# Waters™

# Investigation and Performance Evaluation of a Research Prototype Tool for CCS Prediction

Yu Yanling, Hans Vissers, Kate Yu

Waters Corporation

This is an Application Brief and does not contain a detailed Experimental section.

# Abstract

This application brief demonstrates the CCS values of hundreds of pesticide compounds which were predicted using CCSondemand.

## Benefits

- SYNAPT HDMS, an ion mobility-high resolution tandem mass spectrometry system pioneers a new era of qualitative analysis with four-dimensional data

- With its ease-of-use, flexibility, and speed, CCS prediction has been shown to be a highly useful tool to provide reference and guidance for qualitative analysis

- Based on data independent analysis MS[E] technology, which ensures comprehensive information acquisition without losing even minor or narrow peaks, HDMS[E] enables greater mobility separation for improved spectral quality and significantly increased analytical capacity by performing comprehensive high-definition tandem MS data acquisition

# Introduction

Ion mobility-mass spectrometry (IM-MS) is an innovative 2D detection technique that integrates ion mobility separation (IMS) and mass spectrometry (MS). This technology has been maturing over decades of development and is quickly becoming one of the forefront fields of modern analytical chemistry. Collision cross section (CCS) is an important parameter for discriminating compounds in IM-MS. CCS values remain constant, regardless of instrumentation and experimental conditions, exhibiting unique advantage in qualitative analysis. However, in the absence of CCS reference values for unknowns, the question arises whether a prediction of the CCS value for an unknown can add value to an analysis.

Several CCS prediction tools have emerged in the public domain. Most of them are machine learning based, which train a model using a large number of experimental CCS values, combined with the structural properties of the compounds. By establishing the relationships between the structure and CCS values of compounds, the CCS values can be predicted for a specified structure.

In this application brief, the CCS values of hundreds of pesticide compounds were predicted using CCSondemand. To evaluate the accuracy of CCSondemand, the predicted values were compared with the mean CCS values observed from multiple experimental analyses; in addition, a public CCS prediction tool was used to predict the CCS values of the same compound data set. Finally, evaluation of the two prediction tools was undertaken in terms of prediction accuracy and overall performance. The models provided useful outcomes comparable with those obtained from the publicly available model.

As one of the key pioneers in IM-MS, Waters has demonstrated that it continues to be a technical leader in this cutting-edge analysis realm with its strong, enduring, and comprehensive innovation capabilities in science and technology.

# Experimental

## Test Samples and Overall Experimental Design

A standard mixture containing 20 pesticides (P/N: 186006348), diluted to a desired concentration with 50% acetonitrile, was used as a quality control sample to monitor and adjust retention time prior to data acquisition of a real sample. The real sample was a mixed solution containing hundreds of pesticides, hereafter referred to as the "pooled standard sample". The screening workflow in UNIFI Scientific Information System was used to analyze the acquired data. The screening workflow was based on a self-built high resolution tandem MS spectra library consisting of over 500 pesticide compounds for positive ion mode and over 100 pesticide compounds for negative ion mode. The compound attributes used in the qualitative screen included exact mass, retention time, and structural formulas of precursor/fragment ions, as well as the theoretical fragment ions (generated automatically by UNIFI based on the compound structure).

For the compounds consistently detected in all runs, the mean value and relative standard deviation (RSD) of CCS obtained in multiple injections were calculated. The mean CCS values of compounds having an RSD less than 1% were subsequently used as reference criteria for evaluating the accuracy of the CCS prediction tools.

## Chromatographic Separation

| | |
|---|---|
| Separation system: | ACQUITY UPLC I-Class |
| Columns: | XSelect HSS T3, 2.5 µm, 2.1 × 100 mm, (P/N: 186006151) |
| | CORTECS T3, 2.7 µm, 2.1 × 100 mm, (P/N: 186008484) |
| Temperature: | 45 °C |
| Mobile phase: | Aqueous phase (A) 5 mM $NH_4FA$ in Water + 0.1% FA organic phase (B) 50% methanol + 50% acetonitrile |

## Gradient:

| Time (min) | Flow rate (mL/min) | %A | %B | Curve |
|------------|--------------------|----|----|-------|
| 0.0 | 0.35 | 98 | 2 | – |
| 0.5 | 0.35 | 98 | 2 | 6 |
| 4.0 | 0.35 | 55 | 45 | 6 |
| 13.0 | 0.35 | 15 | 85 | 6 |
| 15.5 | 0.35 | 2 | 98 | 1 |
| 18.0 | 0.35 | 98 | 2 | 1 |

## MS Conditions

MS systems:

SYNAPT G2-S$i$ HDMS

SYNAPT XS HDMS

System resolution:

Sensitivity mode

Resolution mode

System calibration:

Mass axis calibration: Sodium formate solution

CCS calibration: Major mix (P/N: 186008113)

Real-time calibration (CCSLockmass): Leucine enkephalin solution

Data acquisition:

HDMS$^E$

Mass range: 50–1000 Da

HDMS$^E$ Collision energy: Low energy off; High energy 10–40 eV

# Results and Discussion

## Selection of CCS Prediction Tool

· Public CCS prediction tool

A number of research teams have focused their R&D efforts on CCS prediction and as a result multiple prediction tools are available online. The required input for compounds requiring CCS prediction varies in each of the tools. Some tools require many types of information to afford predictions for each compound; others are less readily accessible online and require some computing knowledge, or are not openly available. For this study, a public CCS prediction tool (anonymous, hereafter referred to as the "public CCS prediction tool") was used. This prediction tool was developed based on a machine learning approach and trained on a large number of experimental CCS values obtained by drift tube ion mobility and traveling wave ion mobility techniques. It predicts the CCS values of multiple ion adducts, such as $[M+H]^+$, $[M+Na]^+$, $[M+NH_4]^+$, and $[M-H]^-$, from the structure of the compound.

· CCSondemand

Waters Corporation has developed a research prototype application for CCS prediction, called CCSondemand, which also uses a machine learning based approach and it's model trained using with experimental CCS values from different ion adducts generated from a great number of compounds obtained on multiple instruments (Vion and SYNAPT). To use this tool, you need only to provide the chemical structure of a compound of interest. The tool is compatible with multiple chemical structure formats, such as mol, sdf, inchi, and SMILE, and is able to predict the CCS values of five ion adducts, namely, $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M-H]^-$, and $[M+HCOO]^-$. The user interface of CCSondemand is shown in Figure 1.

*Figure 1. User interface of CCSondemand.*

## HDMS$^E$ Acquisition of the Pooled Standard Sample

Data acquisition of the pooled standard sample (containing hundreds of small molecule pesticide compounds) was performed multiple times. As can be seen from the injection information shown in Table 1 (positive ion mode) and Table 2 (negative ion mode), despite the number of injections varying somewhat due to study objective limitations, a large number of experimental variables, *i.e.* MS instrument, MS resolution, chromatographic column, and data of acquisition were purposely varied to demonstrate that CCS is a robust and reproducible analytical parameter.

| n | MS equipment | MS resolution | Column | Date acquisition | Injection No. |
|---|---|---|---|---|---|
| 1 | SYNAPT XS | Sensitivity mode | XSelect HSS T3 | Apr. 28, 2020 | 1 |
| 2 | SYNAPT XS | Sensitivity mode | XSelect HSS T3 | Apr. 28, 2020 | 2 |
| 3 | SYNAPT XS | Resolution mode | XSelect HSS T3 | Apr. 28, 2020 | 1 |
| 4 | SYNAPT XS | Sensitivity mode | XSelect HSS T3 | May 20, 2020 | 1 |
| 5 | SYNAPT XS | Sensitivity mode | XSelect HSS T3 | May 20, 2020 | 2 |
| 6 | SYNAPT XS | Resolution mode | XSelect HSS T3 | May 20, 2020 | 1 |
| 7 | SYNAPT XS | Resolution mode | XSelect HSS T3 | May 20, 2020 | 2 |
| 8 | SYNAPT G2-S*i* | Sensitivity mode | CORTECS T3 | May 22, 2020 | 1 |
| 9 | SYNAPT G2-S*i* | Sensitivity mode | CORTECS T3 | May 22, 2020 | 2 |
| 10 | SYNAPT G2-S*i* | Resolution mode | CORTECS T3 | May 22, 2020 | 1 |
| 11 | SYNAPT G2-S*i* | Resolution mode | CORTECS T3 | May 22, 2020 | 2 |
| 12 | SYNAPT G2-S*i* | Sensitivity mode | XSelect HSS T3 | Jun. 17, 2020 | 1 |
| 13 | SYNAPT G2-S*i* | Sensitivity mode | CORTECS T3 | Jun. 18, 2020 | 1 |
| 14 | SYNAPT G2-S*i* | Resolution mode | CORTECS T3 | Jun. 18, 2020 | 1 |
| 15 | SYNAPT G2-S*i* | Sensitivity mode | XSelect HSS T3 | Jun. 19, 2020 | 1 |
| 16 | SYNAPT G2-S*i* | Resolution mode | XSelect HSS T3 | Jun. 19, 2020 | 1 |

*Table 1. Injection information for data acquisition of the pooled standard sample in positive ion mode, n = 16.*

| n | MS equipment | MS resolution | Column | Date acquisition | Injection No. |
|---|---|---|---|---|---|
| 1 | SYNAPT XS | Sensitivity mode | XSelect HSS T3 | May 20, 2020 | 1 |
| 2 | SYNAPT XS | Sensitivity mode | XSelect HSS T3 | May 20, 2020 | 2 |
| 3 | SYNAPT G2-Si | Sensitivity mode | CORTECS T3 | May 22, 2020 | 1 |
| 4 | SYNAPT G2-Si | Sensitivity mode | CORTECS T3 | May 22, 2020 | 2 |
| 5 | SYNAPT G2-Si | Sensitivity mode | CORTECS T3 | Jun. 18, 2020 | 1 |
| 6 | SYNAPT G2-Si | Resolution mode | CORTECS T3 | Jun. 18, 2020 | 1 |
| 7 | SYNAPT G2-Si | Sensitivity mode | XSelect HSS T3 | Jun. 19, 2020 | 1 |
| 8 | SYNAPT G2-Si | Resolution mode | XSelect HSS T3 | Jun. 19, 2020 | 1 |

Table 2. Injection information for data acquisition of the pooled standard sample in negative ion mode, n = 8.

## Data Analysis and Summary

More than 500 compounds from the self-built library were used to screen the data obtained in positive ion mode. 339 compounds were detected in 16 screening injections. For negative ion mode, 92 compounds were detected in 8 screening injections, using over 100 compounds from the library. According to the analysis results, the measured CCS values for some injections of 31 compounds were not correctly assigned due to signal overloading. These CCS values were therefore not considered for analysis, with the number of rejected values varied between 1 and 6. The RSD of the remaining measured CCS values for each compound was typically less than 1% with the distribution of the RSD values shown in Figure 2.
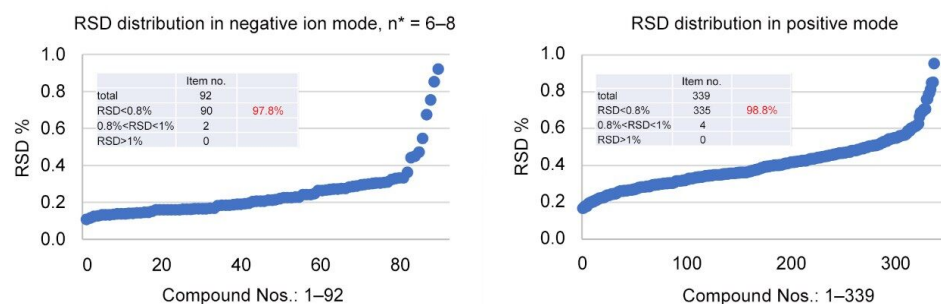


Figure 2. RSD distributions of measured CCS values observed in multiple injections of the pooled standard sample.

Standard deviations (SDs) are known to be very sensitive to extremely large or small intra-group errors and can be used as a measure for the uncertainty and degree of dispersion of observed values within a group, in light of the diversity of the experimental design. If the RSD of the observed CCS values observed in multiple runs is small (<1%), then the mean CCS obtained from multiple runs is expected to be close to the true value. Based on this assumption, the mean measured CCS value of all compound having an RSD <1% were retained; hereafter referred to as the "experimental reference CCS value" (positive ion mode: 339 compounds; negative ion mode: 92 compounds)

served as the reference criterion for subsequent evaluation and analysis of CCS prediction accuracy.

## Performance Evaluation and Analysis of the Two CCS Prediction Tools

· Evaluation of CCS prediction accuracy in positive ion mode

CCSondemand and the public CCS prediction tool were used to predict CCS values of $[M+H]^+$ ions for the 339 compounds detected in positive ion mode. Each predicted CCS value was compared with the experimental reference value, and the relative deviation calculated. Table 3 lists the standard deviations (SDs) of the relative deviations in CCS that were predicted by the two prediction tools, as well as the corresponding probability levels. Their normal distributions are shown in Figure 3. The probability level of the values included within the normal distribution analysis falling within the mean ±1 x SD equals 68%, i.e., 68% of values in the array are within mean ±1 x SD. Similarly, 95% of the values are within the mean ±2 x SD, and 99.7% of the values are within the mean ±3 x SD.

|  | Public CCS prediction tool | CCSondemand | Probability level (%) |
|---|---|---|---|
| 1 x SD | 3.3% | 1.4% | 68.0 |
| 2 x SD | 6.6% | 2.8% | 95.0 |
| 3 x SD | 9.9% | 4.3% | 99.7 |

*Table 3. SDs and probability levels of deviations in CCS predictions in positive ion mode, n = 339.*
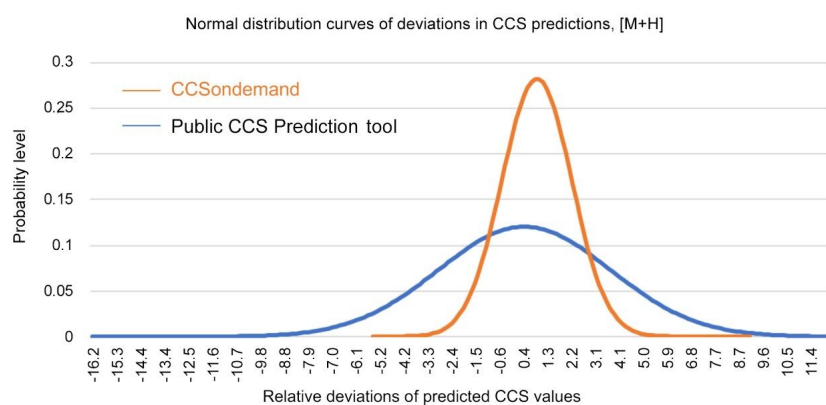


*Figure 3. Normal distribution curves of relative deviations in CCS predictions (positive ion mode).*

In summary, it can be concluded, from the results shown in Table 3 and the normal distributions in Figure 3 that: 1) in positive-ion ionization mode, most of the CCS values predicted by either of the

prediction tools were accurate enough to serve as reference and guidance for the qualitative analysis to some extent; 2) in the case of the public CCS prediction tool, predicted CCS values of 99.7% of the compounds had relative deviations within ±10% compared to the experimental reference values, while in the case of CCSondemand, 99.7% of the compounds had relative deviations within ±5%. In other words, CCSondemand showed relatively small error in CCS prediction vs. experimental values, as demonstrated by the relative deviations distributions in shown in Figure 4.
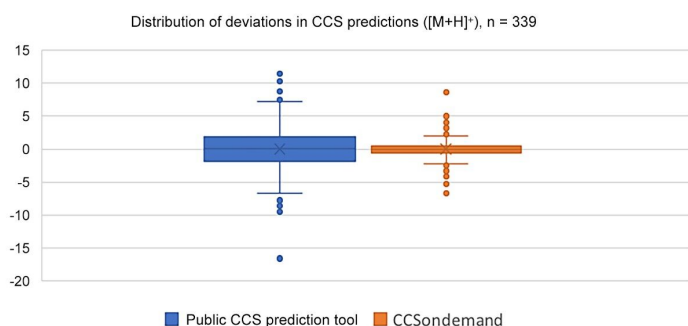


Figure 4. CCS prediction error distributions in positive ion mode.

In addition, linear regression analysis was performed between the predicted CCS values of the 339 compounds and the experimental reference values. The regression curves and corresponding linear correlation coefficients are shown in Figure 5. The distribution of the data points and the linear regression coefficients further indicated good fit between the predicted CCS values and the experimental reference values.
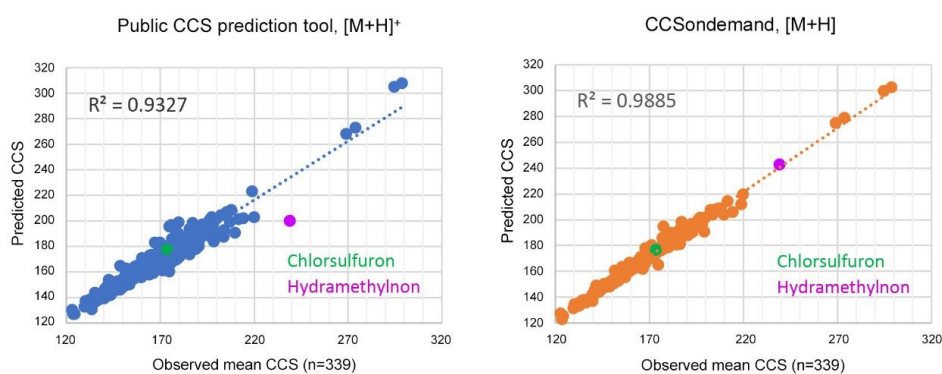


Figure 5. Linear regression curves of predicted CCS values versus experimental reference values (positive ion mode).

Two compounds, chlorsulfuron and hydramethylnon, which were detected in both positive ion mode

and negative ion mode, are highlighted using green and purple dots respectively. These examples are discussed in more detail in the section 'Examples of Compound Data Analysis'.

· Evaluation of CCS prediction accuracy in negative ion mode

Likewise, an overall analysis of the predicted CCS values, experimental reference values, and relative deviations in predictions was performed for the 92 compounds detected in negative ion mode. Figure 6 shows linear regression curves of the predicted CCS values vs. experimental reference values in negative ion mode, illustrating similar performance metrics for both models compared to the positive ionization mode results.
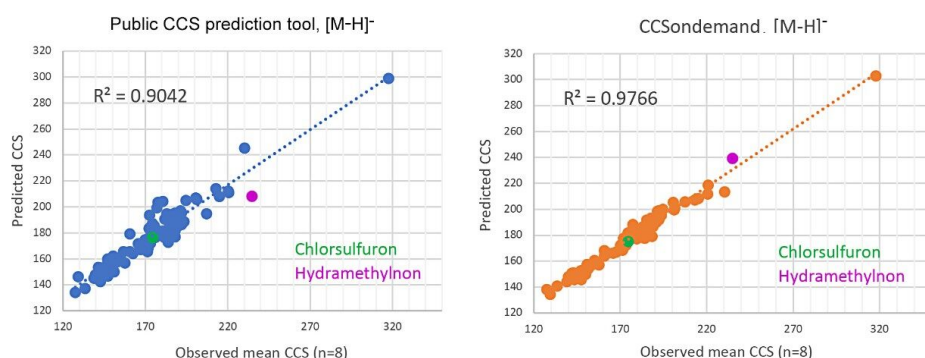


*Figure 6. Linear regression curves of predicted CCS values versus experimental reference values (negative ion mode).*

## Examples of Compound Data Analysis

· Chlorsulfuron

Confirmation of compound identification was achieved by reviewing raw data and the qualitative UNIFI analysis results of chlorsulfuron (green dot Figures 5 and 6) in both positive and negative ion modes from the same analysis batch [exemplified by run No. 16 in Table 1 and run No. 8 in Table 2]. All data shown were acquired in resolution mode on a SYNAPT G2-S*i* instrument using an XSelect HSS T3 Column. Figure 7 shows the chromatograms (left) and arrival time distribution (ATD) (right) of chlorsulfuron in positive and negative ion modes. Since the same column and chromatographic conditions were used, the chromatographic behavior of the compound in positive ion mode and negative ion mode are expected to be similar. This assumption was verified by the identical retention time and peak shape in both ionization modes, further confirming the correct detection of this compound. The CCS values differ slightly due to the different ion adducts.
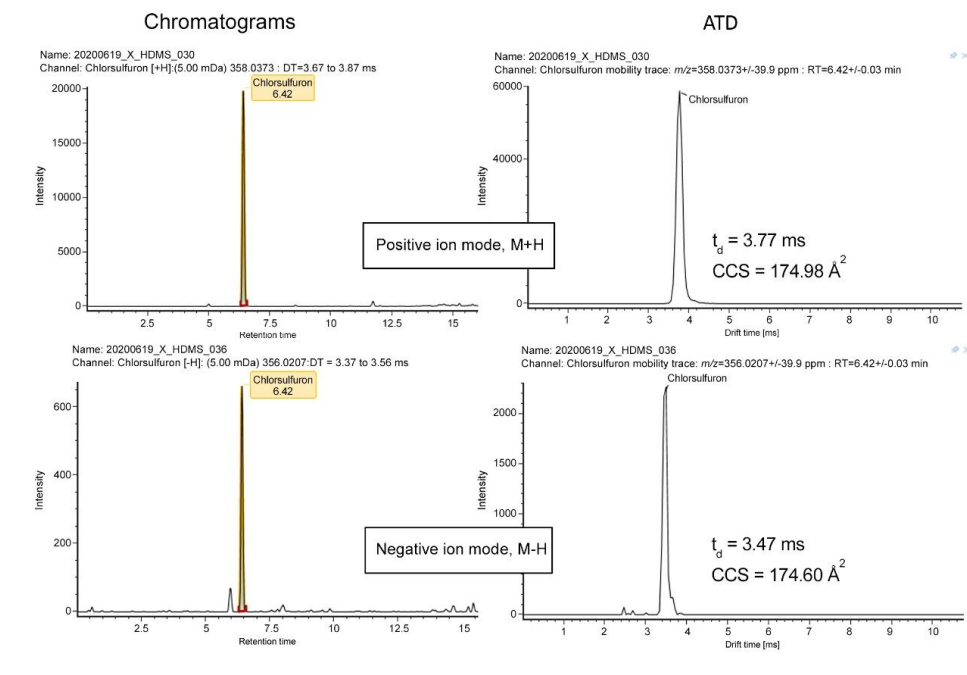
Figure 7. Chromatograms (left) and ATDs (right) of chlorsulfuron.

Table 4 lists the CCS values of chlorsulfuron from multiple injections observed in positive ion mode on the left, and those obtained in negative ion mode on the right. The values in blue font are mean measured CCS values from multiple analyses, *i.e.*, the aforementioned experimental reference values.

| Positive ion mode | | | Negative ion mode | | |
|---|---|---|---|---|---|
| Injection No. | Injection ID | Measured CCS value | Injection No. | Injection ID | Measured CCS value |
| 1 | 0617_X_S | 174.80 | 1 | 0520_X_S1 | 174.64 |
| 2 | 0619_X_S | 174.93 | 2 | 0520_X_S2 | 174.76 |
| 3 | 0619_X_R | 174.98 | 3 | 0619_X_S | 174.60 |
| 4 | 0522_C27_S1 | 173.43 | 4 | 0619_X_R | 174.60 |
| 5 | 0522_C27_S2 | 173.47 | 5 | 0522_C27_S1 | 173.81 |
| 6 | 0522_C27_R1 | 173.75 | 6 | 0522_C27_S2 | 174.07 |
| 7 | 0522_C27_R2 | 173.76 | 7 | 0618_C27_S | 174.83 |
| 8 | 0618_C27_S | 174.76 | 8 | 0618_C27_R | 174.45 |
| 9 | 0618_C27_R | 174.79 | Mean measured CCS, $Å_2$ | | 174.47 |
| 10 | 0428_X_S1 | 174.00 | | | |
| 11 | 0428_X_S2 | 173.86 | | | |
| 12 | 0428_X_R1 | 175.69 | | | |
| 13 | 0520_X_S1 | 174.33 | | | |
| 14 | 0520_X_S2 | 174.38 | | | |
| 15 | 0520_X_R1 | 174.67 | | | |
| 16 | 0520_X_R2 | 174.70 | | | |
| Mean measured CCS, $Å_2$ | | 174.39 | | | |

Table 4. List of measured CCS values of chlorsulfuron in multiple injections.

· Hydramethylnon

Chromatograms, mobility distributions, and measured CCS values of hydramethylnon (purple dot in Figures 5 and 6) are shown in Figure 8 and Table 5, respectively. The data is derived from the same

dataset as the chlorsulfuron analysis. An overlapping peak appeared in the ATD (positive ion mode) as illustrated in Figure 8 and was confirmed to be caused by isomerism.
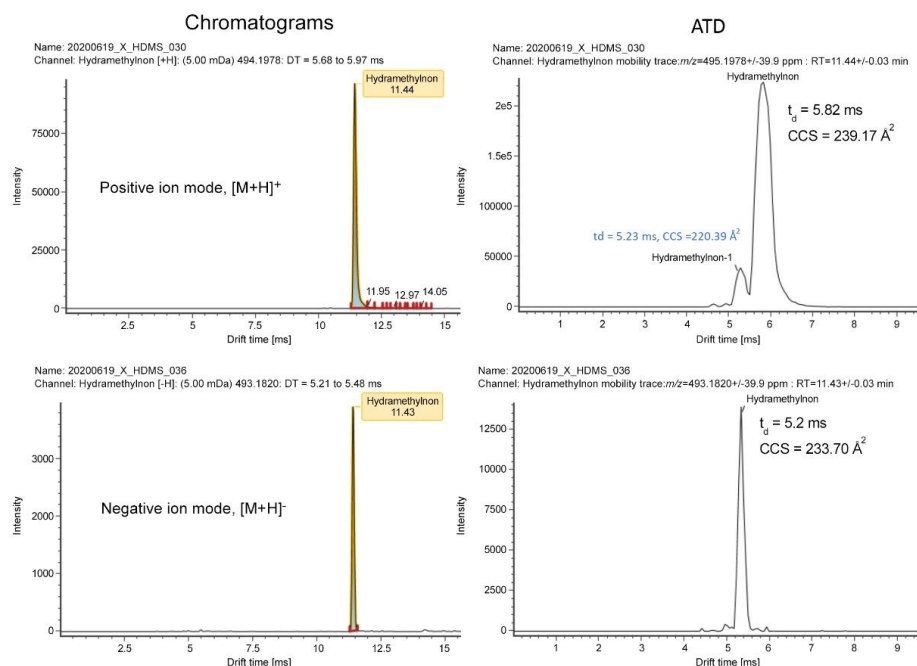


Figure 8. Chromatograms (left) and ATD (right) of hydramethylnon.

| Positive ion mode | | | Negative ion mode | | |
|---|---|---|---|---|---|
| Injection No. | Injection ID | Measured CCS value | Injection No. | Injection ID | Measured CCS value |
| 1 | 0617_X_S | 240.80 | 1 | 0520_X_S1 | 235.00 |
| 2 | 0619_X_S | 241.13 | 2 | 0520_X_S2 | 235.07 |
| 3 | 0619_X_R | 239.17 | 3 | 0619_X_S | 235.79 |
| 4 | 0522_C27_S1 | 239.10 | 4 | 0619_X_R | 233.70 |
| 5 | 0522_C27_S2 | 239.42 | 5 | 0522_C27_S1 | 234.73 |
| 6 | 0522_C27_R1 | 238.43 | 6 | 0522_C27_S2 | 234.84 |
| 7 | 0522_C27_R2 | 238.46 | 7 | 0618_C27_S | 235.95 |
| 8 | 0618_C27_S | 241.76 | 8 | 0618_C27_R | 233.66 |
| 9 | 0618_C27_R | 239.04 | Mean measured CCS, $\text{Å}_2$ | | 234.84 |
| 10 | 0428_X_S1 | 237.72 | | | |
| 11 | 0428_X_S2 | 237.67 | | | |
| 12 | 0428_X_R1 | 237.10 | | | |
| 13 | 0520_X_S1 | 239.10 | | | |
| 14 | 0520_X_S2 | 238.86 | | | |
| 15 | 0520_X_R1 | 238.58 | | | |
| 16 | 0520_X_R2 | 238.54 | | | |
| Mean measured CCS, $\text{Å}_2$ | | 239.06 | | | |

Table 5. List of measured CCS values of hydramethylnon in multiple injections.

The ATD peak widths of this compound in positive-ion and negative-ion ionization modes differ greatly with the ATD peak in positive ion mode evidently broadening. The mobility heatmap (data not shown) confirmed that the broadening was mainly caused by mobility peak tailing due to abundant signals.

· Predictions of CCS values and prediction deviations

CCSondemand and the public CCS prediction tool were used to predict CCS values of [M+H] and [M-H] ions of both compounds under investigation, as shown in Table 6.

| Positive ion mode, M+H | | | | | |
|---|---|---|---|---|---|
| Compound name | CCS Experimental reference value, $Å_2$ | CCSondemand | | Public prediction tool | |
| | | CCS Predicted value | CCS Prediction deviation | CCS Predicted value | CCS Prediction deviation |
| Chlorsulfuron | 174.4 | 176.0 | 0.9 | 175.5 | 0.6 |
| Hydramethylnon | 239.1 | 242.8 | 1.6 | 200.3 | -16.2 |
| Negative ion mode, M-H | | | | | |
| Compound name | CCS Experimental reference value, $Å_2$ | CCSondemand | | Public prediction tool | |
| | | CCS Predicted value | CCS Prediction deviation | CCS Predicted value | CCS Prediction deviation |
| Chlorsulfuron | 174.5 | 175.2 | 0.4 | 176.3 | 1.1 |
| Hydramethylnon | 234.8 | 239.1 | 1.8 | 208.2 | -11.3 |

*Table 6. CCS values predicted by the two prediction tools and their deviations from respective experimental reference values.*

From the data in Table 6 it can be concluded:

· Although both of the CCS prediction tools studied were based on the machine learning principle, the predicted CCS values obtained differed somewhat, which is most likely a direct result of the use of different training data sets.

· CCS values were accurately predicted by CCSondemand for both of the compounds, with prediction deviations being within 2%.

· The public prediction tool exhibited good prediction accuracy for chlorsulfuron but illustrated larger CCS prediction deviations for the [M+H]$^+$ and [M-H]$^-$ ions of hydramethylnon.

In conclusion, for both CCS prediction tools, most predicted CCS values were sufficiently accurate to serve as reference and guidance for the qualitative analysis. However, the accuracy of CCS predictions for some compounds needs further development, and the overall performance of the prediction tools to be modified and perfected, which is expected to occur over time.

## Conclusion

In this study, we performed an overall evaluation of a research prototype CCS prediction tool CCSondemand, in terms of the accuracy of predicted CCS values under variable experimental conditions. The evaluation was based on hundreds of pesticide compounds serving as test sample, and their mean measured CCS values obtained via multiple injections used as reference values, with CCS predictions typically taking <1 s per compound. The results were compared with those obtained with a publicly available CCS prediction tool. The results showed that the CCS prediction accuracy of CCSondemand was good and has potential to improve the characterization of unknowns.

As scientists continuously increase their understanding of the ion mobility principle, hardware innovations are made, and data acquisition methods are constantly updated. As a consequence, analytical experiments with larger data sets are produced. Alongside this CCS prediction techniques will most likely mature, and their prediction accuracy and reliability expected to gradually improve. Eventually, such techniques could become powerful qualitative analytical tools for analytical chemistry researchers.

## Featured Products

ACQUITY UPLC I-Class PLUS System

SYNAPT G2-Si High Definition Mass Spectrometry

SYNAPT XS High Resolution Mass Spectrometer

UNIFI Scientific Information System

720007021, November 2020