC-ESI-MS. A quadrupole time of flight (QTOF) mass spectrometer was operated in the data dependent acquisition (DDA) scan mode in which a method was developed to acquire accurate mass data for thousands of ions detected in MS and MS/MS data for a few hundred precursor ions, all during a single chromatographic run. The integrated strategy for processing spectral data for the structural identification of non-target analytes makes use a number of different methods which can be broadly characterized as either (1) matching experimental data with database spectra; (2) partial structural elucidation to
determine compound class, followed by its confirmation by comparison to previous literature reports, or (3) de novo structural elucidation based on gas phase chemistry of investigated ions. Of the thousands of ions detected in MS mode, a few hundred MS/MS spectra collected by DDA were used to identify 110 major urinary metabolites. One goal of this work is to provide a realistic view of the abundant species detected in urine by LC-ESI-MS but also of the strategy and limitations which has to be considered to carry out such identification of non-target analytes. These results will be of interest to anyone developing targeted methods of analysis in a urine matrix and particularly to those carrying out metabolomics studies who are faced with the challenge of identification of detected compounds, particularly those not present in spectral databases.

2.2. Experimental

Ammonium acetate and HPLC grade water and methanol were purchased from Fisher Scientific (Nepean, Ontario). Sample preparation was intended to desalt urine samples and remove only highly hydrophobic compounds which could interfere with the LC separation or ESI. An overnight urine sample from a healthy volunteer was centrifuged at 26 000 g and aliquots (4 mL) of supernatant were loaded onto an activated 200 mg Oasis HLB glass solid phase extraction cartridge (Waters, UK), washed with 4 mL of water and eluted with 4 mL of 70% methanol in water. The eluent was then evaporated to 1 mL with a gentle stream of nitrogen and analyzed by LC-MS. Injections (3.5 µL) were made onto a Waters nano-Acquity nano-UPLC (Waters, UK) system equipped with a 10 mm x 180 µm, 5 µm C18 Symmetry trapping column (Waters, UK) and a 150 mm x 75 µm Symmetry analytical column (Waters, UK). The aqueous (A) and organic (B) mobile phases used were water and methanol, respectively, each
with 1 mM ammonium acetate. The columns were equilibrated using 98 % A. After sample injection, the trapping column was washed, to waste, for 5 min with 98 % A at 3 µL/min before the flow rate was reduced to 400 nL and the effluent diverted to the analytical column. A linear gradient to 100% B over 40 min was followed by a 20 min isocratic period at 100% B and a 20 min linear gradient back to 98% A and column re-equilibration. While detection was carried for 60 min all analytes eluted in the first 45 min of analysis.

Mass spectral detection was carried out using a QTOF micro quadrupole time of flight mass spectrometer (Waters, UK) equipped with a nano-spray ESI ion source (Waters, UK). Ionization was carried out in positive mode with a capillary voltage of 3700 V, a cone voltage of 14 V, and an extraction cone voltage of 1 V. The desolvation and source temperatures were each 150 °C and 80 °C, respectively. Nitrogen (Linde, Guelph, ON) was used as the cone gas (50 L/hr) and the nebulizing gas (12 psi) and argon (Linde, Guelph, ON) was used as the collision gas. MS and MS/MS data were acquired by data dependant acquisition using the survey scan mode. In this mode, a 1 sec MS survey scan is used to measure masses of all detected ions and determine the three most abundant detected ions which are then subjected alternately to MS/MS experiments for nine seconds each. In MS mode, a collision energy (CE) of 4 V was used to transport ions thorough the collision cell without inducing fragmentation and in MS/MS mode an alternating CE of 15 and 30 V was used in the collision cell to obtain CID spectra. After MS and MS/MS data acquisition, previously selected ions were excluded from DDA for 30 seconds to avoid repetitive selection of the same precursor. Consequently, DDA provides accurate mass MS data for all detected ions as well as MS/MS data from CID for the most abundant components detected in the survey scan. No background subtraction of MS or MS/MS data from the DDA method was carried out.
In addition to DDA, spectral data for some compounds were also acquired using a targeted data acquisition method. This method used the same ions source conditions (cone and extraction cone voltage) and dissociation conditions (collision energies) as the DDA method but only one precursor ion at a fixed nominal $m/z$, corresponding to a known analyte of interest, was subjected to MS and MS/MS analysis during the entire chromatographic run. Spectral data from the targeted data acquisition method was then summed across the entire chromatographic peak in the extracted ion chromatogram and background was subtracted using the same number of summed spectra immediately preceding or following the chromatographic peak.

2.3. Results and Discussion

2.3.1. Detection of Urinary Metabolites Using Data Dependent Acquisition

Figure 2.1 shows spectral data from the LC-ESI-MS/MS analysis of urine. Using data dependent acquisition, over 1000 MS peaks could be detected and hundreds of MS/MS spectra collected for each 45 minute chromatographic separation. Figure 2.1A shows the total ion chromatogram (TIC) of such an analysis as a large hump composed of overlapping peaks of detected compounds eluting over the course of the run. However, because of the relatively high resolution of QTOF-MS, most analytes are resolved from other species detected at the similar mass and retention time. Figure 2.1B illustrates extracted ion chromatograms for several identified urinary metabolites with retention times spanning the entire chromatographic run. Figure 2.1C shows the total ion spectrum (TIS) which represents all urine components detected as positive ions following the LC separation and electrospray ionization. This realistic picture of the complexity of urine samples analyzed using ESI is seldom presented in the literature and illustrates the challenge of chemical identification of detected but unknown metabolites.
Figure 2.1. The most abundant components in urine observed by LC-ESI-MS in positive mode. (A) total ion chromatogram of urine extract; (B) extracted ion chromatograms of several identified metabolites including N1-methylnicotinamide (1), N1-methyl-2-pyridone-5-carboxamide (2), 7-methylguanine (3), 7-methylxanthine (4), N4-methylcytidine (5), N1-methylxanthine (6), proline-alanine diketopiperazine (7), N1-methylinosine (8), N-acetylcystidine (9), theobromine (10), N2,N2,N7-trimethylguanosine (11), paraxanthine (12), theophylline (13), cotinine (14), proline-leucine/isoleucine diketopiperazine (15), acylcarnitines (AC284)a (16), acylcarnitines (AC302B)a (17), acylcarnitines (AC342)a (18), acylcarnitines (AC288A)a (19), acylcarnitines (AC336)a (20), acylcarnitines (AC312B)a (21), dibutylphthalate (22). (C) the sum of all detected ions across the 45 min elution time.

a abbreviated acylcarnitine naming from Table 2.1.
In target analysis, the identification of some of the detected metabolites can be accomplished by selecting a group of metabolites of interest and using their chemical standards to match their retention times and spectral data to those obtained for the metabolites detected in urine. The main drawback of this approach is that some major metabolites will remain unidentified because the scope of the target analysis is always limited to a relatively small number of pre-determined analytes. The objective of this research was to develop and evaluate an alternative non-target analytical strategy for the structural identification of a complex mixture of urinary metabolites detected by LC-ESI-MS. This new approach does not rely on the availability of chemical standards for identification. Non-target analysis is intended to establish structures of investigated chemical species based on MS spectral data and currently available resources for identification of non-target analytes.

Of the hundreds of compounds detected in urine by LC-ESI-MS/MS, the major components detected could be broadly characterized as being from one of the following classes: (1) acylcarnitines, (2) prolyl diketopiperazines, (3) nucleosides and nucleic acid (NA) bases or (4) small endogenous and dietary metabolites. With the exception of some of the modified nucleosides, which have been the subject of previous work in the Gabryelski group, practically none of the detected species were known or anticipated prior to analysis of the samples and processing experimental data. In this type of exploratory analysis many different types of species are detected which can be of varying levels of interest to the researcher. An integrated approach for metabolite identification has been developed which is intended to demonstrate the different levels of depth to which non-target identification can be carried out. In simpler cases, such as many dietary metabolites, experimental MS/MS data were matched to spectra published in databases. In other cases such as prolyl diketopiperazines, for which spectral
data for only a few analogues are available, the information from relevant literature reports was used to facilitate the identification of these metabolites as well as new analogues not previously reported in urine. An in depth understanding of the gas phase ion chemistry of several classes of metabolites has enabled identification of previously unknown abundant metabolites which have been characterized exclusively based on experimental MS/MS spectral data. All the identified compounds are presented in tables which include details of MS and MS/MS analysis and retention time in LC along with intensity relative to a well characterized urinary metabolite, 7-methylguanine.\textsuperscript{116,117}

2.3.2. Identification of Acylcarnitines in Urine

Overall, the most abundant species detected in urine were first recognized as being from the same chemical class based on detection of their abundant MS/MS product ions of $m/z$ 85 and the neutral loss of 59 Da (trimethylamine) as in the MS/MS spectrum in Figure 2.2A. From this point, database and literature searching revealed a class of fatty acid metabolites called acylcarnitines (ACs), an abundant class of urinary metabolites sharing this characteristic fragmentation pattern which have been analyzed in both targeted analytical experiments and fragmentation studies.\textsuperscript{48,77,105,118,119} Data acquired by LC-ESI-MS/MS using DDA, presented in Table 2.1, was then used to identify 49 acylcarnitines in urine.

Acylcarnitines (ACs) are a group of acylated homologues of the endogenous quaternary ammonium compound $l$-carnitine which is responsible for the transport of fatty acids into the mitochondria of cells during lipid metabolism. Levels of urinary free carnitine and total AC are routinely measured in infant screening as biomarkers of metabolic disorders.\textsuperscript{120} However, comprehensive analysis or quantification of ACs as individual species is not routine and
individual isomers varying in acyl chain conformation are often difficult to distinguish.\textsuperscript{105} The gas phase ion chemistry of ACs during CID has been investigated towards the differentiation of isomeric ACs.\textsuperscript{105} This work successfully established the decomposition pathways common to most ACs and structures of some decomposition products have been proposed. It was also feasible to differentiate some chemically modified ACs differing only by acyl chain branching based on differences in the relative intensity of product ions, however only for small ACs with a C4 acyl chains, and only for chemically derivatized analogues.\textsuperscript{105}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.2.png}
\caption{CID spectra of selected acylcarnitines from data dependent LC-ESI-MS/MS analysis of urine extracts. Spectra correspond to entries in Table 2.1 for (A) AC 286A, (B) AC 340, (C) AC 316, (D) AC 356, (E) AC 358A and (F) a glucuronide conjugate of AC286.}
\end{figure}
Table 2.1: Acylcarnitine identified in urine by data dependent LC-ESI-MS/MS.

<table>
<thead>
<tr>
<th>Compound Abbreviation</th>
<th>Fatty Acid Composition</th>
<th>Retention Time (min)</th>
<th>Relative Intensity</th>
<th>Precursor m/z (intensity in MS/MS)</th>
<th>Mass Error (ppm)</th>
<th>Product Ion m/z (% Relative Intensity in MS/MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC246</td>
<td>C4</td>
<td>12.6</td>
<td>0.350</td>
<td>246.173 (53)</td>
<td>11</td>
<td>187 (35) nd 57 (4) 100</td>
</tr>
<tr>
<td>AC266</td>
<td>benzyl</td>
<td>13.9</td>
<td>0.115</td>
<td>266.1397 (42)</td>
<td>4</td>
<td>207 (25) 105 (13) nd 100</td>
</tr>
<tr>
<td>AC284</td>
<td>C8:2</td>
<td>23.8</td>
<td>1.327</td>
<td>284.187 (80)</td>
<td>4</td>
<td>225 (40) nd 100</td>
</tr>
<tr>
<td>AC286</td>
<td>C8:1</td>
<td>28.6</td>
<td>1.186</td>
<td>286.2029 (80)</td>
<td>5</td>
<td>227 (30) 125 (4) 97 (3) 100</td>
</tr>
<tr>
<td>AC286A</td>
<td>C8:1</td>
<td>27.2</td>
<td>3.715</td>
<td>286.2029 (100)</td>
<td>5</td>
<td>227 (32) 125 (11) 97 (11) 100</td>
</tr>
<tr>
<td>AC288</td>
<td>C7:1 + O</td>
<td>12.8</td>
<td>0.055</td>
<td>288.176 (100)</td>
<td>-17</td>
<td>229 (3) 227 (2) nd 13</td>
</tr>
<tr>
<td>AC288A</td>
<td>C8</td>
<td>29.1</td>
<td>1.536</td>
<td>288.2188 (100)</td>
<td>5</td>
<td>229 (36) 127 (8) 99 (8) 52</td>
</tr>
<tr>
<td>AC290</td>
<td>C6:1 + 2O</td>
<td>11.5</td>
<td>0.040</td>
<td>290.1624 (100)</td>
<td>8</td>
<td>231 (27) 129 (27) 101 (27) 64</td>
</tr>
<tr>
<td>AC292</td>
<td>C7 + 2O</td>
<td>16.4</td>
<td>0.165</td>
<td>292.1719 (100)</td>
<td>-13</td>
<td>233 (47) 131 (29) 103 (23) 29</td>
</tr>
<tr>
<td>AC300</td>
<td>C9:1</td>
<td>31</td>
<td>4.891</td>
<td>300.2165 (100)</td>
<td>-3</td>
<td>241 (43) 139 (13) 111 (6) 95</td>
</tr>
<tr>
<td>AC300A</td>
<td>C9:1</td>
<td>29.9</td>
<td>0.455</td>
<td>300.2193 (100)</td>
<td>7</td>
<td>241 (43) 139 (20) 111 (6) 84</td>
</tr>
<tr>
<td>AC302</td>
<td>C8:1 + O</td>
<td>18.3</td>
<td>0.020</td>
<td>302.1946 (100)</td>
<td>-6</td>
<td>243 (33) nd nd nd 100</td>
</tr>
<tr>
<td>AC302A</td>
<td>C8:1 + O</td>
<td>19.5</td>
<td>0.349</td>
<td>302.1972 (100)</td>
<td>2</td>
<td>243 (14) 141 (10) nd 24</td>
</tr>
<tr>
<td>AC302B</td>
<td>C8:1 + O</td>
<td>24</td>
<td>0.415</td>
<td>302.1972 (100)</td>
<td>2</td>
<td>243 (14) 141 (3) nd 6</td>
</tr>
<tr>
<td>AC302C</td>
<td>C9</td>
<td>32.2</td>
<td>3.309</td>
<td>302.2301 (100)</td>
<td>-10</td>
<td>243 (39) 141 (4) 113 (1) 71</td>
</tr>
<tr>
<td>AC302D</td>
<td>C9</td>
<td>31.7</td>
<td>1.218</td>
<td>302.2347 (0)</td>
<td>6</td>
<td>nd nd nd nd 100</td>
</tr>
<tr>
<td>AC310</td>
<td>C10:3</td>
<td>30.4</td>
<td>4.282</td>
<td>310.2001 (100)</td>
<td>-5</td>
<td>251 (22) 149 (11) 121 (8) 100</td>
</tr>
<tr>
<td>AC310A</td>
<td>C10:3</td>
<td>29.3</td>
<td>5.273</td>
<td>310.2004 (100)</td>
<td>-4</td>
<td>251 (19) 149 (8) 121 (7) 85</td>
</tr>
<tr>
<td>AC312</td>
<td>C10:2</td>
<td>31.2</td>
<td>1.018</td>
<td>312.2191 (100)</td>
<td>6</td>
<td>253 (30) 151 (7) 123 (4) 59</td>
</tr>
<tr>
<td>AC312B</td>
<td>C10:2</td>
<td>32.1</td>
<td>0.387</td>
<td>312.2187 (0)</td>
<td>4</td>
<td>nd nd nd nd 123 (8) 100</td>
</tr>
<tr>
<td>AC314</td>
<td>C9:2 + O</td>
<td>13.5</td>
<td>0.042</td>
<td>314.1938 (100)</td>
<td>-9</td>
<td>255 (14) 153 (6) nd 34</td>
</tr>
<tr>
<td>AC314B</td>
<td>C9:2 + O</td>
<td>14</td>
<td>0.055</td>
<td>314.1962 (100)</td>
<td>-1</td>
<td>nd 153 (9) nd nd 39</td>
</tr>
<tr>
<td>AC314C</td>
<td>C10:2</td>
<td>32.5</td>
<td>0.915</td>
<td>314.2347 (0)</td>
<td>5</td>
<td>nd nd nd nd 100</td>
</tr>
<tr>
<td>AC316</td>
<td>C9:1 + O</td>
<td>25</td>
<td>0.636</td>
<td>316.2120 (100)</td>
<td>-1</td>
<td>257 (18) 155 (27) 127 (36) 55</td>
</tr>
<tr>
<td>AC316A</td>
<td>C9:1 + O</td>
<td>21</td>
<td>0.283</td>
<td>316.2111 (100)</td>
<td>-3</td>
<td>nd nd nd nd 34</td>
</tr>
<tr>
<td>Abbreviation of compound name used in text and figures. ACXXXXY where AC = Acylcarnitine XXX = nominal m/z and Y = order of identification used to distinguish acylcarnitines identified at the same nominal mass.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Abbreviation</strong> of compound name used in text and figures. ACXXXXY where AC = Acylcarnitine XXX = nominal m/z and Y = order of identification used to distinguish acylcarnitines identified at the same nominal mass.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Composition of the acyl chain of the acylcarnitine.</strong> Corresponds to the composition of the R group in Scheme 2.3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intensity relative to 7-methylguanine in MS mode</strong> value in brackets represents the % relative spectral intensity of the precursor ion in the product ion spectra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Values represent the % relative spectral intensity of the m/z 85 product ion in the product ion spectra</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The base peak in the product ion spectrum of AC358 is m/z 153 as shown in Scheme 2.4C,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The base peak in the product ion spectrum of AC358A is m/z 95 as shown in Scheme 2.4D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC316B</td>
<td>C10</td>
<td>33</td>
<td>0.144</td>
<td>316.2504 (0)</td>
<td>5</td>
<td>nd</td>
</tr>
<tr>
<td>AC326</td>
<td>C10:3 + O</td>
<td>30.6</td>
<td>0.635</td>
<td>326.1981 (88)</td>
<td>5</td>
<td>267 (47)</td>
</tr>
<tr>
<td>AC328</td>
<td>C10:2 + O</td>
<td>26</td>
<td>1.209</td>
<td>328.2130 (100)</td>
<td>2</td>
<td>269 (8)</td>
</tr>
<tr>
<td>AC328A</td>
<td>C11</td>
<td>34.5</td>
<td>0.068</td>
<td>328.2497 (0)</td>
<td>3</td>
<td>nd</td>
</tr>
<tr>
<td>AC328B</td>
<td>C10:2 + O</td>
<td>27</td>
<td>0.817</td>
<td>328.2133 (100)</td>
<td>3</td>
<td>269 (11)</td>
</tr>
<tr>
<td>AC328C</td>
<td>C10:2 + O</td>
<td>28.2</td>
<td>0.358</td>
<td>328.2136 (100)</td>
<td>3</td>
<td>269 (14)</td>
</tr>
<tr>
<td>AC329</td>
<td>C9:1 + NO</td>
<td>15.7</td>
<td>0.058</td>
<td>329.2065 (100)</td>
<td>-3</td>
<td>270 (10)</td>
</tr>
<tr>
<td>AC330</td>
<td>C10:1 + O</td>
<td>28.6</td>
<td>2.836</td>
<td>330.2276 (100)</td>
<td>-1</td>
<td>271 (8)</td>
</tr>
<tr>
<td>AC330A</td>
<td>C10:1 + O</td>
<td>29</td>
<td>2.836</td>
<td>330.2309 (100)</td>
<td>9</td>
<td>271 (8)</td>
</tr>
<tr>
<td>AC330B</td>
<td>C10:1 + O</td>
<td>25.2</td>
<td>0.909</td>
<td>330.2285 (100)</td>
<td>2</td>
<td>nd</td>
</tr>
<tr>
<td>AC330C</td>
<td>C10:1 + O</td>
<td>26</td>
<td>2.164</td>
<td>330.2309 (100)</td>
<td>9</td>
<td>271 (4)</td>
</tr>
<tr>
<td>AC332</td>
<td>C9:1 + 2O</td>
<td>22</td>
<td>1.382</td>
<td>332.2091 (100)</td>
<td>6</td>
<td>273 (12)</td>
</tr>
<tr>
<td>AC336</td>
<td>C12:4</td>
<td>31.3</td>
<td>0.540</td>
<td>336.2189 (100)</td>
<td>5</td>
<td>277 (25)</td>
</tr>
<tr>
<td>AC338</td>
<td>C12:3</td>
<td>32.6</td>
<td>0.161</td>
<td>338.2354 (0)</td>
<td>7</td>
<td>nd</td>
</tr>
<tr>
<td>AC340</td>
<td>C11:3 + O</td>
<td>23.5</td>
<td>0.535</td>
<td>340.2126 (100)</td>
<td>1</td>
<td>281 (13)</td>
</tr>
<tr>
<td>AC342</td>
<td>C11:2 + O</td>
<td>25</td>
<td>0.973</td>
<td>342.2284 (100)</td>
<td>1</td>
<td>283 (6)</td>
</tr>
<tr>
<td>AC343</td>
<td>C10:1 + NO</td>
<td>25.7</td>
<td>0.991</td>
<td>343.2233 (100)</td>
<td>0</td>
<td>nd</td>
</tr>
<tr>
<td>AC344</td>
<td>C10:2 + 2O</td>
<td>24.4</td>
<td>1.945</td>
<td>344.2408 (100)</td>
<td>3</td>
<td>285 (4)</td>
</tr>
<tr>
<td>AC346</td>
<td>C10:1 + 2O</td>
<td>16.6</td>
<td>0.075</td>
<td>346.2208 (100)</td>
<td>-6</td>
<td>nd</td>
</tr>
<tr>
<td>AC354</td>
<td>C11:4 + 2O</td>
<td>22.5</td>
<td>0.095</td>
<td>354.1925 (100)</td>
<td>3</td>
<td>295 (7)</td>
</tr>
<tr>
<td>AC356</td>
<td>C12:2 + O</td>
<td>31.4</td>
<td>1.791</td>
<td>356.2474 (0)</td>
<td>12</td>
<td>nd</td>
</tr>
<tr>
<td>AC358</td>
<td>C11:2 + 2O</td>
<td>27</td>
<td>0.456</td>
<td>358.2299 (100)</td>
<td>20</td>
<td>299 (12)</td>
</tr>
<tr>
<td>AC358A</td>
<td>C12:1 + O</td>
<td>33.2</td>
<td>0.611</td>
<td>358.2589 (0)</td>
<td>-1</td>
<td>nd</td>
</tr>
</tbody>
</table>

\[a\] Abbreviation of compound name used in text and figures. ACXXXXY where AC = Acylcarnitine XXX = nominal m/z and Y = order of identification used to distinguish acylcarnitines identified at the same nominal mass.

\[b\] Composition of the acyl chain of the acylcarnitine. Corresponds to the composition of the R group in Scheme 2.3.

\[c\] Intensity relative to 7-methylguanine in MS mode

\[d\] Value in brackets represents the % relative spectral intensity of the precursor ion in the product ion spectra

\[e\] Values represent the % relative spectral intensity of the m/z 85 product ion in the product ion spectra

\[f\] The base peak in the product ion spectrum of AC358 is m/z 153 as shown in Scheme 2.4C

\[g\] The base peak in the product ion spectrum of AC358A is m/z 95 as shown in Scheme 2.4D

76
Total concentrations of AC in urine range from between 5 µM to 9 µM in the urine of healthy individuals.\textsuperscript{118} Hence, the concentrations of each abundant AC in the concentrated urine extract are in the mid nM range. These concentrations are in the normal range for detection by LC-ESI-MS, but ACs are detected at high spectral intensity due to high ionization efficiency in electrospray ionization. Their chemical structure (Scheme 2.1) includes a polar head group with a fixed charge on a tertiary ammonium moiety and a hydrophobic fatty acid chain. Thus, ACs act as natural surfactants and very effectively compete for space at the surface of the electrospray droplets and show high analytical sensitivity.

Acylcarnities can be sensitively detected and recognized by their characteristic MS/MS spectra including an abundant \textit{m/z} 85 ion and the abundant neutral loss of 59 Da, most of which closely resembled the spectrum for AC286A in Figure 2.2A. The general reactivity of ACs during CID is shown in Scheme 2.1. The dissociation of ACs during CID has been previously examined using chemical standards of some known ACs and the composition of a few characteristic neutral losses and product ions has been determined, but the mechanisms of most of these dissociation pathways have not been considered.\textsuperscript{36,105} Formation of \textit{m/z} 85 has been proposed to occur by a charge remote McLafferty type elimination of a neutral fatty acid shown in the first step of the reaction in Scheme 2.1A, the product of which readily undergoes elimination of trimethylammonium to produce a cyclic and highly resonance stabilized product at \textit{m/z} 85.\textsuperscript{36} Since \textit{m/z} 144 is detected at low abundance in the spectra of some ACs, this reaction is likely to occur, but may not be the only process yielding abundant product ions with \textit{m/z} 85 in the spectra of all ACs. Charge remote dissociation processes (such as the first step in Scheme 2.1A) are known to occur in CID for ions with a localized charge and little opportunity for conjugation and charge stabilization by resonance.\textsuperscript{121} While AC ions have a fixed charge, the
presence of other heteroatoms and unsaturations in the carnitine and acyl moieties makes more favorable charge directed mechanisms viable, and these are likely to be more favorable than charge remote processes. The elimination of trimethylamine is required in order for charge on the AC ions to become resonance stabilized and this is the first step to the majority of AC reactivity. This characteristic neutral loss of 59 Da can occur by nucleophilic attack of the carbonyl oxygen at the carbon adjacent to the ammonium group as in Schemes 2.1A and 2.1B. Proceeding directly from the precursor as in Scheme 2.1B, this process produces a [M-59]+ ion. As shown in Scheme 2.1B as pathway (a), this product can then undergo proton transfer and elimination of a neutral fatty acid to produce the same m/z 85 product ion as Scheme 2.1A by an exclusively charge directed process. Alternatively, shown as pathway (b), the [M-59]+ ion can react to eliminate the remainder of the carnitine moiety to give [M-161]+ and [M-189]+ product ions also detected as product ions of most ACs. The reactions shown in Scheme 2.1 are the most favorable pathways and produce the most abundant product ions for most ACs. It is these product ions that have been used to detect and identify ACs by ESI-MS/MS in the past and to create sensitive and highly selective MRM methods for their quantification, even in complex urinary samples.

In the time between when the current research was carried out and publication of this thesis, a paper was published by Liang Li’s group from the University of Alberta describing the comprehensive analysis of urinary ACs by LC-ESI-MS/MS with DDA acquisition. The goal of their work was to identify as many ACs as possible from urine after selectively extracting them using analyte specific ion exchange solid phase extraction (reviewed in the introduction in Section 1.3.1). The principal difference between these two studies, aside from the sample preparation method used, is that the MS detection method in their work was optimized for
The carnitine moiety is structurally unchanged in all ACs analogues therefore changes in reactivity of the analogues during CID can exclusively be attributed to their acyl group structure. Differences in relative intensity of fragment ions from observed dissociation pathways can be attributed to particular structural features of the modified acyl groups such as the conjugated system of double bonds and the location of heteroatom(s) with electron lone pairs. The primary driving force for the formation of small ions in CID is stabilization of the developing charge in
the transition state, which can significantly lower activation energy of fragmentation. This concept is discussed in detail in the introduction to this thesis. Such stabilization of the transition state during CID can be approximated by resonance stabilization of the charge in the product ion. In gas phase chemistry of ACs, significant differences in rates of dissociation reactions are observed especially when the charge can be delocalized in a conjugated \( \pi \) system of the acyl chain. For example, AC340 with a C11:3 + O fatty acid composition (13 carbon atoms, 3 double bonds and 1 oxygen atom) exhibits relatively high abundance (compared to most other detected ACs) of the \( m/z \) 179 product ion in its MS/MS spectrum shown in Figure 2.2B. This fragment ion [M-161]\(^+\) originates from the dissociation pathway involving the elimination of the carnitine moiety as illustrated Scheme 2.1B. A significant increase in its abundance can be attributed to the structure of its acyl group because. As shown in Scheme 2.2A, this product ion has a positive charge delocalized between the carbon and oxygen of a carbonyl group and significant additional stabilization of this charge by double bond(s) and/or the oxygen of the chain. This includes at minimum an unsaturation \( \alpha \) to the carbonyl group or possibly a larger conjugated system as shown in Scheme 2.2A for AC340.

Similarly, an unsaturation \( \beta \) to the carbonyl group leads to the preferential formation of the [M-189]\(^+\) product ion at \( m/z \) 127, which is observed at increased relative abundance in the MS/MS spectrum of AC316 in Figure 2.2C. This dissociation product at \( m/z \) 127 could either be formed as a result of sequential elimination of CO from the [M-161]\(^+\) fragment ion at \( m/z \) 155, as shown in Scheme 2.2B for AC316, or a direct decomposition of the precursor ion of AC316 at \( m/z \) 316. An abundant product ion [M-203]\(^+\) at \( m/z \) 127 is observed as the major product in the MS/MS spectrum of AC356 in Figure 2.2D. As shown in Scheme 2.2C, during elimination of ketene (\( H_2CCO \), 42 Da) from the [M-161]\(^+\) of AC356, the developing charge in the product ion
[M-203]$^+$ at $m/z$ 127 becomes resonance stabilized when the carbons $\alpha$ and $\beta$ to the carbonyl group are saturated and the first unsaturation appears at the $\gamma$ position. In this case, a hydroxyl group can also contribute to resonance stabilization of the charge at any position where it is part of the conjugated system of $\pi$ electron density of the product ion.

Finally, when all carbons near the charged carbonyl group are saturated, other charge directed processes take place and involve transfer of the charge to more conjugated sites on the unsaturated portion of the chain. One example where this is particularly evident is AC358A for which none of the processes in Scheme 2.1 yield the base peak in the MS/MS spectrum shown in Figure 2.2E. Here, in addition to the $m/z$ 85.07 ion, the abundant carbocation fragment at $m/z$ 95.09 dominates the spectrum. This ion, with formula C$_7$H$_{11}^+$ determined by accurate mass measurement, occurs as the result of a hydroxyl group at the $\gamma$ position to the carbonyl group and a conjugated system of double bonds as illustrated in Scheme 2.2D. This arrangement leads to formation of the $m/z$ 95 by breaking the acyl chain and eliminating water.

In addition to the identified AC analogous in Table 2.1, one glucuronide conjugate of an acylcarnitine was detected at $m/z$ 462.2 whose MS/MS spectrum is shown in Figure 2.2F. This spectrum shows the characteristic neutral loss of dehydroglucuronic acid moiety (176 Da) to produce an ion with $m/z$ 286, corresponding to an AC with a C8:1 acyl chain composition, which itself was one of the most abundant components detected (AC286) in urine. Glucuronic acid conjugates of acylcarnitines have not previously been reported as urinary metabolites. Based on the characteristic reactivity of protonated acylcarnitines (Scheme 2.1) and their unique reactivity associated with a structure of acyl chain (Scheme 2.2), 49 acylcarnitine analogous were detected in urine extract.
Scheme 2.2. The impact of specific acyl chain modifications of the gas phase chemistry of protonated acylcarnitines during collision induced dissociation.

All identified compounds are presented in Table 2.1, which includes information from MS detection (accurate mass measurement to establish elemental composition of precursor ions), MS/MS detection (accurate mass measurement to establish elemental composition of fragment ions) and retention time form LC along with intensity relative to a well characterized urinary metabolite, 7-methylguanine.\textsuperscript{116,117} Based on this data, the composition of the fatty acid chain could be established for all of the detected ACs and ranged from C4 to C12, from saturated to 3
unsaturations and from zero to two oxygenations on the fatty acid chain. Two acylcarnitines with nitrogenated fatty acid chains (AC329 and AC343) as well as benzoylaclycarnitine (AC226) were also detected. The chain length of AC detected varied between C_4 and C_{12}, while previous reports of urinary AC have found homologues of up to C_{18}. The retention time of the C_{12} homologue was 33 minutes at which time the mobile phase composition corresponds to elution with close to the 70% methanol used in the SPE procedure. This suggests that the current method of sample preparation was effective at eliminating the most hydrophobic species, but of conserving the other small urinary metabolites.

2.3.3 Identification of Prolyl Diketopiperazines

Another class of abundant ions which were readily selected in DDA mode during LC-ESI-MS/MS analysis showed the characteristic low mass fragment ion detected at m/z 70.07. In addition, these species typically displayed neutral losses of 17, 18, 28, and 43 Da, had chromatographic retention times ranging from 9 to 20 min and relative intensities from 0.05 to 8 times that of 7-methyl-guanine. This is a wide range of properties for a class of compounds with analogous reactivity which was attributed to the wide range of properties of amino acids which form a class of small cyclic peptides. Based on literature and database reports relating to detection of peptides, the m/z 70.07 ion along with several other low mass products detected along with it could be associated with immonium product ions of amino acids, which are usually abundant product ions in collision induced dissociation of protonated peptides. The m/z 70.07 fragment is the immonium ion produced in CID of protonated peptides containing proline. The precursor ions of urinary metabolites sharing this product ion are shown in Table 2.2. The mass and elemental composition of each precursor ion corresponds to a dehydration product of a proline containing dipeptide. Ions of this type could originate from two possible sources and
represent either protonated ions of cyclic dipeptides from urine or gas phase dissociation products (b₂ ions) of larger protonated linear peptides which may dissociate in the source of the mass spectrometer. The DDA method used the gentlest source conditions possible to minimize source fragmentation and ensure the detection of intact precursor ions. In no case was the detection of cyclic proline dipeptides accompanied by the presence of larger linear peptides, which suggests the detection of cyclic prolyl dipeptides in urine.

Cyclic dipeptides or 2,5-diketopiperazines (DKPs) are a class of peptide derivatives that are more stable to degradation by peptidases than their linear analogues, thus explaining their presence in human urine. Prolyl DKPs are known products of peptide degradation found in urine and specific targeted species have been investigated as biomarkers. Of particular significance are cyclic proline-hydroxyproline (Pro-Hyp) which has been investigated as a biomarker of collagen metabolism and cyclic proline-histidine (Pro-His) which is known to have antimicrobial properties. Prolyl DKPs also form in food during cooking as the result of the cyclo-condensation reaction of proline containing proteins.¹²⁴⁻¹²⁶ These metabolites also occur endogenously with Pro-His playing an important role in the central nervous system, the brain and the gut.¹²⁷ Although no study has reported the comprehensive detection or identification of this class of compounds in urine by LC-MS, the gas phase ion chemistry of a few protonated DKPs in CID has been reported previously.¹¹² Scheme 2.3 shows dissociation pathways of protonated prolyl DKPs, which are important for the structural identification of this group of metabolites. Scheme 2.3A shows the reaction by which the m/z 70 proline immonium ion is formed from a protonated DKP precursor.¹¹² The amino acid from which an immonium product ion is generated depends on which amino site is protonated before dissociation. While the proline amino group is the lowest energy protonation site (Scheme 2.3A), proton transfer to the non-proline amino group
can occur during collisional excitation as shown in Scheme 2.3B. This proton transfer leads to formation of the non-proline immonium ion.

Table 2.2: Cyclic dipeptides identified in urine by non-target LC-ESI-MS/MS

<table>
<thead>
<tr>
<th>Analyte( ^a )</th>
<th>m/z</th>
<th>Mass error (ppm)</th>
<th>Retention Time (min)</th>
<th>Relative Intensity ( ^b )</th>
<th>Immonium Fragment ( ^c )</th>
<th>Diagnostic Neutral Losses ( ^d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro-Ala</td>
<td>169.0998</td>
<td>15</td>
<td>9.7</td>
<td>0.66</td>
<td>70.07 (Pro)</td>
<td></td>
</tr>
<tr>
<td>Pro-Val</td>
<td>197.1293</td>
<td>4</td>
<td>15.6</td>
<td>0.16</td>
<td>70.07 (Pro), 72.09 (V)</td>
<td></td>
</tr>
<tr>
<td>Pro-Thr</td>
<td>199.1090</td>
<td>6</td>
<td>10.1</td>
<td>0.21</td>
<td>70.07 (Pro), 74.07 (V)</td>
<td></td>
</tr>
<tr>
<td>Pro-Hyp</td>
<td>211.1058</td>
<td>-9</td>
<td>17.3</td>
<td>0.19</td>
<td>70.0676 (Pro), 86.07(Hyp)</td>
<td></td>
</tr>
<tr>
<td>Pro-Leu/Ilu</td>
<td>211.1469</td>
<td>13</td>
<td>19.3</td>
<td>1.65</td>
<td>70.07 (Pro),</td>
<td></td>
</tr>
<tr>
<td>Pro-Gln</td>
<td>226.1210</td>
<td>10</td>
<td>10.0</td>
<td>0.05</td>
<td>70.07 (Pro), 125.07</td>
<td>[M+H-NH₃]⁺ (100), [M+H-NH₂-CO]⁺ (30)</td>
</tr>
<tr>
<td>Hyp-Leu/Ilu</td>
<td>227.1369</td>
<td>-9</td>
<td>14.9</td>
<td>0.10</td>
<td>86.07 (Hyp)</td>
<td></td>
</tr>
<tr>
<td>Pro-His</td>
<td>235.1174</td>
<td>-7</td>
<td>10.4</td>
<td>0.22</td>
<td>70.07 (P), 110.08 (H)</td>
<td></td>
</tr>
<tr>
<td>Pro-Mso</td>
<td>245.0979</td>
<td>10</td>
<td>10.6</td>
<td>0.16</td>
<td>70.07 (P), 125.07</td>
<td>[M+H₂CSOH]⁺ (100)</td>
</tr>
<tr>
<td>Pro-Arg</td>
<td>254.1632</td>
<td>8</td>
<td>16.8</td>
<td>7.82</td>
<td>70.07 (P), 125.07</td>
<td>[M+H-HNC(NH₂)₂]⁺ (100)</td>
</tr>
<tr>
<td>Pro-Tyr</td>
<td>261.1227</td>
<td>-3</td>
<td>17.0</td>
<td>0.05</td>
<td>70.07 (P), 136.08 (T)</td>
<td></td>
</tr>
</tbody>
</table>

\( ^a \) Short form names of amino acids Proline (Pro), alanin (Ala), valine (Val), threonine (Thr), hydroxyproline (Hyp), leucine or isoleucine (Leu/Ilu), glycine (Gln), histidine (His), methionine (Mso), arginine (Arg) and tyrosine (Tyr) making up the cyclic dipeptide with structures from Figure 2.1.

\( ^b \) Intensity relative to the intensity of nucleic acid base 7-methyl-guanine identified in urine.

\( ^c \) m/z of fragments, imine fragments of amino acids show short form name of amino acid from which they originate in brackets.

\( ^d \) m/z of fragments with their relative spectral intensity in brackets.
Scheme 2.3. The gas phase reactivity of protonated prolyl-diketopiperizines.
Based on spectral data from the LC-ESI-MS/MS analysis of urine, 12 unique cyclic dipeptides were identified, all containing either proline or hydroxyproline as shown in Figure 2.3. Table 2.2 provides additional data for their identification including accurate masses of the precursor and of the diagnostic immonium ions, which were used to determine their elemental compositions. Figure 2.4 shows MS/MS spectra for two protonated prolyl DKPs which have been investigated in CID. In the case of Pro-Val (Figure 2.4A) both \( m/z \) 70 and 72 immonium fragments are detected, which is consistent with the two protonation sites of the precursor ion as
shown in Scheme 2.3A. In some cases other unique dissociation reactions were observed, which were found to be specific to the structures of cyclic peptides. One such example is proline-glutamine (Pro-Glu) cyclic dipeptide with its MS/MS spectrum shown in Figure 2.4B. The Pro-Glu DKP has not been detected in urine or previously studied in CID thus its MS/MS spectra are not available in databases. The identification of this metabolite has been accomplished by *de novo* structural elucidation based on dissociation reactions observed in CID.

![Diagram](image)

**Figure 2.4.** Product ion spectra of two urinary cyclic dipeptides identified by data dependent LC-ESI-MS/MS analysis. (A) proline-valine cyclic dipeptide and (B) proline-glutamine cyclic dipeptide.
Scheme 2.3C illustrates dissociation pathways leading to formation of fragment ions in the MS/MS spectrum in Figure 2.4B. The \([\text{M} + \text{H} – \text{NH}_3]^+\) at \(m/z\) 209.0926 is the most abundant product detected and is consistent with the elimination of ammonia from the amide group on the glutamine side chain, which can occur by nucleophilic attack of the cyclic imine nitrogen on the carbonyl carbon of the amide group, at which charge develops as ammonia is eliminated. This mechanism will be favored over direct loss of \(\text{NH}_3\) because of the additional resonance and inductive stabilization of the charge after the nucleophilic attack and \(\text{NH}_3\) elimination. The cyclization step is supported by the fact that the subsequent reaction steps involve initial loss of CO to form the \(m/z\) 181 fragment stabilized by resonance and following elimination of ethane and CO to produce the intermediate ion at \(m/z\) 125 with no charge stabilization, which can either be stabilized by internal nucleophilic attack of the amino nitrogen to from the highly stabilized product ion at \(m/z\) 125 (pathway a) or react further by elimination of HCN and CO to produce the proline immonium ion at \(m/z\) 70 (pathway b).

The goal of de novo identification is to correlate the product ions detected in MS/MS (Figure 2.4B) to the structure of the precursor ion (Scheme 2.3C), which can be established based on its gas phase chemistry. The identification of a new urinary metabolite such as Pro-Glu DKP also provides new information on gas phase chemistry of analogous compounds. For instance, the \(m/z\) 125 fragment ion (Figure 2.4B, Scheme 2.3C) is expected to appear in MS/MS spectra of proline cyclic dipeptides containing asparagine, glutaminic and aspartic acids. The side chain of these amino acids will be involved in similar reactions (protonation, cyclization and neutral losses) as those presented in Scheme 2.3C, ultimately leading to the \(m/z\) 125 fragment ion. Thus, the \(m/z\) 125 is the important diagnostic product ion for cyclic dipeptides with such amino acid compositions. Such de novo identification of non-target metabolites is feasible only
for high quality of MS/MS spectral as shown for relatively abundant DKPs (Figure 2.4). The presence of background or interfering ions in MS/MS spectra, which commonly occurs for less abundant metabolites, makes similar de novo identification of non-target urinary metabolites extremely difficult and often impossible.

2.3.4. Other Small Endogenous and Dietary Metabolites

The current interest in the field of metabolomics has created a demand for methods of identifying compounds found to be important in endogenous metabolic pathways and in the metabolic processing of ingested compounds. The field of metabolomics routinely uses LC-ESI-MS and thus databases containing tandem mass spectra for known metabolites commonly detected in biofluids have been developed and greatly enabled this challenging task. By using two such databases, Massbank\textsuperscript{77} and the Human Metabolome Database\textsuperscript{48}, along with some literature reports on investigation of well known, targeted urinary metabolites it was possible to identify many abundant species by LC-ESI-MS/MS based primarily on spectral matching. Several such analytes were identified by this approach and are listed in Table 2.3 and can be broadly characterized as being small endogenous and dietary urinary metabolites.
<table>
<thead>
<tr>
<th>Analyte</th>
<th>measured m/z (MS/MS intensity)</th>
<th>mass error (ppm)</th>
<th>retention time (min)</th>
<th>intensity relative to 7-MeG</th>
<th>product ions m/z (MS/MS relative intensity)</th>
<th>Metabolite Source</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>urea</td>
<td>61.0411 (0)</td>
<td>25</td>
<td>1.6</td>
<td>0.63</td>
<td>-</td>
<td>amino acid degradation</td>
<td>48</td>
</tr>
<tr>
<td>creatinine</td>
<td>114.0661 (100)</td>
<td>0.9</td>
<td>2.2</td>
<td>1.17</td>
<td>86 (25), 44 (10)</td>
<td>creatine degradation</td>
<td>48</td>
</tr>
<tr>
<td>pyrrolidine-2-carboxamide</td>
<td>115.0857 (10)</td>
<td>-7</td>
<td>2.6</td>
<td>0.22</td>
<td>97 (6), 70 (100)</td>
<td>unknown</td>
<td>48,12</td>
</tr>
<tr>
<td>acetylputrescine</td>
<td>131.1174 (0)</td>
<td>-4</td>
<td>3.8</td>
<td>0.17</td>
<td>114 (100)</td>
<td>amino acid degradation</td>
<td>48</td>
</tr>
<tr>
<td>N1-methylnicotinamide</td>
<td>137.0693 (100)</td>
<td>-12</td>
<td>2.4</td>
<td>1.36</td>
<td>94 (23), 92 (6), 78 (8), 65 (4)</td>
<td>vitamin B3</td>
<td>48</td>
</tr>
<tr>
<td>N1-methyl-2-pyridone-5-carboxamide</td>
<td>153.0672 (100)</td>
<td>9</td>
<td>5.7</td>
<td>1.13</td>
<td>154 (16), 136 (76), 108 (26), 92 (15), 80 (16)</td>
<td>NAD degradation</td>
<td>48</td>
</tr>
<tr>
<td>1-methyl xanthine</td>
<td>167.0571 (100)</td>
<td>4</td>
<td>9</td>
<td>0.29</td>
<td>110 (66)</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>7-methylxanthine</td>
<td>167.0571 (100)</td>
<td>4</td>
<td>7.6</td>
<td>0.28</td>
<td>150 (29), 124 (43), 110 (29)</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>diacetylputrescine</td>
<td>173.1297 (0)</td>
<td>7</td>
<td>8.4</td>
<td>0.09</td>
<td>131 (66), 114 (100), 72 (33)</td>
<td>amino acid degradation</td>
<td>48</td>
</tr>
<tr>
<td>cotinine</td>
<td>177.1015 (100)</td>
<td>-4</td>
<td>15.6</td>
<td>0.26</td>
<td>80 (64), 71 (45), 98 (34), 114 (34), 177 (18), 146 (18)</td>
<td>nicotine</td>
<td>48</td>
</tr>
<tr>
<td>sulforaphan</td>
<td>178.0364 (0)</td>
<td>5</td>
<td>17.9</td>
<td>1.84</td>
<td>72 (100), 114 (81), 55 (25), 119 (16)</td>
<td>cruciferous vegetables</td>
<td>48</td>
</tr>
<tr>
<td>hippuric acid</td>
<td>180.0681 (0)</td>
<td>14</td>
<td>4.9</td>
<td>0.26</td>
<td>105 (100), 77 (31)</td>
<td>phenolic compounds</td>
<td>77</td>
</tr>
<tr>
<td>Paraxanthine (1,7-dimethylxanthine)</td>
<td>181.0739 (100)</td>
<td>10</td>
<td>14.2</td>
<td>1.57</td>
<td>124 (58), 163 (2), 142 (3), 96 (8)</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>Theophyline (1,3-dimethylxantine)</td>
<td>181.0739 (100)</td>
<td>10</td>
<td>14.8</td>
<td>0.30</td>
<td>124 (63), 69 (25)</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>Theobromine (3,7-dimethylxanthine)</td>
<td>181.0739 (100)</td>
<td>10</td>
<td>12.5</td>
<td>0.74</td>
<td>163 (11), 138 (21), 137 (15), 124 (24), 69 (18)</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>caffeine (1,3,7-trimethylxanthine)</td>
<td>195.0868 (100)</td>
<td>-5</td>
<td>16.9</td>
<td>0.20</td>
<td>138 (56), 83 (33), 110 (22),</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>dimethyluric acid (1,3 or 1,7)</td>
<td>197.0685 (100)</td>
<td>8</td>
<td>4.3</td>
<td>0.03</td>
<td>169 (40), 140 (38)</td>
<td>caffeine</td>
<td>110</td>
</tr>
<tr>
<td>tryptophan</td>
<td>205.0994 (9)</td>
<td>11</td>
<td>10.8</td>
<td>0.69</td>
<td>188 (100), 146 (54), 118 (45), 115 (31), 91 (14), 117 (18)</td>
<td>amino acid</td>
<td>77</td>
</tr>
<tr>
<td>zeatin</td>
<td>220.1179 (100)</td>
<td>-6</td>
<td>15.5</td>
<td>0.11</td>
<td>148 (38), 136 (13), 119 (13)</td>
<td>cytokinin</td>
<td>129</td>
</tr>
<tr>
<td>Compounds</td>
<td>Molecular mass (m/z)</td>
<td>Charge</td>
<td>Intensity</td>
<td>Neutral Loss</td>
<td>Spectra</td>
<td>Information</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1-ribosylnicotinamide</td>
<td>255.0935 (0)</td>
<td>-2</td>
<td>2.6</td>
<td>0.07</td>
<td>123 (100), 106 (15), 133 (15), 69 (15)</td>
<td>vitamin B3</td>
<td></td>
</tr>
<tr>
<td>phenylacetylglutamine</td>
<td>265.1208 (0)</td>
<td>9</td>
<td>6.7</td>
<td>0.50</td>
<td>130 (100), 136 (26), 147 (21), 91 (32), 84 (60)</td>
<td>modified amino acid</td>
<td></td>
</tr>
<tr>
<td>dibutylphthalate</td>
<td>279.1573 (0)</td>
<td>-6</td>
<td>39.9</td>
<td>0.19</td>
<td>149 (100), 121 (23), 93 (13), 65 (3), 57 (9)</td>
<td>plasticizer</td>
<td></td>
</tr>
<tr>
<td>riboflavin</td>
<td>377.1480 (100)</td>
<td>6</td>
<td>20</td>
<td>0.16</td>
<td>243 (36), 200 (21), 198 (12), 172 (15), 69 (9)</td>
<td>vitamin B2</td>
<td></td>
</tr>
<tr>
<td>S-adenosyl-methionine</td>
<td>399.1390 (0)</td>
<td>-14</td>
<td>8.1</td>
<td>0.02</td>
<td>250 (100), 298 (60), 264 (40), 136 (40)</td>
<td>endogenous</td>
<td></td>
</tr>
</tbody>
</table>

*a* value in brackets represents the % relative spectral intensity of the precursor ion in the product ion spectra

*b* Intensity relative to 7 - methylguanine in MS mode

*c* no CID spectrum has previously been published for this analyte
Urine contains many compounds ingested as part of diet or lifestyle. Of these, the most numerous and abundant detected in the current urine samples were metabolites of caffeine (1,3,7-trimethylxanthine) which was also detected as an un-metabolized species. The CID reactivity of 15 purine derived caffeine metabolites has recently been reported and was instructive in identifying those isomers of methyl and dimethylxanthine in Table 2.3. Isomeric species included, 1- and 7-methylxanthine, 1,3- and 3,7- dimethylxanthine and were each easily distinguished from one another based on this previous gas phase chemistry study. A compound with a measured accurate mass corresponding to dimethyluric acid showed very low abundance with product ions consistent with both 1,3- and 1,7- dimethyluric acid. However, at the low spectral intensity in MS/MS it was not possible to distinguish the methylation sites. Other exogenous metabolites identified primarily by spectral matching include metabolites of nicotine, phenolic compounds and plant derived metabolites. One such species, zeatin, is a known plant hormone and has previously been analyzed by ESI-MS/MS but has not been reported in urine.

Endogenous metabolites detected and listed in Table 2.3 include amino acids, their degradation products and products of nicotinamide adenosine dinucleotide (NAD\(^+\)) metabolism. One of these, N1-methyl-2-pyridone-5-carboxamide (1-MPCA) is a known urinary metabolite which has been analyzed by LC with optical detection but for which no published CID spectrum could be found. This abundant \(m/z\) 153 ion has been identified based on its MS/MS spectral data (Figure 2.5A) which shows several characteristic product ions consistent with the proposed structure.
Figure 2.5. Collision induced dissociation spectra from non-target LC-ESI-MS/MS analysis of urine for identified metabolites N1-methyl-2-pyridone-5-carboxamide (A), diacetylputrescine (C), pirrolidine-2-carboxamide (D) and literature spectrum\textsuperscript{77} for acetylputricline (B) used in the identification of diacetylputrescine.
The proposed dissociation pathway for 1-MPCA is shown in Scheme 2.4A. The most favorable position for protonation of 1-MPCA in the gas phase is likely the amide carbonyl group because of the high potential for resonance stabilization of the charge from both amino nitrogens through the conjugated system of $\pi$ electron density. Upon excitation, proton transfer to the amide nitrogen leads to ammonia elimination which forms the most abundant product ion in the MS/MS spectrum in Figure 2.5A. The charge on the dissociation product at $m/z$ 108 form elimination of ammonia followed by CO is also resonance stabilized. Further elimination of CO from $m/z$ 108 can produce a highly stabilized ion detected at $m/z$ 80. An alternative pathway (Scheme 2.4B) accounts for the fragment ion detected at $m/z$ 92 (Figure 2.5A) and involves the sequential elimination of isocyanic acid (HNCO) and water. This process occurs by elimination of the carbamoyl side group forming an unstable intermediate which undergoes elimination of water to form the stabilized ion at $m/z$ 92.

Another diagnostic spectral feature in the MS/MS spectrum of 1-MPCA (Figure 2.5A) is the presence of a product ion at the mass of 1 Da higher than the precursor ion. This product ion can be attributed the ion-neutral reaction between the $m/z$ 136 fragment ion and residual water present in the collision cell as shown in Scheme 2.4C. Similar ion-neutral reactions between product ions with a water or methanol molecules are reported in the recent literature, but overall occur for a very small proportion of ions. Various mechanisms for this water adduction are proposed but little rationale has been provided for the selectivity of the reaction. In most cases examined, the common structural feature of ions reacting with residual solvent molecules is the ketene group, as shown for 1-MPCA in Scheme 2.4C. The ketene group bears a high positive charge density on the carbonyl carbon which is a target for nucleophilic attack by a
water molecule. The product of this ion-neutral reaction is the highly resonance stabilized carboxylic acid detected at \( m/z \) 154 in Figure 2.5A.

**Scheme 2.4.** The gas phase chemistry of protonated N1-methyl-2-pyridone-5-carboxamide during collision induced dissociation.
N-acetylputrescine (NAP) is an endogenous metabolite of putrescine, a small polyamine found in all cells as well as at high levels in foods. The CID spectrum of this well characterized endogenous and dietary metabolite is available in metabolite databases. One such database MS/MS spectrum is shown in Figure 2.5B. NAP was detected and identified in urine based on its accurate mass and the detection of a single, low abundance, product ion at \( m/z \) 114 corresponding to loss of ammonia, and the database spectrum in Figure 2.5B.\(^{77}\) In addition to NAP, its structural analogue, \( N,N'\)-diacetylputrescine (DAP), has also been identified in urine. This compound is a known metabolite but no CID spectrum for this compound has been reported.\(^{137}\) The database spectrum of NAP\(^ {48,128}\) (Figure 2.4B) was initially used to determine CID reactions of protonated NAP, which are presented in Scheme 2.5A. Charge redistribution from the likely O-protonation site and proton transfer to the terminal amino group leads to ammonia elimination and the formation of the resonance stabilized ion detected at \( m/z \) 114 as shown in Scheme 2.5A. A critical step in gas phase ion reactivity of heteroatom containing even electron species during CID is their tendency to tautomerize during collisional excitation prior to dissociation.\(^ {94,95}\) An alternative process of ammonia elimination is shown in Scheme 2.5B and is analogous to Scheme 2.5A, but proceeding from the acetylamino tautomer. This process yields a less stable product ion which reacts further to eliminate neutral ketene (-42 Da) to form the ion detected at \( m/z \) 72 (Scheme 2.5B). The reaction of ketene elimination proposed in Scheme 2.5B could proceed through an ion-neutral complex between a neutral imine and a charged acetyl group followed by proton abstraction to form a protonated imine.
Based on the reactivity established for NAP (Schemes 2.5A and 2.5B), its diacetylated analogue DAP has been detected and identified in urine. The carbonyl-O protonated protomer of DAP has a potential for intra-ionic interaction between the two terminal acetylamino groups through hydrogen bonding (Scheme 2.5C). Similar intra-ionic charge stabilization by heteroatomic functional groups separated by a hydrocarbon chain has been reported previously in the reactivity of deprotonated di-carboxylic acids.\textsuperscript{138} This stabilizing interaction facilitates proton transfer to the carbonyl oxygen of the non-charge bearing acetyl group during elimination of
ketene from the charge bearing acetyl group to form an ion with \( m/z \) 131 which is identical in structure and reactivity to protonated NAP.

Also detected in non-target experiments was an ion at \( m/z \) 115.0857 which corresponds to an ionic formula of \( \text{C}_5\text{H}_{13}\text{N}_2\text{O}^+ \) (10 ppm) for which no match could be found in database or literature spectra. The MS/MS spectrum acquired by DDA for this ion is presented in Figure 2.5C and shows neutral losses corresponding to elimination of both ammonia and CO (- 45 Da), elimination of water and a low abundance peak for ammonia loss. The composition of the most abundant product ion at \( m/z \) 70.07 is the same as the diagnostic product for proline containing cyclic peptides described in section 3.3. Unmodified protonated proline has a molecular formula of \( \text{C}_5\text{H}_{13}\text{NO}_2^+ \) and a nominal \( m/z \) of 114 and would be expected to show loss of water as well as loss of water and CO to form the product ion at \( m/z \) 70.07, which is the cyclic immonium ion of proline. Thus, the ion detected at \( m/z \) 115.0857 is the amide derivative of proline called prolamide (pyrrolidine-2-carboxamide). Scheme 2.6 illustrates the dissociation reactions of protonated prolamide. Despite the primary amino group in its structure, ammonia elimination from prolamide is less favorable than water elimination. This is because of the difference in the stabilization of the developing charge during dissociation between the two tautomers which lead to ammonia and water elimination as shown in Scheme 2.6A and 2.6B, respectively. The product of ammonia elimination (Scheme 2.6A) undergoes a further elimination of CO to produce the abundant ion at \( m/z \) 70. Prolamide is neither a known urinary metabolite nor otherwise reported in the literature as an individual modified amino acid.
2.3.5. Identification of Nucleic Acid Bases and Nucleosides

The detection of urinary modified nucleosides and NA bases has been the focus of many years of research, beginning in 1898 when Kruger and Salomon first isolated 7-methylguanine (7-MeG) after fractionating 10 000 L of urine.¹¹⁶,¹¹⁷ Nucleosides and NA bases are present in urine either as degradation products of functionally modified ribonucleic acids (RNA)¹³⁹ or as products of spontaneous depurination or enzymatic repair of chemically damaged DNA¹¹⁷. Urinary products of RNA degradation are of interest as possible biomarkers of cancer,¹¹⁷ metabolic disorders¹³⁹ and chemical damage to RNA¹⁴⁰. Urinary products of DNA damage are of interest as biomarkers of exposure to or effect of known mutagens whose mechanism of action is through chemical damage to DNA.¹¹⁷ Since this early work, targeted analytical methods have detected many nucleoside adducts in in vitro studies of various carcinogens. More recent
attempts have also been limited to the identification of well-established functional modifications such as base methylation of RNA nucleosides.\textsuperscript{130,141,142}

Towards identification of new modified analogues, a large amount of work has gone into establishing the gas phase reactivity of unmodified\textsuperscript{93,106–108} and modified\textsuperscript{93,94,106–109} NA bases and nucleosides during collision induced dissociation (CID). An understanding of this reactivity can be used to fully elucidate the structure of a modified nucleoside, including the nature and site of chemical modification. Conjugated heterocyclic compounds such as NA bases are known to have particularly complex reactivity during CID, which has made them the subject of much study in this area.\textsuperscript{93,106–108,143} For some NA bases, such as adenine and cytosine, isotopic labeling studies have allowed for the relatively straightforward elucidation of these pathways, most of which often result in multiple isomers of product ions with the same elemental composition at each \textit{m/z}. In other cases such as guanine and uracil, even extensive isotopic labeling CID studies have been unable to completely elucidate dissociation pathways, which are just now beginning to be better understood.\textsuperscript{95,144}

In addition to the characteristic sugar loss pathway, which is detected at low collision energy (CE), spectral detail required to identify the site of base modification can only be obtained at higher collision energy. The data dependent LC-ESI-MS/MS method developed here addresses this by utilizing an alternating CE of 15 and 30 V. Both these pathways could be observed using these alternating collision energies. Thus it was possible to detect several urinary nucleosides as well as nucleic acid bases, elucidate the position of most of their modifications dependent on ion intensity as well as identify nucleosides with sugar modifications. A large body of literature including gas phase ion chemistry studies and targeted urine analysis allowed us to identify many of the detected species once they were identified as being part of this class of
compounds. Many of the nucleic acid related compounds identified have been previously reported either in urine\textsuperscript{130,141,142} or in cellular RNA\textsuperscript{139}. Details of the identification of these compounds are presented in Table 2.4, along with references to published spectra where available.

In addition to nucleosides with modified base moieties, several RNA nucleosides with modified ribose moieties were also detected based on the neutral loss of a modified sugar. These include sugar methylation products of O’-methylguanosine, O’-methylcytosine, O’-methyladenosine, O’-methylinosine, O\textsuperscript{5’}-methylthioadenosine (MTA) and the sugar dehydration product 5’-dehydro-2’-deoxyinosine (O’dh-dI). By CID of protonated nucleosides, it has not been possible to differentiate between the site of protonated sugar methylation products of nucleosides with neutral loss of methylribose (-146 Da) being the only diagnostic fragmentation process reported involving the sugar.\textsuperscript{109} Under the current dissociation conditions, protonated dehydrated methyl ribose at \textit{m/z} 147 was observed as a minor product ion along with ions at \textit{m/z} 129 and \textit{m/z} 115 from its fragmentation, but it was not possible to distinguish isomers based on this reactivity.
<table>
<thead>
<tr>
<th>Analyte</th>
<th>measured m/z (MS/MS intensity)</th>
<th>mass error (ppm)</th>
<th>retention time (min)</th>
<th>relative intensity</th>
<th>Product Ion m/z (MS/MS relative intensity)</th>
<th>Spectrum Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>methylcytosine (3 or 5)c</td>
<td>126.0661 (100)</td>
<td>-0.8</td>
<td>2.8</td>
<td>0.12</td>
<td>109 (80), 81 (20)</td>
<td>107</td>
</tr>
<tr>
<td>1-methyladenine</td>
<td>150.0794 (100)</td>
<td>13.3</td>
<td>4</td>
<td>0.05</td>
<td>119 (10), 109 (6), 94 (6), 92 (6), 82 (6)</td>
<td>108</td>
</tr>
<tr>
<td>7-methylguanine</td>
<td>166.0711 (100)</td>
<td>-7.2</td>
<td>7.1</td>
<td>1.00</td>
<td>149 (42), 124 (32), 96 (11), 69 (9), 55 (4)</td>
<td>136</td>
</tr>
<tr>
<td>?'-dehydro-?'-deoxyadenosine</td>
<td>250.0898 (20)</td>
<td>-14.8</td>
<td>14.4</td>
<td>0.11</td>
<td>136 (100), 119 (20), 232 (13)</td>
<td></td>
</tr>
<tr>
<td>5O'-dehydro-2'-deoxyinosine</td>
<td>251.0807 (14)</td>
<td>12.7</td>
<td>11.3</td>
<td>0.17</td>
<td>137 (100), 119 (14), 110 (14), 97 (7), 94 (7)</td>
<td>142</td>
</tr>
<tr>
<td>N4'-methylcytidine</td>
<td>258.1091 (66)</td>
<td>2.3</td>
<td>7.8</td>
<td>0.29</td>
<td>126 (100)</td>
<td>107</td>
</tr>
<tr>
<td>N3'-methylcytidine</td>
<td>258.1105 (0)</td>
<td>7.7</td>
<td>5.9</td>
<td>0.07</td>
<td>126 (100), 109 (9), 95 (3), 223 (5)</td>
<td>107</td>
</tr>
<tr>
<td>O'-methylcytidine</td>
<td>258.1158 (0)</td>
<td>-14.3</td>
<td>2.7</td>
<td>0.01</td>
<td>112 (100), 241 (22), 94 (22)</td>
<td>109</td>
</tr>
<tr>
<td>adenosine</td>
<td>268.1059 (28)</td>
<td>7.1</td>
<td>12.1</td>
<td>0.21</td>
<td>136 (100)</td>
<td>48</td>
</tr>
<tr>
<td>?'-methyladenosine</td>
<td>282.1208 (19)</td>
<td>3.9</td>
<td>10.2</td>
<td>0.17</td>
<td>150 (100), 100 (5), 85 (5)</td>
<td>108</td>
</tr>
<tr>
<td>N1'-methyladenosine</td>
<td>282.1220 (13)</td>
<td>8.2</td>
<td>10.6</td>
<td>0.70</td>
<td>150 (100), 133 (2), 109 (1), 94 (1)</td>
<td>108</td>
</tr>
<tr>
<td>N1'-methyl inosine</td>
<td>283.1058 (4)</td>
<td>5.7</td>
<td>10.8</td>
<td>0.46</td>
<td>151 (100), 137 (6), 110 (8), 94 (3), 124 (2), 100 (2), 97 (2),</td>
<td>141</td>
</tr>
<tr>
<td>O'-methylinosine</td>
<td>283.1058 (100)</td>
<td>5.7</td>
<td>12.8</td>
<td>0.14</td>
<td>137 (50), 109 (16), 266 (16), 84 (14),</td>
<td>142</td>
</tr>
<tr>
<td>xanthosine</td>
<td>285.0830 (0)</td>
<td>0.0</td>
<td>10.2</td>
<td>0.03</td>
<td>153 (100), 110 (100)</td>
<td>141</td>
</tr>
<tr>
<td>N-acetylcytidine</td>
<td>286.1067 (19)</td>
<td>11.5</td>
<td>11.7</td>
<td>0.29</td>
<td>112 (100), 154 (81), 94 (11), 69 (7)</td>
<td>141</td>
</tr>
<tr>
<td>O',?'-dimethyladenosine</td>
<td>296.1384 (50)</td>
<td>10.1</td>
<td>17.4</td>
<td>0.19</td>
<td>150 (100), 110 (20)</td>
<td>-</td>
</tr>
<tr>
<td>Compound</td>
<td>Mass (m/z)</td>
<td>m/z 0.1</td>
<td>m/z 0.5</td>
<td>m/z 1.0</td>
<td>Precursor Ions</td>
<td>Notes</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>-------------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>N3-methylguanosine</td>
<td>298.1170 (3)</td>
<td>6.7</td>
<td>11.1</td>
<td>0.37</td>
<td>166 (100), 149 (9), 135 (5), 114 (3), 110 (2)</td>
<td></td>
</tr>
<tr>
<td>O'-methylguanosine</td>
<td>298.1170 (3)</td>
<td>6.7</td>
<td>11.2</td>
<td>0.30</td>
<td>152 (4)</td>
<td></td>
</tr>
<tr>
<td>N1-methylguanosine</td>
<td>298.1179 (20)</td>
<td>9.7</td>
<td>11.7</td>
<td>0.15</td>
<td>166 (100), 149 (30), 100 (20)</td>
<td></td>
</tr>
<tr>
<td>5'-deoxy-O5'-methylthioadenosine</td>
<td>298.0999 (44)</td>
<td>10.4</td>
<td>18.5</td>
<td>0.84</td>
<td>136 (100), 145 (22), 97 (22)</td>
<td></td>
</tr>
<tr>
<td>dimethyl guanosine</td>
<td>312.1328 (10)</td>
<td>8.0</td>
<td>12.2</td>
<td>0.17</td>
<td>180 (100), 136 (5), 135(5), 110 (7)</td>
<td></td>
</tr>
<tr>
<td>N2,N2-dimethylguanosine</td>
<td>312.1328 (6)</td>
<td>8.0</td>
<td>13.4</td>
<td>0.77</td>
<td>180 (100), 137 (3), 135 (3), 110 (2), 85 (1)</td>
<td></td>
</tr>
<tr>
<td>N2,N2,N7-trimethylguanosine</td>
<td>326.1468 (3)</td>
<td>2.8</td>
<td>17.7</td>
<td>6.93</td>
<td>194 (100), 124 (4), 149 (3), 165 (1), 151 (1), 142 (1)</td>
<td></td>
</tr>
<tr>
<td>trimethylguanosine</td>
<td>326.1472 (6)</td>
<td>4.0</td>
<td>17.3</td>
<td>0.06</td>
<td>194 (100), 124 (3), 149 (3), 165 (3), 179 (3), 167 (3)</td>
<td></td>
</tr>
</tbody>
</table>

a value in brackets represents the % relative spectral intensity of the precursor ion in the product ion spectra
b Intensity relative to 7-methylguanine in MS mode
c Identified as N3-methylcytocine based on nontarget ESI-FAIMS-MS work in section 4 of this thesis.
d methylation site established is targeted experiment in Figure 2.6
Also detected were two nucleosides with modified sugars that have not previously been reported in urine. These are close analogues to previously reported species and were identified based on their MS/MS spectra which showed analogous fragmentation patterns to those of previously reported close analogues. Adenosine methylated on both the sugar and on the base is a known modification in functional RNA but has not been reported as a urinary nucleoside. The dehydration product of adenosine (dh-Ado), the adenosine analogue of the previously reported O’dh-dI was also detected and has not previously been reported as a NA modification. The spectral intensity and number of product ions detected for dh-Ado was not sufficient to establish the site of modification, and targeted sample preparation or analysis would be required in order to obtain better quality spectral data that could do this. Upon closer examination of the previous report of O’dh-dI identification it is evident that the position of deoxygenation (2’) and dehydration (5’) reported was speculative and not based on spectral data. Since it is likely that these nucleosides represent the dehydration products of RNA nucleosides rather than the dehydrogenation product of a 2’-deoxynucleosides from DNA, the position of the sugar modifications remains unassigned.

As shown in Table 2.4, the intensities of most detected nucleosides and NA bases was very low compared to most of the other species identified. These ion intensities extended to and below the levels where MS/MS spectral data suitable for unknown identification could be collected by the data dependent LC-ESI-MS/MS method. Even using alternating collision energies, relative intensity of peaks arising from base fragmentation of nucleosides were in some cases low compared to the sugar loss peak. This meant that the short length of time a precursor ion was selected for dissociation in MS/MS by the DDA method was not always sufficient to provide the spectral detail for elucidation of the site of base modification. The limit on analysis
time imposed by the width of the eluting chromatographic peak is a fundamental limitation to sensitivity in LC-ESI-MS. The non-target DDA method samples each selected ion for 3 seconds. This allows for a large number of analytes to be selected for MS/MS analysis and as demonstrated here many of the most abundant ions can be identified. As shown in Figure 2.1B, the width of the chromatographic peaks using this method was on the order of 1 min. Thus by extending the sampling time it is possible to obtain better quality MS/MS spectra but for a smaller number of analytes. To demonstrate the implication of sampling time on spectral quality, results from the DDA method were compared to those from a targeted data acquisition method developed for the analysis of modified nucleosides for which spectral quality was too low to determine their structural features such as the site of methylation using DDA. This targeted data acquisition method operates at a fixed precursor $m/z$ for the whole chromatographic run and allows for data acquisition across the entire chromatographic peak but only for one preselected $m/z$ corresponding to the target precursor ion. This represents the upper limit to intensity of MS/MS spectral data that can be acquired using the current instrumentation from a given sample without resorting to highly targeted sample preparation.

Figure 2.6 shows MS/MS spectra from LC-ESI-MS/MS for data dependent and targeted data acquisition methods during the detection of base methylated cytidine and guanosine detected at $m/z$ 258 and $m/z$ 298, respectively. There are two important differences in the spectral data from the DDA method (Fig 6A and C) and those from the targeted method (Figure 2.6B and 2.6D). First, the spectral intensity is around 50 fold greater using the targeted method than in the DDA method, which corresponds approximately to the increase in analysis time for each compound. More important for identification than absolute spectral intensity was the significant increase in the number of product ions that is observed with increasing spectral intensity. Thus,
elucidation of structural features of methylated nucleosides is more feasible using the more detailed MS/MS spectral data in Figures 2.6B and 2.6D. The second difference between MS/MS spectral data presented in Figure 2.6 is the high level of noise which can be observed using the targeted method as in the spectra in Figures 2.6B and 2.6D. This high background, which was below the detection limit of the DDA method (Figs. 2.6A and 2.6C), consists of a poorly characterized mixture of sample components, buffer, solvents and system contaminants as well as their fragment ions, adducts and products of electrochemistry during ESI. This relatively high background is characteristic for liquid separation and ESI and represents another fundamental limitation of LC-ESI-MS/MS for non-target analysis. In MS mode this results in detectable signal at every nominal mass which, as can be seen in Figure 2.6 results in the detection of product ion background in MS/MS spectra. For identification of unknown species based on MS/MS spectral data this product ion interference can represent a significant challenge because identification relies on high quality of spectral data from MS/MS.

**Figure 2.6.** Comparison of MS/MS spectra from data dependent acquisition (A, C) and targeted data acquisition (B, D) LC-ESI-MS/MS method analysis of N4-methylcytidine (4-MeCyt) (A, B) and N1-methylguanosine (1-MeGuo) (C, D) in urine. Protonated nucleosides of m/z 258 for 3-MeCyt and m/z 298 for 1-MeGuo were selected as precursor ions and readily undergo loss of 132 Da deoxyribose moiety.
Based on the MS/MS spectral data from the targeted method (Figure 2.6B and 2.6D), the methylation site of methylcytosine (Figure 2.6B) and methylguanosine (Figure 2.6D) could be established. Spectral data for isomers of methylated cytosine has been published and can be used for spectral matching. Section 3.3 of this thesis presents a detailed description of the gas phase ion chemistry involved in the differentiation of isomers of methylated cytosine detected in urine. It is known that the pathway leading to the elimination of water is significantly suppressed in 3-methylcytosine due to the change in the conjugated system of $\pi$ electron density compared with cytosine and N$^4$-methylcytosine. For O$^2$-methylated cytidine, water elimination is also not observed but a methanol loss occurs. For C-5 or C-6 methylated pyrimidines, HNCO loss is prominently observed compared with H$_3$CNCO loss. The MS/MS spectrum in Figure 2.6B shows product ions corresponding to loss of water at $m/z$ 108 and loss of H$_3$CNCO at $m/z$ 95 from the precursor ion, which indicates the unknown species is N$^4$-methylcytidine. N$^4$-methylcytidine is a less abundant metabolite than its isomer N3-methycytidine (Table 2.4) which was detected at a different retention time and identified solely based on the DDA data due to absence of a product ion at $m/z$ 108 in its MS/MS spectrum. In the case of methylated guanosine (Figure 2.6D), the MS/MS spectrum from targeted data acquisition contains product ions at $m/z$ 109 and $m/z$ 110 which correspond to elimination of CH$_3$NCO and methylated CH$_3$NCNH from the pyrimidine ring. It is known that scrambling of N1 and N$^2$ in methylated guanosine occurs and could lead to elimination of methylamine from N1-methylguanosine. However, the very low relative abundance of this product ion at $m/z$ 135 (< 0.1 \%) in the spectrum in Figure 2.6D combined with the detection of $m/z$ 109 and $m/z$ 110 indicates the detection of 1-methylguanosine.
It is important to realize that the structure of these two isomers could only be established based on MS/MS spectral data from the targeted method because their MS/MS spectral data from CID are available from previous studies using authentic standards.\cite{106,107} Otherwise, the high background observed in the spectra would have made determination of the methylation site nearly impossible. In de novo identification the spectral quality is particularly important because it is not possible to distinguish between product ions from the precursor and background ions from ubiquitous noise. In Chapter 3 of this thesis the use of an alternative separation technique to LC called high-field asymmetric waveform ion mobility (FAIMS) is investigated which offers a potential solution to the limitations of LC in non-target analysis identified here. Understanding the reactivity of NA bases is crucial for the identification of modified product of nucleic acids. Chapter 4 of this thesis presents extensive study of another NA base, uracil, whose reactivity during CID was still relatively poorly understood.

4.0 Conclusion

This work has demonstrated the capabilities and limitations of using LC-ESI-MS/MS with data dependent acquisition for non-target analysis of a complex urine sample. In a single chromatographic run, this technique is capable of providing accurate and tandem MS data allowing for the identification of over one hundred distinct chemical species in the sample without the use of authentic chemical standards. Instead, metabolite identification can be carried out based solely on spectral data from mass spectrometry using an integrated approach which involves (a) searching spectral databases, (b) determining analyte class followed by literature confirmation (c) de novo identification based on manual interpretation of spectral data. The biggest drawback to this identification approach is the massive amount of time and expertise it
currently requires. As better tools are developed for spectral interpretation and database searching of spectra, this type of non-target identification based on MS spectral data will only be facilitated.

The non-target instrumental method that was developed used DDA as a way of collecting accurate and tandem mass data for a large number of the most abundant ions during a chromatographic run. For the more abundant ions detected, such as acylcarnitines, this method provided high quality spectral data suitable for de novo structural elucidation. Other less abundant ions such as some nucleosides were selected for MS/MS sampling but did not provide sufficient spectral quality to allow for their complete structural elucidation. This can be attributed to the limited length of time available for MS analysis imposed by the width of an eluting peak in LC, a fundamental limitation of column chromatography in non-target analysis. Targeted data acquisition methods were able to provide higher intensity spectral data which allowed for the identification of some of these compounds, but at the cost of the broad specificity of the DDA method. Another limitation to unknown identification by LC-ESI-MS was found to be the high ESI background which leads to product ion interference in MS/MS spectra that can hinder identification, especially for low abundance metabolites. An alternative method of analytical separation which overcomes these limitations for non-target analysis is explored in Chapter 3 of this thesis.

This work represents the most comprehensive report to date on the detection and identification of urinary metabolites using LC-ESI-MS. This information will be useful for all those developing methods for urine analysis using similar techniques as it provides valuable information about the abundant compounds present in this complex and frequently analyzed biological matrix. The integrated approach for identification of non-target metabolites based
solely on their MS spectral data which was presented will be instructive to those in the field of metabolomics in search of new approaches for identifying the species they detect.
Chapter 3

Non-Target Analysis of Urine by Electrospray Ionization - High Field Asymmetric Waveform Ion Mobility - Tandem Mass Spectrometry

Abstract

Nearly a decade after first commercialization, high field asymmetric waveform ion mobility spectrometry (FAIMS) has yet to find its place in routine chemical analysis. Prototypes have been used to demonstrate the utility of this separation technique combined with mass spectrometry (MS). Unfortunately, first generation commercial FAIMS instruments have gone practically unused by early adopters. Here, this is shown to be due in part to poor ion transmission in the FAIMS-MS source interface. This chapter presents simple instrumental modifications and optimization of experimental conditions to achieve good sensitivity from the first generation commercial FAIMS device (the Ionalytics Selectra) coupled to a relatively high resolution QTOF-MS. Using nanospray ionization, the non-target analysis of urine by FAIMS with minimal sample preparation is demonstrated for the first time. The unique suitability of electrospray ionization (ESI)-FAIMS-MS for identification of low abundance species such as urinary biomarkers of damage of nucleic acids in a complex biological matrix is shown. The elimination of electrospray background and matrix components by FAIMS and the continuous flow of analytes through FAIMS for accurate and tandem mass analysis produce high quality

\(^{a}\) Chapter 3 is based on the article “Non-Target Analysis of Urine by Electrospray Ionization – High Field Asymmetric Waveform Ion Mobility – Tandem Mass Spectrometry (ESI-FAIMS-MS/MS)” by Daniel G. Beach and Wojciech Gabryelski published in Analytical Chemistry.
spectral data suitable for structural identification of unknowns. These characteristics make ESI-FAIMS-MS suitable for non-target identification, even when compared to LC-ESI-MS.

3.1. Introduction

High field asymmetric waveform ion mobility spectrometry (FAIMS) is an atmospheric pressure gas phase ion separation technique that operates as an ion filter between an electrospray ionization (ESI) source and a mass spectrometry (MS) detector. FAIMS was the first differential mobility separation technique to be combined with MS in the 1990s at the National Research Council of Canada (NRC). The utility of the device as a method of chemical analysis was successfully demonstrated by the NRC team and their collaborators using prototypes of the FAIMS instrument. Unique applications of FAIMS include separations of difficult to resolve isomers or enantiomers as well as analysis of complex environmental samples and mixtures of biomolecules. The most often cited strengths of FAIMS are its orthogonality to the m/z separation in MS, its fast analysis times, and its ability to improve selectivity and detection limits when compared with traditional ESI-MS and LC-ESI-MS methods.

The first commercial FAIMS instrument was marketed as the Ionalytics Selectra almost a decade ago. Despite the potential of this technology, which had been demonstrated with the prototypes, there has been a relative lack of success with this first generation of commercial instruments. The only work to be published using this first generation of commercial instruments involved the analysis of large peptides or the use of a different FAIMS-MS source interface than that provided for interfacing the Ionalytics Selectra to Waters MS instruments used in this study. Since 2005, the FAIMS device has been redesigned by
Thermo for use with Thermo Ion Trap Mass Spectrometers. The availability of these second generation commercial instruments has resulted in a recent resurgence of FAIMS research.\textsuperscript{152–154} This work has shown FAIMS to be a valuable tool for improving detection limits by reducing chemical interference in LC-ESI-MS experiments, particularly in highly targeted MRM methods. This approach, using low resolution MS detection, aims to sensitively and selectively detect a relatively small number of target analytes in complex mixtures. Very recently, and mostly since the publication of this work,\textsuperscript{29} FAIMS has been coupled to ultra-high resolution Orbitrap MS for the analysis of complex mixtures of isolated peptides including difficult to detect post translational protein modifications.\textsuperscript{10,62,72}

This work focuses instead on the use of FAIMS as a primary separation tool for non-target analysis. Here, sensitivity and selectivity for known analytes of interest are still crucial but, as is the case with global metabolomics approaches,\textsuperscript{155} the ideal method should also exhibit specificity for as large a number of analytes as possible. In addition, such techniques need to supply the best possible spectral data to allow for the identification of unknown species. Using this approach, not only analytes of known importance are detected but others that would go unnoticed by target approaches can also be discovered. The best illustration of this approach is the non-target identification of highly polar disinfection byproduct in drinking water by ESI-FAIMS-MS/MS. In addition to known contaminants of concern, such as haloacetic acids, a large number of other small organic acids were identified including abundant species which had not previously been considered as drinking water contaminants. This chapter details the investigation of whether the same strategy could be useful for the even more challenging matrix of a real biological sample. The success of previous ESI-FAIMS-MS/MS in this type of non-target identification was due to a large extent to the coupling of FAIMS to high resolution MS capable
of providing accurate mass measurements for all precursor and product ions.\textsuperscript{30,102} Recently, the majority of FAIMS research involves low resolution MS, and the combination of a commercial FAIMS device with a high resolution mass spectrometer capable of tandem MS is rarely seen in recent publications.\textsuperscript{10,72,156}

The often described merits of FAIMS: resolution, analysis time, and selectivity, are important for any analytical technique, but are arguably not unique to FAIMS. However, there are two characteristics that are unique to FAIMS which make it suitable for the identification of non-target analytes in complex sample matrixes, a formidable challenge for any analytical technique. Chemical background is a significant source of interference in any ESI-MS experiment and occurs as a result of the formation of cluster ions of mobile phase components and system contaminants present at the time of ionization, their electrochemical by-products in ESI, and the products of their dissociation in the MS source.\textsuperscript{3,20,21} In FAIMS, this ESI background is separated from analytes before MS detection, resulting in generally higher signal-to-noise ratio in both MS and MS/MS modes than that of either ESI-MS or LC-ESI-MS.\textsuperscript{157} Unlike column chromatography separations where spectral data has to be acquired during the short elution time of an analyte, the FAIMS analyzer acts as an ion filter, transmitting a continuous flow of analyte ions for mass and tandem mass analysis for a time limited only by sample volume. This more flexible time scale also allows for more detailed investigations of the gas phase chemistry of unknown ions than is possible with LC-MS/MS including optimization of source and CID conditions or conducting energy resolved MS/MS experiments.

The current research interest involves a comprehensive investigation of urinary metabolites including products of nucleic acid (NA) modification. Such biomarkers indicate specific modifications of NA in living organisms and are excreted in urine in the form of
modified nucleosides and modified NA bases.\textsuperscript{140} The central part of this work is developing an analytical strategy for the identification of non-target urinary biomarkers of NA damage, which would rely solely on spectral data from accurate and tandem MS to elucidate their molecular structures. LC coupled to high resolution MS represents the most suitable standard technique available for this type of analysis.\textsuperscript{142,155} However, due to the high background of ESI, LC-MS was found in Chapter 2 to be most suitable only for the identification of the most abundant components of urine. In order for non-target identification of lower abundance species to be feasible, a more suitable technique that will allow for acquisition of higher quality spectral data for these analytes is required.

This chapter presents the simple modifications and optimizations required to achieve good sensitivity for the first time for small analytes of the Ionalytics Selectra FAIMS device coupled to a QTOF-MS using the standard commercial interface. This work demonstrates that any prior lack of success in this area is due not to any fundamental limitations of FAIMS but rather to poor engineering of the FAIMS-MS interface. It is shown for the first time that ESI-FAIMS-MS can be used for the analysis of a complex biological sample, urine, with minimal sample preparation. Also in this chapter, the most critical advantage of ESI-FAIMS-MS/MS over LC-ESI-MS/MS in non-target analysis are identified as the quality of the spectral data that can be acquired for low abundance analytes. This work will be of interest to those in the metabolomics field who are in search of better tools for identification of unknown metabolites

3.2. Experimental

Ammonium acetate and HPLC grade water and methanol were purchased from Fisher Scientific (Nepean, Ontario). The standard solution used for intensity optimization consisted of
~ 500 nM of 1-methylguanine, 9-ethylguanine, 2'-deoxyguanosine, 2'-deoxyinosine, N²-methylguanosine (Sigma-Aldrich, Oakville, ON) and O²'-methylcytidine (R.I. Chemicals, Inc., Orange, CA) in 9:1 methanol:water with 0.1 mM ammonium acetate. An overnight urine sample was collected from a healthy volunteer and centrifuged at 26 000 g. Aliquots (4 mL) of supernatant were loaded onto an activated 200 mg Oasis HLB glass solid phase extraction cartridge (Waters, UK), washed with 4 mL of water and eluted with 4 mL of 70% methanol in water.

For FAIMS analysis, a 10 µL subsample of the eluent was diluted to 1 mL in 9:1 methanol/water with 0.1 mM ammonium acetate. Injections were made using a 50 µL sample loop into a 400 nL min⁻¹ flow of the same composition as the dilution buffer. Ionization was carried out in positive mode using a nano-electrospray source with a spray voltage of 5000 V (4000 V relative to FAIMS curtain plate). Ions were then separated using an Ionalytics Selectra FAIMS analyzer with 20% CO₂ in N₂ as a carrier and desolvation gas and a dispersion voltage of 3900 V. Detection was carried out using a Micromass QTOF micro mass spectrometer (Waters, UK) either in full scan (MS) mode over a range of CV values or in tandem (MS/MS) mode at a particular CV value. In MS mode a cone voltage of 14 V and a collision energy of 4 V were used to minimize ion dissociation in the MS source after FAIMS separation, as well as a scan time of 15 sec. In MS/MS mode, a cone voltage of 25 V and a scan time of 5 sec were used and collision energy (CE) was optimized and is presented in figure captions. External mass calibration was carried out using Glu-Fibrinopeptide and internal mass calibration was carried out using the well characterized ions of known composition detected in urine. Mass accuracies in MS mode were < 10 ppm and those in MS/MS mode ranged from 1 – 120 ppm with a typical value around +/-20
ppm. These represent realistic mass accuracies for the instrument employed and were sufficient to determine the elemental composition of the product ions presented.

The LC-ESI-MS/MS method used for comparison of the quality of spectral data included a nano-flow capillary reversed phase separation with online pre-concentration using the nano-Acquity LC system and MS detection using the same nano-ESI source and QTOF-MS as the FAIMS experiments. The urine sample was the same as the sample used for FAIMS analysis but was 400 times more concentrated. Injections of 3.5 µL were made onto a 10 mm x 180 µm, 5 µm C18 trapping column where it was washed with aqueous mobile phase before being eluted onto a 150 mm x 75 µm, 3 µm analytical column with a 45 min linear gradient to 100% organic mobile phase. This method used data dependent acquisition (DDA) to acquire MS data for all eluting compounds and MS/MS spectra for a few hundred of the most abundant species. Conditions in MS mode were the same as those described above but with a capillary (nanospray) voltage of 3700 V. A 1 sec MS scan was used to determine the three most abundant ions at a particular retention time. These were then automatically selected for MS/MS detection and sampled alternately for a total of 9 sec at a CE alternating between 15 and 30 V. These energies represent conditions optimized to show abundant protonated bases as product ions of modified nucleosides (15V) as well as product ions of base fragmentation useful for establishing the site of base modification (30V). After MS/MS acquisition, an ion was excluded from selection for 30 sec to avoid continuous selection of the same precursors.

3.3. Results and Discussion

When the Ionalytics Selectra was coupled to the QTOF using the supplied source interface and manufacturer recommended specifications practically no ion current was detected.
from a mixture of modified nucleoside and NA base standards. Trace 1 in Figure 3.1 shows the extremely low intensity of total ion current for the standards that was detected when the compensation voltage (CV) was scanned. This trace, referred to as a total ion compensation voltage spectrum, was collected by continuously scanning CV from 0 to 35 V at a rate of 0.4 V min$^{-1}$ with MS acquisition at 0.1 V steps. Even at this low ion current, it was evident that a CV separation of analytes was occurring but analytical sensitivity of the method was too low to carry out any practical analysis. Using the same mixture of standard compounds, optimization of ESI-FAIMS-MS with respect to ion transmission and reproducibility of CV separation was carried out.

The purity of the carrier gas in FAIMS has a significant effect on the instrument’s performance. Even at small amounts, changing concentrations of water vapor has a significant impact on the reproducibility of CV separation because of interactions between ions and water molecules that change their ion mobility at either high or low field strength. Three measures were taken to minimize the impact of residual water vapor on the FAIMS operation in the relatively humid climate of South-Western Ontario. To minimize the water vapor contribution from the carrier gas, two hydrocarbon/moisture traps were installed in series (1:1 by volume molecular sieves: activated carbon). This was critical for stable operation of FAIMS and achieving reproducible CV separations. Second, the configuration of the nano-electrospray source which originally had the spray tip located only a few mm from and in direct alignment with the orifice in the FAIMS curtain plate was modified. This direct mode of spraying resulted in fast contamination of the FAIMS electrodes and the introduction of droplets and solvent vapors into the carrier gas. By installing the sprayer off-axis by 20°, solvent droplets and neutrals strike the curtain plate to one side of the inlet and are not introduced between the FAIMS
electrodes. This modified off-axis configuration resulted in the modest signal increase seen in trace 2 in Figure 3.1 but effectively eliminated contamination problems. This allowed the total carrier gas flow (used in main part for ion desolvation and in small part for ion transport in FAIMS) to be decreased from 2 to 1 L min\(^{-1}\) producing another increase in the intensity of detected ions (trace 2 in Figure 3.1). Finally, including a relatively large proportion of CO\textsubscript{2} (20\%) as a component of the carrier gas gave an increase in signal intensity (trace 4 in Figure 3.1) but also significantly improved reproducibility of CV separations. This can be attributed to the fact that an excess of polar CO\textsubscript{2} will out-compete trace amounts of water vapor for polar interactions with analyte ions, increasing the threshold where water impacts the separation. Together, these modifications have increased the time for which the CV value of an analyte remains constant from just minutes with the default configuration to approximately 5 days, even when using much less expensive liquid nitrogen boil-off.

**Figure 3.1.** Additive effects of sequential optimization of signal intensity for a mixture of standard nucleosides and bases from the manufacturer recommended configuration (1) to an off-axis ESI (2), a carrier gas flow of 1 L/min (3), addition of 20\% CO\textsubscript{2} to the carrier gas (4), and installation of an additional o-ring around the MS-sampling cone (5).
The most significant gain in detection sensitivity (trace 5 in Figure 3.1) was obtained by improving ion transmission at the interface between the FAIMS device and the inlet of the mass spectrometer. Figure 3.2 illustrates the configuration of the commercial interface which consists of the interface lens (a) with a concave orifice on both the side of the FAIMS electrodes (b) held in place by a PEEK spacer (c) and the inner MS sampling cone (d). The lens (a) and the inner cone (d) are both held at the same voltage controlled by the MS software. This interface has been designed with the purpose of analyzing ions of large biomolecules so even at the optimal experimental conditions the ion transmission of small species is very low (trace 4 in Figure 3.1). Several different modifications of the interface were attempted and a simple solution to poor ion transmission has been developed. Plugging the 0.3 mm gap between the interface lens and the inner cone with a 1.5 mm ID x 1 mm Viton O-ring (e) (Able Seal & Design, Concord, ON) has greatly improved transmission of small ions (trace 4 to trace 5 in Figure 3.1). Poor ion transmission of small ions in the commercial source is due to a disruption of the ion beam from the FAIMS outlet (f) to the entrance of the sampling cone (d).

Gas flow in the gap between the interface lens and the sampling cone causes turbulence at the entrance of the sampling cone and significant loss of ions. This is supported by the fact that ion transmission deteriorates significantly even with small increases in gas flow, especially during sampling of smaller ions which exhibit a small momentum during ion transport. Based on the work outlined in this chapter it is possible that the poor transmission of small ions from the FAIMS device into the mass spectrometer is part of the reason for the temporary stagnation of FAIMS research. The simple modification of the commercial interface and the intensity optimization of experimental conditions detailed here provide an increase in signal intensity of
over two orders of magnitude as illustrated in Figure 3.1. Most importantly, with these modifications an investigation of standard compounds and real urine samples was possible.

Figure 3.2 shows the first FAIMS separation of electrospray-generated, protonated ions of standard nucleosides and NA bases at optimized conditions. Such ions exhibit larger ion mobility at a high electric field and are transmitted at positive values of CV. Modified bases with a more rigid and compact structure are separated at higher positive CV than modified nucleosides. The narrow range of CV where modified nucleosides are transmitted suggests that the dangling sugar moiety represents the dominant feature in their mobility separation. Separation properties of nucleosides and bases in FAIMS are important because they can be used for selecting the most appropriate CV conditions at which modified nucleosides, modified bases and other similar metabolites are transmitted during ESI-FAIMS-MS analysis of a real urine sample.

Figure 3.3. Cross section of the ESI-FAIMS-MS interface. Labeled components are (a) FAIMS interface lens, (b) FAIMS electrodes, (c) PEEK spacer, (d) MS sampling cone, (e) added 1.5 mm ID 1 mm O-ring, and (f) FAIMS outlet.
Sample preparation for urine was intended to make it amenable for direct ESI. Centrifugation and SPE were successful at desalting the sample and removing particulate, protein material, and other species that might interfere with the ESI process. In addition, the eluate from SPE was diluted 100:1 with ammonium acetate methanol buffer. The dilution step was introduced to improve ionization efficiency of low abundance sample components. Nanospray has a limited ionization capacity which is accommodated by the most abundant components of the undiluted urine extract. After dilution, the same ionization capacity can also be accommodated by less abundant sample components. Dilution is a viable option in ESI-FAIMS-MS because any loss in intensity resulting from dilution can be easily compensated for by extending the period of data acquisition.

Figure 3.4A shows the total ion current compensation voltage (CV) spectrum for a diluted urine sample analyzed by ESI-FAIMS-MS where protonated ions are transmitted through
FAIMS at CV values between -15 and 35 V. The spectrum in Figure 3.4B shows all ions detected over the CV separation range from 0 to -35 V where most small urinary metabolites identified in Chapter 2 were detected. Notably absent at positive CV values were the abundant acylcarnitines detected previously. The spectrum in Figure 3.4C shows all of the ions transmitted at negative CV values and includes all acylcarnitines identified in Chapter 2 as well as a high chemical background compared to positive values of CV. These spectra give a good indication of the sample’s complexity and the detection capabilities of the technique. Many hundreds of MS peaks have been detected in a urine sample after minimal sample preparation using ESI-FAIMS-MS. The current approach for identification of these ions differs from the typical approach of acquiring a large number of chemical standards and comparing their properties to the ions detected. The spectral data acquired using ESI-FAIMS-MS/MS is of sufficient quality to identify unknowns, often without the need for chemical standards. This is done by combining knowledge from previous publications dealing with the detection or dissociation of targeted analytes, spectral databases, and manual interpretation of MS/MS data as well as by carrying out supplemental gas phase ion chemistry studies when of interest.
Figure 3.4. Urine sample analyzed by ESI-FAIMS-MS. (A) The total ion current CV spectrum from -15 V to 35 V. MS spectra in panels D-G were acquired at the CVs shown in A. (B) All ions detected in the CV range from 0 V to 35 V. (C) All ions detected in the CV range from 0 V to -15 V. The MS spectra at specific CV values showing some of the most abundant metabolites detected including N1-methylnicotinamide at CV = 12.8 V (C), N1-methyl-2-pyridone-5-carboxamide at CV = 11 V (D), creatinine CV = 23.6 V (E), and methyladenosine CV = 3.1 V (F).
Using this approach, it was possible to identify a large number of compounds in urine which represent largely the same species identified in Chapter 2 of this thesis. Many of the most abundant ions correspond to well characterized urinary metabolites that were easily identified by comparing their spectral data to relevant databases and literature reports. The most intense CV peak in Figure 3.4A, centered at CV = 11.0 V, corresponds to protonated N1-methylnicotinamide which was detected at m/z 137.0701 (Figure 3.4D). Another abundant metabolite of nicotinamide adenine dinucleotide (NAD$^+$), N1-methyl-2-pyridone-5-carboxamide, was transmitted through FAIMS at CV = 12.8 and detected at m/z 153.0542 (Figure 3.4E). The most abundant species detected was creatinine. Its protonated ion at m/z 114.0678 (Figure 3.4F) is transmitted in FAIMS as a broad peak at CV = 23.6 V. The detection of creatinine is useful since it serves as a natural internal standard in urine analysis. The most abundant modified nucleoside detected was methyladenosine at m/z 282.1161 (Figure 3.4G) separated in FAIMS at CV= 3.1 V (Figure 3.4A). All ions identified in Figure 3.4 were examined in ESI-FAIMS-MS/MS, and their structures confirmed by MS/MS.

Unlike products of well characterized metabolic processes such as NAD$^+$ metabolism, modified nucleosides such as methyladenosine (Figure 3.4G) or modified NA bases, can appear in various isomeric forms and MS/MS is indispensable in determining their structure. The spectral data for the identification of some of these species is presented in Figure 3.5. The separation of isobaric m/z 298 ions S-methyl-50-thioadenosine (MTA) at CV = 4.0 V and N1-methylguanosine at CV = 6.0 V is shown in Figure 3.5A. MTA is an example of a nucleoside modified at the sugar moiety. Its structure has been determined by accurate mass measurements for the precursor and product ions as well as neutral losses detected in the MS/MS spectrum in Figure 3.5D. CID of protonated MTA involves highly diagnostic product ions of the S-methyl-
5’O-thiobase moiety as well as an abundant protonated adenine peak. N1-Methylguanosine is an example of a nucleoside modified at the base moiety. The MS/MS spectrum of the precursor ion detected at m/z 298.1118 (Figure 3.5E) is characteristic for N1-methylguanosine. The methylation site can be assigned on the basis of the presence of [M+H-132-CH$_3$NCO]$^+$ at m/z 109.0498 and the absence of a [M+H-132-HNCO]$^+$ peak at m/z 123.

Figure 3.5. Identification of modified nucleosides and bases in urine. Extracted ion CV spectra for m/z 298 (A), m/z 312 (B), and m/z 166 (C) showing peaks for four ions identified by accurate and tandem mass spectra including S-methyl-50-thioadenosine at collision energy (CE) = 35 V (D), N1-methylguanosine at CE = 40 V (E), N$^2$,N$^2$-dimethylguanosine at CE = 40 V (F), and N7-methylguanine at CE = 23 V (G).

It is important to note that in order to determine fine structural details of a nucleoside such as a base methylation site, relatively high collision energy (CE) is required to induce base
fragmentation which results in complete precursor dissociation. N²,N²-Dimethylguanosine, separated at CV = 5.4 V (Figure 3.5B) is an example of a nucleoside modified by two methyl groups at the base. On the basis of the MS/MS spectrum of the precursor detected at \( m/z \) 312.1285 (Figure 3.5F), this compound can be distinguished from other isomers by ions \([M + H-132-HNCO]^+\) at \( m/z \) 137.0803 and \([M + H-132-(CH_3)_2NH]^+\) at \( m/z \) 134.0276, which would not both be detected for other isomers of dimethylguanosine.

Figure 3.5C illustrates the separation of isobaric \( m/z \) 166 metabolites which have been identified as N7-methylguanine (CV = 20.1 V), phenylalanine (CV = 16.0 V), and N1-methylguanine (CV = 6.0 V), the latter occurring because of source fragmentation of N1-methylguanosine detected at the same CV value. The MS/MS spectrum of N7-methylguanine (Figure 3.5G) shows a large number of product ions. The most important spectral features for distinguishing this compound from other isomers is the absence of \([M + H – CH_3NH_2]^+\) at \( m/z \) 135 which would be detected for any isomer with a pyrimidine ring methylation site and detection of product ions such as \( m/z \) 96 representing the imidazole ring containing a methyl group at N-7. The fragments detected at \( m/z \) 167.0490 in Figure 3.5E,G are the result of ion-neutral reaction of reactive product ions of methylguanine with residual solvent molecules in the collision cell.

The number of identified compounds presented here is a reflection of the scope of the current work, to demonstrate the unique strengths of FAIMS in non-target analysis, not the limit of our ability to identify unknown species which is described in much more detail in Chapter 2. Using the FAIMS analyzer, high quality accurate and tandem mass data can be collected for any of the many hundreds of ions detected in MS mode, regardless of their intensity in the exploratory CV scan. As can be seen in Figure 3.5D – G, the length of time and CID conditions
required to acquire a high quality MS/MS spectrum is highly variable and depends on the abundance, stability, and reactivity of the precursor ion selected. The tremendous advantage of the FAIMS analyzer is its ability to isolate protonated metabolite ions from the chemical background produced by ESI. This background, composed of cluster ions of mobile phase components, matrix, and system contaminants\textsuperscript{20,21,153} is transmitted at negative CV values and therefore does not account for any interference with ions transmitted at positive values. In direct ESI-MS or LCESI-MS, chemical background at every nominal \( m/z \) limits selectivity. Under these experimental conditions in ESI-FAIMS-MS, the background level has been effectively reduced to the electrical noise of the MCP detector. These background ions are separated at negative CV values in Figure 3.4 while most small urinary metabolites are separated at positive values. This results in a very high signal-to-noise ratio even for low abundance ions and means that very long analysis times can be effectively used to probe these species. This concept and the use of dilution as a method of sample preparation is explored in more detail in Chapter 5 of this thesis. In general, this also leads to better mass accuracies in MS mode when using FAIMS compared to those observed using LC-MS.

In Chapter 2, nanoflow capillary LC coupled to high resolution QTOF-MS using nano-ESI was implemented for the non-target identification of urinary metabolites. This technique represents the most suitable standard instrumentation for non-target analysis of components of complex biological samples. The current work with ESI-FAIMS-MS provides a unique opportunity to compare these two techniques with respect to the quality of spectral data that can be obtained using the same urine sample, the same sample preparation, and the same ionization source and MS detector but two different separation methods. Figure 3.6 shows spectral data that have been acquired for a low abundance \( m/z \) 126 compound in urine by LC-ESI-MS (Figure
In LC-ESI-MS, all sample components were 400 × more concentrated than in the diluted sample used in ESI-FAIMS-MS. The highly polar species of interest is separated in reverse phase LC near the solvent front and is detected at \( m/z \) 126.0675 in MS mode (Figure 3.6A) at moderate intensity along with other coeluting compounds including N1-methylnicotinamide at \( m/z \) 137.0755. When \( m/z \) 126 was selected by the DDA method for MS/MS analysis, the spectrum in Figure 3.6B was acquired. The measured accurate mass and MS/MS data strongly suggest the unknown species is an isomer of methylated cytosine. However, since all known isomers of methylcytosine exhibit loss of ammonia (− 17.03 Da) and carbon monoxide and ammonia (− 45.02 Da), the LC-MS/MS data was not sufficient to rule out even a single isomer.

By scanning CV in ESI-FAI-MS/MS of a diluted subsample of the same urine, the MS peak for this same ion can be detected at CV= 22.0 V and an MS spectrum can be acquired (Figure 3.6C). Even though the intensity of the signal (per unit time) is significantly lower than it was for the concentrated sample analyzed by LC, it is still possible to obtain a more intense MS/MS spectrum because of the unique ability of the FAIMS analyzer to transmit an isolated precursor ion for a practically unlimited length of time. The spectrum collected using FAIMS for methylated cytosine (Figure 3.6D) is the sum of 120 scans over about 10 min when CV was parked at 22.0 V. The FAIMS-MS/MS spectrum shows better intensity, better mass accuracy, and detection of more product ions than did the spectrum collected using LC which represents 3 sec of data acquisition time (Figure 3.6B).
Figure 3.6. Comparison between spectral quality in LC-ESI-MS/MS and ESI-FAIMS-MS/MS showing (A) the LC-MS spectrum of compounds in a urine sample eluting at 2.8 min over 30 s, (B) the MS/MS spectrum of ion B by LC-MS/MS at a collision energy (CE) alternating between 15 V and 30 V for 3 s, (C) the ESI-FAIMS-MS detection of m/z 126 in urine at compensation voltage (CV) = 22 V for 5 min, (D) the MS/MS spectrum of ion D identified as N3-methylcytosine at CE = 18 V and CV = 22 V for 10 min by ESI-FAIMS-MS/MS.

It should be highlighted that MS/MS acquisition conditions, particularly collision energy (CE) for Figure 3.6B,D, differ slightly. A key feature of ESI-FAIMS-MS/MS is the ability to optimize all MS source and dissociation conditions, even for low abundance unknowns in
MS/MS mode. This is not possible using LC-MS without the use of standards. The CE values used in the non-target LC-MS method were optimized using nucleoside and NA base standards to give the best possible signal for protonated bases as product ions of modified nucleosides as well as the best possible spectral detail for fragmentation of protonated modified bases. The nature of non-target analysis by LC-MS limits the amount which a method can be optimized for a specific analyte rather than an entire class of compounds of interest. With the extended analysis time of FAIMS, optimization of all MS/MS conditions was possible even for a low abundance unknown precursor. Such an optimization would not have been possible using only LC-MS. While it would now be possible to create a targeted MS/MS method for the detection of 3-methylcytosine using LC-MS using these MS conditions, this is only the case because of the results acquired using FAIMS.

Each product ion detected in Figure 3.6D represents the product of a unique chemical reaction that can be used to elucidate the structure of the unknown precursor, and the detection of more products is equivalent to more structural information. Equally important is confidence that all detected product ions originate from the precursor ion of interest and not from co-eluting compounds or chemical background. It is common in LC-MS, especially for low abundance precursors sampled for a short period of time, to observe only partial MS/MS spectra and product ions from unidentified interfering species. Identification of an unknown is based on correlation between spectral data and chemical structure of the investigated compound. For positive identification, product ions in tandem MS must be explained in terms of the expected reactivity of the identified compound.
Scheme 3.1. The elimination of water from protonated 3-methylcytosine.

The most important aspect of reactivity of 3-methylcytosine is the absence of a product of water loss. Water elimination from unmodified cytosine (Scheme 3.1A) occurs also for all known methylcytosine isomers except O\textsuperscript{2}-methylcytosine which loses methanol (−32 Da) by an analogous pathway (also not observed) and N3-methylcytosine. The reaction involves initial keto-enol tautomerization and proton transfer from N1 to O\textsuperscript{2} during ion activation. A reactive intermediate ion with a positive charge localized at O\textsuperscript{2} dissociates by heterolytic C2-O\textsuperscript{2} bond cleavage which involves neutral loss of a water molecule and developing a positive charge on C2. The positive charge on C2 can be delocalized throughout the entire ion structure through a conjugated system of double bonds which provides resonance stabilization for the transition state and the product of water elimination. A high resonance stabilization of the transition state during the bond cleavage lowers its activation energy and makes the reaction of water loss favored kinetically. In contrast, methylation at N3 alters the conjugated system of double bonds of the intermediate precursor ion shown in Scheme 3.1B so that the π electrons of N\textsuperscript{4} and all double bonds no longer offer any stabilization to developing charge on C2. The elimination of water
from N3-methylcytosine would have to go through a transition state with highly localized charge making the reaction unfavorable compared to that of unmodified cytosine. Also of great importance are the extra product ions detected in Figure 3.6D compared to Figure 3.6B. Using FAIMS to eliminate this interference in MS/MS mode, more confidence can be had in the significance of low abundance product ions, enabling the *de novo* identification process.

In addition to the absence of diagnostic product ions for water and methanol loss described in Scheme 3.1A, all other product ions for 3-methylcytosine in Figure 3.5D can also be rationalized using principles of gas phase ion chemistry. The gas phase ion chemistry of cytidine has previously been investigated using several isotopically labeled and structurally modified analogues.\(^{142}\) Gas phase reactions of protonated N3-methylcytosine are quite complex because during collisional activation tautomerization, proton transfer, and charge redistribution reactions can occur, all of which are directed by the positive charge. For brevity, only tautomeric forms of precursor ions leading to specific dissociation products are shown in Scheme 3.2. Scheme 3.2 illustrates proposed reactions describing the formation of all product ions detected in Figure 3.5D using similar charge directed mechanism to that of water loss from unmodified cytosine in Scheme 3.1A.
Scheme 3.2. Proposal for formation of product ions detected for 3-methylcytosine.

Scheme 3.2A illustrates formation of the major CID products corresponding to loss of ammonia \((m/z\ 109)\) and methylamine \((m/z\ 95)\) detected in Figure 3.5D. Ring opening at the C2-N3 bond, charge redistribution from N\(^4\), and proton transfer from N\(^4\) to N3 lead to elimination of methylamine \((-31\ \text{Da})\) and formation of the highly resonance stabilized ion detected at \(m/z\ 95\). The analogous reaction (not shown) involving the ring opening, charge redistribution from N3 and proton transfer from N3 to N\(^4\) leads to elimination of ammonia \((-17\ \text{Da})\) and formation of a
methyl analogue of the \textit{m/z} 95 ion detected at \textit{m/z} 119. Formation of the \textit{m/z} 82 fragment (Scheme 3.2B) is associated with a proton transfer to exocyclic \textit{N}^4 without ring opening which results in concerted loss of ammonia (\textendash 17 Da) and hydrogen cyanide (\textendash 27 Da). The charge on the product ion of this reaction could be stabilized by both resonance from N and O lone pairs but also by homoconjugation\textsuperscript{159} through space from \(\pi\) electron density in the C-C triple bond. The \textit{N}3-imino tautomer shown in Scheme 3.2C generates an intermediate ion in which a proton transfer from C6 to \textit{N}^4 through a 5-membered ring arrangement assists in elimination of CO (\textendash 28 Da) and spontaneous loss of ammonia to produce the highly resonance stabilized product detected at \textit{m/z} 81. An analogous reaction (not shown) to Scheme 3.2C proceeding from the N3 protonated \textit{N}^4-imino tautomer leads to ring opening at the C2-N3 bond, proton transfer from C6 to N3 and CO elimination followed by spontaneous loss of methylamine to form the product ion detected at \textit{m/z} 67. These two reaction are fully supported by a previous isotope labeling study of another pyrimidine base, uracil,\textsuperscript{93} were the analogous [M+H-CO-NH\textsubscript{3}]\textsuperscript{+} ion is produced by loss of a proton from C6, as proposed here. Chapter 4 describes a detailed study of the reactivity of Uracil. Scheme 3.2D shows reactions of the \textit{N}^4 protonated \textit{N}^4-imino tautomer leading to formation of the fragment ion detected at \textit{m/z} 69. The intra-molecular nucleophilic attack by \textit{N}1 at C4 forms a reactive bicyclic intermediate ion which eliminates methyl isocyanic acid (\textendash CH\textsubscript{3}NCO, \textendash 57 Da) to produce the highly resonance stabilized product. Each of the reactions presented in Scheme 3.2 help confirm the structure of the investigated ion. The fact that all product ions detected originate from 3-methylcytosine demonstrates the strength of FAIMS to separate analytes of interest from unresolved electrospray background ions. Even in ESI-MS/MS or LC-ESI-MS/MS experiments involving pure standards, product ions are detected that are not
associated with the targeted precursor. This problem becomes much more severe in complex mixtures and when low abundance precursors are selected, as described in Chapter 2.

3.4. Conclusions

The lack of published results using the first generation of commercial FAIMS instrument, the Ionalytics Selectra, can now be partially attributed to poor ion transmission through the source interface region of the instrument. Simple modifications have been presented which provide good ion transmission when combining FAIMS to a relatively high resolution QTOF instrument. This has made it possible to carry out the non-target analysis of urine by ESI-FAIMS-MS/MS after minimal sample preparation. It has been demonstrated that FAIMS has unique advantages for the identification of low abundance unknown species in complex mixtures such as urine. These include the separation of ESI background from analytes of interest and the continuous filtering ability of FAIMS which allows for much longer MS analysis times of an unknown than any traditional column separation. The elimination of chemical background by FAIMS means that even low abundance species can be detected at high S/N values in both MS and MS/MS mode. Using this method, high quality accurate and tandem MS data can be acquired for any of the many hundreds of MS peaks detected in a urine sample after minimal sample preparation. This high quality data makes FAIMS more suitable for de novo identification of unknown species at low abundance in complex mixtures than either ESI-MS/MS or LC-ESI-MS/MS. It is hoped that this work will help demonstrate to the analytical community that FAIMS devices designed as a method of primary separation combined with high resolution MS are of great value and should not be abandoned because of any early setbacks in the commercialization of the instrument.
Chapter 4

Gas Phase Ion Chemistry of Uracil During Collision Induced Dissociation

Abstract

During the *de novo* identification of urinary metabolites in Chapter 2 and Chapter 3 it was determined that previous work describing the dissociation of NA bases was not adequate to fully explain their observed reactivity. Here the gas-phase chemistry of protonated uracil (U) during collision induced dissociation (CID) is revisited using two modern tandem mass spectrometry techniques; quadrupole ion trap (QIT) and quadrupole time of flight (QTOF). Detailed mechanistic proposals are presented which account for all observed products of the current experiments and from previous isotope labeling data and that are supported by previous ion spectroscopy results and theoretical work. The diverse product-ions of U cannot be explained adequately by only considering the lowest energy form of protonated U as a precursor. The possible tautomers adopted by U during collisional excitation make it possible to relate the complex reactivity observed to reasonable mechanistic proposals and feasible product-ion structures for this small highly conjugated heterocycle. These reactions proceed from four different stable tautomers, which are excited to a specific activated precursor from which dissociation can occur via a charge-directed process through a favorable transition state to give a

---

*Chapter 4 is based on the article “Revisiting the Reactivity of Uracil During Collision Induced Dissociation: Tautomerism and Charge-Directed Processes” by Daniel G. Beach and Wojciech Gabryel'ski published in Journal of the American Society for Mass Spectrometry.*
stabilized product. Understanding the gas phase ion chemistry of uracil at this level will facilitate the identification of new modified uracil derivatives in biological samples based solely on their reactivity during CID. The integrated approach presented for describing ion dissociation is widely applicable to other NA bases and similar classes of biomolecules.

4.1. Introduction

In the non-target identification of urinary metabolites in Chapters 2 and 3, several products of nucleic acid degradation were identified.\textsuperscript{29} These compounds, which are detected as modified nucleosides or bases in urine, can provide important information about DNA\textsuperscript{160} or RNA damage\textsuperscript{140} that has occurred in the cell. The identification of low abundance unknown species such as these in a complex biological sample matrix remains a significant challenge for any analytical technique. High resolution tandem mass spectrometry (MS/MS) with a suitable analytical separation has a realistic possibility of addressing this challenge, and Chapters 2 and 3 show some success in this direction.\textsuperscript{29,30,102} However, in addition to the instrumental challenges associated with acquiring high quality spectra for metabolites in urine,\textsuperscript{29} the interpretation of MS/MS data for the purpose of unknown identification remains a significant challenge. Such \textit{de novo} identification aims to correlate observed spectral features in MS/MS to the structures of investigated species based on established reactivity for a class of compounds.\textsuperscript{35,89,102,113,161} The idea of using MS/MS as stand-alone identification tool is of increasing interest and is also being addressed using other more general approaches intended to generalize dissociation pathways,\textsuperscript{88} to automate data analysis,\textsuperscript{162} or to compile spectra in databases.\textsuperscript{48,77}

All nucleosides and NA bases as well as a large number of their isotopically labeled analogues and modified species have been studied by CID.\textsuperscript{93,94,106–108} These reports have allowed for the facile and selective detection of those target analytes in biological samples as well as
some other compounds that showed similar fragmentation patterns. Knowing a compound’s fragmentation pattern allows one to recognize modified species when they show analogous dissociation patterns. Understanding a compound’s gas-phase ion chemistry on a mechanistic level is a much more powerful tool in *de novo* identification. This understanding allows both analogous and unique fragmentation patterns of derivatives to be correlated to their chemical structures. For example, the identification of pyrimidine ring methylated guanosines in biological samples is relatively routine. This is because these ions exhibit pyrimidine ring fragmentation pathways matching those of guanine whose mechanisms were fully described. However, even minor modifications of nucleobases can drastically alter their fragmentation patterns. The best example is hydroxylation of 2’-deoxyguanosine at the C-8 position to produce 8-oxo-2’-deoxyguanosine. This modification not only suppresses the elimination of the deoxyribose moiety, which is used for recognizing deoxynucleoside ions, but it also changes the mechanism of pyrimidine ring decomposition. It can easily be envisioned that without prior knowledge, identification of this DNA oxidation product and its derivatives in a real sample might not be trivial. It is, therefore, important to understand and interpret data from tandem mass spectrometry in terms of the actual chemical reactions occurring in the gas-phase so that spectral features can better be correlated to the structure of unknown species.

The dissociations of conjugated heterocycles including NA bases are known to be particularly complex, and their mechanisms are of fundamental interest. Ions with a fixed charge site or little conjugation can undergo charge-remote dissociation that is similar in mechanism to thermal degradation of neutral molecules. In contrast, the small, highly conjugated ions of NA bases typically react through charge-mediated processes. Intramolecular chemical reactions of precursor ions are initiated by a charge site, and the stability of specific
conformations of precursor ions and product-ions are determined by inductive effects, steric effects and resonance in the conjugated system of double bonds. These chemical effects and their influence on reactivity of the ions of NA bases and their derivatives can be understood, predicted, modeled and used for the structural elucidation of unknown compounds. A key factor in the gas-phase ion chemistry of NA bases and their derivatives is their tautomerism. Interconversion between the lowest energy tautomer and higher energy tautomeric forms occurs during collisional activation. This adds to the complexity of the gas-phase chemistry of this class of compounds because multiple dissociation pathways involve different tautomeric forms of precursor and product-ions. However, the link between proposed or observed tautomers and detected product-ions has rarely been made.  

While a great deal of work has previously been done to investigate the CID of NA bases, major gaps remain and there is a lack of a unified simple model that would be able to explain the reactivity of highly conjugated positive and negative ions of NA bases. In this chapter the gas-phase ion chemistry of several NA bases is re-examined in order to establish a suitable model that can explain all their MS/MS spectral features in terms of specific chemical reactions and accounts for changes in reactivity observed in modified analogs. Interpretation of the reactivity of radical cations in electron ionization (EI) has been developed to the point where spectral interpretation based on simple reaction mechanism is now sophomore lecture material. A better fundamental understanding of CID mechanisms is the first step towards this type of interpretation of MS/MS data in the future.

Based on the hundreds of reaction mechanisms that have been examined in recent published and unpublished studies in the Gabryelski group as well as the fundamentals of ion dissociation, a unique approach has been developed for describing
MS/MS reactions of small conjugate ions during CID. In general: (1) Reaction steps can be described as charge-directed processes.\textsuperscript{163} (2) Collisionally excited ions can adopt higher energy conformations that are the precursors to dissociation.\textsuperscript{94,96,163} (3) The activated precursor offers stabilization to the charge as it is developing during the dissociation step. (4) The product of dissociation is stabilized by delocalization of the charge as much as possible given the structure of the ion.\textsuperscript{86}(5) Ions with a localized charge are transient species prone to rapid fragmentation and are often not detected in MS/MS spectra. (6) Bond making (nucleophilic attack) or non-dissociative bond breaking (ring opening) occurs prior to elimination of a neutral, which is the final step of dissociation forming a stable product.

In this chapter, the gas-phase ion chemistry of protonated U is revisited. Uracil and several of its isotopologues have been previously analyzed by Fast Atom Bombardment ionization with low energy CID.\textsuperscript{93} However, the reactivity of this fundamental component of RNA is not yet fully understood. In this study, a quadrupole time of flight (QTOF) capable of acquiring accurate mass MS/MS data and the quadrupole ion trap (QIT) capable of providing sequential tandem MS data were used. Accurate mass of product-ions were used to confirm their elemental composition while sequential dissociation of product-ions provided information about their structure. By combining these results with previously published data from isotopic labeling\textsuperscript{93}, ion spectroscopy\textsuperscript{96} and theoretical\textsuperscript{96,164,165} studies, a more detailed description of the reactivity of U during CID than existed previously is presented. This work will be of interest to those studying the CID of nucleic acids and those wishing to use tandem mass spectrometry for non-target identification. The integrated approach presented here for describing charge-directed reaction mechanisms is widely applicable and could equally be applied to other classes of molecules of interest.
4.2. **Experimental**

Ammonium acetate, uridine (99%), HPLC grade water and methanol were purchased from Fisher Scientific (Nepean, Ontario). Uracil (≥ 99%) was purchased from Sigma Aldrich (Oakville, Ontario). Standard compounds were dissolved in a solution of 90:10 methanol:water with 0.1 mM ammonium acetate at an approximately 5 µM concentration.

The sequential dissociation study was carried out on a Thermo LCQ Deca ion trap mass spectrometer where standard was infused at a rate of 5 µL/min. Spray voltage was 5 kV, capillary and tube lens offset voltages were both set to -10 V and helium was used as a collision gas (Linde, Guelph). For MS² and MS³ experiments, the qₓ value was set to 0.25, and MS⁴ experiments were carried out with qₓ = 0.35. For MS³ and MS⁴ experiments the instrument’s “low mass” setting was used along with automatic gain control. All collision energy (CE) values are presented in figure captions. In order to carry out the dissociation of protonated uracil at m/z 113, protonated uridine (m/z 245), which is known to produce abundant protonated uracil as its primary MS² dissociation product, was infused and selected in MS². Protonated uracil at m/z 113 was then selected as the MS³ precursor and its dissociation products selected as precursors in MS⁴.

Accurate mass of product-ions was acquired using a QTOF Micro quadrupole time of flight mass spectrometer (Waters, UK) equipped with a nano-spray ESI ion source (Micromass, UK). A Selectra (Ionalytics Corp, Ottawa) high field asymmetric waveform ion mobility spectrometer (FAIMS) was used to separate protonated uridine from a background ion detected at m/z 113. The parameters used for FAIMS analysis of nucleosides were those developed during the non-target analysis of urine by ESI-FAIMS-MS in Chapter 3 where modified nucleosides represented some of the most abundant ions detected. Ionization was carried out in positive mode.
using a nano-electrospray source with a spray voltage of 5000 V (4000 V relative to FAIMS curtain plate). A FAIMS carrier gas of 20% CO₂ in N₂ and desolvation gas and a dispersion voltage of 3900 V were used. By this method, an abundant source fragment corresponding to pure protonated uracil at \( m/z \) 113 could be continuously transmitted to the MS detector for single stage quadrupole dissociation in MS². Injections were made into a 400 nL/min flow of 0.1 mM ammonium acetate using the pump and auto-sampler of a nanoAcquity nano-flow LC system (Waters, UK). A cone voltage of 25 V, and an extraction cone voltage of 1 V were used along with Argon (Linde, Guelph) as the collision gas.

4.3. Results and Discussion

The first attempt to analyze a uracil standard by quadrupole time of flight (QTOF) and the quadrupole ion trap (QIT) showed a number of CID product-ions that had not been previously reported for uracil. This was the result of an interfering isobaric \( m/z \) 113 ion contaminant present in the analyzed samples as shown in Figure 4.1A for QTOF. To overcome this, uridine (UR) was chosen, which produces protonated U by elimination of a ribose moiety. For QIT it was possible simply to select \([\text{UR} + \text{H}]^+\) at \( m/z \) 245 in MS² as shown in Figure 4.1D in order to obtain \([\text{U} + \text{H}]^+\) at \( m/z \) 113 as a precursor in MS³. The product-ion spectrum in MS³ of protonated uracil by QIT-MS is shown in Figure 4.2A and shows abundant products of the loss of ammonia (- 17), water (- 18) and isocyanic acid (- HNCO, - 43) as well as an ion-neutral reaction with residual water (+ 18) present as an impurity in the collision gas of the QIT.

For a QTOF, which is only capable of single stage MS/MS, source fragmentation was used to produce an abundant \([\text{U} + \text{H}]^+\) ion at \( m/z \) 113.0346 as shown in Figure 4.1.C. However,
the isobaric background ion, which could be resolved at \( m/z \) 113.0952 in MS mode by the TOF mass analyzer (Figure 4.1.A) was still selected for dissociation by the quadrupole mass analyzer. High field asymmetric waveform ion mobility spectrometry (FAIMS) was implemented to separate these ions, which operates as an ion filtering device between the ESI source and QTOF detector.\(^{29,58}\) In Chapter 3 this technique was shown to be capable of separating modified nucleosides and NA bases for MS/MS analysis even in the complex biological matrix of a urine sample.\(^{29}\) For a standard solution, FAIMS was easily able to separate \([\text{UR}+\text{H}]^+\) from the \( m/z \) 113 background ions before UR was fragmented in the source to form protonated U. In this way, it was possible to deliver a continuous flow of pure protonated uracil as shown in Figure 4.1B for dissociation in the quadrupole of the QTOF and obtain the product-ion spectrum in Figure 4.2B.

This spectrum is comparable to previously published spectra with the absence of a few low abundance ions previously reported using different instrumentation.\(^ {93}\) The results of accurate mass measurements annotated in Figure 4.2B agree with previously published data from isotope labeling studies carried out on U.\(^ {93}\) This labeling data, which was originally only discussed in terms of \(^{15}\text{N},^{13}\text{C}\) and \(^{18}\text{O}\) labeling,\(^ {93}\) has been re-interpreted to include the results from dissociation of \([5,6-\text{D}_2]\text{uracil}\) that had not previously been explained but are critical in explaining several dissociation pathways. Based on this interpretation, the composition and atom position of each product-ion and all neutral losses are presented in Table 4.1, which is the starting point for spectral interpretation.
Figure 4.1. MS analysis of protonated uracil ([U+H]^+). (A) Interference from background ions in the analysis of [U+H]^+ using ESI-QTOF MS. (B) [U+H]^+ signal isolated using ESI-FAIMS-QTOF MS. (C) Source fragmentation of protonated uridine [UR+H]^+ forming an abundant product of [U+H]^+ detected in MS mode using QTOF. (D) CID of [UR+H]^+ at m/z 245 in MS^2 at normalized collision energy = 17 V using a QIT.
Figure 4.2. Product-ion spectra of protonated uracil at $m/z$ 113 obtained by quadrupole ion trap MS at normalized collision energy $= 32$ V (A) and the product-ion spectrum of protonated uracil obtained by quadrupole-time-of-flight MS at collision energy $= 24$ V (B).
Table 4.1: Interpretation of data from the CID of stable isotope labeled uracil in Nelson and McCloskey, 1994.\textsuperscript{93}

<table>
<thead>
<tr>
<th>m/z</th>
<th>Product-ion\textsuperscript{a}</th>
<th>Atoms Lost\textsuperscript{b}</th>
<th>Relative Abundance of Product-Ion (%) \textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>[M + H - NH\textsubscript{3}]\textsuperscript{+}</td>
<td>N-1, H\textsuperscript{5/6}, H\textsubscript{2}\textsuperscript{x}</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-3, H\textsubscript{3}\textsuperscript{x}</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-3, H\textsuperscript{5/6}, H\textsubscript{2}\textsuperscript{x}</td>
<td>30</td>
</tr>
<tr>
<td>95</td>
<td>[M + H - H\textsubscript{2}O]\textsuperscript{+}</td>
<td>(\text{O}^\text{2}, \text{H}_2\textsuperscript{x})</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\text{O}^\text{4}, \text{H}_2\textsuperscript{x})</td>
<td>50</td>
</tr>
<tr>
<td>70</td>
<td>[M + H - HNCO]\textsuperscript{+}</td>
<td>N-1, C-2, (\text{O}^\text{2}, \text{H}^\text{x})</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-3, C-2, (\text{O}^\text{4}, \text{H}^\text{x})</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-3, C-2, (\text{O}^\text{4}, \text{H}^{5/6})</td>
<td>40</td>
</tr>
<tr>
<td>68</td>
<td>[M + H - NH\textsubscript{3}-CO]\textsuperscript{+}</td>
<td>N-3, C-4, (\text{O}^\text{4}, \text{H}_3\textsuperscript{x})</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N-3, C-2, (\text{O}^\text{4}, \text{H}_3\textsuperscript{x})</td>
<td>10</td>
</tr>
<tr>
<td>67</td>
<td>[M + H - H\textsubscript{2}O-CO]\textsuperscript{+}</td>
<td>(\text{C}^\text{2}, \text{O}^\text{2}, \text{O}^\text{4}, \text{H}_2\textsuperscript{x})</td>
<td>100</td>
</tr>
<tr>
<td>53</td>
<td>[M + H - NH\textsubscript{3} - HNCO]\textsuperscript{+} and [M + H - HNCO - NH\textsubscript{3}]\textsuperscript{+}</td>
<td>N-1, C-2, (\text{O}^\text{2}, \text{N}-3, \text{H}^{5/6}, \text{H}_3\textsuperscript{x})</td>
<td>100</td>
</tr>
<tr>
<td>43</td>
<td>[M + H - HNCO - HCN]\textsuperscript{+}</td>
<td>N-1, C-2, (\text{O}^\text{2}, \text{N}-3, \text{C}-6, \text{H}^{5/6}), (\text{H}^\text{x})</td>
<td>100</td>
</tr>
<tr>
<td>42</td>
<td>[M + H - HNCO - CO]\textsuperscript{+}</td>
<td>C-2, (\text{O}^\text{2}, \text{N}-3, \text{C}-4, \text{O}^\text{4}, \text{H}^\text{x})</td>
<td>100</td>
</tr>
<tr>
<td>40</td>
<td>[M + H - NH\textsubscript{3} - CO - HCN]\textsuperscript{+}</td>
<td>C-2, (\text{O}^\text{2}, \text{N}-3, \text{C}-4, \text{O}^\text{4}, \text{H}^{5/6}), (\text{H}_4\textsuperscript{x})</td>
<td>100</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Product-ion origin determined from sequential ion trap dissociation.

\textsuperscript{b} Determined from published data for CID of stable isotope labeled derivatives of uracil.\textsuperscript{93} \textsuperscript{H}_n\textsuperscript{x} refers to the number of exchangeable protons lost, whose ring position cannot be specifically assigned by isotope labeling.
4.3.1. Ionization, Tautomerization and Collisional Activation of Uracil

In solution, uracil is predominantly found in its di-keto form shown in Scheme 4.1 and the most basic site on uracil, and hence the most likely to be protonated under electrospray conditions, is the keto oxygen $O_4$. $^{93,94}$ Once in the gas-phase, ground state proton affinities of $O^2$ and $O^4$ become much closer in energy than in solution and tautomerization and proton transfer are much more feasible. $^{93}$

The importance of tautomeric form for the CID of a modified nucleoside was recently shown by the Gabryelski group. $^{94}$ This work also demonstrated that different reaction pathways of a nucleoside proceed exclusively from different tautomers that are formed during collisional activation. $^{94}$ Heterocyclic and highly conjugated uracil can be represented by eight different tautomeric forms each with four different protonation sites for a total of at least 32 possible precursors to dissociation. The relative energies of all stable uracil tautomers and protonation sites have been previously determined theoretically, revealing the N-1 protonated di-enol tautomer (I, all in Scheme 4.1) to be the lowest energy with the $O^4$ enol form II only slightly more energetic ($\sim 5$ kJ mol$^{-1}$).$^{96,164,165}$ Energy barriers for inter-conversion from I to the other stable tautomers range from 5 to 221 kJ mol$^{-1}$. $^{165}$ Thus, regardless of the tautomeric form and protonation site of the original solution species, once protonated, the gas-phase ion is free to undergo proton transfer and tautomerization reactions as it is excited collisionally. Spectroscopic study of protonated uracil during infrared multiphoton dissociation (IRMPD) have confirmed that multiple tautomers of protonated uracil are present in ion trap at energies consistent with dissociation. $^{96}$ These results show at least two tautomers present at a threshold energy where the first fragment ions begin to appear. It is reasonable to assume that at higher excitation energies more energetic tautomers as well as their ring-opened products could be easily accessible and
that any of the 18 stable tautomers reported can represent a viable precursor in CID. During excitation, some of these tautomers will adopt conformations from which simple charge-directed mechanisms can lead to dissociation upon further excitation. Scheme 4.1 shows those five stable tautomers (III – VI) that can account for all detected product ions of protonated uracil.

Scheme 4.1. The protonated of uracil during positive electrospray and the tautomerization during collision induced dissociation of protonated uracil.

For small, highly conjugated ions such as uracil, the charge site mediates all reaction steps including proton transfer, charge redistribution, tautomerization and ultimately dissociation. Charge-directed dissociations proceed through favorable (stabilized) transition states to produce stable neutral molecules and product-ions. While compounds less readily able to interconvert between multiple tautomeric forms might dissociate through a single dissociation pathway, the multiple tautomers adopted by U in CID result in multiple dissociation pathways depending on the conformation of the precursor ion during excitation. This is an important factor in describing CID reactivity that is often overlooked. The ion structures presented in Scheme 4.1 are the stable precursors of uracil that dissociate to form the detected product-ions when excited.
For simplicity and space considerations, the dissociation mechanisms proposed in subsequent sections of this work begin from these precursors and do not show each step in tautomerization from the ground state ion I.

These tautomers provide the critical link between the complex dissociation pathways observed for uracil and a simple model that can describe dissociation via charge-directed mechanisms \(^{163}\) to form products with a stabilized charge.\(^{86}\) By taking tautomerization into account all observed experimental data from an extensive isotopic labeling study, accurate mass measurements and sequential tandem MS can now be explained. The proposed reaction mechanisms are intended to show intramolecular interactions and stabilization of charge at critical points during dissociation. The proposed structures might not represent their lowest energy conformers, which could be calculated theoretically but rather demonstrate the potential for charge stabilization at the transition state and in the product-ion, important factors in determining which dissociation processes are rational.

4.3.2. Ammonia Elimination from Protonated Uracil

The primary dissociation product of uracil in Figure 4.2 is the product of ammonia loss detected at \(m/z\) 96. The results of isotope labeling (Table 4.1) show that at least three separate processes are occurring that result in elimination of ammonia. Based on all available data, three charge-directed processes for the elimination of ammonia from \([U + H]^+\) are proposed in Scheme 4.2. Contrary to previous interpretation of ammonia loss from pyrimidine rings,\(^{93,107}\) Scheme 4.2 accounts for ammonia loss from N-3 with either exchangeable (Scheme 4.2A) or carbon bound (Scheme 4.2B) protons lost as well as ammonia loss from N-1 (Scheme 4.2C). For tautomer IV, ring opening at the C2-N3 bond (Scheme 4.2A) and proton transfer through a favorable 6-membered ring conformation yield a quaternary ammonium precursor with charge localized at
N-3. During dissociation of this transient species, charge developing at C-4 is stabilized by electron density from the conjugated system of double bonds and non-bonding electrons on hetero-atoms. The product of this reaction that, from Table 4.1, represents about 60% of the total signal detected at \( m/z \) 96, is also highly resonance stabilized and could be represented by six resonance structures as shown in Scheme 4.3. The final structure shown in Scheme 4.2A represents an overlay of these resonance structures with dashed bonds representing some contribution of \( \pi \) electron density from at least one resonance structure as demonstrated in Scheme 4.3. The same approach is used throughout the manuscript to show the potential for resonance stabilization of the charge. A high degree of resonance stabilization can be shown for all abundant product-ions of uracil, demonstrating the importance of this factor in determining which product-ions are detected.
Scheme 4.2. Gas phase ion chemistry processes leading to the formation of $m/z$ 96 product ion by elimination of ammonia from protonated uracil.
Scheme 4.3. Dashed bond shorthand used to demonstrate the contributions of multiple resonance structure to overall charge stabilization.

Ring opening of tautomer VI at the N3-C4 bond would induce a similar mechanism (Scheme 4.2B) where proton transfer would occur from C-5 that can account for 30% of the intensity of m/z 96 ion based on isotope labeling data (Table 4.1). Elimination of ammonia containing a C-5 bound proton could also occur after ring opening of tautomer V at the N-1/C-2 position (Scheme 4.2C). This reaction accounts for both the 10% of intensity of the m/z 96 product ion attributed to N-1 loss as well as the remaining 10% of ammonia loss including a C-bound proton (Table 4.1). The transfer of H⁺ to form a quaternary ammonium ion before ammonia elimination differs significantly from proton scrambling prior to dissociation. This phenomenon, that was previously reported for the nucleobase guanine[^94] and considered as a possible explanation for puzzling results for the CID of [5,6-D₂]Uracil,[^93] can be completely ruled out based on experimental data in the case of Uracil.

Dissociation of the m/z 96 ion in MS⁴ mode (Figure 4.3A) supports the structures proposed in Scheme 4.2. The m/z 96 ions dissociate by elimination of CO to produce an
abundant ion with m/z 68, elimination of HNCO to produce an abundant ion with m/z 53, and elimination of ethyne (-HC≡CH) to produce a minor product ion with m/z 70. Based on labeling data (Table 4.1) and on the precursor ions proposed in Scheme 4.2 these second generation products have been assigned to a specific m/z 96 precursor in the pathways in Scheme 4.4.

4.3.3. Elimination of Water from Protonated Uracil

Like ammonia loss, previous proposals for the mechanism of water loss from pyrimidine bases have involved ring opening followed by proton transfer to an exocyclic oxygen. This proposal does not fit experimental data. Isotope labeling data in Table 4.1 show a ratio of 1:1 for loss of water from O² and O⁴, which is significant in that it suggests equivalent mechanisms for each process. Also important is the fact that neither of these processes shows proton transfer from C-5 or C-6 to O before water elimination as was the case for ammonia elimination in Schemes 4.2B and 2C. After ring opening at any position, O² and O⁴ will not be equivalent with respect to proton transfer reactions and the rate of water loss. Furthermore, reasonable charge-directed mechanisms involving initial ring opening followed by water loss would involve proton transfer from C-5 or C-6, which based on ²H labeling data in Table 4.1, does not occur. However, by taking into account all possible tautomers it is evident that tautomer III (Scheme 4.1) is a reasonable precursor from which water can be lost without ring opening and where O² and O⁴ are nearly equivalent.
Figure 4.3. Sequential dissociation of major products of protonated uracil in MS$^4$ using QIT-MS. (A) CID of the m/z 96 product of ammonia loss at NCE = 32 V. (B) CID of the m/z 95 product of water loss at NCE = 29 V. (C) CID of the m/z 70 product of isocyanic acid loss at NCE = 32 V.
<table>
<thead>
<tr>
<th>MS²</th>
<th>MS³</th>
<th>MS⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Scheme 4.4.** Hierarchy of product ions of collision induced dissociation of protonated uracil considering tautomerization prior to dissociation.
Proposals that account for both the equivalent loss of O$_2$ and O$_4$ as well as product-ions of subsequent stages of sequential tandem mass spectrometry are presented in Schemes 4.5A and 4B. Water elimination from either of these precursors is favorable because of resonance stabilization of the developing charge at either C-4 or C-2. Further, stabilization of the charge on the product-ions is significant and nearly equivalent between the two structures.

Dissociation of the m/z 95 ion in MS$^4$ mode (Figure 4.3B) supports the structures proposed in Scheme 4.5. The m/z 95 ions dissociate by elimination of hydrogen cyanide (-HCN) to produce an abundant ion with m/z 68, elimination of CO to produce a product ion with m/z 67 and elimination of ketene (-H$_2$CCO) to give a product ion with m/z 53. Based on labeling data (Table 4.1) and on the precursor ions proposed in Scheme 4.5 these second generation products have been assigned to a specific m/z 95 precursor in the pathways in Scheme 4.4.

Scheme 4.5. The elimination of water from protonated uracil.
4.3.4. Elimination of Cyanic and Isocyanic Acid from Protonated Uracil

Based on the isotope labeling data in Table 4.1, at least three different mechanisms exist for elimination of isocyanic acid (-HNCO, -43) from protonated uracil during CID. The primary isomer of m/z 70 involves loss of O^2 and N-3 and retention of O^4 and N-1 with loss of either an exchangeable (N or O bound) proton or a C-5/C-6 bound proton at a ratio 5:4. The previously proposed retro-Diels Alder (RDA) reaction \(^{93,106,107}\) of the O^4 protonated di-keto tautomer shown in Scheme 4.6A is unsatisfactory for several reasons. This reaction cannot occur in a single step as a charge-directed process, but rather would need to first undergo tautomerization or else ring opening at C-2/N-1 before elimination of isocyanic acid could occur. Even if this were to occur, the ion shown in Scheme 4.6A does not represent a stable product and would represent a lower abundance product. Charge is localized on the O^4 carbonyl group and not stabilized by the abundant \(\pi\) electron density in the rest of the ion. If formed, this ion would represent a reactive intermediate that could be expected to immediately eliminate ketene (-H\(_2\)CCO, -42) by the mechanism shown in Scheme 4.6B to form protonated hydrogen cyanide at m/z 28, which was not detected as a first generation product-ion of protonated uracil in Figure 4.2, but has been reported previously at low abundance.\(^{93}\)

Scheme 4.6C shows a proposal for an alternative charge-directed process, also induced by proton transfer to O^4 in the N-3 protonated di-keto tautomer IV that would lead to the abundant elimination of HNCO from O^2, C-2 and N-3. Instead of the RDA reaction previously proposed,\(^{93}\) intramolecular nucleophilic attack by N-1 at C-4 creates an activated precursor from which isocyanic acid can be eliminated to produce an m/z 70 product-ion with charge delocalized throughout the ion. According to Table 4.1, approximately 50% of the detected abundance at m/z 70 originates from dissociation through this pathway, without loss of H^5 or H^6.
Scheme 4.6. The elimination of isocyanic acid and cyanic acid from protonated uracil.
About 40% of the $m/z$ 70 product-ion in Figure 4.2 originates from a pathway that does involve loss of a proton from C-5 or C-6 during dissociation. This reaction can be rationalized through a similar mechanism to those leading to ammonia loss in Scheme 4.2 with abstraction of a proton from C-5. In Scheme 4.6D, after ring opening of the $O^4$ enol tautomer V at N-1/C-2, an intermediate identical to that shown in Scheme 4.2C as a precursor to ammonia loss is formed. If the C-5 proton is abstracted by $O^4$ as shown in Scheme 4.6D or N-3 (not shown) as opposed to N-1, charge becomes localized on the other side of the molecule leading to elimination of cyanic acid (NCOH) or isocyanic acid (not shown), respectively. Table 4.1 also shows minor contributions to $m/z$ 70 from loss of HNCO from N-3/C-4/$O^4$ could occur by abstraction of H-5 by $O^2$ after ring opening at C-2/N-3 as shown in Scheme 4.6E.

Dissociation of the $m/z$ 70 ion in MS$^4$ mode (Figure 4.3C) supports the structures proposed in Scheme 4.6. The $m/z$ 70 ions dissociate by elimination of ketene to produce an abundant ion with $m/z$ 43, elimination of ammonia to form a product ion with $m/z$ 53, and elimination of CO to produce a minor product ion with $m/z$ 42. Based on labeling data (Table 4.1) and on the precursor ions proposed in Scheme 4.6 these second generation products have been assigned to a specific $m/z$ 70 precursor in the pathways in Scheme 4.4.

### 4.4. Conclusion

By combining accurate mass and sequential tandem MS results with previously published data from isotope labeling, theoretical and ion-spectroscopy studies it was possible to describe the reactivity of uracil during CID in much greater detail that was previously possible. The very complex reactivity of this heterocyclic ion was shown to be the result of its propensity to undergo tautomerization in the gas-phase before dissociation so that multiple precursor ions are
present for a single species. This leads to numerous dissociation pathways possible for each observed neutral loss.

All reactions proposed for protonated uracil also have important implications for the tandem mass spectrometry identification of modified forms of uracil in biological samples. As a result of chemical modification of uracil, reactions established for the unmodified base can be suppressed or enhanced or new reactions can be induced. This valuable information can be used for structural elucidation of derivatives of uracil now that mechanisms of its fundamental reactions are better understood.

In total, 20 reactions are examined in the current work as well as the hundreds of others in previous published,\textsuperscript{29,30,89,94,102,113} and unpublished studies in the group. Based on these reactions, it was possible to make some generalizations about what makes a mechanistic proposal suitable for describing the dissociation of small even electron ions during CID. In general: (1) All reaction steps can be described as charge-directed processes. (2) Collisionally excited ions can adopt higher energy conformations that are the precursors to dissociation. (3) The activated precursor offers stabilization to the charge developing during the dissociation step. (4) The product of dissociation is stabilized by delocalization of the charge as much as possible given the structure and formula of the ion. (5) Ions with a localized charge are transient species prone to fragmentation and are generally not detected as abundant products in MS/MS spectra. (6) Any bond making (nucleophilic attack) or non-dissociative bond breaking (ring opening) occurs prior to elimination of a neutral, which is the final step of reaction forming a stable product.

By using these simple generalizations as a starting point for describing reactivity, spectral differences observed between related species can usually be described and understood based on chemical principles such as steric or electronic effects. The conformations generated from these
mechanisms can serve as a reasonable starting point for more detailed theoretical studies of CID reaction mechanism. Further, this type of an understanding of reactivity can be used to develop structural proposals for unknown molecules during non-target analysis based on the known reactivity of compounds with structural similarities.
Chapter 5

Linear and Nonlinear Regimes of Electrospray Signal Response in Analysis of Urine by ESI-FAIMS-MS

Abstract

Quantitative non-target analysis is intended to provide a measurement of concentration of newly identified components in complex biological or environmental samples for which authentic or labeled standard do not exist. Electrospray ionization-high field asymmetric waveform ion mobility spectrometry-mass spectrometry (ESI-FAIMS-MS) has unique advantages that enabled the development of a novel approach for quantification of non-target analytes. In the non-target analysis of urinary metabolites by ESI-FAIMS-MS, the analysis of highly diluted urine samples is found to be practical and beneficial. It is shown that urine extracts can be analyzed directly at very high dilutions (up to 20 000 times) by extending MS analysis times during slow FAIMS scanning. The effect of sample dilution on ionization efficiency and ionization suppression in direct electrospray of complex sample matrixes is explored. Two distinct regimes in ESI operation can be observed which are related to the limited ionization capacity of this method. In the linear dynamic concentration range below the limiting ionization capacity, the analytical sensitivity of an analyte is constant and does not depend on matrix composition and concentration. Once the capacity of ESI is exceeded, all species exhibit log–log linearity in signal response. Two different approaches for quantification are proposed, one for

---

Chapter 5 is based on the article “Linear and Nonlinear Regimes of Electrospray Signal Response in Analysis of Urine by ESI-FAIMS-MS and Implications for Non-target Quantification” by Daniel G. Beach and Wojciech Gabryelski published in Journal of the American Society for Mass Spectrometry.103
analytes which can be detected in the linear regime and another for those only detected in the suppression regime that overcomes the effects of ionization suppression. The new insight into ionization suppression effects in ESI presented here is of broad interest to anyone using ESI as an ionization technique for the MS analysis of complex samples.

5.1. Introduction

The identification and quantification of low abundance unknown analytes in a complex sample matrix remains one of the most significant challenges in analytical chemistry. When coupled to a suitable analytical separation, mass spectrometry (MS) represents the best tool available for this type analysis, but several challenges remain to be addressed in both the development of instrumental methods\(^\text{29}\) and in the interpretation of MS data for unknown identification.\(^\text{95}\) Non-target analysis can be described as an analytical strategy which aims to acquire quantitative and qualitative information for a large number of sample components without prior knowledge of sample composition.\(^\text{29,102}\) As in global metabolomics, non-target methods must selectively produce an analytical signal for as large a number of analytes as possible, preferably across several compound classes. In addition to reliable quantitative data, non-target methods must be capable of acquiring the best possible spectral data for any sample component to allow its structural identification. Liquid chromatography (LC) coupled to MS through electrospray ionization (ESI) can provide spectral data for a large number of analytes, but non-target analysis of complex samples can be problematic, especially at a trace level.\(^\text{29}\) The chemical background produced by ESI generally shows some contribution to signal at all nominal masses, limiting the low end of the dynamic range of LC-ESI-MS methods.\(^\text{3,8,20}\) Increasing the mass resolving power of MS can improve selectivity in full scan MS mode.
However, precursor ion selection windows in high resolution MS/MS are never as narrow as the detection resolution of an instrument,\textsuperscript{18} and in tandem mass spectrometry (MS/MS), background ions are selected for dissociation along with the precursor of interest. This leads to the detection of product ion interference\textsuperscript{10,29} which can significantly hinder identification based on MS/MS data.

High field asymmetric waveform ion mobility spectrometry (FAIMS) is a separation technique that operates between an ESI source and MS detection to separate gas phase ions at atmospheric pressure.\textsuperscript{51} FAIMS acts as an ion filter which separates ions based on their differential ion mobility at high and low electric field strength. A DC voltage called compensation voltage (CV) is used to selectively transmit ions of a particular differential mobility characteristic from ESI to MS. In chapter 3 it is shown that ESI-FAIMS-MS is suitable for the identification of non-target analytes due to its unique ability to separate analytes of interest from chemical background produced by ESI.\textsuperscript{29,30,102} Since FAIMS acts as a continuous ion filter, it offers a flexible time scale to carry out MS and MS/MS experiments. By extending the time of data acquisition and optimizing source and dissociation conditions, this has the desirable effects of enhancing signal-to-noise and improving spectral quality.\textsuperscript{10,29,157} For example, the number of peptides identified by an automated database search was increased by 35\% when product ion interference was minimized by using FAIMS.\textsuperscript{10} These advantages of FAIMS relate primarily to the qualitative aspects of non-target analysis, but quantitative aspects of such an experiment are important to consider as well. Here, it is shown that ESI-FAIMS-MS can address some quantitative challenges related to the ESI of complex samples. Electrospray-MS is considered a poor technique for absolute quantitative analysis and is often characterized by observed nonlinear signal response as a result of ionization suppression or detector saturation.
Ionization suppression arises from a limited ionization capacity of ESI when sample matrix and buffer components compete with the analyte for charge or surface area on the ESI droplet during ionization. Suppression can be observed, differentially depending on the properties of the analyte and the sample matrix, any time the limited ionization capacity of ESI is exceeded. Below this capacity, typically cited as $\sim 10^{-5} \text{ M}$ total concentration of ionizable species, all analytes exhibit linear response. Suppression is also often described as an attenuation of sensitivity without deflection from linearity, but this is only reported over narrow concentration ranges.

Ionization suppression when analyzing complex samples by ESI can be addressed either with sample preparation or with a liquid separation that ensures analytes are suitably separated from matrix suppressants. Another common approach for minimizing ESI suppression is dilution of a sample, commonly referred to as the “dilute and shoot” method of sample preparation. This approach can be successful in mitigating serious suppression and contamination which hinders the operation of ESI. However, it suffers from the limitation that the dilution required to eliminate suppression can often result in analytes of interest in a sample being diluted to concentrations below their limits of detection.

The only method of true quantitative analysis in ESI-MS uses an isotopically labeled standard for each analyte. By this approach, analyte and standard show identical sensitivity regardless of matrix effects and linearity of the absolute signal is not required, provided detector saturation does not occur. For an analyte just identified by non-target analysis, authentic standards are usually unavailable and analytical sensitivity remains unknown until one can be acquired. Thus, an alternative approach is required that could provide quantitative information for all analytes detected without using their authentic standards. Here, another compound with
structural similarities could be used as an approximation of the analytical sensitivity of the analyte in an approach referred to as semi-quantification. Toward the goal of quantification without authentic standards, work has been done to develop ESI ionization efficiency (IE) indices for metabolites and other small analytes. A challenge in this work has been that suppression effects differ significantly between different instruments due to differing ion source designs. Another significant challenge in this area is that of variable IE and suppression effects in liquid column separations. The absolute IE of an analyte depends heavily on the source conditions and the composition of the solvent and buffer in which it is analyzed. While an effective IE index can be developed under one set of conditions, extrapolating these results to the constantly changing ionization conditions of a gradient LC elution will be challenging. Further, background or sample components leading to ionization suppression elute at different points during the separation leading to highly variable suppression effects. Methods exist to evaluate ionization suppression of targeted analytes in LC-MS, but these are not applicable in non-target analysis as they rely on availability of standards.

Electrospray-MS alone does not have the required selectivity for analysis of complex biological or environmental samples. In Chapter 3 it is shown that dilute urine samples could be analyzed by ESI-FAIMS-MS using extended run times. Here, the quantitative capabilities of this method are explored in non-target analysis of urinary metabolites and unique insight is gained into the signal response of analytes during ESI of complex samples.

5.2. Experimental

Ammonium acetate, uridine (99%), L-alanine (99%), HPLC grade water, and methanol were purchased from Fisher Scientific (Nepean, Ontario). 2’-Deoxyinosine (98%), 5’-
methyl-5′-thioadenosine (≥98%), guanosine (98%), 1-methylguanine (Fluka, ≥ 97%), 7-deazaguanine (95%), N2,N2-dimethylguanine (98%), and palmitoyl-L-carnitine chloride (98%) were purchased from Sigma Aldrich (Oakville, Ontario). Urine sample preparation was carried out as described in Chapter 2. Briefly, an overnight urine sample from a healthy volunteer was centrifuged at 26 000 g. Aliquots (4 mL) of supernatant were loaded onto an activated 200 mg Oasis HLB glass solid phase extraction cartridge (Waters, UK), washed with 4 mL of water, and eluted with 4 mL of 70% methanol in water. All samples were made up in solutions of 9:1 methanol/ water with 0.5 mM ammonium acetate and infused using the fluidics system of a nanoAcquity LC (Waters, Cambridge, MA) to a nanospray source (Micromass, Manchester, UK) at a flow rate of 400 nL min⁻¹.

The nanospray source was positioned 20° off axis of the orifice of a Selectra (Ionalytics, Ottawa) FAIMS device with a modified source interface described in detail in Chapter 3. The FAIMS device was operated at a dispersion voltage of 4000 V, outer electrode offset of 150 V, a curtain plate voltage of 1000 V, and a buffer gas of 1 L min⁻¹ of 20% CO₂ in dry N₂ (Linde, Guelph). Spray voltages between 2500 and 3500 V (relative to the FAIMS curtain plate) were optimized to give stable electrospray current in cone jet mode. Mass spectrometry data was carried out using a QTOF-MS “Micro” (Micromass, Manchester, UK) using gentle source conditions of 14 and 1 V for the outer and inner cone voltages, respectively, and with MS scan time equivalent to the time needed to scan the FAIMS CV by 0.1 V. Error bars in and uncertainties on quantitative values represent standard deviation for replicate analyses of the same sample with the number of replicates quoted in each case.
5.3. Results and Discussion

5.3.1. Signal to Noise in ESI-FAIMS-MS Using Extended Data Acquisition Time in MS

The separation parameter in FAIMS is a compensation voltage (CV) which can be scanned to produce a CV spectrum with a CV peak for each analyte that has been detected by MS at its corresponding accurate \( m/z \). The CV peak represents the analytical signal, and its area or height can be used for quantification. Thus, the CV can either be scanned over a range of values to obtain the CV peak area or held constant (parked) at the maximum transmission CV of the analyte to assess the height of its CV peak. Unlike liquid chromatography where the length of data acquisition in MS for an analyte is limited to the width of its eluting peak, the time of data acquisition in a FAIMS separation can be set arbitrarily depending on the required sensitivity and throughput. For quantitative work using FAIMS, the scanning mode has been chosen because it can compensate for variability in CV, asymmetry of CV peaks, and CV-dependent peak width during the FAIMS separation. In such an experiment, the scan rate (\( V \text{ min}^{-1} \)) can be easily adjusted to control the length of data acquisition at each interval of the CV scan.

The main factor limiting the low end of the dynamic range in ESI-MS is chemical background from ESI\(^3,8,20\). In a typical ESI-MS experiment, increasing the data acquisition time increases the intensity of both signal (S) and noise (N) leading to S/N being proportional to the square root of data acquisition time (\( t \)) as shown in eq 1:

\[
\frac{S}{N} \propto \sqrt{t} \tag{1}
\]

Because FAIMS is able to effectively eliminate ESI chemical background from analytes of interest prior to MS analysis, a more efficient increase in S to N is observed when compared
to direct ESI-MS and S/N increases linearly with t as shown in eq 2:

\[
\frac{S}{N} \propto t \tag{2}
\]

**Figure 5.1.** Linear relationship between signal and scan time using ESI-FAIMS-MS. (A) A 400 nM 2'-deoxyinosine (dI) standard (error bars show SD, n = 5) with changing data acquisition time. (B) Buffer blank, 200 pM dI, and 1 nM dI standards analyzed at 40 min \( V^{-1} \) CV scan time and detected as their sugar loss fragments at \( m/z \) 137.0458. Total summed MS data in each spectrum is 60 min with background at most \( m/z \) equal to 1 spectral count without smoothing or background subtraction.

Figure 5.1A shows results of analysis of a 400 nM solution of the nucleoside standard 2'-deoxyinosine (dI) by ESI-FAIMS-MS where the CV scan rate was varied to alter the MS data acquisition time. This experiment shows an excellent linear correlation \((R^2 > 0.999)\) between S/N for dI and the data acquisition time. One of the limitations of using QTOF-MS in any
quantitative application is the relatively narrow dynamic range limited by multichannel plate
detector saturation. Although S/N can be increased by extending data acquisition time, the total
ion current detected per unit time does not change and hence never exceeds the detector
saturation point. Linearity in signal response with data acquisition time also means that analysis
time and scan rate can be varied throughout a set of experiments and results compared by using a
normalized (counts × V × min⁻¹) scale. This has important practical implication because
experiment times required to operate in the sub nM range and in complex samples can be up to 2
h and being able to shorten this time for more concentrated samples is highly desirable.

The linear relationship in Figure 5.1A also holds for very long data acquisitions required
to analyze a compound near its limit of detection. The MS spectra in Figure 5.1B illustrate the
detection of dI standard near its instrumental detection limit at 200 pM and 1 nM dI as a sugar
loss (−116 Da) fragment ion at m/z 137.0458. The CV scan in this case was carried out at 40 min
V⁻¹, the slowest investigated. At this rate, the 1.5 V wide CV peak of dI was scanned over 60
min with continuous MS data acquisition. The MS spectra in Figure 5.1B show the extremely
low chemical background signal that can be acquired with ESI-FAIMS-MS when the blank
dilution buffer is analyzed under identical conditions and for the same length of time as the
standard without additional background correction or smoothing. The solvent blank analysis over
60 min of MS data acquisition produces a background representing very sharp spikes at spectral
intensity of only 1–2 counts. This characteristic noise can be easily distinguished from chemical
background species which should be detected as MS peaks with a peak width corresponding to
the resolution of the TOF-MS. This is a significant difference when compared to direct ESI-MS
where chemical background peaks on the order of 300–1200 spectral counts can be observed for
buffer blanks analyzed at these experimental conditions.
When chemical background is effectively eliminated, the electrical noise of the QTOF’s microchannel plate detector is the factor limiting S/N, even for very long data acquisition times. This random electrical noise is equal to 1 count at any random $m/z$ (Figure 5.1B). Only for very long acquisitions times, greater than 60 min, would noise spikes of 2–3 counts begin to be observed regularly at $m/z$ values where no specific chemical species is detected. However, it is important to realize that background signal can originate from specific chemical species in blanks and samples (Figure 5.1B) and, as in any separation, care must be taken to resolve analyte from interfering species for the above advantages to be observed. The main advantage of using ESI-FAIMS-MS when compared to ESI-MS is the reduction in background and corresponding increase in S/N which has been well documented. However, this is accompanied with an overall decrease in signal intensity which, as shown in Figure 5.1, can be at least partially compensated for by using extended FAIMS runs and MS data acquisition times.

5.3.2 Ionization Efficiency and Ionization Suppression in ESI

Section 5.3.1 describes how the bottom end of the dynamic range, limited by chemical background of ESI, can be extended by using FAIMS. The upper end of the linear dynamic range is limited by the ionization capacity of ESI. A great deal of attention has been devoted to studying the ionization efficiency of analytes and the magnitude of ionization suppression under various conditions in ESI. However, the concentration at which suppression begins to be observed has been much less carefully studied, especially for complex samples. It is not clear from the literature whether all analytes in complex mixtures exhibit ionization suppression at the same total sample concentration or whether the physicochemical properties of analyte and suppressant determine at which concentration it will be suppressed.
To evaluate the nature of the matrix effects in urine analysis using direct ESI, the ionization properties of metabolites from classes which were identified in urine by LC-ESI-MS/MS in Chapter 2 and ESI-FAIMS-MS/MS in Chapter 3 were examined. Palmitoyl-L-carnitine (PC) represents a large group of acylcarnitines which were observed as some of the most abundant ions detected in the analyzed urine. Their ion abundance in ESI-MS is due in part to very high ESI efficiency of this class of compounds, arising from their surfactant like properties of their chemical structure (Scheme 2.1). Alanine (ALA), on the other hand, represents a compound with relatively low ionization efficiency compared to the other classes of compounds investigated. N$_2$$^2$,N$_2$$^2$-Dimethylguanine (DMG) is one of modified nucleic acid bases identified in urine in Chapter 2, and dI is also a modified nucleoside but is not detected in urine samples; it finds utility as an internal standard in subsequent sections of this chapter.

Figure 5.2 shows the log–log plot of signal response of four metabolites in ESI-FAIMS-MS in an equimolar mixture of these compounds from 100 to 6000 nM for each analyte. In such a plot, linear signal response is observed at analyte concentrations below the limiting capacity of electrospray as a slope of 1, with the y intercept corresponding to the log of ionization efficiency or the slope of the zero-intercept linear calibration curve (eq 3) for each analyte:

$$S = IE(c)$$

(3)

where $S$ is signal intensity, $IE$ is the ionization efficiency, and $c$ is concentration. The log–log form of this equation that is plotted in Figure 5.2 is therefore:

$$\log(S) = \log(c) + \log(IE)$$

(4)
Beyond the ionization capacity of ESI, the slope of the log–log curve is generally considered to exhibit nonlinearity quickly approaching zero and becoming negative. However, a region of log–log linearity for analyte concentrations is consistently observed above the limiting capacity of ESI as in Figure 5.2. This can be observed in previous reports, but has rarely been discussed. In this region, termed the suppression regime, the slope is between 0 and 1 which corresponds to the power $n$ of the curve with the equation:

$$S = IE_{sup}c^n$$

where $IE_{sup}$ is the suppressed ionization efficiency. The equation for signal response in the suppression regime of the log–log plot in Figure 5.2 is therefore:

$$\log(S) = n\log(c) + \log(IE_{sup})$$

In Figure 5.2, linear signal response for each sample component is observed up until around a concentration of $10^{-6}$ M, which correspond to a total concentration of $4 \times 10^{-6}$ M. It is significant that all sample components show a transition from linear to suppression regime at the
same total sample concentration. This demonstrates that it is the total capacity of ESI that determines the total sample concentration at which linear signal response can be observed for all analytes. Regardless of the different ionization properties of sample components, provided their total concentration is below the ionization capacity of ESI, any analyte above its limit of quantification can be analyzed using its linear signal response.

Even though a FAIMS separation is being carried out following ESI and prior to MS analysis, the relative analytical sensitivity of metabolites investigated is determined by their IE in ESI. For example, in Figure 5.2, the ratio of sensitivities between PC and ALA is the same as to the ratio of sensitivities for PC to valine, another small amino acid, reported previously with a completely different MS system.172

The most critical factor for practical analysis remains whether it is possible to dilute a urine sample sufficiently so that the total capacity of ESI is not exceeded but at the same time diluted analytes of interest could still be detected and quantified. To address this, a urine extract was spiked with dI and sequentially diluted extending to dilutions as high as 20 000 times. The concentration of dI in the diluted urine samples ranged from 3.4 nM (20 000 times dilution) to 6750 nM (10 times dilution). Figure 5.3A shows that linearity of the dI signal can be observed in the urine extract diluted more than 1000 times and a dI concentration below around 65 nM. Figure 5.3B shows the signal response of selected urinary metabolites in the linear and suppression regimes from the same dilution experiment. As was the case with the dI standard (Figure 5.3A), the selected metabolite ions and all other species detected in the urine sample show a deflection point between the linear and suppression regime that occurs at the same total sample concentration, equivalent to the dilution of urine extract by about 1000 times. A broad range of log–log slopes, from 0.1 to 0.9, in the suppression regime can also be observed when all
detected ions are examined. Some examples of ions detected in urine with a wide range of
behaviors in the suppression regime are shown in Figure 5.3B, all with good linearity ($R^2 \geq 0.99$). Figure 5.3B also demonstrates that ions with a wide range of ionization properties all show transition between linear and suppression regimes at the same sample concentration. Above about 10 times dilution, many ions show a decrease in intensity.

**Figure 5.3.** Sequential dilution of 2'-deoxyinosine (dI) spiked urine by ESI-FAIMS-MS. (A) A log–log plot of intensity of dI signal VS dilution (bottom axis) with intercepts from eqs 7 and 8 shown on the right and concentration (top axis) with intercepts from eqs 4 and 6 shown on the
left. (B) A log–log dilution plot of selected urinary metabolite ions in the linear and suppression regimes.

In Chapter 3, qualitative analysis of urine extracts diluted 100 times was effective at reducing ionization suppression. Signal intensities of urine metabolite at this dilution factor of 0.01 were more representative of analyte concentration than in more concentrated samples, but here it is determined that linearity of signal response does not occur until even higher dilution factors of 0.001. The ionization capacity of the ESI source and dilution buffer corresponds to around 4 µM total concentration of ionisable components. From this value, it is possible to estimate the total concentration of all ionizable species in undiluted urine extract at around 4 mM.

Unlike the calibration plots in Figures 5.2 and 3A, Figure 5.3B is a sequential dilution plot which does not have concentration on the x-axis. It is therefore impossible to obtain concentration directly from the plot in Figure 5.3B since each analyte has its own unknown IE. Instead, signals in the linear regime of sequential dilution plots (Figure 5.3B) are a combination of metabolite concentration and IE. The relationship between a dilution and a calibration plot is demonstrated graphically in Figure 5.3A. This difference between a calibration plot and a dilution plot can be introduced into eqs 4 and 6 by replacing $c$ with the unknown concentration of the analyte in the undiluted sample ($c_o$) times the dilution factor ($d$) to give an expression for the dilution curve in the linear regime.

$$\log(S) = \log(d) + \log(IEc_o)$$ (7)

and in the suppression regime.

$$\log(S) = n\log(d) + \log(IE_{sup}c_o^n)$$ (8)
Equation 7 can be used along with a measured value of IE from an internal or external standard to obtain $c_0$ and Equation 8 finds further utility in the subsequent sections of this chapter when an analyte cannot be detected in the linear regime.

5.3.3. Suppression and Linearity in Matrix Matched Calibration Curve

Figures 5.2 and 5.3 show that ionization suppression in complex samples is observed in the nonlinear region as a log–log slope between 0 and 1. It was important to establish whether ionization efficiency and consequently analytical sensitivity in the linear regime (log–log slope = 1) is prone to any additional matrix effects in a complex urine sample. In order to clarify the effect of matrix composition on ionization efficiency in the linear ESI region, matrix matched calibration curves using dI standards were constructed and are shown in Figure 5.4. These curves show dI calibration in three different concentrations of the same urine corresponding to the dilution of urine by 10 000, 5000, and 1000 times as well as a neat dI standard. Figure 5.4A confirms that, in the suppression regime, the dI signal is suppressed more significantly at higher matrix concentrations. However, Figure 5.4B, an enlarged plot of the region between 0 and 100 nM dI, clearly shows that in the linear regime the sensitivity of dI is identical for all matrix concentrations. What changes significantly between the different matrix concentrations is the dI concentration at which nonlinearity can be observed. For the 1000 times diluted matrix deflection form, linearity occurs at around 50 nM dI in Figure 5.4B. For the 5000 times dilution of urine extract, the dI signal exhibits linearity until above 500 nM while the 10 000 times dilution and the neat standard exhibit linearity above 1 µM, all in Figure 5.4A. Figure 5.4 demonstrates clearly that, below the ionization capacity of ESI, the analytical sensitivity of a given analyte is constant, regardless of matrix concentration. By comparing analytical sensitivities of dI from the matrix matched calibration curves (Figure 5.4), the spiked urine
sample (Figure 5.3), and mixed standard calibration (Figure 5.2), it can be calculated that the variability in IE between different matrix compositions and across a two week analysis time is also low at 8 % RSD (n = 5). The fact that the slope of the calibration curve in the linear regime of ESI is constant regardless of matrix composition has important implications for the quantitative analysis of urine samples by ESI-FAIMS-MS.

5.3.4 Quantitative Non-target Analysis in the Linear and Suppression Regimes of ESI

A simple approach for non-target semi-quantification has been developed using ESI-FAIMS-MS which can be used for any class of analytes in a real complex sample. By spiking a standard representative of IE of a class of analytes of interest into a sample prior to sequential dilution into the linear regime, the concentration of all analytes of that class that are identified can be estimated. As long as at least three dilution points showing signal linearity are established, they can then be extrapolated to the concentration of the analyte in the undiluted sample using eq 7. Figure 5.3 shows the sequential dilution plot from a urine sample spiked with dI (which is a nucleoside standard not detected in the urine samples prior to spiking). This standard was added at a level 67.5 µM to the urine extract which was then diluted past the suppression regime into the linear range.

This non-target quantification approach was used to determine the concentrations of three modified nucleosides in Chapter 2 and Chapter 3, that could be detected in the linear regime at dilution factors below 0.001 in a real urine sample. By using eq 7 with the measured IE of the internal nucleoside standard dI (Figure 5.3A), identified nucleosides can be quantified in the undiluted urine sample using their intensity in each of the dilution samples in the linear range. This analysis gave values of 5 ± 1 µM 1-methylguanosine (3 linear points, R² > 0.99), 7 ± 1 µM
1-methylinosine (4 linear points, $R^2 > 0.9$), and $4.0 \pm 0.9 \mu M$ N2,N2-dimethylguanosine (3 points, $R^2 > 0.9999$) in urine.

![Matrix matched standard calibration curves](image)

**Figure 5.4.** Matrix matched standard calibration curves of 2'-deoxyinosine (dI) as a neat standard (curve a), in 10 000 × diluted urine matrix (curve b), in 5000 × diluted urine matrix (curve c), and in 1000 × diluted urine matrix (curve d). (A) Full calibration range and (B) low dI concentrations.

The largest source of uncertainty in such an approach when compared to targeted methods of quantitative analysis using isotopically labeled authentic standards is the difference in ionization efficiency between the standard and the analyte. Thus, some knowledge of the range of ionization efficiencies within a class of interest is required in order to determine the range of uncertainty of the method. The ESI-FAIMS-MS sensitivities of ten standards of urinary metabolites from classes identified in Chapter 2 were established by analysis of neat 400 nM standards. They show (Table 1) that nucleosides and nucleic acid bases have analytical
sensitivities within 2-fold for nucleosides and 4-fold for bases. This is a relatively narrow range when it is considered that the relative IE for analytes with a wider range of ionization properties can vary by a factor of over $10^7$. It is also possible to estimate ranges of IE from the literature values for other classes of small biological molecules as a factor of 2 for triglycerides ($n = 23$) and a factor of 5 and 100 for acylcarnitines ($n = 6$) and modified amino acids ($n = 13$), respectively.

Table 5.1: Relative sensitivities of urinary metabolite standards.

<table>
<thead>
<tr>
<th>Analyte Class</th>
<th>Analyte</th>
<th>Relative Sensitivity</th>
<th>% RSDa</th>
<th>m/z</th>
<th>Compensation Voltage (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>nucleosides</td>
<td>dI</td>
<td>1.74</td>
<td>3</td>
<td>253</td>
<td>4.7 V</td>
</tr>
<tr>
<td></td>
<td>UR</td>
<td>0.79</td>
<td>12</td>
<td>245</td>
<td>5.3 V</td>
</tr>
<tr>
<td></td>
<td>MTA</td>
<td>0.88</td>
<td>5</td>
<td>298</td>
<td>4.5 V</td>
</tr>
<tr>
<td></td>
<td>GUA</td>
<td>1.00</td>
<td>6</td>
<td>268</td>
<td>4.4 V</td>
</tr>
<tr>
<td>bases</td>
<td>1-ME-G</td>
<td>0.75</td>
<td>12</td>
<td>166</td>
<td>9.0 V</td>
</tr>
<tr>
<td></td>
<td>7-DAZ-D</td>
<td>3.23</td>
<td>4</td>
<td>151</td>
<td>13.5 V</td>
</tr>
<tr>
<td></td>
<td>N$^2$-DMG</td>
<td>2.63</td>
<td>7</td>
<td>180</td>
<td>9.8 V</td>
</tr>
<tr>
<td>amino acid</td>
<td>Alanine</td>
<td>0.37</td>
<td>5</td>
<td>90</td>
<td>22.5 V</td>
</tr>
<tr>
<td>acylcarnitine</td>
<td>p-carnitine</td>
<td>14.90</td>
<td>4</td>
<td>400</td>
<td>-7.5 V</td>
</tr>
</tbody>
</table>

a % relative standard deviation, $n = 5$

The primary limitation of using dilution to mitigate ionization suppression in ESI, regardless of the separation method used, is that many analytes will be diluted below their limit of detection. The urine samples analyzed in this work contain many hundreds of detectable species when analyzed at moderate dilution factors of 0.1 to 0.01 as in Chapter 3. At very high dilutions where linear response of all analytes is observed (Figure 5.3B), less than 100 of the most abundant ions are typically detectable for quantification. Here, a solution to this problem is proposed based on the concept of extrapolative dilution, previously proposed for use in targeted LC-ESI-MS experiments. In this previous approach, the ratio of measured to known
concentrations in the suppression regime is acquired at a number of dilution factors and extrapolated to infinite dilution. By this method, analytes which could not be detected in samples sufficiently dilute to eliminate suppression were quantified using an approximation of their response in the linear regime. However, extrapolative dilution as described previously could not be applied in non-target analysis without authentic standards because ionization efficiency of the analyte remains unknown.

Here, a novel extrapolative approach is proposed that allows for similar information to be obtained from analytes detected only in the suppression regime in non-target ESI-FAIMS-MS experiments. Critical to this approach is the log–log linearity of dilution curves in the suppression regime. The log–log response plot has traditionally been exploited in studies of ESI fundamentals to show response over a wide range of concentrations, but it is found to have additional utility in non-target quantification, where it can be used to relate the response of an analyte in the suppression regime to its response in the linear regime. Since in non-target analysis the IE of an analyte remains unknown until it is identified and a standard acquired, it is impossible to construct the suppression VS dilution curves used previously for extrapolative dilution in targeted LC-MS. However, in the unique case of ESI-FAIMS-MS, where all analytes are ionized under identical spray and solution conditions, all detected analytes show transition from linear to suppression regimes at the same dilution factor, as shown in Figures 5.2 and 5.3. The extrapolation of the log–log slope in the suppression regime to this deflection point (d_{lin}) can be used to acquire the response of an analyte in the linear regime. At this deflection point, the log–log plots of the linear regime and the suppression regime intersect. By extrapolating the log–log plot in the linear regime from dint to d_o = 0, the signal for the analyte in an undiluted sample absent the effects of suppression can be acquired and used for
quantification. This novel extrapolative approach is demonstrated graphically in Figure 5.5 for the quantification of 1-methylcytidine from dI spiked urine. At the dilution factor where deflection from linearity occurs ($d_{int}$), the fit lines of the linear and suppression regimes intersect (as in Figure 5.3) and $S_{int}$ is equal for equations 7 and 8, allowing them to be equated.

$$\log(c_oIE) + \log(d_{int}) = \log(IE_{sup}c_o^n) + n\log(d_{int})$$ (9)

The virtual signal of the suppression regime in the undiluted sample ($S_{o sup}$) is an easily measurable experimental value of the log of the y-intercept of the suppression curve in Figure 5.5 and can be defined as

$$S_{o sup} = c_o^n IE_{sup}$$ (10)

Combining eqs 9 and 10 gives a simplified expression for analyte concentration.

$$c_o = \frac{S_{o sup} \times d_{int}^{n-1}}{IE}$$ (11)

Equation 11 uses only experimental values from the suppression regime of the analyte being measured along with the IE from a spiked standard and $d_{int}$ measured from any ion in the sample detected in the linear regime.
Figure 5.5. Quantification of urinary 1-methylcytidine by extrapolation of suppression regime sensitivity to the linear regime followed by extrapolation to point $S_o$, the virtual signal intensity in an undiluted sample absent ionization suppression.

By this extrapolative method, quantitative non-target analysis can be carried out on any ion in a sample as long as at least one species in the sample (even the spiked standard) can be detected in the linear regime after dilution to acquire $d_{\text{int}}$ for that sample. This process has been used for the estimation of the concentrations of modified nucleosides identified in Chapter 2 by LC-ESI-MS/MS and Chapter 3 by ESI-FAIMS-MS, but which could not be detected in the linear regime in highly diluted samples in the current work. Using the measured IE of $dI$ and experimental values of $d_{\text{int}}$ and $n$, values of 0.63 $\mu$M for 1-methylcytidine, 0.21 $\mu$M for 1-methyladenosine, and 0.52 $\mu$M for adenosine are obtained. The concentration of $dI$ standard (67.5 $\mu$M) spiked into the urine sample can also be calculated using both external calibration in the linear regime and the extrapolative approach in the suppression regime. External calibration gave a value of 68 $\pm$ 7 $\mu$M for an average of 4 points in the linear regime while the extrapolative approach gave 55 $\mu$M as measured only from points in the suppression regime. It is important to note however that this is significantly better estimation than using any point from the suppression
regime to perform external calibration which gives results ranging from 3 to 30 µM. These preliminary results show that non-target semi-quantification can be carried out in both the linear and the suppression regimes depending on the detected ion intensity. While solution and ionization conditions in the current work were optimized for sensitive detection of modified nucleosides and bases in urine samples, future work will be directed toward optimization of these conditions for quantitative analysis in the nonlinear regime and further validation of the new extrapolative approach. Of particular interest is whether any relationship exists between the sensitivities of analytes in the linear and suppression regimes, which could be exploited to carry out non-target quantification of analytes prior to their identification.

5.4. Conclusions

In this chapter, several questions relating to the principal challenge of using ESI-FAIMS-MS for the quantitative non-target analysis of complex samples have been addressed. Because of the elimination of chemical background from ESI by FAIMS, it is shown that extended analysis times can be efficiently used to increase S/N in linear proportion to analysis time. This can be exploited either to analyze purified analytes at low pM concentrations or to analyze highly diluted complex samples with true linear signal response.

These experiments demonstrate clearly that ionization suppression in ESI of complex samples such as urine manifests itself only as nonlinearity of signal response beyond the certain total concentration of sample components due to the limited ionization capacity of ESI. Below this total sample concentration, all analytes exhibit linear signal response which is independent of matrix concentration or composition. In this linear regime, changes in matrix concentration and composition affect only the concentration range of linear response of analytes but do not
affect their analytical sensitivities (ionization efficiencies). Above the limiting ionization
capacity of ESI, suppression of all analytes is observed as log–log linear signal response which is
highly dependent on the nature of all species in solution. Because FAIMS separation occurs after
ESI, it has no impact on ionization and, thus, these findings apply equally to other hyphenated
MS techniques that include ESI of complex samples.

Quantitative non-target analysis is specifically intended for applications where many
often unexpected analytes of a complex biological or environmental sample need to be identified
and an estimation of concentration is used to determine their potential importance. Here, it has
been demonstrated that ESI-FAIMS-MS can be used in both the linear and suppression regimes
of a sequential dilution experiment to obtain quantitative data during non-target analysis in
addition to the qualitative data necessary for identification described in Chapter 2 and Chapter 3.
Chapter 6

Summary of Conclusions

In the introduction to this thesis the challenges and current capabilities of chemical analysis for the identification of non-target analytes in complex samples were outlined. These include limitations to instrumental methods for comprehensive analysis of complex samples, unknown identification from MS spectral data and quantification of analytes for which no chemical standards are available. In this thesis, the capabilities and limitations of two analytical separation techniques coupled to MS, one widely used in trace analysis and one relatively new emerging technique, were explored for the non-target analysis of urinary metabolites. The current approaches for quantifying and identifying unknown species have also been evaluated and a contribution has been made to the understanding of gas phase ion chemistry during collision induced dissociation of small even electron ions.

It was demonstrated that LC-ESI-MS/MS with data dependent acquisition was capable of providing high quality MS and MS/MS data suitable for identification of the most abundant species detected in urine samples. Identification of 110 urinary metabolites was carried out based solely on MS spectral data without the use of chemical standards and by using all available resources for identification. This integrated approach included (a) searching spectral databases, (b) elucidation of the class of an analyte followed by literature confirmation of the structure, and (c) de novo identification based on the principals of gas phase ion chemistry. The biggest drawbacks to this identification approach were the time and expertise it currently requires. As better tools are being developed for spectral interpretation and database searching of spectra, this
type of non-target identification based on MS spectral data will only be facilitated. This work also identified some fundamental limitations of using LC-ESI-MS for the identification of non-target analytes. These include a limit to the length of time available for MS analysis imposed by the width of an eluting peak in LC, and the high chemical background produced by ESI during ionization of complex samples. This work represents the most comprehensive report to date of the identification of the most abundant metabolites detected in ESI-MS and LC-ESI-MS analysis of urine. This information will be useful for all those developing methods for urine analysis using these techniques as it provides valuable information about the abundant compounds present in the complex matrix.

The limitations of LC-ESI-MS for identification of non-target analytes in this thesis were addressed by the investigation of an alternative separation method, high field asymmetric waveform ion mobility spectrometry (FAIMS). A simple modification of the FAIMS-MS source interface region was proposed and implemented, which allows for the sensitive analysis of small ions which were not transmitted to MS by the original interface. This technical limitation accounts in part for the lack of published results using the first generation of commercial FAIMS instruments. Some unique capabilities of FAIMS for the detection and identification of non-target metabolites in urine were demonstrated. The key advantages of using FAIMS for non-target analysis were the separation of ESI background from analytes of interest prior to MS analysis and the practically unlimited MS analysis times that were possible. The challenge of carrying out direct ESI of urine samples was initially seen as the primary limitation of using ESI-FAIMS-MS for analysis of a complex biological fluid. This was overcome for urine samples by using sample dilution as a method of sample preparation. Consequently, the ESI-FAIMS-MS technique could be used for first time in the analysis of a complex biological sample.
In general, a more detailed understanding of the gas phase ion chemistry than is typically available in the literature was required to allow for de novo identification of unknown metabolites using experimental MS data. During the identification of non-target analytes in this thesis, models for the reactivity of important classes of metabolites in collision induced dissociation (CID) were proposed based on principles of gas phase ion chemistry. Of these, the most detailed was a study of the CID of protonated uracil for which extensive isotope labeling, ion spectroscopy and theoretical data was available in the literature but had remained practically un-interpreted with respect to dissociation mechanisms. Modern MS/MS techniques providing accurate mass and sequential MS/MS were combined with available literature to determine the mechanisms of dissociation of protonated uracil. The most unique part in the dissociation of highly conjugated heterocyclic species such as uracil was found to be their tautomerization prior to dissociation, which occurs in the gas phase after ionization and during collisional excitation. This work represents the only example of a direct link being made between known tautomers detected by ion spectroscopy and the complex reactivity of gas phase ions of heterocyclic species. The tendency for conformational changes such as tautomerization during CID is one of several novel generalizations proposed in this thesis which can be of broader utility for the interpretation of the reactivity of other small, organic even electron ions.

Finally, the implications of using dilution as a method of sample preparation in ESI-FAIMS-MS were investigated. The chief drawback of using sample dilution for sample preparation has been the decrease to sensitivity of the analysis. Because of the elimination of chemical background from ESI by FAIMS, it was shown that analysis time could be efficiently extended to provide a linear increase in signal to noise. The nature of ionization suppression in ESI of complex samples such as urine was also investigated. Ionization suppression was
consistently observed as nonlinearity of signal response beyond the certain total sample concentration due to the limited ionization capacity of ESI. Below this total sample concentration, all analytes exhibit absolute linear signal response which is independent of matrix concentration or composition. Changes in matrix concentration and composition affect only the concentration range of linear response but not the analytical sensitivity. Above the limiting ionization capacity of ESI, suppression of all analytes is observed as log–log linear signal response which is highly dependent on the nature of all species in solution. These two distinct signal response regimes were exploited to analyze metabolites in highly diluted urine samples with absolute linearity of signal response. The above description represents a significant contribution to the understanding of ionization suppression in ESI of complex samples, which had primarily limited to the study of pure standards. Based on these findings, approaches were proposed for the semi-quantification of non-target metabolites newly identified in urine samples. Because of the unique properties of ESI-FAIMS-MS, where all species are ionized under identical solution conditions, it was shown for the first time that it was possible to semi-quantify non-target metabolites in urine samples without using their chemical standards even despite ESI suppression.

The introduction to this thesis describes a number of current examples of research which represent the growing interest in developing a non-target analytical strategy. This thesis represents a number of important steps towards this goal and the results of this work can help guide future research in the field. The identification of unknown analytes from MS and MS/MS spectral data still suffers from significant limitations that need to be addressed by future research. In particular, the utility of spectral libraries of MS/MS data for ESI generated ions is currently very limited compared with those for electron ionization. The adoption of universal standards for
acquiring high quality MS/MS spectra for such databases, significantly expanding the numbers of spectra in such databases and developing new search algorithms will all increase the utility of these databases in the future. This could significantly simplify the process of identification of previously identified unknowns and better enable new workflows which include such identification in the automated data-processing of existing analytical methods.

This thesis demonstrates the possible advantages in MS/MS data quality which can be observed when using FAIMS, an emerging separation technique, as a stand-alone separation tool in place of LC. There are currently a number of limitations to the robustness and reproducibility of this technique which would need to be resolved in order for it to see widespread use as a separation tool in analytical mass spectrometry. As with any emerging analytical technique, it is expected that these factors will be improved in the engineering of successive generations of commercial instruments, ultimately improving the utility of the technique as it is presented here. This thesis addressed a significant problem in the engineering of the source interface for transmitting ions from the outlet of FAIMS to the inlet of MS. There also exists a fundamental limitation to the overall transmission through the FAIMS device due to diffusion. The factors leading to separation of analytes in FAIMS, particularly closely related structural isomers, remains relatively poorly understood and difficult to model. While it is possible to use retention time in LC or migration time in capillary electrophoresis to obtain structural information of unknown species and aid in their identification, this is currently not possible using FAIMS. Improvements in the sensitivity and fundamental understanding of theory of the technique will further enable its utility in non-target analysis.

This thesis represents a significant advancement in the understanding of ESI suppression effects in the direct analysis of complex samples which was required for quantitative non-target
analysis of urine by ESI-FAIMS-MS. The approaches outlined in this thesis currently represent the best possible way of acquiring quantitative data for newly identified species in non-target analysis but with future work significant improvements in this could be possible. Once suppression effects are mitigated using the approaches outlined in this thesis, the biggest limiting factor in obtaining quantitative information for non-target analytes are the large differences which are observed in ESI sensitivity between analytes. Possible future improvement could be made by finding a way to determine the ionization efficiency of an analyte directly from experimental data. As described in Chapter 5, the signal response in the suppression regime of sequentially diluted urine samples show log-log linearity with a slope which is highly dependent on the physicochemical properties of an analyte. With further study, it may be possible to correlate observed non-linearity for an analyte in the suppression regime to the analytical sensitivity of that analyte in the linear regime. Such a correlation would allow for quantification of non-target analytes without a standard, even prior to unknown identification.

This thesis has outlined the challenges, current capabilities and future directions of chemical analysis for the quantification and identification of trace non-target analytes in complex samples. From the very recent literature reviewed, it is clear that this type of analysis is gaining recognition. Research towards enabling the non-target analytical approach is becoming more and more prevalent in fields of mass spectrometry and analytical separations instrumentation, data processing and gas phase ion chemistry. The research in this thesis was carried out and published simultaneously with many such studies and has contributed to the overall advancement of non-target analysis. With further developments in this field, a universal analytical strategy can easily be envisioned in the next decade. This would involve a powerful platform consisting of a global separation technique with ultra-high performance mass spectrometry detection combined with
automated database searching for previously identified species and computational structural elucidation of species not yet present in databases.
References


(58) Kolakowski, B. M.; Mester, Z. Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS). *The Analyst* 2007, 132, 842.


(103) Beach, D. G.; Gabryelski, W. Linear and Nonlinear Regimes of Electrospray Signal Response in Analysis of Urine by Electrospray Ionization-High Field Asymmetric...
Waveform Ion Mobility Spectrometry-MS and Implications for Nontarget Quantification.  


(106) John M. Gregson; James A. McCloskey Collision-induced dissociation of protonated guanine.  


(108) Nelson, C. C.; McCloskey, J. A. Collision-Induced Dissociation of Adenine.  

(109) Zhang, Q.; Wang, Y. Differentiation of 2′-O- and 3′-O-methylated Ribonucleosides by tandem mass spectrometry.  

(110) Bianco, G.; Abate, S.; Labella, C.; Cataldi, T. R. I. Identification and fragmentation pathways of caffeine metabolites in urine samples via liquid chromatography with positive electrospray ionization coupled to a hybrid quadrupole linear ion trap (LTQ) and Fourier transform ion cyclotron resonance mass spectrometry and tandem mass spectrometry.  

(111) Guo, Y. Fragmentation of deprotonated cyclic dipeptides by electrospray ionization mass spectrometry.  


(113) Baker, M.; Gabryelski, W. Collision induced dissociation of deprotonated glycolic acid.  


(115) Witham, A. A.; Beach, D. G.; Gabryelski, W.; Manderville, R. A. Hydroxyl Radical-Induced Oxidation of a Phenolic C-Linked 2′-Deoxyguanosine Adduct Yields a Reactive Catechol.  


(117) Shuker, D. E.; Farmer, P. B. Relevance of Urinary DNA Adducts as Markers of Carcinogen Exposure.  


(119) Morrow, R. J.; Rose, M. E. Isolation of acylcarnitines from urine: a comparison of methods and application to long-chain acyl-CoA dehydrogenase deficiency.  
*Clinica Chimica Acta* 1992, 211, 73–81.


