A Comparison of Collision Cross Section Values Obtained via Travelling Wave Ion Mobility-Mass Spectrometry and Ultra High Performance Liquid Chromatography-Ion Mobility-Mass Spectrometry: Application to the Characterisation of Metabolites in Rat Urine

Leanne C. Nye[1,2,Δ], Jonathan P. Williams[1,Δ], Nyasha C. Munjoma[1,2,Δ], Marine P.M. Letertre[2,Δ], Muireann Coen[2,3], Robbin Bouwmeester[4,5], Lennart Martens[4,5] Jonathan R. Swann[2], Jeremy K. Nicholson[6], Robert S. Plumb[7], Michael McCullagh[1,Δ], Lee A. Gethings[1], Steven Lai[8], James I. Langridge[1], Johannes P.C. Vissers[1,*], Ian D. Wilson[2,*]

1. Waters Corporation, Wilmslow, SK9 4AX, UK
2. Section of Computational and Systems Medicine, Imperial College, London, SW7 2AZ, UK
3. Safety and ADME Translational Sciences, Drug Safety and Metabolism, IMED Biotech Unit, AstraZeneca, Unit 310, Cambridge Science Park, Milton Road, Cambridge, CB4 0WG, UK
4. VIB-UGent Center for Medical Biotechnology, VIB, Ghent, Belgium
5. Faculty of Medicine and Health Sciences, Department of Biochemistry, Ghent University, Ghent, Belgium
6. Australian National Phenome Centre, Murdoch University, Harry Perkins Building, Perth, WA
7. Waters Corporation, Milford, MA, USA
8. Waters Corporation, Beverly, MA, USA

[Δ] contributed equally

* to whom correspondence should be addressed:

i.wilson@imperial.ac.uk

hans_vissers@waters.com

Keywords: Ion Mobility, Collision Cross Section, Metabolomics, Metabolites

Abstract

A comprehensive Collision Cross Section (CCS) library was obtained via travelling wave ion guide mobility measurements through direct infusion (DI). The library consists of CCS and Mass Spectral (MS) data in negative and positive ElectroSpray Ionisation (ESI) mode for 463 and 479 endogenous metabolites, respectively. For both ionisation modes combined, $^{TW}CCS_{N2}$ data were obtained for 542 non-redundant metabolites. These data were acquired on two different ion mobility orthogonal acceleration QToF MS systems in two different laboratories, with the majority of the resulting $^{TW}CCS_{N2}$ values (from detected compounds) found to be within 1% of one another. Validation of these results against two independent, external $^{TW}CCS_{N2}$ data sources and predicted CCS values indicated be within 1-2% of these other values. The same metabolites were then analysed using a rapid reversed-phase ultra (high) performance liquid chromatographic (U(H)PLC) separation combined with IM and MS (IM-MS) thus providing retention time (t_r), *m/z* and $^{TW}CCS_{N2}$ values (with the latter compared with the DI-IM-MS data). Analytes for which $^{TW}CCS_{N2}$ values were obtained by U(H)PLC-IM-MS showed good agreement with the results obtained from DI-IM-MS. The repeatability of the $^{TW}CCS_{N2}$ values obtained for these metabolites on the different ion mobility QToF systems, using either DI or LC, encouraged the further evaluation of the U(H)PLC-IM-MS approach via the analysis of samples of rat urine, from control and methotrexate-treated animals, in order to assess the potential of the approach for metabolite identification and profiling in metabolic phenotyping studies. Based on the database derived from the standards 63 metabolites were identified in rat urine, using positive ESI, based on the combination of t_r, $^{TW}CCS_{N2}$ and MS data.

Keywords: Ion mobility spectrometry, collision cross section, metabolic phenotyping, metabonomics, metabolomics.

1. Introduction

There is increasing interest in the application of ion mobility spectrometry (IMS) in metabolic phenotyping (metabolomics/metabonomics) [1-3] and lipidomic studies [4-12] where it is widely recognized that both the separation and identification of the many hundreds/thousands of metabolites present in samples can represent a major problem for the investigator seeking to find mechanistic "biomarkers". For the detection and identification of analytes present in complex mixtures, such as biological fluids, both the additional separation afforded by IM and the collision cross section (CCS) areas obtained can potentially be of great benefit. Thus, the extra orthogonal separation that can be added by IMS to both DI and chromatographic analysis can enable the resolution of co-eluting species via differences in analyte mobility rather than *e.g.*, differences in LogP as used in *e.g.*, reversed-phase LC separations. This separation not only increases the number of "features" (mass/retention time pairs) detected using MS, but also can provide better MS data by reducing spectral overlap. The orthogonality between separation methods has been discussed previously by e.g., Frahm *et al* [13] and Rodriguez-Suarez *et al.* [14]. Thus the available *m/z* space that can be occupied by a mass analyser, and the consequences of this are on total system peak capacity when one and two-dimensional separation methods are included in the analysis, are discussed conceptually in [13]. Rodriguez-Suarez *et al.* [14] determined empirically that the number of precursor and product ions detected increased on average by a factor of 1.3 when IM was added to either a 1DLC-IM-MS or 2DLC-IM-MS schema. A similar increase in the number of deconvoluted features was noted by Rainville *et al.* [15] when comparing the addition of ion mobility to reversed-phase separations of metabolites in urine using different gradient and column lengths. Moreover, as the CCS value is a physicochemical measurement, it can greatly aid identification, or confirmation of identity. As a result, LC-based applications that include IMS for metabolomics/metabonomics area have begun to grow in number [15-18] and there is also interest in employing IMS in capillary electrophoresis [19]. Previously we have shown that an increased number of features can be detected using U(H)PLC-IM-MS compared to U(H)PLC-MS alone, with various combinations of column length and gradient time [15]. Here, as part of ongoing studies looking to the application of IMS in metabolic phenotyping, $^{TW}CCS_{N2}$ measurements have been measured in triplicate for a library of some 614 metabolites, on each of two different ion mobility-enabled mass spectrometers located in different laboratories. The test metabolites (IROA Mass Spectrometry Metabolite Library of Standards) were first analysed individually using DI on the two ion mobility-enabled mass spectrometers (DI-IM-

MS) and the results additionally validated using external resources. The same set of compounds was then analysed, in duplicate, by microcolumn reversed phase U(H)PLC-IM-MS, using a previously described rapid gradient ("RAMMP LC") method [20], to obtain retention time (t_r) data to further aid metabolite identification in a test set of rat urine samples [21] run under the same conditions.

2 Materials and Methods

2.1 *Chemicals and reagents*

Solvents were of LC-MS grade with water containing 0.1% formic acid (v/v), or LC-MS grade water, both purchased from Fisher Scientific (Loughborough, UK); LC-MS grade acetonitrile containing 0.1% formic acid (v/v) and LC-MS grade methanol and LC-MS grade ethanol were purchased from Sigma Aldrich (Gillingham, Dorset, UK), whilst formic acid was purchased from Fisher Scientific (Loughborough, UK). The test metabolites (the Mass Spectrometry Metabolite Library of Standards (MSMLS), in 96 well plates, (5 µg sample per well) were obtained from IROA Technologies, (Bolton, MA). Samples in plates one to four were dissolved in 400 µL water, those in plate five were dissolved in 400 µL 40% methanol in water (v/v), whilst for plate six, some standards were dissolved in 400 µL 40% methanol in water and some in 400 µL water, and for plate seven the standards were dissolved in 400 µL of a 1:1 v/v water:ethanol mixture. The resulting analyte concentrations were 12.5 µg/mL for each compound. For a full compound listing, plate positioning, molecular formula and molecular mass see **Table S1** in the supplementary data.

For the calibration of the IMS a mixture of compounds was used, covering a range of *m/z* values and selected for use in either positive of negative ESI modes. These calibrants, and their sources, are listed in **Table** 1 below, along with their source. The Major Mix IMS/ToF Calibration Kit was obtained from Waters Corporation, Wilmslow, UK. Compounds with the source Sigma Aldrich were purchased from Sigma-Aldrich, Gillingham, Dorset, UK. Compounds marked with Alfa Aesar were purchased from Alfa Aesar, Heysham, Lancashire, UK.

For mass calibration, a 0.5 M sodium formate solution in water:isopropanol 1:1 v/v was used(Alfa Aesar) . This solution was infused directly into the mass spectrometer, creating a range of sodium formate clusters (when measuring between *m/z* 50 - 1200), with *m/z* values ranging from 91.0 to 1178.8 in positive-ion mode, and from 113.0 to 1132.8 in negative-ion mode, which were compared to reference values to calibrate the time-of- flight tube analyser of the mass spectrometers.

2.2 *Direct infusion (DI) two site determination of $^{TW}CCS_{N2}$ values*

Two sites and two different mass spectrometers were used to determine $^{TW}CCS_{N2}$ values via DI-IM-MS using an HDMS acquisition method. The sites and equipment involved in this study were:

1. A Synapt G2-S Ion Mobility QToF mass spectrometer (Wilmslow, UK) located in the Section of Computational and Systems Medicine within the Department of Surgery and Cancer at Imperial College, London, UK.
2. A Synapt G2-Si Ion Mobility QToF mass spectrometer located at Waters Corporation, Wilmslow, Cheshire, UK.

The two Synapt G2-S and G2-Si platforms hold the same travelling wave ion guide configuration, but the ionization source and ion transfer optics of the G2-Si version of the instrument were changed to improve overall ion transmission through the IM-MS instrument. Prior to use the ion mobility cell settings were standardised for both systems by setting the following values for each instrument: 2 mL/min gas flow for the Trap cell, 90 mL/min gas flow for the mobility (IMS) cell and 180 mL/min for the helium cell, with ramping the mobility cell velocity from 1100 m/s to 300 m/s with a constant wave height of 40 V.

The pressure in the helium cell was approximately 91 bar and the pressure in the IMS cell 0.23 bar. Calibration of the drift cell was achieved using the same combination of compounds, as detailed in the materials and methods section (**Table 1**), and reference values for both instruments. The calibration details are provided in the following paragraph. The instruments were then finally mass calibrated with sodium formate prior to use. LockSpray data were recorded for single point calibration during acquisition.

The mass spectrometers were operated in positive or negative ESI modes with the resolving quadrupole set to a wide pass mode, with a low energy function in continuum mode acquired at a scan rate of 500 ms with an interscan time of 15 ms over the *m/z* range 50 to 1200. The instrument was operated in electrospray mode with LockSpray enabled. The capillary voltage was set to 3 kV in positive-ion mode and 0.5 kV in negative-ion mode; the following conditions were the same for both ion modes: cone voltage 30 V, source temperature 120 ºC. For positive-

ion mode, desolvation gas temperature was 300 ºC with a flow rate of 600 L/hr, and in negative-ion mode the desolvation gas temperature was 450 ºC with a flow rate of 1000 L/hr. Leucine enkephalin was employed as the LockSpray solution at a concentration of 200 pg/µL acquired every quarter of a minute with 3 scans averaging to provide a single point mass and $^{TW}CCS_{N2}$ calibration.

2.2.1 *Travelling-Wave Ion Mobility CCS ($^{TW}CCS_{N2}$) Calibration Procedure*

The Travelling-Wave ion mobility CCS calibration procedure has been described in detail previously [22-24]. The methodology adopted for metabolite CCS measurement was calibration of T-Wave mobility with ionic species of known CCS (Ω) using standard drift tube instruments, whilst the IM-MS definitions employed were taken from Gaelica et al [25]. The T-Wave ion mobility drift times were calibrated using singly charged ions formed from Major Mix IMS/Tof compounds complemented with polyalanine, UltraMark 1621 and several organic acids (**Table 1**). Protonated and deprotonated species of these molecular species provided a mobility calibration over the CCS range from 130-306 $Å^2$ (*m/z* 151-1921) and from 117-367 $Å^2$ (*m/z* 118-1967) for positive and negative ESI, respectively. The normalized collision cross section (Ω') values [(Ω' = ($\Omega_{published}$ x (√µ/z), where µ = reduced mass and z = charge, were fitted against corrected drift time (t'_d) values of the calibrant ions using power regression. The derived calibration coefficients were used to calculate the $^{TW}CCS_{N2}$ of the individual metabolites following measurement of their individual drift times (t'_d) by the analysis software.

2.2.2 *DI-IM-MS*

Loop injections (5 µL in positive-ion mode and 10 µL in negative-ion mode) of each of the individual standards from the IROA MSMLS metabolites were made in triplicate (on the same day) using an AQUITY UPLC I-Class sample manager (Waters Corporation, Milford, MA) at each site, in both positive and negative electrospray (ESI) ion-mode, resulting in up to 12 measurements per analyte. Samples were infused for 60 s via the LC system, with the LC column replaced with 100 cm of 0.004 inch i.d. PEEK tubing to ensure that the sample produced a signal for approximately 40 s, using a solvent of water:acetonitrile (1:1 v/v), containing 0.1% formic acid, at a flow rate of 0.05 mL/min. Each sample was injected individually to ensure no suppression of signal. Detection was considered when a compound

was measured in at least two technical replicates and confirmed when detected at one and measured at the other site.

2.2.3 *U(H)PLC-IM-MS*

This part of the study was performed on the Synapt G2-S Ion Mobility QToF instrument at Imperial College with the chromatographic separation performed using an AQUITY UPLC I-Class system. The U(H)PLC conditions were those previously described [24] as the "RAMMP" method and employed a short reversed-phase gradient (2.5 min including re-equilibration) using a 1 x 50 mm HSS T3 1.8 µm column to provide a high throughput chromatographic separation. For the U(H)PLC-IM-MS experiments, the MSMLS test metabolites were each injected individually (5 µL), in duplicate, at the same concentration as for the DI-IM-MS experiments (12.5 µg/mL) in order to obtain retention time, $^{TW}CCS_{N2}$ and MS data. Duplicates for positive and negative-ion mode data were acquired.

2.2.4 *Application of U(H)PLC-IM-MS to Rat Urine Samples*

A set of ten rat urine samples, obtained from 5 vehicle dosed control and 5 treated animals following a single oral dose (40 mg/kg) of methotrexate (MTX), were used to evaluate the IMS system. These samples, taken 48-hours after the commencement of the study, represent a subset of those obtained during an investigation by the University of Arizona in collaboration with Imperial College London (ICL) on the effects of MTX on healthy animals and in NASH (non-alcoholic steatohepatitis) models to explore the association between liver pathology and adverse drug reactions (ADRs) induced by MTX treatment [21]. The study approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Arizona. The work was undertaken in with the NIH guidelines on the care and use of experimental animals as previously described [21].

The rat urine samples were prepared as described previously [24]. Briefly, 20 µL of each urine sample was mixed with 60 µL of MeOH and stored at -20˚C overnight for protein removal. The samples were then centrifuged (15,000 g, 5 min, 4˚C). From the supernatants, 25 µL was transferred into a 350 µL 96 well-plate and 225 µL of water was then added to each sample. The plates were centrifuged (700 g, 5 min) and placed into the auto sampler at 4 ˚C. The samples were analysed in positive ESI-ion mode using a sample volume of 5 µL under the

same conditions as those used to acquire the data for the standards apart from the MS scan rate, which was set at 100 ms and the desolvation gas temperature, which was set at 450 ºC.

2.2.5 *Machine Learning Procedure*

$^{TW}CCS_{N2}$ predictions were obtained with a model trained with machine learning. The approach is similar to the method of Zhou et al. [26] but is trained with $^{TW}CCS_{N2}$ to fit an appropriate model. The $^{W}CCS_{N2}$ data that were internally acquired with IMS-Q-oaToF and Q-IMS-oaToF geometries and covers a wide range of polarity and a large number of chemical classes. For each compound, 196 chemical descriptors were extracted [27,28] and a model was trained with a gradient boosting algorithm [29]. $^{TW}CCS_{N2}$ predictions for compounds described here were obtained using a nested 10-fold cross-validation strategy [27,30]. This means that the compounds and their $^{TW}CCS_{N2}$ predictions were not part of the training set used for optimizing the model parameters and hyperparameters. Prediction method and model details are presented elsewhere [30].

2.2.6 *Data Analysis*

The DI-IM-MS acquired data were processed, including 4D peak detection, lock mass and drift lock correction, and targeted screening, using UNIFI Scientific Information System software 1.8.2.0 (Waters Corporation, Milford, MA). The LC-IM-MS data were processed with UNIFI and Progenesis QI (Nonlinear Dynamics, Newcastle upon Tyne, UK). In positive-ion mode, the protonated, sodiated and potassiated adducts were searched for, and in negative-ion mode, only deprotonated adducts, using a 5 ppm mass tolerance. Multivariate analysis was conducted with SIMCA-P+ v11 (Umetrics, Malmö, Sweden). Chemical classification of the content of the MSMLS library was conducted with ClassyFire [32] and the Medical Subject Headings (MeSH) controlled vocabulary (https://pubchem.ncbi.nlm.nih.gov/classification) applied for application centric classification of the compounds.

3. Results and Discussion

3.1 *Direct infusion IM-MS Studies*

As indicated in the introduction, the application of IMS in *e.g.*, LC-MS-based metabolic phenotyping studies offers both an orthogonal, and rapid, separation in addition to that provided by chromatography or electrophoresis. IMS also provides a degree of separation in DI-MS-based assays. For both DI-MS and LC-MS-based analyses IMS can, in addition, also be used to acquire ion-specific physicochemical measurements under a set of given experimental conditions in the form of the CCS values. Whilst MS-dependent data, such as *e.g.* the molecular mass of a compound can be determined with high accuracy using ion mobility enabled MS acquisitions, and is readily transferable between instruments and laboratories, it is less clear that CCS measurements are as robust. We therefore undertook a two site investigation using two similar, but not identical, IMS-capable MS instruments to determine both the within and between site repeatability of CCS determinations. At this stage of the analysis, the identification of a compound was accepted when it was detected at both sites using the criteria specified in the Experimental Section. In total, in positive-ESI mode, IM-MS data for 510 compound measurements were initially obtained and $^{TW}CCS_{N2}$ values and mass spectra recorded. Following careful review of the data and the removal of adducts and the duplicates within the MSMLS set, $^{TW}CCS_{N2}$ values and mass spectra for a total of 463 unique compounds were obtained by combining the data from both sites. However, differences were noted between sites in terms of unique compounds detected by MS. Consequently, 404 of the MSMLS standards were detected at site 1, whilst another 404 were detected at site 2, with some compounds (118) only detected on one IMS-MS system or the other instruments, but not both. As a result, the detection and $^{TW}CCS_{N2}$ values of only 345 compounds were common to both instruments. Such differences are not unexpected given that differences in ionisation efficiency between different instruments are frequently observed. The formed adducts were broadly the same for both instruments; however, for a subset of 36 compounds, different adducts (H^+, Na^+, or K^+) were detected. For those compounds observed in positive ESI by both sites, with the same adduct, a maximum relative $^{TW}CCS_{N2}$ difference of 2% was found between the averaged $^{TW}CCS_{N2}$ values from the two sites. The detected $^{TW}CCS_{N2}$ values for DI-IM-MS using +ve ESI MS are summarised in supplementary **Tables S2a**.

Similarly, in negative ESI mode, IM-MS results were obtained for a total of 479 unique compounds. Once again, not all compounds were detected using both instruments, with 418 for site 1 and 437 detected by site 2. Overall, 376 compounds were detected by both instruments at both sites, with a further 103 compounds only detected by one site or the other. In negative ESI mode, all compounds measured by both sites were well within 2% of the averaged $^{TW}CCS_{N2}$ values. The detected $^{TW}CCS_{N2}$ values for DI-IM-MS with negative ESI MS are given in supplementary **Table S2b**). The measurement error distributions, *i.e.* the % differences between the average $^{TW}CCS_{N2}$ values from the two sites for both positive and negative ESI, following curation as described in the following two sections and centring as expected at means of 0%, are shown in **Figure 1.**

The 79 compounds in the MSMLS test set that were detected in neither positive nor negative ESI mode are listed in the supplementary **Table S3**. The majority of these non-detected metabolites (47) had a neutral molecular mass of below *m/z* 150 and, whilst such compounds may be often detectable, the $^{TW}CCS_{N2}$ accuracy obtained could be lower. Improved detection below m/z 150 can be achieved by optimization of the gas pressure and ion optics voltage in certain regions of the instrument. This was however not considered since CCS would add little additional specificity to the detection of such low molecular mass compounds. Moreover, these would not represent typical high-throughput metabolomics settings, i.e. experimental MS conditions. A small number of the non-detected metabolites had a relatively high molecular mass, with 5 compounds having a molecular mass greater than 600 Da, the largest being vitamin B12 at 1354.6 Da.

The results of the DI-IM-MS experiments, following further curation and statistical outlier removal, either the inter or intra sample standard deviation, or both, exceeding a 95% confidence level, are graphically summarised in **Figures 1** to **4**. The $^{TW}CCS_{N2}$ and MS/MS coverage and intersection for both ionisation modes are shown in **Figure 2**. Site two also collected accurate mass CID MS/MS data [33] that can additionally be used in the identification schema as demonstrated in supplementary **Figure S1**. Since, as previously mentioned, some MSMLS compounds are represented more than once, in different (adducted) forms or as duplicates, within the library, experimental $^{TW}CCS_{N2}$ data were obtained for a total of 568 species/compound forms. The results shown in **Figure 1** demonstrate both inter and intra site precision and provide a comparison with external data available in the public domain [34]. Note that the latter IMS data were acquired using a previous generation travelling wave enabled

IM-MS instrument (Synapt HDMS Q-TOF Mass Spectrometer, Waters Corporation) and used a different informatics analysis tool, comprising another peak detection algorithm, for data processing. Therefore, slightly higher deviations are to be expected but the reported values were still found to be within the expected error distribution range. However, the intra-site precision results illustrate the robustness/repeatability of IMS measurements of modern IM-MS instruments since the data was, in part, *i.e.* the Paglia *et al.* data set [34] *vs.* site 1, 2 (and 3) results, acquired more than five years apart. A recent study demonstrated similar inter and intra-day precision figures of merit using a different IMS analyser type [35, 36]. Both external data sets are provided as a reference in supplementary **Table S4**. A comparison of the obtained *vs.* predicted $^{TW}CCS_{N2}$ values, including an assessment of the accuracy of predictions using publically available CCS data [26] and instrument specific training data sets, is provided in **Figure 3**. The prediction errors for most analytes are within 2 %, and thus show that the predicted $^{TW}CCS_{N2}$ values are highly correlated with the experimentally observed values. Lastly, the detected *vs.* non-detected MSMLS library compound distributions and the relative class specific detected *vs.* non-detected class specific ratio for both ionisation modes are shown in in **Figure 4**, illustrating that three chemical classes in particular were underrepresented in negative electrospray mode compared to positive electrospray ionization, namely organometallic, represented by a mere two compounds with the MSMSL library as a whole, organic nitrogen compounds, represented by 22 compounds, and benzenoids, represented by 50 compounds. The curated and comparative results are summarised numerically in **Table 2**.

As can be observed from **Figures 1 to 4** and **supplementary Tables 2a** and **2b,** the $^{TW}CCS_{N2}$ values from the different sites show slight, but very minor biases. In addition, some compounds were detected in the protonated form at one site, and as sodiated adducts at the other, or *vice versa*, possibly caused by differences in glassware and/or solvents used. As adduct formation is an in-source phenomenon, and instrument and experimental settings dependent, and as the $^{TW}CCS_{N2}$ values differed slightly, despite having, as far as possible, identical settings, and being calibrated in the same way, it may be that, at least in the short term, databases involving CCS information would need to be tailored for individual instruments and/or geometries. Lastly, as illustrated in **Figure 5**, it was noted that the observed $^{TW}CCS_{N2}$ data correlated well with class-specific *m/z* values, which has been previously reported for lipids and other compound classes [6,11,12, 37].

3.2 *Comparison of DI-IM-MS vs. U(H)PLC-IM-MS*

The addition of a chromatographic LC separation step prior to IM-MS resulted in a reduction in the number of detected metabolites. This reduction in detectability was probably due, at least in part, to compounds either not being retained, or not eluted, from the column, and eluting within the solvent front. Poor chromatographic properties and ion suppression may also have been contributing factors. Nevertheless, a total of 301 compounds were detected with the rapid U(H)PLC-IM-MS method in positive-ion mode providing a complete set of $^{TW}CCS_{N2}$ values, MS and t_r data for these metabolites (the complete list is provided in supplementary **Table S5**). Comparison of the LC-derived $^{TW}CCS_{N2}$ values against the averaged values found from direct infusion showed that 270 (89.7%) of the LC-IM-MS-detected compounds were within ±1% and 292 (97.0%) were within ±2% of the values obtained from DI-IM-MS. The remaining nine compounds showed somewhat larger differences for the $^{TW}CCS_{N2}$ values, possibly because of low abundance/challenged ion statistics, with the sodiated adduct of gamma-linolenic acid showing the largest deviation at 4.7%. A comparison of the obtained average values for DI-MS and LC-MS/MS is provided in **Figure 6**.

3.3 *Identification of metabolites in rat urine by RP-RAMMP-U(H)PLC-IM-MS/MS*

The identification of metabolites in biological samples such as urine provides a particular problem for metabolomic/metabonomic studies because of the complexity of the mixtures of metabolites encountered and their structural diversity. In **Figure 7** the type of data that can be obtained for a sample of rat urine using the combination to U(H)PLC-IM-MS is shown.
The MSMLS sample set contains a wide range of metabolites, not all of which are likely to be present in samples such as urine (*e.g.*, NAD, fructose 1,6-biphosphate, 2'-deoxyguanosine 5'-triphosphate, *etc.*), even if such compounds were retained under the reversed-phase LC conditions employed here. However, the data for the ten rat urine samples (five from control and five from methotrexate-dosed animals) analysed with the RAMMP method in positive ESI mode were searched against the database created from the $^{TW}CCS_{N2}$ values obtained from the direct infusion experiments of the whole MSMLS sample set. To be considered positive any putative database identification from this compound search had to meet the dual criteria of having a $^{TW}CCS_{N2}$ value within 2% of the mean DI-IM-MS value and an *m/z* within 10 ppm of that of the suggested metabolite. The resulting list of potential urinary-excreted metabolites indicated as being present by this search was then further assessed with respect to the RAMMP-

derived t_r, which had to be to be within 0.2 min of the library value, thereby providing three points of contact for each metabolite, and clearly indicating where further confirmation by reanalysis after spiking in an authentic standard would be appropriate.

Of the nearly 300 compounds from the MSMLS sample set detectable in positive ESI using the RP-LC RAMMP-IM-MS method the putative identification of 63 metabolites (some of which were detected in only 1 sample) was supported by all three criteria with the $^{TW}CCS_{N2}$ values reported in **Tables S2a** and **S2b**, and the retention time data provided in the supplementary information (**Table S6**). The repeatability of the retention time data obtained for these analytes is also indicated by the data provided in **Table S6.**

Whilst this represents a relatively small proportion of the total number of ~ 2,700 deconvoluted compounds, *i.e.* the grouping of the features of multiple adduct forms and charge states into a single neutral mass, present in these samples it nevertheless represents a promising start and, as compound collections/databases increase in size, should help to reduce the bottleneck of compound identification for relatively common (and therefore commercially available) metabolites that currently plagues metabolic phenotyping studies. However, in our view (at least for the foreseeable future) MS and CCS values alone will still be insufficient for unequivocal identification and chromatographic retention time data will remain key to discriminating between isomeric compounds (and should there be no chromatographic separation in the LC system used one would have to be specifically developed). Such compounds represent a significant subset of important metabolite classes (*e.g.*, monosaccharides) and unequivocal identification may still require comparison with an authentic standard on the same system. The results shown top right in **Figure 6** graphically illustrate the precision of the $^{TW}CCS_{N2}$ measurements for urine, from both the control and the MTX-dosed animals, as a function of identification frequency. The average % CV values *vs.* the library complement were 0.5% and 0.6% for the control and MTX-dosed rat urine samples, respectively. Note that these relative values are derived from native urine metabolite ion detections and can therefore be of low abundance, thereby potentially challenging ion statistics and the subsequent *m/z*, $^{TW}CCS_{N2}$ and intensity read-outs. Shown bottom right in **Figure 6** are the unsupervised PCA results of the tentatively identified metabolites, illustrating that differentiation of the samples was feasible based on a subset of identified compounds for *m/z*, t_r, $^{TW}CCS_{N2}$, and abundance.

As is clear from the data provided in **Table 3**, there are several compounds detected in the samples from all, or the majority, of animals, such as *e.g.*, 3-(2-hydroxyphenyl)propanoate, 3',5'-cyclic AMP, 5-hydroxy-L-tryptophan, 5'-methylthioadenosine, adenine, adenosine, creatine, creatinine, D-pantothenic acid, lumichrome (6,7-dimethylalloxazine), riboflavin, sphinganine, suberic acid, sucrose, trans-cinnamate, urocanate, xanthine and xanthurenic acid whilst others were found less frequently, including some such as urate and taurine which were detected only in a single sample.

Clearly, had this been a full-scale metabolic phenotyping exercise, with the aim of biomarker discovery, rather than a more limited "proof of concept" exercise, metabolite identification would have been concentrated on those compounds that discriminated between the control and MTX-treated groups. The use of rotationally averaged collisional cross sections as a means for identification and confirmation of compound identity is an intriguing prospect, providing a physicochemical supplement to retention time and tandem MS information. The limiting factor is currently the lack of CCS measurements populated into metabolite compound libraries, and the lack of computational tools to rapidly generate theoretical, either calculated using molecular modelling approaches [38] or predicted with machine learning based tools [26,39], CCS values from compound structure. Although there has been development, across multiple areas of research, in terms of empirical library development [40-43], including pesticides, veterinary drugs, mycotoxins, metabolism, steroids and steviol glycosides, undoubtedly further progress will make the routine use of IMS information for identification purposes a powerful analytical complement. On the basis of this investigation it is however clear that rapid U(H)PLC-IM-MS shows promise when applied to biological samples.

4. Conclusions

The $^{TW}CCS_{N2}$ values for a range of metabolite standards, determined in triplicate using standardised settings, measured on two separate instruments (located on different sites) were found to be very similar. For DI-IM-MS it was possible to obtain $^{TW}CCS_{N2}$ values that were well within 1% of each other on two different sites and within 1-2% of two external reference sources. Application of U(H)PLC separation prior to IM-MS, gave results for 292 (97.0%) metabolites that were within ±2% of those measured with DI-IM-MS, with some 270 (89.7%). within ±1%. Analysis of rat urine samples by U(H)PLC-IM-MS enabled 65 compounds to be

identified using the combination retention time, $^{TW}CCS_{N2}$ and MS data. This demonstrates the potential utility of adding IMS with CCS values to metabolite identification studies.

5. Acknowledgements

Chris Hughes, Mark Towers, David Eatough, and Richard Tyldesley-Worster are kindly acknowledged for their assistance with data acquisition and processing. ML is supported as part of the STRATIGRAD program at IC. RB is part of MASSTRPLAN and receives funding from the Marie Sklodowska-Curie EU Framework for Research and Innovation Horizon 2020, under Grant Agreement No. 675132.

5. References

[1] T. Mairinger, T.J. Causon, S. Hann, The potential of ion mobility–mass spectrometry for non-targeted metabolomics, Curr. Opinion Chem. Biol. 42 (2018) 9-15.

[2] X. Zhang, K. Quinn, C. Cruickshank-Quinn, R. Reisdorph, N. Reisdorph, The application of ion mobility mass spectrometry to metabolomics. Curr. Opinion Chem. Biol. 42 (2018) 60–66.

[3] C. M. Nichols, J. N. Dodds, B. S. Rose, J. A. Picache, C. B. Morris, S. G. Codreanu, J. C. May,S. D. Sherrod, J. A. McLean, Untargeted molecular discovery in primary metabolism: collision cross section as a molecular descriptor in ion mobility-mass spectrometry. Anal. Chem. in press DOI: 10.1021/acs.analchem.8b04322

[4] C. Hinz, S. Liggi, J. L. Griffin, The potential of Ion Mobility Mass Spectrometry for high-throughput and high-resolution lipidomics, Curr. Opinion Chem. Biol. 42 (2018) 42–50.

[5] G.Paglia, G. Astarita, Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. Nature Protocols. 12 (2017) 797-813.

[6] J.E. Kyle, X. Zhang, K.K. Weitz, M.E. Monroe, Y.M. Ibrahim, R.J. Moore, J. Cha, X. Sun, E.S. Lovelace, J. Wagoner, S.J. Polyak, T.O. Metz, S.K. Dey, R.D. Smith, K.E. Burnum-Johnson, E.S. Baker, Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry, Analyst. 141 (2016) 1649-1659.

[7] V. Shah, J.M. Castro-Perez, D.G. McLaren, K.B. Herath, S.F. Previs, T.P. Roddy, Enhanced data-independent analysis of lipids using ion mobility-TOFMSE to unravel quantitative and qualitative information in human plasma, Rapid Commun. Mass Spectrom. 27 (2013) 2195-2200.

[8] G. Paglia, M. Kliman, E. Claude, S. Geromanos, G. Astarita, Applications of ion-mobility mass spectrometry for lipid analysis, Anal. Bioanal. Chem. 407(2015) 4995-5007.

[9] G. Paglia, P. Angel, J.P. Williams, K. Richardson, H.J. Olivos, J.W. Thompson, L. Menikarachchi, S. Lai, C. Walsh, A. Moseley, R.S. Plumb, D.F. Grant, B.O. Palsson, J. Langridge, S. Geromanos, G. Astarita, Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification, Anal. Chem. 87 (2015) 1137–1144.

[10] K.M. Hines, J. Herron, L. Xu, Assessment of altered lipid homeostasis by HILIC-ion mobility-mass spectrometry-based lipidomics, J. Lipid Res. 58 (2017) 809-819.

[11] M. Kliman, J.C. May, J.A. McLean, Lipid analysis and lipidomics by structurally

selective ion mobility-mass spectrometry, Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. 1811(2011) 935-945.

[12] K.M. Hines, J.C. May, J.A. McLean, L. Xu, Evaluation of collision cross section calibrants for structural analysis of lipids by traveling wave ion mobility-mass spectrometry, Anal. Chem. 88 (2016) 7329–7336.

[13] J. Frahm, B.E. Howard, S. Heber S, D.C. Muddiman. Accessible proteomics space and its implications for peak capacity for zero-, one- and two-dimensional separations coupled with FT-ICR and TOF mass spectrometry, J Mass Spectrom. 41(2006)281-288.

[14] E. Rodriguez-Suarez, C. Hughes, L. Gethings, K. Giles, J. Wildgoose, M. Stapels, E. Fadgen, J. Keith, S. Geromanos, J. Vissers, F. Elortza, J. Langridge, An Ion Mobility Assisted Data Independent LC-MS Strategy for the Analysis of Complex Biological Samples. Curr. Anal. Chem. 9 (2013) 199-211.

[15] P.D. Rainville, I.D. Wilson, J.K. Nicholson, G. Isaac, L. Mullin, J.I. Langridge, R.S. Plumb, Ion mobility spectrometry combined with ultra performance liquid chromatography/mass spectrometry for metabolic phenotyping of urine: Effects of column length, gradient duration and ion mobility spectrometry on metabolite detection, Anal. Chim. Acta. 982 (2017) 1–8.

[16] E.L. Harry, D.J. Weston, A.W.T. Bristow, I.D. Wilson, C.S. Creaser, An approach to enhancing coverage of the urinary metabonome using liquid chromatography-ion mobility-mass spectrometry, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 871 (2008) 357-361.

[17] A. Malkar, N.A. Devenport, H.J. Martin, P. Patel, M.A. Turner, P. Watson, R.J. Maughan, H.J. Reid, B.L. Sharp, C.L.P. Thomas, J.C. Reynolds, C.S. Creaser, Metabolic profiling of human saliva before and after induced physiological stress by ultra-high performance liquid chromatography-ion mobility-mass spectrometry, Metabolomics. 9 (2013) 1192–1201.

[18] X. Zhang, K. Kew, R. Reisdorph, M. Sartain, R. Powell, M. Armstrong, K. Quinn, C. Cruickshank-Quinn, S. Walmsley, S. Bokatzian, E. Darland, M. Rain, K. Imatani, N. Reisdorph, performance of a high-pressure liquid chromatography-ion mobility-mass spectrometry system for metabolic profiling, Anal. Chem. 89 (2017) 6384–6391.

[19] N. Drouin, J. Pezzatti, Y. Gagnebin, V. González-Ruiz, J. Schappler, S. Rudaz, Effective mobility as a robust criterion for compound annotation and identification in metabolomics: Toward a mobility-based library, Anal. Chim. Acta. 1032 (2018) 178-187.

[20] N. Gray, K. Adesina-Georgiadis, E. Chekmeneva, R.S. Plumb, I.D. Wilson, J.K.

Nicholson, Development of a rapid microbore metabolic profiling ultraperformance liquid chromatography-mass spectrometry approach for high-throughput phenotyping studies, Anal. Chem. 88(2016) 5742-5751.

[21] M. Kyriakides, R.N. Hardwick, Z. Jin, M.J. Goedken, E. Holmes, N.J. Cherrington, M. Coen, Systems level metabolic phenotype of methotrexate administration in the context of non-alcoholic steatohepatitis in the rat, Toxicol. Sci. 142 (2014) 105–116

[22] B.T. Ruotolo, J.L. Benesch, A.M. Sandercock, S.J. Hyung, C.V. Robinson. Ion mobility-mass spectrometry analysis of large protein complexes. Nat. Protoc. 3(2008)1139-1152.

[23] J.G. Forsythe, A.S. Petrov, C.A. Walker, S.J. Allen, J.S. Pellissier, M.F. Bush, N.V. Hud, F.M. Fernández. Collision cross section calibrants for negative ion mode traveling wave ion mobility-mass spectrometry. Analyst. 21;140 (2015):6853-6861.

[24] Q. Duez, F. Chirot, R. Liénard, T. Josse, C. Choi, O. Coulembier, P. Dugourd, J. Cornil, P. Gerbaux, J. De Winter. Polymers for Traveling Wave Ion Mobility Spectrometry Calibration. J Am Soc Mass Spectrom. 28 (2017)2483-2491.

[25] V. Gabelica, A.A. Shvartsburg, C. Afonso, P. Barran, J.L.P. Benesch, C. Bleiholder, M.T. Bowers, A. Bilbao, M.F. Bush, J.L. Campbell, I.D.G. Campuzano, T. Causon, B.H. Clowers, C.S. Creaser, E. De Pauw, J. Far, F. Fernandez-Lima, J.C. Fjeldsted, K. Giles, M. Groessl, C.J. Hogan Jr, S. Hann, H.I. Kim, R.T. Kurulugama, J.C. May, J.A. McLean, K. Pagel, K. Richardson, M.E. Ridgeway, F. Rosu, F. Sobott, K. Thalassinos, S.J. Valentine, T. Wyttenbach. Recommendations for reporting ion mobility Mass Spectrometry measurements. Mass Spectrom. Rev. 2019. https://doi.org/10.1002/mas.21585

[26] Z. Zhou, X. Shen, J. Tu, Z.J. Zhu, Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry, Anal. Chem. 88 (2016) 11084–11091. 11084–11091

[27] R. Bouwmeester, L. Martens, S. Degroeve. Comprehensive and empirical evaluation of machine learning algorithms for small molecule LC retention time prediction. Anal. Chem., 91(2019), 3694–3703.

[28] G. Landrum. The RDKit Documentation — The RDKit 2016.09.1 documentation. (2016). Retrieved January 31, 2017, from http://www.rdkit.org/docs/

[29] T.Chen, C. Guestrin. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (2016) pp. 785–794.

[30] Wessels, L. F. A., Reinders, M. J. T., Hart, A. A. M., Veenman, C. J., Dai, H., He, Y. D.,

& van't Veer, L. J. (2005). A protocol for building and evaluating predictors of disease state based on microarray data. Bioinformatics, 21(19), 3755–3762.

[31] R Bouwmeester, L Martens, S Degroeve, K Richardson, JPC Vissers, Predicting Ion Mobility Collision Cross Sections by Combining Conventional and Data Driven Modelling. in: ASMS Proc., 2019: p. submitted.

[32] Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner, D.S. Wishart, ClassyFire: automated chemical classification with a comprehensive, computable taxonomy, J. Cheminform. 8 (2016) 61.

[33] J.P. Williams, D. Eatough, L.A. Gethings, C.J. Hughes, M. Towers, L. Nye, S. Lai, R. Tyldesley-Worster, J.P. Vissers, S. Dhungana, Automatic CCS and MS/MS Library Creation and Application for Large Scale Metabolic Profiling, in: ASMS Proc., 2016: p. MP 258.

[34] G. Paglia, J.P. Williams, L. Menikarachchi, J.W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, A. Moseley, D. Grant, J. Langridge, B.O. Palsson, G. Astarita, Ion mobility derived collision cross sections to support metabolomics applications, Anal. Chem. 86 (2014) 3985–3993.

[35] F. Zhang, S. Guo, M. Zhang, Z. Zhang, Y. Guo, Characterizing ion mobility and collision cross section of fatty acids using electrospray ion mobility mass spectrometry. J Mass Spectrom. 50 (2015) 906-913.

[36] S.M. Stow, T.J. Causon, X. Zheng, R.T. Kurulugama, T. Mairinger, J.C. May, E.E. Rennie, E.S. Baker, R.D. Smith, J.A. McLean, S. Hann, J.C. Fjeldsted. An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. Anal Chem. 89 (2017) 9048-9055.

[37] X. Zheng, N.A. Aly, Y. Zhou, K.T. Dupuis, A. Bilbao, V.L. Paurus, D.J. Orton, R. Wilson, S.H. Payne, R.D. Smith, E.S. Baker, A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry, Chem. Sci. 8 (2017) 7724-7736.

[38] S.M. Colby, D.G. Thomas, J.R. Nuñez, D.J. Baxter, K.R. Glaesemann, J.M. Brown, M. Pirrung, N. Govind, J.G. Teeguarden, T.O. Metz, R.S. Renslow, ISiCLE: A molecular collision cross section calculation pipeline for establishing large in silico reference libraries for compound identification, (2018). https://arxiv.org/abs/1809.08378.

[39] M. Heinonen, H. Shen, N. Zamboni, J. Rousu, Metabolite identification and molecular fingerprint prediction through machine learning, Bioinformatics. 28 (2012) 2333–2341.

[40] M. McCullagh, D. Douce, E. Van Hoeck, S. Goscinny, exploring the complexity of steviol glycosides analysis using ion mobility mass spectrometry, Anal. Chem. (2018) 90 (2018) 4585–4595.

[41] L. Righetti, M. Fenclova, L. Dellafiora, J. Hajslova, M. Stranska-Zachariasova, C. Dall'Asta, High resolution-ion mobility mass spectrometry as an additional powerful tool for structural characterization of mycotoxin metabolites, Food Chem. 245 (2018) 768–774.

[42] M. Hernández-Mesa, B. Le Bizec, F. Monteau, A.M. García-Campaña, G. Dervilly-Pinel, Collision Cross Section (CCS) Database: an additional measure to characterize steroids, Anal. Chem. 90 (2018) 4616–4625.

[43] E. Sinclair, K.A. Hollywood, C. Yan, R. Blankley, R. Breitling, P. Barran, Mobilising ion mobility mass spectrometry for metabolomics, Analyst. 143 (2018) 4783–4788.

Figure Captions

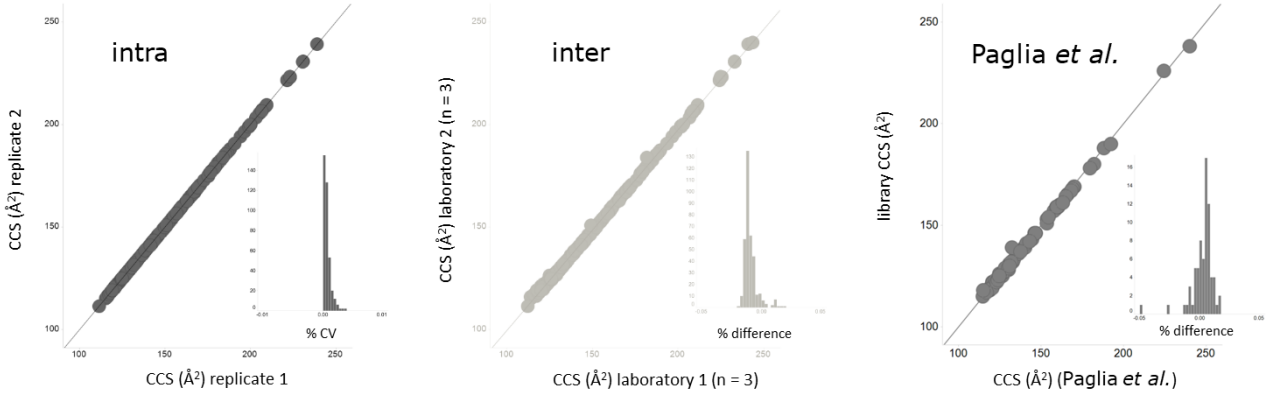

Figure 1. Intra and inter (internal and external Paglia et al. [24]) DI-IM-MS TWCCSN2 measurement precision.

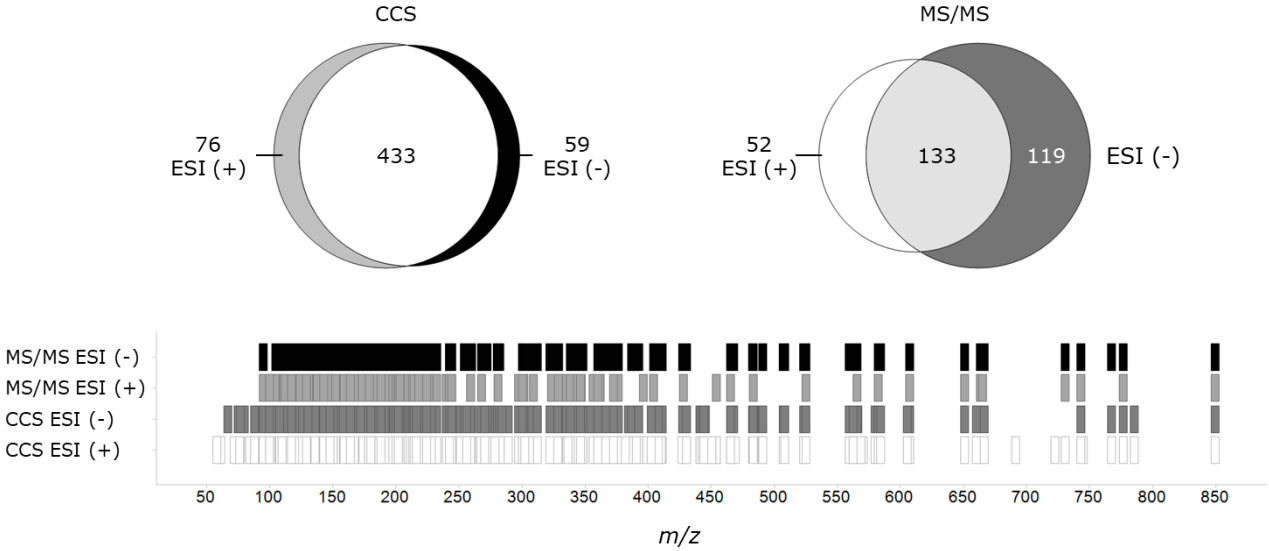

Figure 2. CCS and MS/MS library coverage (CID fragmentation data/results not shown (available for download at http://nonlinear.com/progenesis/qi/v2.2/download/ccs-libraries/ and https://marketplace.waters.com/apps/177290/metabolic-profiling-ccs-library#!overview).

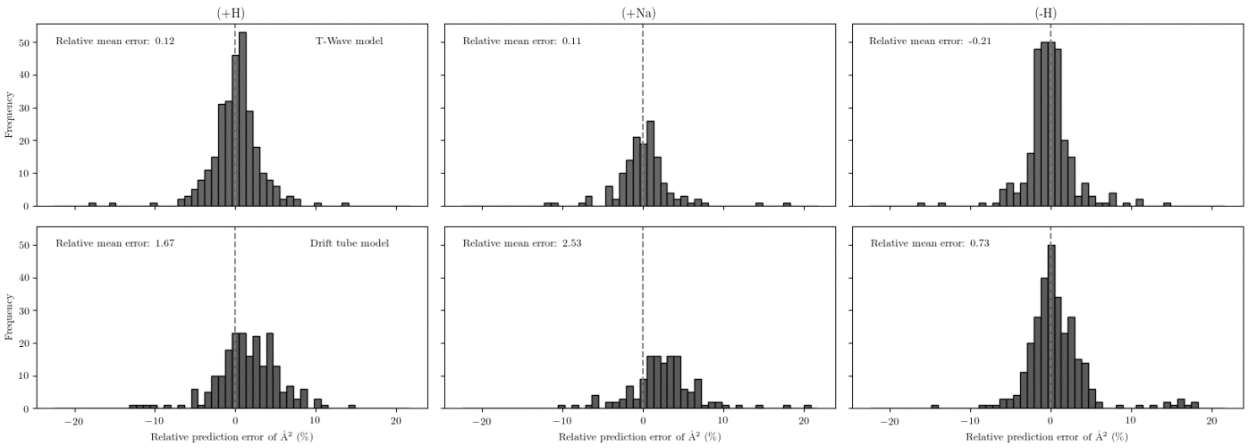

Figure 3. TWCCSN2 machine-learning prediction (blue) and MetCCS [20] DTCCSN2 based prediction (green) values vs. observed TWCCSN2 data.

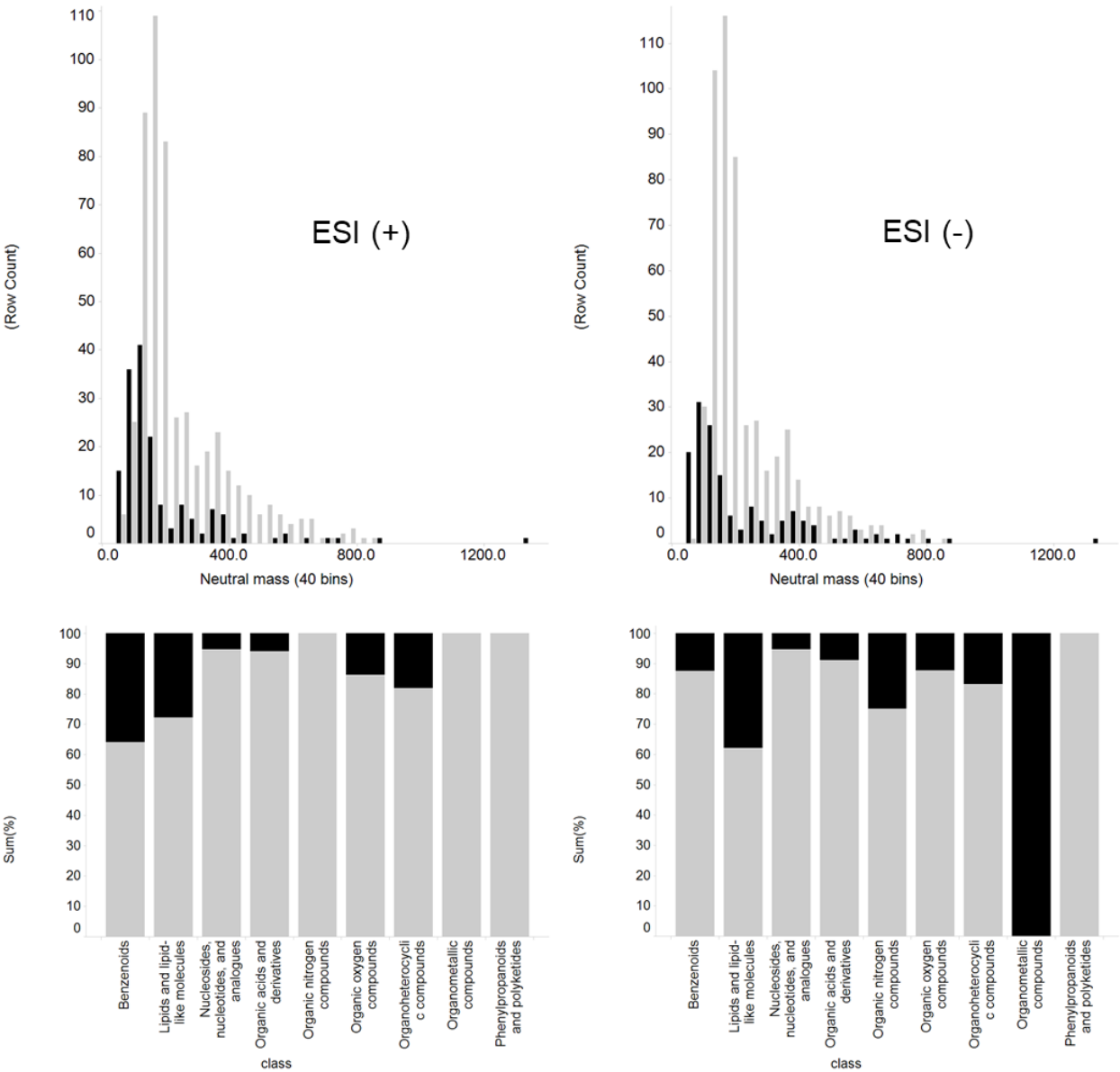

Figure 4. Detection frequency as a function of m/z (top) and relative (%) chemical class annotation of the library compounds (m/z 150 - 800) (bottom). Blue = detected; red = not detected.

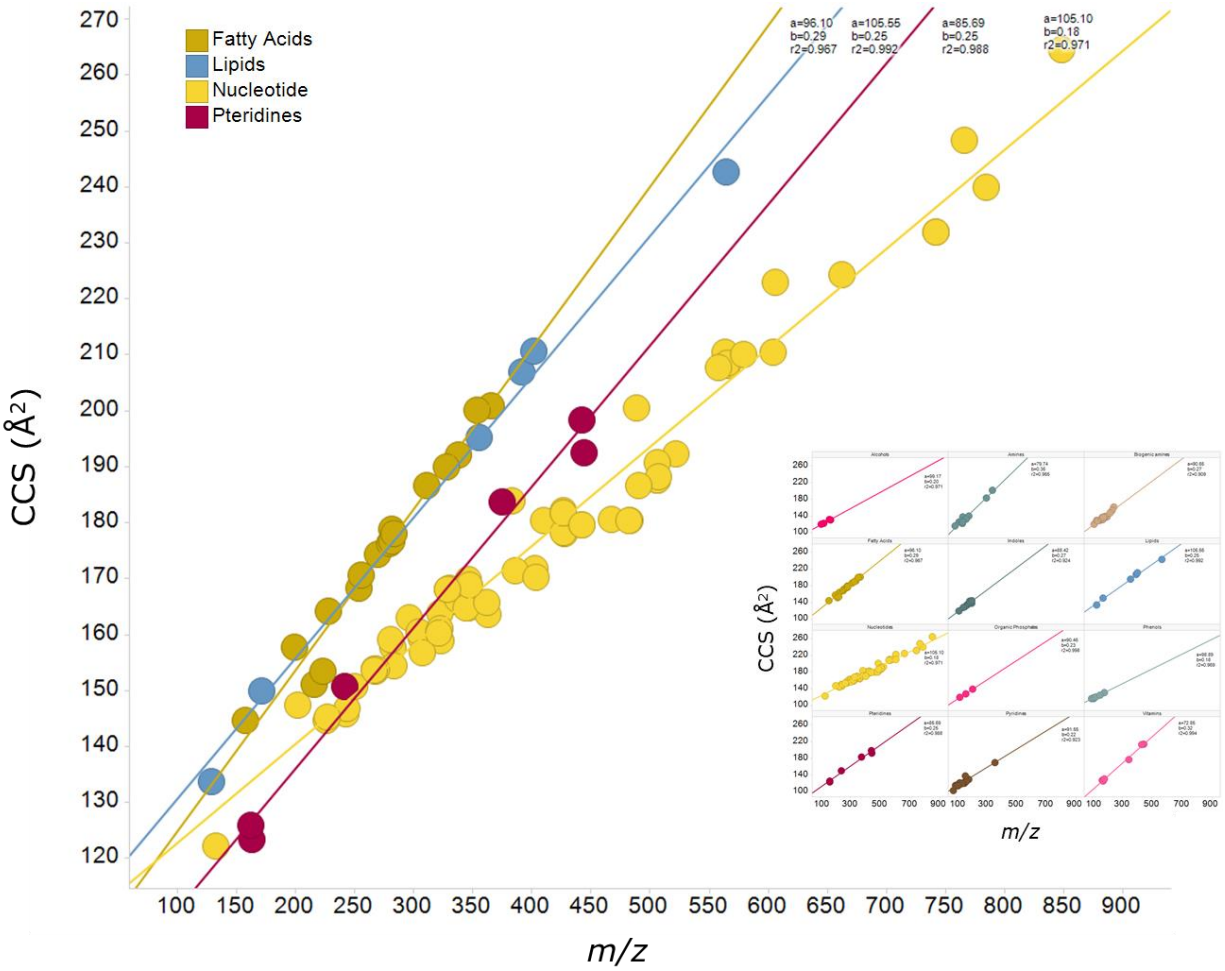

Figure 5. Class centric m/z vs. CCS relationship based on MeSH (Medical Subject Headings) classification.

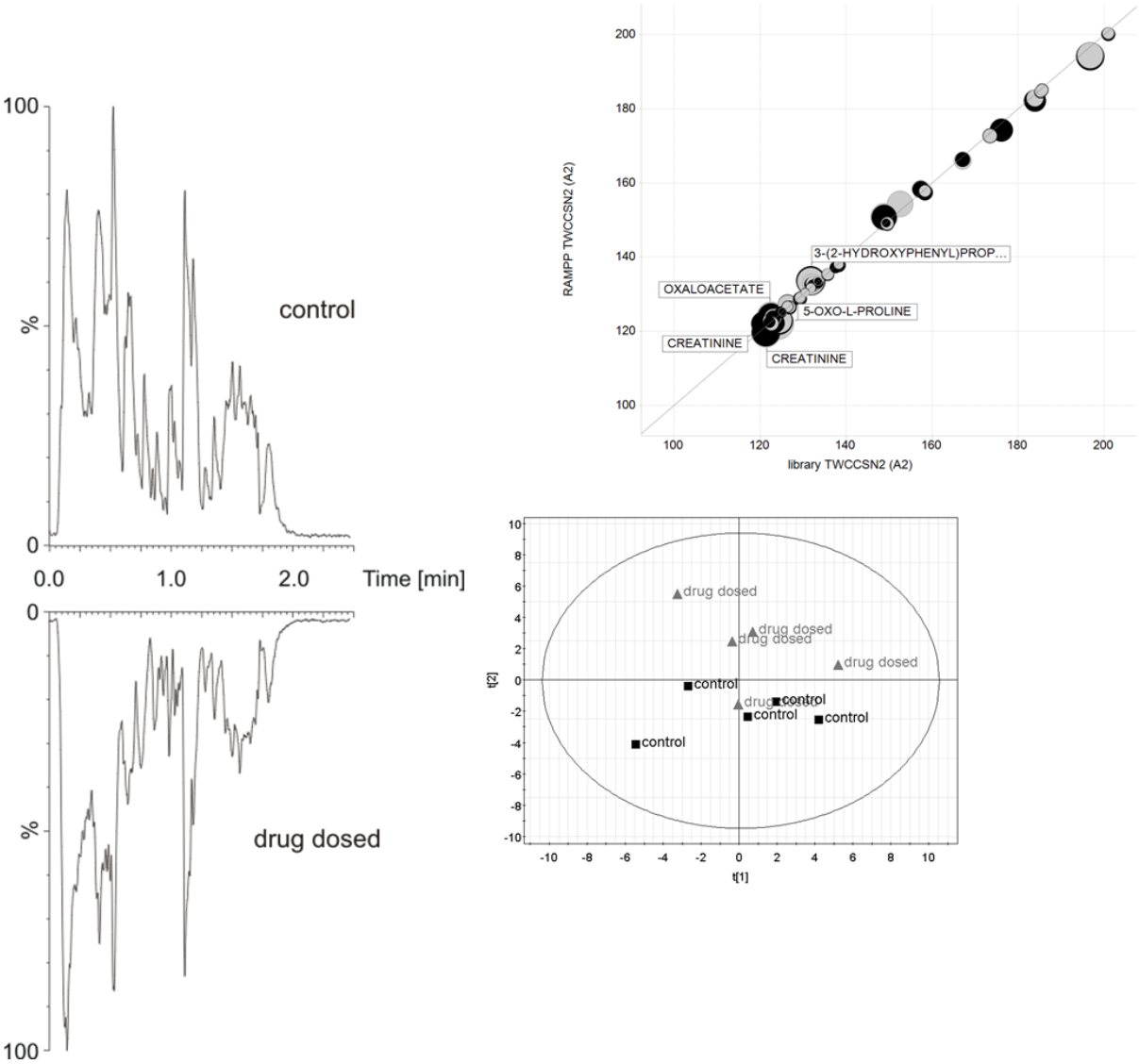

Figure 6. RAMMP LC-IM-MS BPI chromatograms (left), LC-IM-MS-derived vs. DI-IM-MS TWCCSN2 values (top right; the biological replication rate is represented by size and the sample by colour, blue = control, red = drug dosed) and unsupervised PCA on the abundances of the detected metabolites in control and MTX dosed rat urine. The numerical TWCCSN2 values, together with the retention times are listed in Table 3 and supplementary Table S6, respectively.