Determining Molecular Physicochemical Properties Using Differential Mobility Spectrometry

By

Zachary Alan Bowman

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Chemistry (Nanotechnology)

Waterloo, Ontario, Canada, 2019

Author’s Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 3 examines the separation of derivatized amphetamine and methamphetamine isomers. The experimental work in this chapter was carried out by collaborators at SCIEX, while the computational work was carried out by me, along with an analysis of the combined data. A manuscript of this work is currently under review:

Reference: Kalfe, Amol; Bowman, Zack; Le Blanc, J.C. Yves; Liu, Chang; Hopkins, W.; Campbell, J Separating Chiral Isomers of Amphetamine and Methamphetamine using Chemical Derivatization and Differential Mobility Spectrometry. *Anal. Chem. (In review).*

Chapter 4 uses experimental data provided by David Ruskic and Prof. Gérard Hopfgartner at the University of Geneva that was published this year in Analytical chemistry. Further computational and machine learning work was carried out alongside analysis of results to complete this chapter.

Reference: Ruskic, D.; Hopfgartner, G. Modifier Selectivity Effect on Differential Ion Mobility Resolution of Isomeric Drugs and Multidimensional Liquid Chromatography Ion Mobility Analysis. *Anal. Chem.* **2019**. *91*(18). 11670-11677. https://doi.org/10.1021/acs.analchem.9b02212

Chapters 4 and 5 make use of computational and experimental data gathered in a manuscript published last year in Nature Communications. Further experimental, computational, machine learning work and analysis was carried out to complete chapter 5.

Reference: Walker, S. W. C.; Anwar, A.; Psutka, J. M.; Crouse, J.; Liu, C.; Le Blanc, J. C. Y.; Montgomery, J.; Goetz, G. H.; Janiszewski, J. S.; Campbell, J. L.; et al. Determining Molecular Properties with Differential Mobility Spectrometry and Machine Learning. *Nat. Commun.* **2018**, *9* (1). 1-7. https://doi.org/10.1038/s41467-018-07616-w.

Abstract

This thesis aims to explore the usage of differential mobility spectrometry (DMS) and quantum chemical calculations to separate and identify drug compounds, as well as the use of machine learning (ML) to predict physicochemical properties such as the collision cross section (CCS) and single solvent binding energies (BEs) using experimental DMS data. Chapter 3 shows the ability of DMS to separate derivatized amphetamines and methamphetamine isomers and uses calculated BEs to identify the separated isomers by their DMS behavior. Chapter 4 demonstrates the separation of (+)-ephedrine and (+)-pseudoephedrine as well as three groups of sulfonamide isomers using a variety of gas modifiers within the DMS. CCS are used to describe behavior in pure N_2 gas, while BEs are used to predict the ordering of the separation voltage at the compensation voltage minima (SV at CV_{min}) values within different isomeric groups. By making use of data gathered by Walker *et al.*[1] ML models were generated for prediction of CCS as well as H_2O and MeOH BEs and tested using the isomers studied in this chapter. In order to demonstrate the need for the model to train on similar compounds to those being tested, several of the test compounds were moved into the training set of the ML model and the change in predictivity observed. In the final chapter of this work, chapter 5, the CCS ML database was expanded using a sizable number of compounds with varied structure and functionalities in order to increase the predictivity of the ML model for new compounds. The effectiveness of incorporating additional compounds was evaluated by the creation of learning curves for the ML training and test sets. These projects ultimately show the capability of DMS in making separations of isomeric compounds, as well as its potential for use in the prediction of CCSs and BEs through ML modelling.

Acknowledgements

I would like to thanks my supervisor, Dr. Scott Hopkins, for giving me the opportunity to grow as a graduate student within his lab group; it has provided invaluable experience in both experimental and computational techniques that will serve me well in future work. His insight and expertise have proven to be invaluable in guiding me along the right track as I've worked on various projects and particularly this thesis.

I would also like to thank the members of the Hopkins group for their assistance over my time as a graduate student, and for making me feel right at home in Waterloo. In particular, I would to extend my thanks to Dr. Jeff Crouse for his assistance with machine learning work, Ce Zhou and Josh Featherstone for their programming knowledge and the various scripts that they have developed that make my life easier on a daily basis, Dr. Pat Carr and Christian Ieritano for serving as a sounding board for many of my more foolish questions, Nour Mashmoushi for her tireless efforts in making sure that the DMS instruments are behaving as intended, and Suzy Lim for always finding a way to make me laugh.

Finally, I would like to say my thanks to my family back home, and my friends both near and far to being there when I've needed them over the years. I wouldn't have gotten this far without them!

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

ACE	Acetone
AcOEt	Ethyl acetate
AMP	Amphetamine
ANN	Artificial neural networks
BE	Binding energy
BH	Basin hopping
CC	Coupled cluster
CCS	Collision cross section
CapE	Capillary electrophoresis
CE	Collision energy
CEC	Capillary electrochromatography
CID	Collision induced dissociation
CIP	Cahn-Ingold-Prelog
CV	Compensation voltage
DFT	Density Functional Theory
DMS	Differential Mobility Spectrometry
DP	Declustering Potential
EPI	Enhanced product ion
ESI	Electrospray ionization
EtOH	Ethanol
FAIMS	High-field asymmetric waveform ion mobility spectrometry
GC	Gas chromatography
HDX	Hydrogen deuterium exchange
HF	Hartree-Fock
HPLC	High performance liquid chromatography
IMS	ion mobility spectrometry

IPA	Isopropyl alcohol
LC	Liquid chromatography
LJ	Lennard-Jones
MAE	Mean absolute error
MeAMP	Methamphetamine
MeOH	Methanol
ML	Machine learning
MPE	Mean percent error
MPI	Message passing interface
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass to charge ratio
NDDO	Neglect of diatomic differential overlap
PES	Potential energy surface
PrOH	Propanol
QML	Quantum machine learning
QSAR	Quantitative structure activity relationships
QSPR	Quantitative structure property relationships
RF	Random forest
RMSE	Root mean square error
SFC	Supercritical fluid chromatography
SV	Separation voltage
SVM	Support vector machine
Tol	Toluene
TPC	*N*-trifluoroacetyl prolyl chloride

TWIMS	Traveling wave ion mobility spectrometry
UHPSFC	Ultra-high performance supercritical fluid chromatography
UPLC	Ultra-performance liquid chromatography
vdW	van der Waals

Chapter 1 : Introduction

In recent years it has become increasingly important to be able to distinguish between different isomeric compounds for the purposes of drug development and medicinal chemistry. One reason for this is the increased importance of enantiomerically pure drugs.[2] Some drugs, such as ibuprofen (Figure 1.1) have even been switched from a racemate or diastereomeric mixture to a pure composition (in this case S-ibuprofen) due to various preferable properties. By utilizing a single enantiomer rather than a racemic mixture, improvements can be made in therapeutic properties (i.e., increasing potency and selectivity, reducing side effects), the time before onset of effects can be decreased, and the chances of drug-drug interactions occurring can be lowered.[3] Another example of a compound which is now being produced as a single enantiomer is the local anesthetic bupivacaine, which had its composition changed from a racemic mixture to purely its *S*(−)-enantiomer because it was shown that the R(+) isomer was the cause of the racemic mixture's cardiotoxic effects.[4]

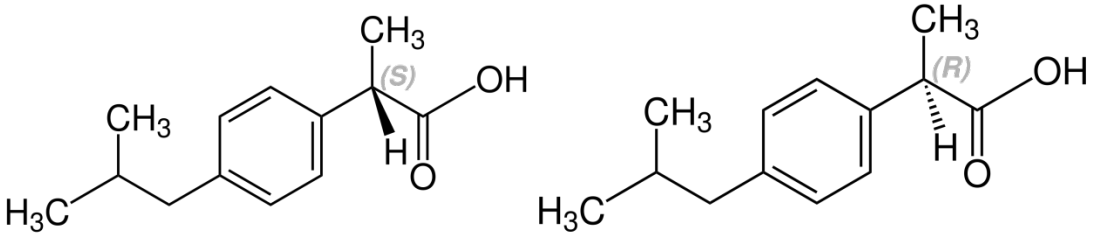


Figure 1.1: Structures of S-ibuprofen (right) and R-ibuprofen (left)

Structural isomers are also often observed in the drug development process; they are commonly observed in drug metabolite studies[5] and are used in the process of creating new drugs. By modifying the location of a functional group within a drug candidate the potency of beneficial effects can increased or the potency of detrimental effects can be decreased. One such example is the case of paracetamol (acetaminophen) and metacetamol (See Figure 1.2), which vary by the placement of a hydroxyl group at the para or meta positions along a phenyl ring. Metacetamol has been shown in some studies to be

significantly less hepatotoxic.[6] A similar effect can be observed between valproic acid, an anticonvulsant/bipolar disorder medication with teratogenic properties, and 2-isopropylpentanoic acid, which maintains the mood stabilizing effects while reducing the potency of the detrimental side effects.[7] Due to the variance in effects that these minor substitutions can produce, it is important to use analytic tools that can separate and quantify these different kinds of isomeric compounds.

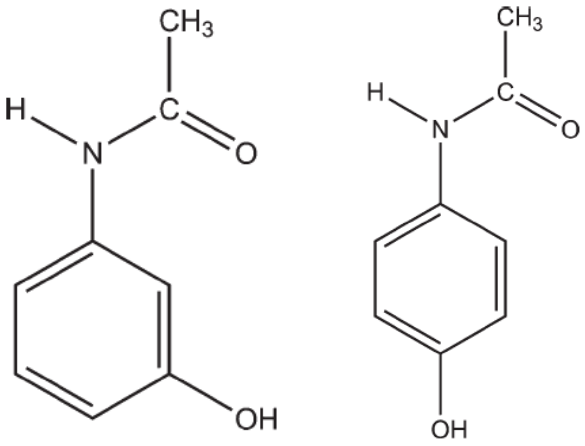


Figure 1.2: Structures of metacetamol (left) and paracetamol (right)[6]

Mass spectrometry (MS) is a commonly used technique for analytical applications due to its sensitivity and high throughput capacity.[2] However, it is typically unable to distinguish between different isomeric compounds due to their identical masses and fragmentation. Several techniques exist for separating chiral isomers using MS, including hydrogen deuterium exchange (HDX) or collision induced dissociation (CID). In the case of HDX, one of the chiral analytes is selectively deuterated so that it and its isomer can be distinguished by their now-differing masses.[8] Alternatively, a chirally selective molecule can be added to solution to form diastereomeric complexes. These complexes can then be broken down by CID and differentiated by differences in their fragmentation patterns.[9] Other methods for isomeric separation require that an orthogonal technique be added to the MS workflow. Some common techniques include Liquid Chromatography-Mass Spectrometry (LC-MS)[10,11], Gas Chromatography-Mass Spectrometry (GC-MS)[12,13], Capillary Electrophoresis-Mass Spectrometry (CapE-MS)[14,15], Capillary

Electrochromatography-Mass Spectrometry (CEC-MS)[16,17], Supercritical Fluid Chromatography-Mass Spectrometry (SFC-MS)[18,19] as well as Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS)[20,21].

Differential Mobility Spectrometry (DMS) is a relatively new technique that can be used for separating various isomeric species including structural isomers[22] as well as various tautomeric species of protonated nucleobases.[22–24] This thesis will further explore the DMS as a method for separating isomeric species by examining a variety of molecular systems. Furthermore, if the separation of isomers can be predicted computationally, it should also be possible to use the experimental DMS data in order to predict the properties which were used to make these predictions. In order to facilitate this, machine learning (ML) techniques will be used.

In chapter 2, the computational and experimental techniques used in this thesis are described. The computational techniques include basin hopping, quantum chemistry methods, MobCal-MPI, and ML methods. This is followed by the eexperimental techniques that were used (i.e., QTRAP 5500 setup). Electrospray ionization is used to introduce analytes into the instrument, which then are analyzed and potentially separated in the DMS portion of the instrument, before being distinguished by their charge to mass ratio by the tandem mass spectrometer. A section on the analysis of the DMS data is also included. Chapter 4 focuses on the separation of amphetamine and methamphetamine isomers via DMS after a derivatization reaction with (*S*)-*N*-trifluoroacetyl prolyl chloride. These results are further backed by computational work comparing the binding energies of solvent-ion clusters. Chapter 5 expands upon the work done in Chapter 4, attempting DMS separation for four groups of compounds including ephedrine and its diastereomer pseudoephedrine as well as three groups of sulfonamide structural isomers. These experimental results are accompanied by computational work for the binding energies of these ions with a variety of chemical modifiers that are used in DMS and are used to attempt the prediction of DMS behavior. Furthermore, machine learning is used to predict the collision cross section of ions in pure N_2 and binding energies are predicted for ions with H_2O and MeOH modifiers. Finally, Chapter 5 is focused on expanding the machine learning database with a number of compounds with different activities and the

DMS behaviors so that the machine learning model can be used to make predictions about new compounds.

Chapter 2 : Methods

2.1 Computational Methods

2.1.1 Basin Hopping

Basin hopping (BH) is a search algorithm that is employed to find low energy isomer structures for analyte compounds by thoroughly exploring their potential energy surfaces (PES).[25] This allows for the identification of the global minimum isomer structures and the determination of which isomers are relevant at a selected temperature. Figure 2.1 shows a simplified version of the BH process. The user imposes restrictions on parameters such as the acceptable range of energies, the geometric distortions as well as the temperature. A modified Monte Carlo simulation guides the system towards a minimum via a series of random distortions.[26] Distortions include rotations through dihedral angles, bonding angle bends, as well as translations of atoms. (Figure 2.2) Low energy conformers obtained from the BH routine are optimized using the Gaussian quantum chemical package[27] and then kept or discarded based upon their energies. If the energy is lower than the "global" minimum it is retained and set as the new global minimum as well as the starting structure for the next BH iteration. If the energy of a structure is not below that of the current global minimum, it is kept or rejected according to a thermal Boltzmann probability distribution window, which is displayed as Eq. 2.1:

$$\exp\left(\frac{-(E-E_{GM})}{k_B T}\right) > a \qquad \text{Eq. 2.1}$$

In which E is the energy of the current structure, E_{GM} is the current global minimum energy, k_B is the Boltzmann constant and T is the temperature of the simulation. In typical BH searches the temperature is varied such that 50% of examined structures are accepted. "*a*" is a number chosen between 0 and 1 that gives the strictness of Boltzmann acceptance criteria for new structures. If the result of the Boltzmann probability is greater than a, the structure is kept and used for the next BH iteration, while if the result is less than a, the structure is discarded, and the most recently accepted geometry is used (i.e., from the previous step). This process continues until a pre-set number of BH iterations are completed.

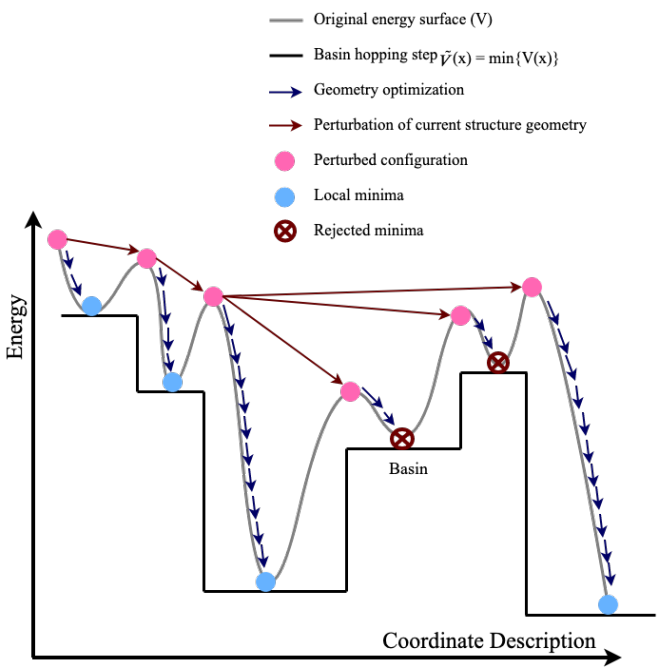


Figure 2.1: Schematic diagram of BH algorithm process for several steps, see text for details.

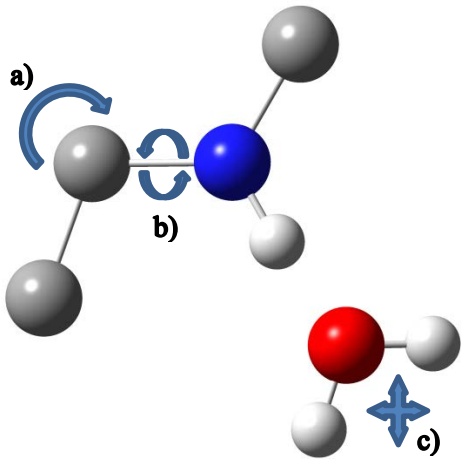


Figure 2.2: Different distortions available for use in basin hopping: a) Angle bend b) Rotation about dihedral c) translation of atoms (i.e., movement of free atoms/molecules)

As the BH algorithm will typically undergo tens of thousands of iterations per simulation, it is important to choose an efficient level of theory which also adequately describes the potential energy landscape for

the structural optimizations. The BH algorithm employed by the Hopkins group makes use of molecular mechanics force fields, such as UFF[28] and AMBER[29], due to the speed of the calculations that are produced. At this level of theory compounds are treated classically with atoms being treated as point charges and masses with their vibrations being modelled harmonically. Total energy is then calculated as the sum of the bonding interaction energies (bond stretching, angle bending and angle torsion) and the nonbonding interaction energies. [van der Waals (vdW) and Coulombic][29] Though the energies and structures produced at this level of theory are not particularly accurate, they do serve as candidate structures for further optimization. All resulting unique structures from BH are reoptimized at higher levels of theory in order to obtain more accurate energies and structures.

2.1.2 Quantum Chemistry Methods

Using the results of the BH algorithm as a base, structures, energies, and other chemical properties in this work were further refined using the PM7 and density functional theory (DFT) methods. PM7 is a semi-empirical method which employs the neglect of diatomic differential overlap (NDDO) approximations.[30] NDDO reduces the complexity of the two electron integrals in the Hartree-Fock (HF) equations. The PMX methods, of which PM7 is the latest, replace the two electron integrals with functions that are fit to known experimental and empirical quantities rather than calculating them directly, thus greatly reducing computational time.[31]

DFT is one of the most popular quantum chemistry methods for determining ground state properties of systems due to its ability to provide relatively accurate results at low computational expense. This is achieved by using electron density functionals rather than the wavefunction, as is done in HF or coupled cluster (CC) theory.[32] Additionally, each of the different DFT functionals (such as B3LYP, PBEO, etc,) use different empirical functions to approximate the many particle interactions or the exchange correlations.[33,34]

2.1.3 MobCal-MPI

Collision cross sections (CCSs) were generated via the use of the MobCal-MPI software[35] that was developed by Ieritano *et al.* at the University of Waterloo based on the original MobCal implementation.[36,37] It uses the trajectory method for calculating CCS which is commonly used in similar software. This method computes CCS through numerically approximating the momentum transfer integrals in Eq. 2.2:

$$\Omega_{avg} = \frac{1}{8\pi^2}\int_0^{2\pi} d\theta \cdot \int_0^{\pi} \sin\varphi \, d\varphi \cdot \int_0^{2\pi} \frac{\pi}{8}\left(\frac{\mu}{k_b T}\right)^3 d\gamma \cdot \int_0^{\infty} \mathrm{g}^5 \cdot \exp\left(-\frac{\mu g^2}{2k_b T}\right) dg \cdot \int_0^{\infty} 2b(1-\cos\chi(\theta,\varphi,\gamma,g,b))db \quad \text{Eq. 2.2}$$

Here, θ, φ, and γ are used to define the orientation of the ion relative to the center of mass axis of the ion-collision gas, g gives the relative velocity, b gives the impact parameter, μ gives the reduced mass of the ion-collision gas pair, and χ gives the angle of scattering after buffer gas collision with the ion. These integrals are averaged over many permutations of velocity and geometry for the given choice of buffer gas and analyte ion. Mobcal-MPI also improves upon its predecessor MobCal in several ways. The standard Lennard-Jones (LJ) 12-6 potential (Eq. 2.3) used in determining the vdW interactions was replaced with the Exp-6 potential (Eq. 2.4) which has been shown to be more accurate when calculating CCS.[38]

$$V_{vdW}(r_i) = \sum_{i=1}^{n} \varepsilon_i \left[\left(\frac{r_i^*}{r_i}\right)^{12} - 2\left(\frac{r_i^*}{r_i}\right)^{6}\right] \quad \text{Eq. 2.3}$$

$$V_{vdW}(r_i) = \sum_{i=1}^{n} \varepsilon_i \left[1.84\times10^5 \exp\left(\frac{12r_i}{r_i^*}\right) - 2.25\left(\frac{r_i^*}{r_i}\right)^{6}\right] \quad \text{Eq. 2.4}$$

For these equations, ε_i gives the depth of the potential well, r_i gives the distance between buffer gas and its interaction partner, while $r_i{}^*$gives equilibrium distance between the same. Parameters were fit for

different polarization states of atoms and cover a wider chemical space than previously, adding P, S, Cl, Br, and I potentials to the original C, H, O, N, F list. Additionally, calculations have been sped up significantly through parallelization using message passing interface (MPI).

2.1.4 Machine Learning

The Orange software suite was utilized to perform machine learning predictions in the given work.[39] Built using the python scripting language, Orange makes use of a visual programming GUI and groups low level procedures (e.g., feature scoring and data filtering) into larger algorithms in order to simplify ML for end users. This allows for easy creation of workflows and for various ML methods to be examined at once. (Figure 2.3)

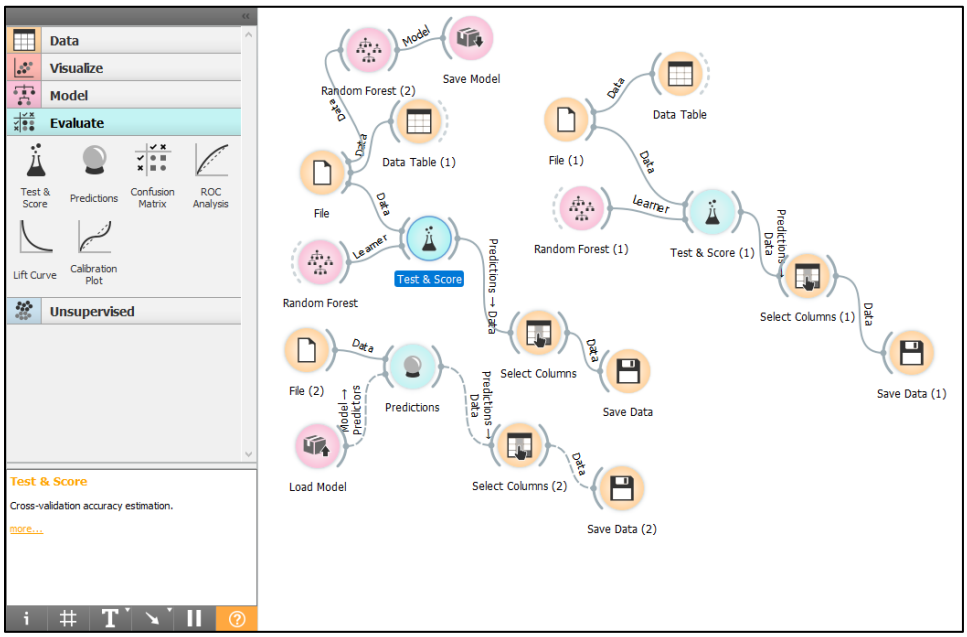


Figure 2.3: Example Orange worksheet, showing the evaluation of two different sets of parameters for predicting the CCS for the same set of compounds. The set of parameters that scored more accurately on the training set had its ML model save, which was then used to predict the CCS of a test set of compounds.

Random forest (RF)[40] is a ML algorithm that has been shown to perform exceptionally well in predicting physicochemical properties through quantitative structure activity relationships (QSAR).[41] It performs well without requiring a high degree of input data tuning or pruning, can handle large numbers of

descriptors, and also copes well with poorly chosen descriptors; such as those that are redundant or irrelevant. This algorithm has seen previous use in the Hopkins group for predicting various physicochemical properties such as the CCS, pK_b, LogD, EPSA and cell permeability in a small set of different drug candidate compounds.[1]

The RF algorithm (Figure 2.4) uses a large (~100-1000) number of decision trees, each of which generates a numeric prediction or classification. Numeric predictions can then be averaged to yield a prediction for the “forest” of decision trees. RF is a supervised learning method which uses a training set of data to construct a model to be employed for unknown data predictions. Each node on a tree, shown in Figure 4 as circles in each tree, is created to sort data based upon a random feature (such as CV at given SV, m/z, charge dipole, etc) within the training set. As a sample molecule progresses through the nodes within the tree (path is shown in yellow) it becomes grouped with molecules which are increasingly similar to itself. Using this grouping and the known data within the training set, classifications such as chemical grouping or numeric predictions like CCS are made. Below, a classification result is depicted between red and green boxes in each tree. In order to mitigate the effect of any individual input from having a disproportionate effect on the algorithm, each tree only examines a portion of the training data. Some entries will be used multiple times in the construction of a decision tree, while some will not be used at all. This latter case will happen for each “training" compound in around 37% of all trees.[42] By utilizing a sufficient number of trees in order to come to a consensus, the overall result converges very close to the true answer of the classification/regression query, and is relatively insensitive to additional features which do not correlate strongly to the target variable.

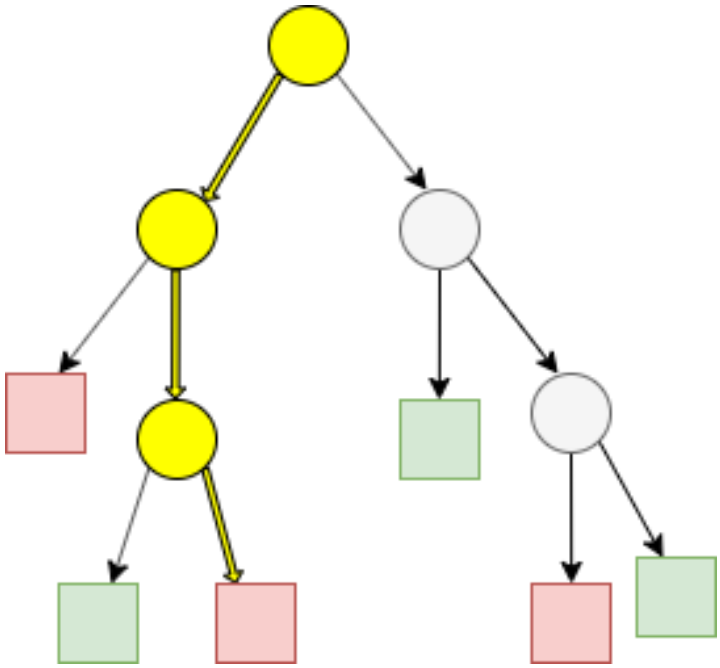


Figure 2.4: Example RF decision tree. [42]

A key part of ML is to be able to predict unknown values.[42] In order to test the predictive ability, the input data is broken up into test and training sets, and a k-fold cross-validation approach is used In k-fold cross validations the data is split into k equal sets (or folds). One set is used for testing, while the remaining (k-1) sets are used for training. Sometimes one additional fold is removed from the training set, the internal validation set, to further elucidate which parameter values influence predictivity the most. A visual of one such k-fold cross validation approach is shown in Figure 2.5. A predictive model is generated by teaching the ML algorithm with the training set to infer relationships/trends between selected features and the target. The generated model is used for target prediction on the data within the test set, and its error is evaluated through the root mean square error (RMSE) or the average absolute error (AAE) between predicted results and the known target values. The total cross validation error is evaluated by iterating the test set over each of the k folds, training on the remaining (k-1) (or k-2 with an internal validation set) folds in each case, The final cross validation error results from an average over the k model errors. Typically, 5 to 10-fold cross validations are used to test generated RF models.

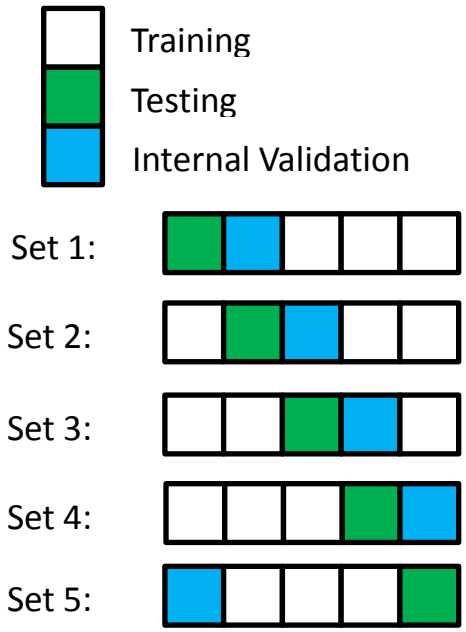


Figure 2.5: 5-fold cross validation approach.[42]

2.2 Experimental Methods

2.2.1 Electrospray Ionization

In a standard DMS setup the analyte is injected into an electrospray ionization (ESI) source, where cations or anions are pumped by a syringe pump through a charged capillary and released as a fine spray of charged droplets. The applied capillary voltage helps generate the ions, often alongside an acid or base within the sample solution. Positive and negative ions can be selected by placing positive potential either at the capillary or at the counter-electrode. The droplets are guided along a pressure and potential gradient towards the DMS cell, and are reduced in size through a combination of collisions with a nebulizing gas (i.e., N_2), release of smaller droplets, and thermal evaporation (see Figure 2.6).[43,44] The droplets shrink until they eventually reach a critical size called the Rayleigh stability limit. Here, the repulsive electrostatic forces within the droplet exceed the surface tension of the droplet, which will subsequently break apart into smaller droplets in a process called Coulomb explosion.[44] When the droplet becomes

sufficiently small, excess charge is removed by inducing a phase change into the gas state rather than breaking apart further.[45]

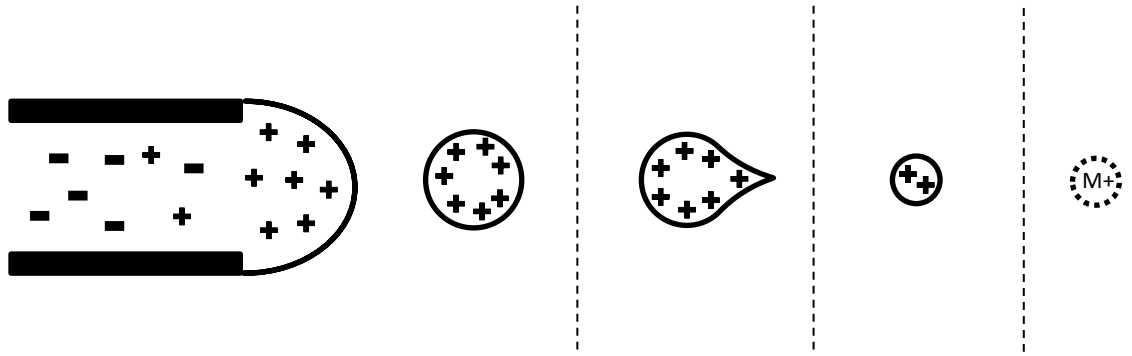


Figure 2.6: Display of the ESI process showing the initial spray, reduction of size and deformation before Coulomb explosion into gas phase .[43] The dotted lines represent time gaps in the process, showing different stages.

2.2.2 Differential Mobility Spectrometry

Inside the DMS cell (See Figure 2.7) ions at atmospheric pressure are forced to oscillate perpendicular to the longitudinal axis by an applied asymmetric waveform called the separation voltage (SV). An atmospheric pressure of a background gas (N_2) is present in the cell and collides with the ions as they traverse the cell. Eq. 2.5 describes the changes in ion mobility as a function of the electric field due to changes in the apparent collision gas viscosity.[46]

$$K\left(\frac{E}{N}\right) = K(0)\left[1 + \alpha\left(\frac{E}{N}\right)\right] \qquad \text{Eq. 2.5}$$

E/N is the field per volume density of neutral particles (in Td), K(0) is mobility at low field, and α(E/N) is a function that gives mobility change at different E/N values. The differential mobility which results from the asymmetric waveform potential, as depicted in Figure 2.7, will cause ions to move towards the electrodes .[47] A DC voltage, called the compensation voltage (CV), is then applied in order to direct ions back along the longitudinal axis and towards the mass spectrometer. Ions with differing zero field mobilities and α values will behave differently as they pass through the DMS cell, thus requiring a different CV value to guide them through the cell. This difference of CV value allows for these ions to be

distinguished in a method orthogonal to that of a mass spectrometer. The selectivity of DMS is further increased by the usage of solvent modifiers (e.g., acetonitrile, water, etc.) that are seeded in the background gas that can modify the mobility of the analyte ions.[48] In low field conditions the ions can become solvated, reducing their mobility due to increased ion-solvent interactions. This decrease in mobility is caused by the ions' apparent size being increased by the bound solvent molecules. Under high field conditions the ions are desolvated and their mobility increased as they are accelerated through the cell. This dynamic clustering process can generate significantly different clustering profiles for isobars with only minute differences in their binding energies (BEs) and thus enabling separation.[47] As the analyte leaves the DMS cell, a potential difference is applied which accelerates the ion toward the mass spectrometer and removes excess solvent molecules leaving the bare ion for detection by MS. The potential difference is controlled by the declustering potential (DP).

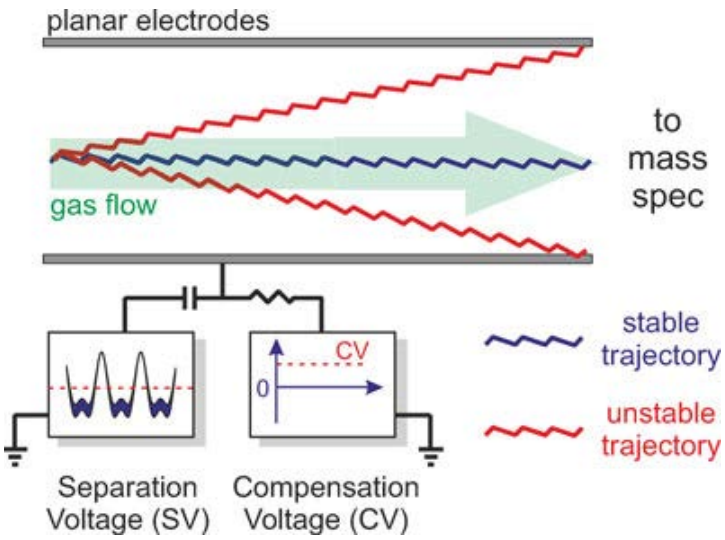


Figure 2.7: Schematic diagram of the DMS cell. The SV controls the magnitude of the migration toward the electrodes due to differing high and low field mobility of ions, and the CV directs ions with a specific mobility into the MS.[47]

2.2.3 Tandem Mass Spectrometry

Analyte ions pass through a system of four parallel metal rods, called a quadrupole mass analyzer (see Figure 2.8), where the pairs of diagonally aligned rods have a radio frequency AC voltage applied,

along with a DC voltage offset. This configuration of voltages creates an electric field which pushes the ions along parallel to the rods while causing the ions to oscillate perpendicularly to their movement. This spiraling movement is described using Eq. 2.6, Mathieu's differential equation[49]:

$$\frac{d^2u}{d\xi^2} + (a_u - 2q_u cos2\xi) * u = 0 \qquad \text{Eq. 2.6}$$

U represents the direction along the coordinate axes, ξ is a dimensionless parameter equal to *Ωt/2* where Ω is the radial frequency of the rf voltage and t is time, and a_u and q_u are the dimensionless trapping parameters. This movement allows only ions with a m/z corresponding to a specific AC/DC voltage pairing to pass through the quadrupole. By choosing different of AC and DC voltage pairs, ions of a specific m/z can be allowed to pass through to the detector while others are filtered out.[50]

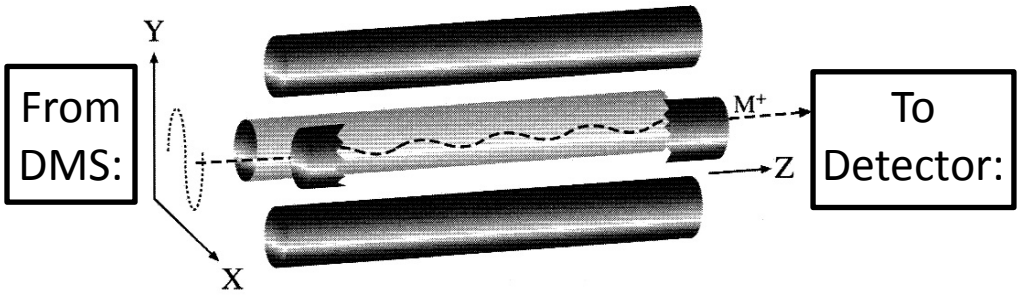


Figure 2.8: Oscillation of an ion as it moves from source to detector within a quadrupole mass analyser. Adapted from Ho et al.[43]

The utility of the MS is further improved by using multiple quadrupoles in series, resulting in the tandem MS (MS/MS) system. The triple quadrupole MS variant, as is used in the Hopkins lab, is a common example of this kind of setup. The first quadrupole (Q1) is used to select for an ion of interest, which is accelerated into the second quadrupole (Q2) through application of a potential difference called the collision energy (CE). In Q2 ions collide with N_2 and fragment according to the CE applied in a process

known as collision induced dissociation (CID). The product ions can then be examined in the third quadrupole (Q3), giving additional structural information about the parent.[43] (See Figure 2.9)

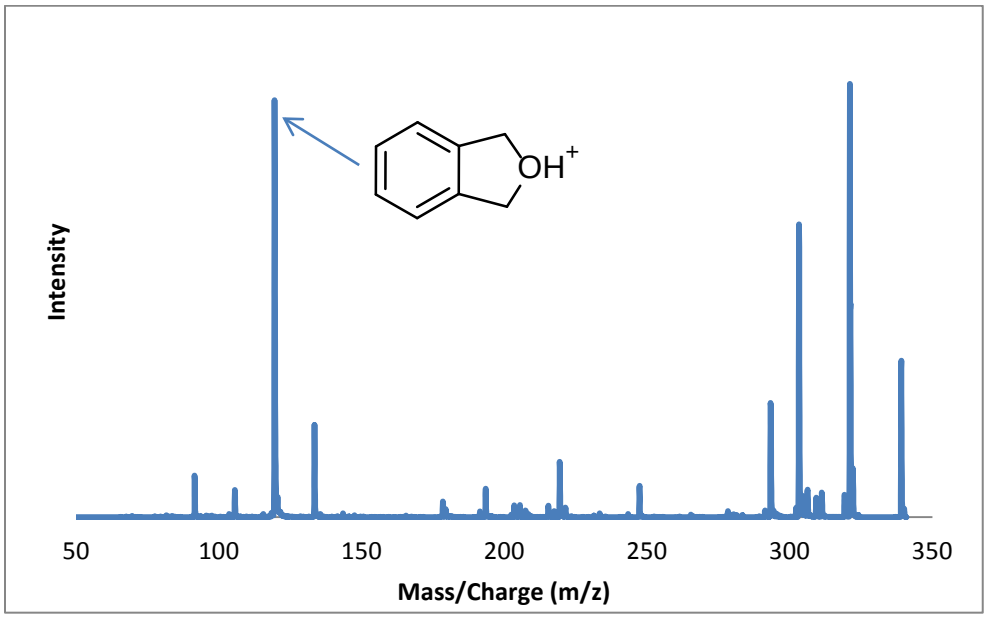


Figure 2.9: Example enhanced product ion (EPI) scan spectra of an unknown 339.3 m/z cluster that contains protonated phthalane (depicted in figure) at 121.1 m/z.

2.3 Data Analysis

Dispersion plots are generated to examine and characterize the DMS behavior of analyte ions. Peak CV values are taken from ionograms, which are plots of signal intensity versus CV for a specific SV (see Figure 2.10). A dispersion plot is generated by plotting optimal CV for ion transmission as a function of SV. Dispersion plots display three general types of behavior: Type A, B, and C.[51] Type A behavior demonstrates strong ion-solvent clustering, in which ions exhibit increasingly negative CV values as the SV increases. In Type B behavior, ions exhibit weak clustering with the solvent and CV initially decreases with SV, before eventually increasing towards positive values. Lastly, Type C behavior is indicative of only hard sphere interactions with buffer gas molecules, and CV strictly increases with increasing SV.[48] Figure 2.11 provides example dispersion plots showing these different behaviors. In order to achieve separation of similar compounds, care must be taken to find a modifier that strikes a balance between the different DMS behaviors. In general, pure N_2 generates the least amount of

clustering, while acetone (ACE) provides the strongest clustering. Typically, ideal separation occurs by picking a modifier that clusters moderately well with analyte ion, displaying type B behavior. Using a modifier that provides weak clustering behavior with a given analyte allows for greater resolution of differing ion behavior. In practice, a grouping of ions that displays purely type A or type C behavior is less likely to be separable than a grouping of ions that displays purely type B behavior or some mixture of the behaviors.

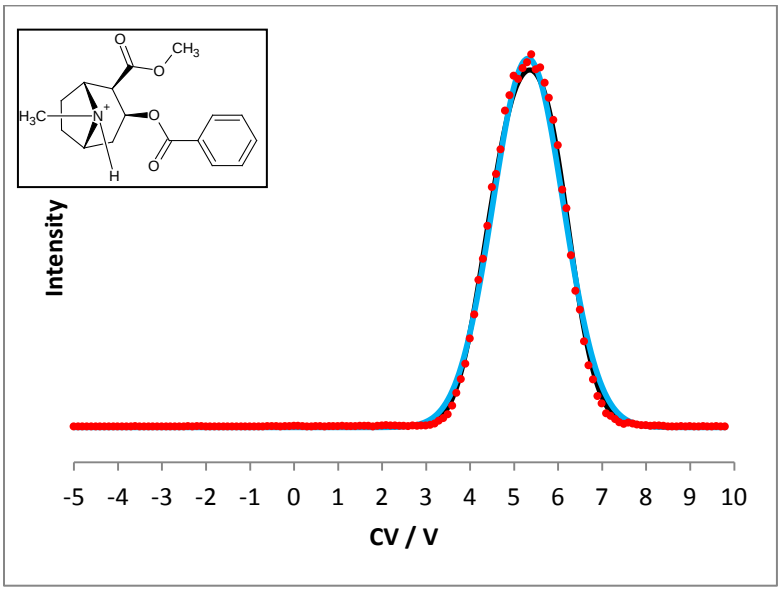


Figure 2.10: Ionogram generated for protonated cocaine ions at a SV of 3250 V and temperature of 150 □C.

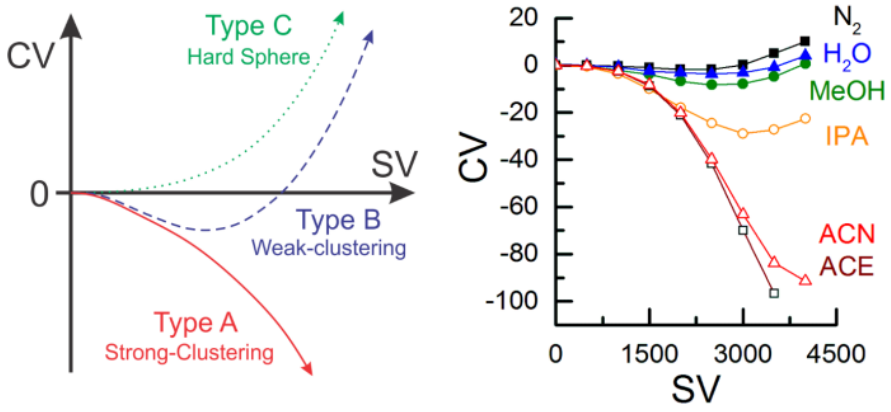


Figure 2.11: (Left) General types of DMS behavior. (Right) Dispersion plots for $(CH_3)_4N^+$ in $N_2$with various solvent modifiers added at 1.5% (v/v).[47]

Chapter 3 : Separating Chiral Isomers of Amphetamine and Methamphetamine using Chemical Derivatization and Differential Mobility Spectrometry

Chapter 3 examines the separation of derivatized amphetamine and methamphetamine isomers. The experimental work in this chapter was carried out by collaborators at SCIEX, while the computational work was carried out by me, along with an analysis of the combined data. A manuscript of this work is currently under review:

Reference: Kalfe, Amol; Bowman, Zack; Le Blanc, J.C. Yves; Liu, Chang; Hopkins, W.; Campbell, J Separating Chiral Isomers of Amphetamine and Methamphetamine using Chemical Derivatization and Differential Mobility Spectrometry. *Anal. Chem. (In review).*

3.1 Introduction

Amphetamine (AMP) and methamphetamine (MeAMP) are a similar pair of stimulants that each appear as enantiomeric mixtures, with a single chiral center adjacent to the methyl and amine groups (Figure 3.1). AMP is primarily used as a treatment for attention deficit hyperactivity disorder (ADHD), and has seen use in treatment for mild depression, narcolepsy and various other disorders.[52] The more potent methylated derivative, MeAMP, is also used for treatments of ADHD, though as a second line treatment, as well as for obesity and other conditions.[53] Typically, only the S isomers are used for medicinal purposes with a few notable exceptions; such as the 3:1 R:S AMP isomer ratio used in Adderall[52], and R-MeAMP as an over the counter decongestant.[54,55] However, the R and S isomers of both compounds are restricted due to their potential for abuse, particularly the S isomer, due to its faster and more complete metabolization relative to the R isomer.[56] Illegal use is typically also centered on the more potent S-isomer[52,57], so the ability to differentiate these isomers is an invaluable tool in examining seized illicit materials. In fact, by US sentencing guidelines a greater punishment is recommended for possession of a

>80% mixture of S-MeAMP than for lower amounts, making the sensitivity and specificity of detection especially important.[57]

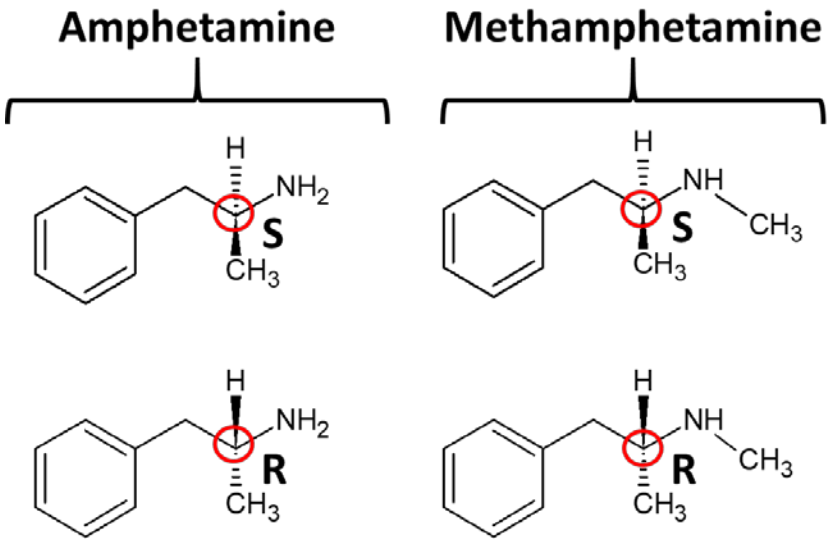


Figure 3.1: Structures for AMP and MeAMP with the chiral center indicated by the red circle. Designation of chirality is made using the Cahn-Ingold-Prelog (CIP) convention.[58]

Various methods have been used previously in order to discriminate between the AMP and MeAMP enantiomeric pairs. Liquid chromatography (LC) techniques have been used to separate AMP and MeAMP enantiomers by using a chiral stationary phase of compounds such as vanomycin and β-cyclodextrin (i.e., large compounds with many chiral centers).[56,59] Comparable compounds, such as HP-β-cyclodextrin, have also been used as chiral selectors in capillary electrophoresis (CapE) experiments to a similar degree of success.[60] Alternatively, the separations are produced in the gas phase using methods such as gas chromatography (GC).[54,61] Here, however, additional preparations must be made in the form of a chemical derivatization step. By using (*S*)-*N*-trifluoroacetyl prolyl chloride (TPC) to convert these enantiomers into larger diastereomers the structure becomes less polar and more thermally stable than the original compounds. Additionally, the physicochemical properties of the enantiomers are modified relative to each other, becoming distinct, and the GC sampling and separation is improved.

Gas phase chemistry can also be employed alongside ion mobility spectrometry (IMS) techniques in order to separate chiral compounds.[62] Previous works have shown that compounds which cannot be resolved using only nitrogen drift gas can be separated using a gas dopant that is either chiral, such as (*S*)-2-butanol,[63] or achiral, such as 1-octanol.[64] One IMS method that has shown particular promise in distinguishing different isomers is DMS[24,32,65–68], which separates ions based upon high and low field differences in the CCS rather than purely the low field CCS used in drift-time or travelling wave IMS. This can allow for separations that might not be possible in only low field conditions. Amphetamine has been previously examined with DMS in order to separate it from isobaric contaminants. A variant of DMS, high-field asymmetric waveform ion mobility spectrometry (FAIMS)[51], has previously been used to separate chiral compounds.[69,70] However, these separations required metal cores and reference ligands to be used in the experiment to create separable clusters, and required high concentrations of solution to create clusters. Additionally, these separations were not observed to follow predictable trends.

Taking advantage of the previous literature results, this experiment will make use of the TPC derivatization strategy used in gas chromatography experiments, as well as making use of gas modifiers to separate and identify the derivatized analyte diastereomers. Using DMS as a method for separating chiral compounds is desirable due to shorter run times and less complicated experimental setups compared with GC or LC methods.[71,72]

3.2 Experimental Methods

Reagents and Chemicals: Pure species and racemic mixtures of amphetamine (AMP) and methamphetamine (MeAMP) were purchased from Cerilliant Corporation and the (S)-N-trifluoroacetyl prolyl chloride (TPC) was purchased from Sigma-Aldrich. Each compound was used as received.

Derivitization reaction (Figure 3.2): 12.5 μL of sample was added to a 500 μL capped Eppendorf tube and dried down using nitrogen gas flow. 100 μl of hexane and 100 μL of TPC were added to reconstitute the solution before it was sealed and vortexed at 500 rpm for 10s to thoroughly mix the reactants. The

mixture was then heated for 15 min at 70 °C, shaking at 500 rpm before being dried and reconstituted with 250 μL of 50/50 water/acetonitrile along with 0.1% formic acid by volume. A 30 μL sample of this solution was diluted with 1470 μL of the same solvent to form a stock solution of 1000 ng/mL concentration.

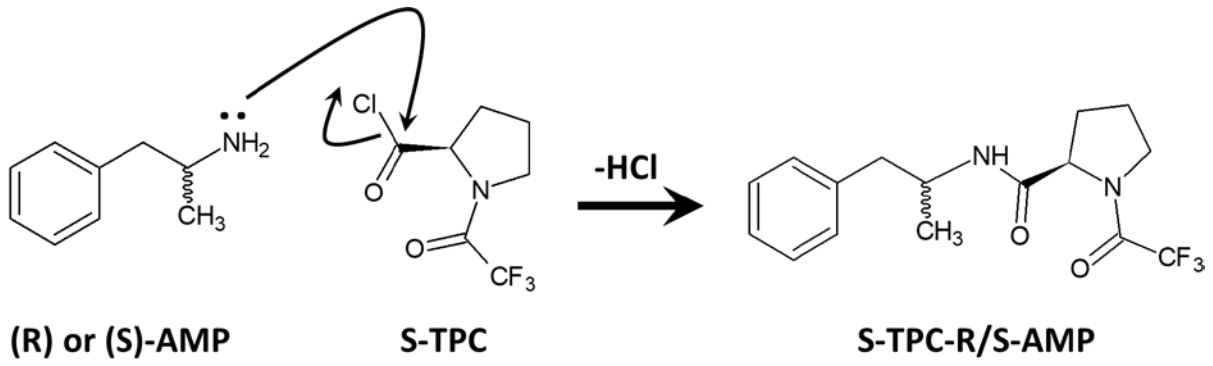


Figure 3.2: Reaction of AMP and TPC to form TPC-AMP

DMS Conditions: The DMS was mounted at atmospheric pressure between an ESI source (5500 V) and a hybrid triple quadrupole MS. For this experiment, the DMS was set at 150 °C with a nitrogen curtain gas of 10 psi and a declustering potential of 100 V. (For further details on the DMS system, see section 2.2.2) The SV was held constant at 4000 V in order to allow for maximum separation of the analytes, while the CV was varied across a wide range in order to ensure that all relevant isomers were observed. An external pump provided the gas modifiers in a 1.5% V/V of the total curtain gas flow. Each component of the experimental setup was controlled using the Analyst® software (v. 1.6.2). The enhanced product ion (EPI) mode was used in order to observe the fragmentation of mass peaks in the DMS. EPI mode selects for a parent mass of interest in Q1 and scans over a user selected range of m/z in Q3 in order to observe all fragment peaks from the selected parent ion. This allows for the removal of isobaric interferences due to differences in the fragmentation pattern.

3.3 Computational Methods

For each compound of interest, initial neutral structures were generated using chemical intuition. These structures were then geometrically optimized at the PM7 level of theory to give approximated structures to use as a starting point for further calculations.[30] These preliminary geometries were first used to generate various prototropic isomers, which were optimized at the B3LYP/6-31++G(d,p) level of theory and examined by comparison of their relative energies.[73] The lowest energy protomers were then carried forward to be manipulated using the BH algorithm[27,32,74–76] which interfaced with Gaussian09 software for structural optimization steps. As discussed in Section 2.1.1, this ensures a thorough investigation of the PES to identify candidate structures for low energy isomers. For this work, the BH routine used the AMBER molecular mechanics forcefield, and partial charges calculated using the CHelpG method for the input structure.[29,77] Approximately 20,000 geometries were sampled across the PES by manipulating the dihedral angle of selected bonds by a random rotation of $-10° \leq \theta \leq 10°$ in each protonated structure. The unique structures were then pre-optimized at the semi-empirical PM7 level of theory. This resulting set of unique structures was reoptimized at the B3LYP/6-311++G(d,p) level of theory. The DFT-optimized isomers were ranked based upon their electronic energies to determine which species were most likely to be observed at the internal DMS temperature (150 °C). In order to ensure that each structure was a local minimum on the PES (rather than a transition state), and to obtain thermochemical corrections for the energies, normal mode analyses were performed.

The MobCal-MPI code was used to calculate the orientationally-averaged CCS (Ω_{N2}) for the various isomers of amphetamine and methamphetamine with molecular nitrogen as the collision partner.[35–37] The trajectory method is employed in order to generate these theoretical Ω_{N2} values.[32] For a given structure, 10 cycles of mobility calculations are completed, using 512 points of impact parameter integration and 48 points of velocity integration. (See Eq. 2.2)

Minima structures from BH were also used to generate ion-solvent structures. For initial guess structures, the electrostatic potentials of the protonated ions were examined for areas with a particularly high positive/negative partial charge, as these are the likely sites of bonding (see Figure 3.3). The solvent molecules are positioned and oriented depending on the polarity of their partial charge and then optimized in order to find a minima cluster structure. For example, in the case of ACN where the negative partial charge is most accessible, in front of the triple bonded nitrogen along the principal axis, it is placed near the most positive areas on the analyte molecule. Optimized geometric structures and thermochemical corrections were calculated using the B3LYP/6-311++G(d,p) level of theory. These structures are assumed to be the global minima ion solvent structures in this case.

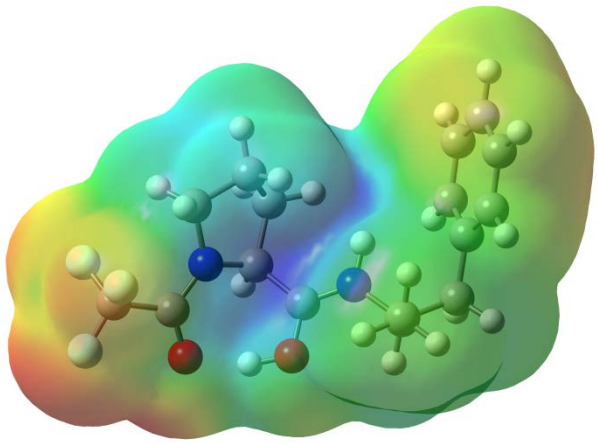


Figure 3.3: Partial charge for one *S*-TPC-*R*-AMP isomer. Blue indicates an area of relatively high positive partial charge and red indicates an area of relatively high negative partial charge.

3.4 Results & Discussion

3.4.1 Separation of (S,R/S,S)- AMP/MeAMP

Initially, an attempt was made to observe separation of racemic AMP and MeAMP using DMS alone. Unfortunately, the peaks for both enantiomers lie at the same CV value, and so are not separable. In order

to increase separation, a derivatization reaction between the AMP/MeAMP isomers and (S)-N-trifluoroacetyl prolyl chloride (*S*-TPC) was performed in order to create greater differentiation between the chiral groups.

The TPC-derivatized compounds were initially examined using a pure nitrogen environment with no additional gas modifiers. Unfortunately, in this case no separation was observed between the ions of interest. Therefore, additional treatment was required for the isomers to be resolved. Two methods were examined in order to make a distinction in the isomers' DMS behavior: i) Supplying a throttle gas to the DMS cell in order to extend the DMS flight time and ii) supplementing the cell's nitrogen environment with a modifier gas to enhance separation through modifier-ion clustering.

Increasing the cell retention time with a throttle gas (@10 psi in this case) causes minute differences in structure to be exacerbated and allow for separations that would not occur under standard conditions, as is shown in Figure 3.4 for the TPC-AMP derivatives. Since there is very little binding observed in a pure N_2 environment, it is assumed that the CCS is the main contributor to the different mobilities in order to determine which peak in the ionograms corresponds with which given isomer. *S*-TPC-*R*-AMP was calculated to have a CCS of 165.2 $Å^2$, while *S*-TPC-*S*-AMP had a calculated CCS of 172.3 $Å^2$. Due to the significant difference between these values it is plausible that these compounds could be separable. It is expected that more positive CV values will be observed for analytes with smaller calculated CCSs, as qualitatively this will lead to larger differential mobilities between high and low field conditions in the hard sphere limit. Using this as an "eyeball test" it would be expected that the CV = 11.7 V peak would correspond to *S*-TPC-*R*-AMP and the CV = 9.9 V peak would correspond to *S*-TPC-*S*-AMP. In the case of the MeAMP chiral isomers, a CCS of 170.2 $Å^2$ was calculated for *S*-TPC-*R*-MeAMP and a CCS of 171.6 $Å^2$ was calculated for *S*-TPC-*S*-MeAMP. Despite the similarity of the CCS values, falling with the 96% confidence range (~±2 $Å^2$) of each other (as determined by calculated MobCal error), these compounds also separated in the N_2 gas environment with resolving gas (Figure 3.5) with a similar degree of separation as the *S*-TPC-AMP diastereomers. Due to this overlap it is hard to make an authoritative

judgment as to the identity of which isomer is generating each peak. However, due to the high fields present; the CCS is not the only property that determines separation. In order to more conclusively identify the identity of the *S*-TPC-AMP and MeAMP diastereomers using theory, another experimental method must be used.

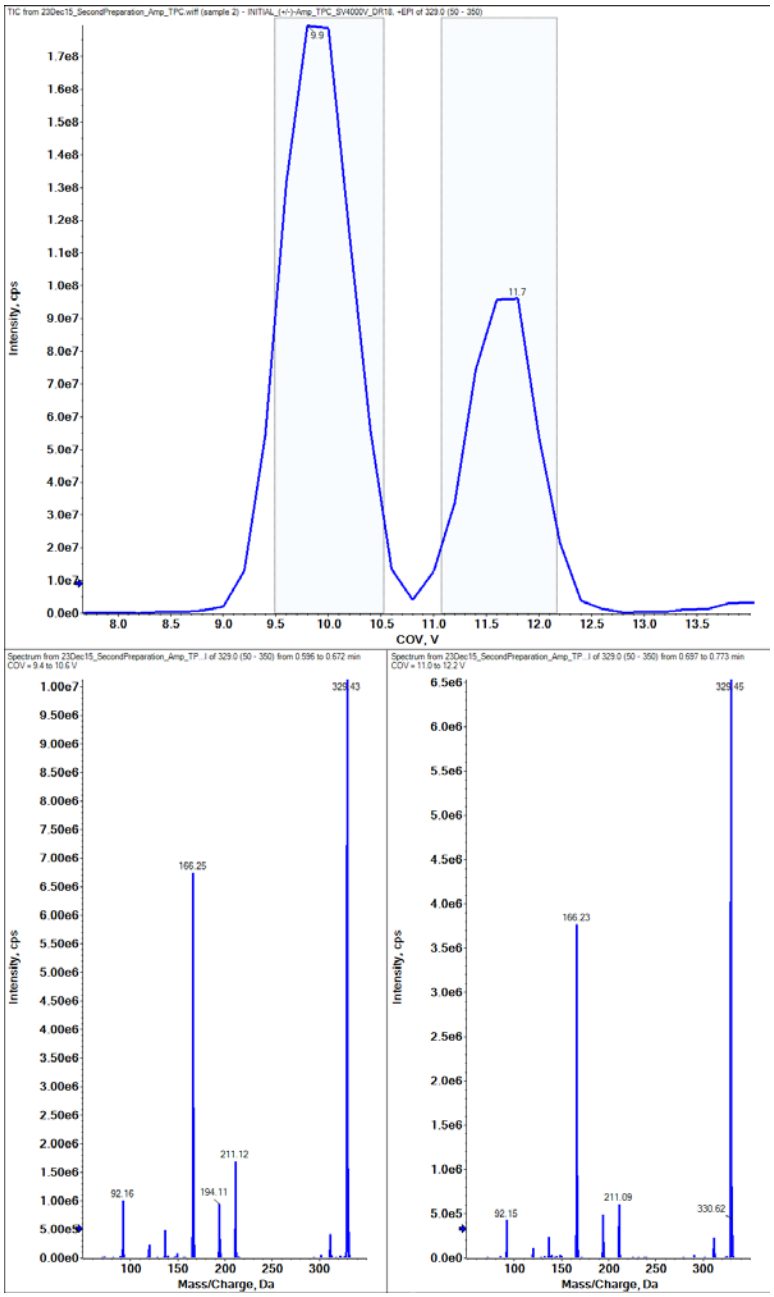


Figure 3.4: (Top) Experimental DMS data showing separation of racemic mixture of *S*-TPC- *S*/*R*-AMP in pure N_2 gas modifier. (Bottom) Fragmentation of each parent ion peak, showing the strong similarities between them.

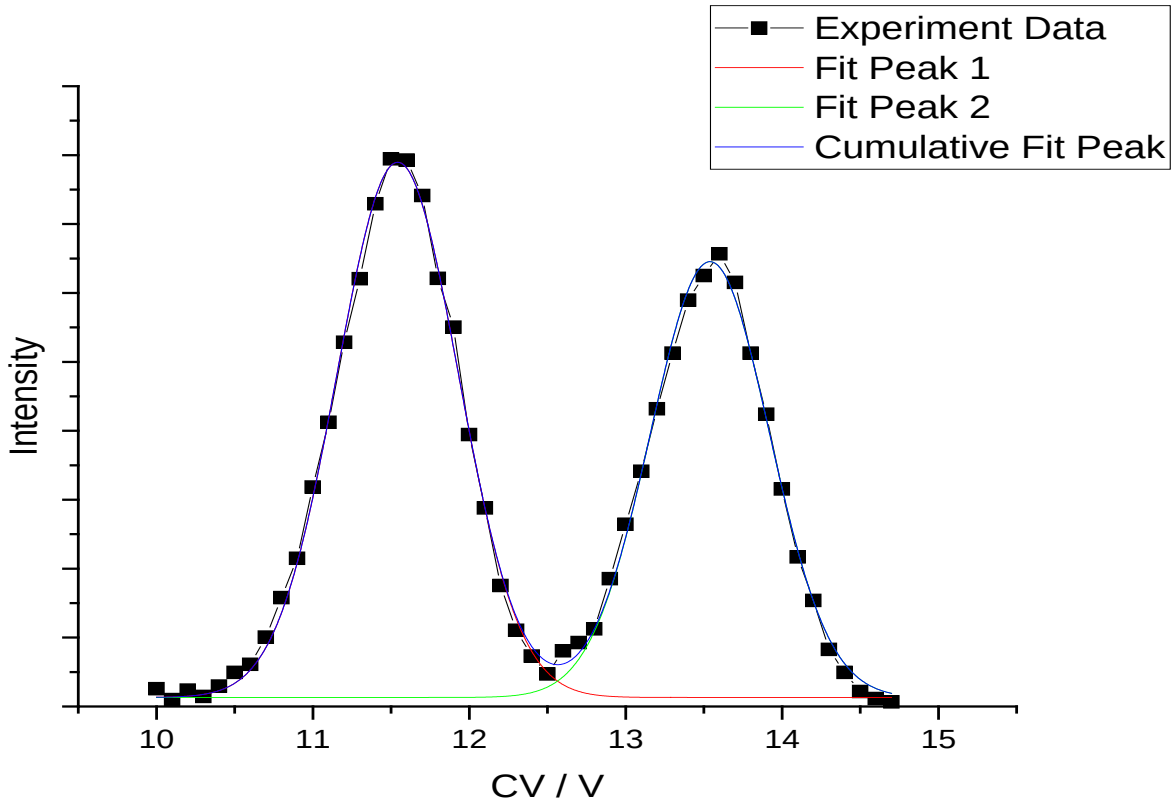


Figure 3.5: Experimental DMS data showing separation of racemic mixture of *S*-TPC-*S*/*R*-MeAMP in pure N_2 gas modifier.

Seeding the buffer gas with modifier enhances differential mobilities due to differences in modifier-ion clustering that can vary even in very similar structures. When ACN was used as a modifier gas, separation was observed between both TPC AMP/MeAMP pairs. (Figure 3.6) Peak assignments were established by examining each diastereomer separately using the DMS. From the separate diastereomer experiments *S*-TPC-*R*-AMP was assigned to the peak at a CV of −19.6 V, while *S*-TPC-*R*-AMP was assigned to the peak at −13.3 V. The MeAMP pair had less separation than the AMP pair but the two isomers were still distinguishable from each other. The peak at −6.3 V was assigned to *S*-TPC-*S*-MeAMP while the peak at −3.5 V was assigned to *S*-TPC-*R*-MeAMP. One additional peak at a CV of −9.5 V was believed to originate from the MeAMP isomers, likely resulting from a cluster of *S*-TPC-*R*- and *S*-TPC-*S*-MeAMP due to the appearance of this peak in both chiral isomer scans. By performing this experiment in multiple reaction monitoring (MRM) mode potential isobaric interferences can be removed, such as one observed

in a full scan MS/MS analysis. In this case a 343 m/z isobar that was observed in EPI mode, (Figure 3.7) was omitted from in Figure 3.6 because it did not have a 166 m/z or 91 m/z fragment that is used for signal identification in MRM mode when scanning for the TPC-MeAMP derivatives.

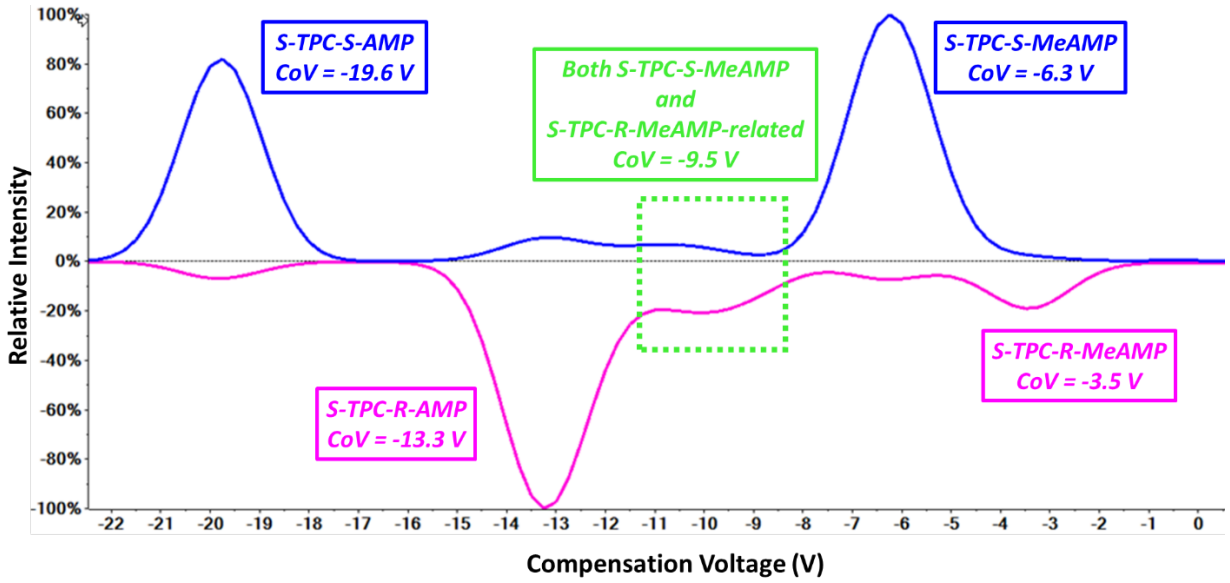


Figure 3.6: Peak separation of the (S/R) pairs of *S*-TPC-AMP and *S*-TPC-MeAMP using ACN modifier gas. The *S*-isomers are shown in blue on the positive intensity portion of the y-axis, while the *R* isomers are shown in pink on the negative intensity portion of the y-axis.

Figure S1

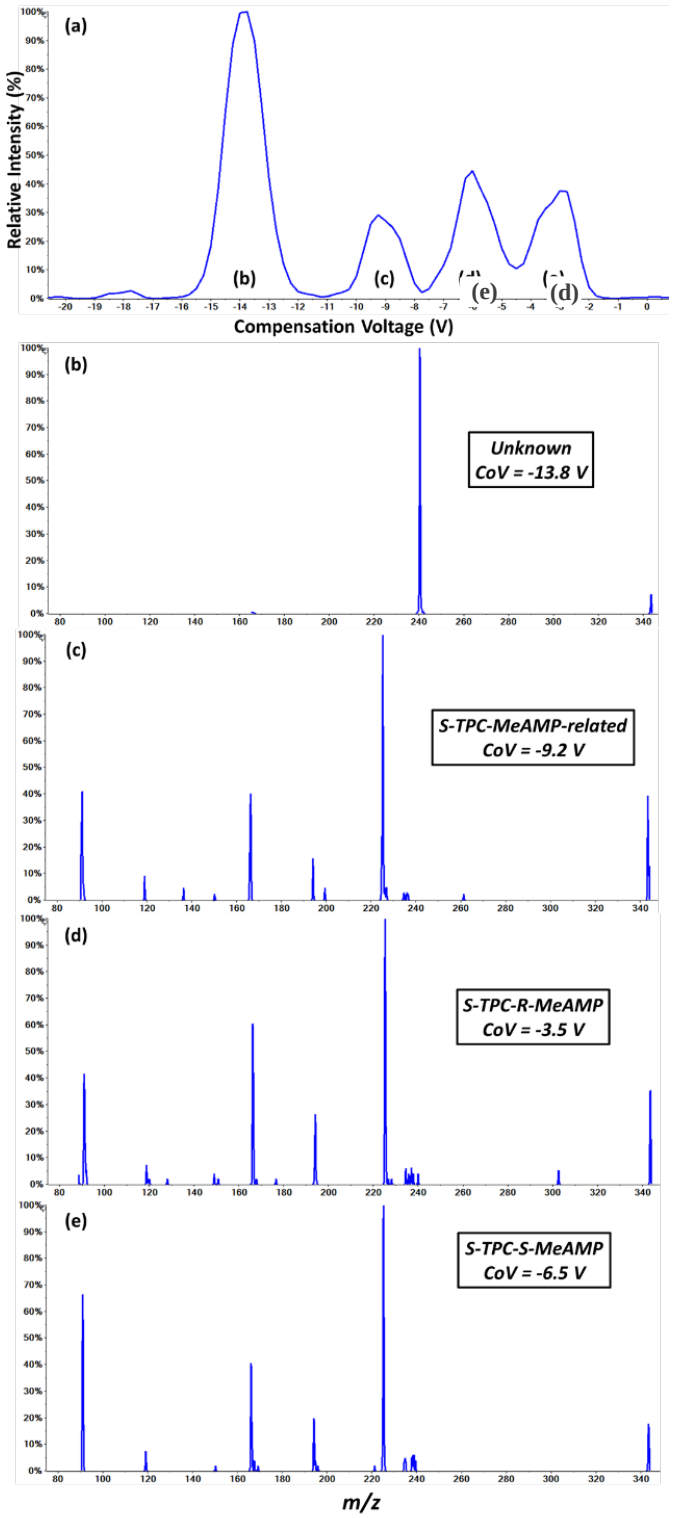


Figure 3.7: EPI scan of *S*-TPC-MeAMP racemic mixture and fragmentation and observed peaks. Plot a) gives the overall ionogram for the mixture at SV = 4000V, while plots b-d) give the fragmentation patterns observed for each ionogram peak as well as their identifications.

3.4.2 Comparison of computational and experimental results:

While in this case the TPC-AMP/MeAMP peak assignments in the ionograms were made by examining individual TPC-AMP/MeAMP derivatives separately and comparing to the racemic mixtures, this is not always possible. Sometimes the peaks can be identified by their fragments, such as in the case of the isobar observed in EPI mode. However, this is not always the case, especially in such similar compounds as diastereomers. An alternative method is to use the empirical understanding of DMS alongside computational modeling in order to compare theoretical parameters to the behavior observed in DMS experiments. The TPC-AMP/MeAMP system detailed thus far serves as an example for this method of understanding DMS separations.

The BH algorithm was utilized in order to determine the global minimum TPC-AMP and MeAMP unsolvated structures, which were then used to determine ACN clustered structures and energies. The energy differences between the solvated and unsolvated structures are used to determine the binding energy for each of the TPC derivative species with ACN. The energy of binding is presumed to play a major role in determining DMS behavior in the presence of a solvent modifier. Compounds which display a stronger affinity for binding with the modifier gas (i.e., a larger BE) should display stronger type A behavior and therefore require a more negative CV to be applied in order to be detected.

The BH script (and subsequent optimizations) identified three, two, four and four low energy isomers for *S*-TPC-*R*-AMP, *S*-TPC-*R*-MeAMP, *S*-TPC-*S*-AMP and *S*-TPC-*S*-MeAMP respectively within ΔG= 3.59 kcal/mol of the global minimum structure. (See Figure 3.8 below for global minimum structures) The amphetamine derivatives displayed an OH•••O sharing of the charge carrying proton, while the methamphetamine derivatives exhibited an OH•••N H-bonding motif with the ring bound nitrogen.

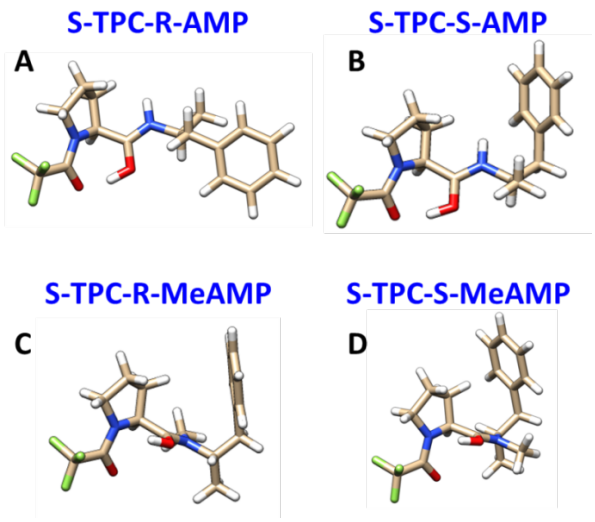


Figure 3.8: Global minimum structures generated at the B3LYP/6-311++G(d,p) level of theory for (A) *S*-TPC-*R*-AMP, (B) *S*-TPC-*S*-AMP, (C) *S*-TPC-*R*-MeAMP and (D) *S*-TPC-*S*-MeAMP.

ACN-ion complexes were generated using the unsolvated compounds as a starting point, resulting in the creation of 10 *S*-TPC-*R*-AMP, 12 *S*-TPC-*S*-AMP, 2 *S*-TPC-*R*-MeAMP and 9 *S*-TPC-*S*-MeAMP structures for an overall total of 33 candidate structures. The global minimum structures that were generated for *S*-TPC-*R*-AMP and *S*-TPC-*S*-AMP had solvent binding energies of 7.43 kcal/mol and 7.91 kcal/mol, respectively while those generated for *S*-TPC-*R*-MeAMP and *S*-TPC-*S*-MeAMP had solvent binding energies of 8.09 kcal/mol and 8.32 kcal/mol, respectively.(See Figure 3.9 below) While the differences between these BEs are small they still should be seen as significant. Error in the binding energies should be in the range of 0.25-0.5 kcal/mol, though due to the similar structures of the isomers the errors should occur in the same direction and reduce the relative error observed in the calculations. The TPC-MeAMP isomers both share the same location for the ACN-ion interactions, with the ACN molecule interacting with the proton that is engaged in an OH•••N H-bonding with the ring nitrogen. Very little structural change was observed in the ion upon the addition of ACN. The ordering of the binding energies agreed with the experimental results; the more strongly bound cluster (*S*-TPC-*S*-MeAMP) is found to have the more negative peak CV value of the diastereomer pair. Conversely, the TPC-AMP diastereomers did not share the same location for their interactions with the ACN molecule, nor the same H-bonding motifs. *S*-

TPC-*R*-AMP contains an OH•••O motif and the ACN interacts with the hydrogen bonded to by the chain nitrogen, while *S*-TPC-*S*-AMP exhibits an NH•••O motif and the ACN clusters with the protonated carbonyl group. In either case, the resulting structures draw the phenyl ring close to the area of positive charge surrounding the IMHB, bringing its face almost parallel with that of the five membered ring. These compact structures are each more favorable than the alternative, where the phenyl ring is located on the opposite side of the backbone chain of the structure and cannot compact itself as well, nor can it access the positively charged areas due to the curve in the backbone caused by the IMHB. Just like for the *S*-TPC-MeAMP diastereomer pair, the more strongly bound *S*-TPC-AMP diastereomer (*S*-TPC-*S*-AMP) produced the more negative peak CV value in DMS experiments. Note that the argument does not hold between TPC-AMP and TPC-MeAMP as peak CV values were less negative for the more strongly bound MeAMP diastereomers. This highlights the complex behavior underlying observed differential mobility and the need for further study. An explanation for this discrepancy could come from the size differences in *S*-TPC-AMP versus *S*-TPC-MeAMP. Due to the larger size of the *S*-TPC-MeAMP species, the ΔCCS between bare ions versus ion-solvent clusters should be lower relative to the *S*-TPC-AMP species, resulting in a less negative overall CV shift.

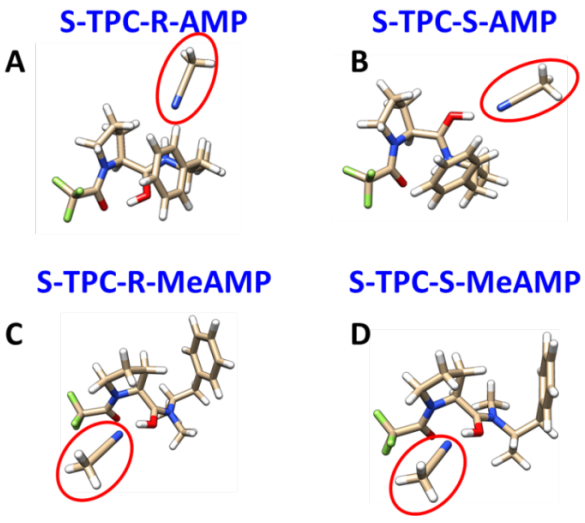


Figure 3.9: Global minimum structures for single ACN molecule – analyte clusters of (A) *S*-TPC-*R*-MeAMP, (B) *S*-TPC-*S*-MeAMP, (C) *S*-TPC-*R*-AMP and (D) *S*-TPC-*S*-AMP. The ACN molecules have been highlighted by red circles.

3.5 Conclusion:

Nitrogen only DMS experiments on AMP and MeAMP were unable to separate their chiral pairs, so a derivatization reaction with *S*-TPC was used to make the compounds easier to resolve. Using nitrogen modifier these diastereomers were not able to be separated until a 10 psi throttle gas was applied within the DMS. However, the TPC-MeAMP isomer pair was indistinguishable based upon a theoretical prediction of CCS for the two diastereomers. Therefore, a different experimental method was followed, in which DMS experiments were performed using ACN as a gas modifier. It was demonstrated that the *S*-TPC-*S* pairs for AMP and MeAMP displayed stronger Type-A (more negative CVs) DMS behavior then the *S*-TPC-*R* pairs. Computational results confirm that the stronger solvent binding *S*-TPC-*S* pairs demonstrate stronger Type-A behavior as compared with the S-TPC-R variants. While this does allow for some understanding of the separations between the diastereomer pairs, in which most properties are similar if not equal, the overall trend is not predicted for all the compounds. Following the order of peaks CVs from most positive to most negative the trend is *S*-TPC-*R*-MeAMP, *S*-TPC-*S*-MeAMP, *S*-TPC-*R*-AMP, *S*-TPC-*S*-AMP, while the order of BE from smallest to largest: *S*-TPC-*R*-AMP, *S*-TPC-*S*-AMP, *S*-TPC-*R*-MeAMP, *S*-TPC-*S*-MeAMP. Although the calculated BEs for the MeAMP derivatives are greater than those calculated for the AMP derivatives, the AMP derivatives show significantly more negative peak CV values. This demonstrates that DMS behavior is a complex mixture of ion-solvent CCS and bare ion CCS which depends on ion-solvent interactions, which BE trends only can capture for similarly structured species.

Chapter 4 : Separation of Ephedrine Diastereomers and Sulfonamide Structural Isomers Using Differential Mobility Spectrometry

Experimental work for this chapter was performed within the lab of Prof. Gérard Hopfgartner at the University of Geneva. All other work included in this chapter is my own.

Reference: Ruskic, D.; Hopfgartner, G. Modifier Selectivity Effect on Differential Ion Mobility Resolution of Isomeric Drugs and Multidimensional Liquid Chromatography Ion Mobility Analysis. *Anal. Chem.* **2019**. *91*(18). 11670-11677. https://doi.org/10.1021/acs.analchem.9b02212

4.1 Introduction:

Ephedrine and pseudoephedrine are a pair of diastereomers (Figure 4.1) that share a variety of medical properties, including diaphoretic, antipyretic, respiratory, antitussive and antiasthmatic effects.[78] Both compounds have two chiral centers: in ephedrine there are the 1*R*,2*S* (−) or 1*S*,2*R* (+) chiral configurations (*i.e.* the hydroxyl and methyl groups face the same directions) and in pseudoephedrine there are the 1*R*,2*R* (−) and 1*S*,2*S* (+) configurations (*i..e.*, the hydroxyl and methyl groups face opposite directions). Both ephedrine and pseudoephedrine find use in Chinese traditional medicine as an active ingredient in the plant *Ephedra sinica*, also known as "Ma-Huang".[79] Additionally from the perspective of, law enforcement, (−)-ephedrine and (+)-pseudoephedrine are commonly used as precursors in the creation of illicit methamphetamine, which differs only by the substitution of the hydroxyl group for a hydrogen atom.[80] A concern is the false identification of methamphetamine usage in urine sampling, which can be caused by overconsumption of ephedrine and/or pseudoephedrine. In such cases the correct identification between can only be provided by the measurement of the enantiomeric composition; methamphetamine usage will show majority (−)-ephedrine and (+)-pseudoephedrine while medical usage should show a more even distribution of isomers.[80]

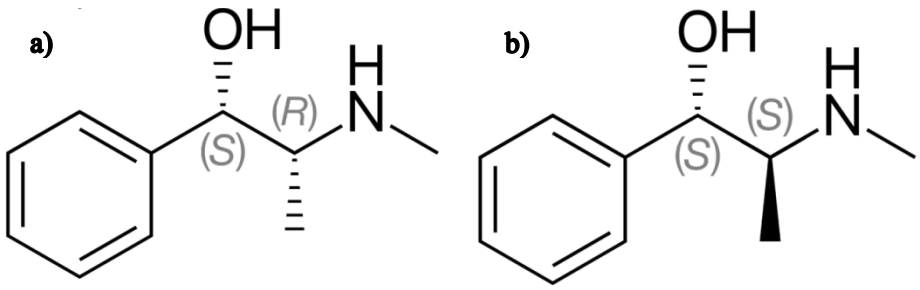


Figure 4.1: Structural representations of a) (+)-ephedrine (1S,2R) and b) (+)-pseudoephedrine (1S,2S) labeled with their chiralities as per Cahn-Ingold-Prelog (CIP) convention.

Ephedrine and pseudoephedrine have previously been examined with a variety of experimental techniques. McCooeye *et al.*[81] used FAIMS in 2003 to separate ephedrine and pseudoephedrine from each other as well as several associated metabolites that were obtained from weight loss products. Specific enantiomeric designations were not given in this case. More recently in 2012, Holness *et al.*[64] used IMS with achiral solvent modifiers in order to separate (−)-ephedrine and (+)-pseudoephedrine, showing a 1.34% resolution between these diastereomers when using *n*-octanol as the modifier gas. Other previous experimental techniques include GC-MS/MS[82], LC-MS/MS[83], CE[78] and high performance liquid chromatography (HPLC).[84]

Sulfonamides are another interesting group of compounds that contain many different isomers. They are predominately used in veterinary medicine for the treatment and prevention of bacterial infections, and as additives to animal feeds to assist in growth.[85–87] Sulfonamides are derived from sulfanilamide (Figure 4.2), by replacing the one of the hydrogen atoms on the sulfonamide nitrogen with another group, often featuring a nitrogen containing ring (See Figure 4.2 for an example of a sulfonamide). One area of concern is that the half-life of sulfonamides within animal tissue is long enough that they are still likely present when the meat is prepared for human consumption. Additionally, sulfonamides have been shown to appear in animal byproducts like milk[85] and honey,[88] as well as the runoff water from agricultural operations[89] which has the potential for creating antibiotic resistances to the sulfonamides Furthermore, some of these drugs (e.g., sulfadimidine) have been shown to potentially cause cancer and allergic

reactions in humans.[85] Many countries have placed restrictions on the amount of sulfonamides that can be present as contaminants in animal-derived foods, with sulfadimidine and sulfadimethoxine in particular having stricter limits. As different sulfonamides have different limits it is important to be able to distinguish between different sulfonamides, particularly those that share the same m/z value, which can cause false positive detection for the more strictly regulated compounds.

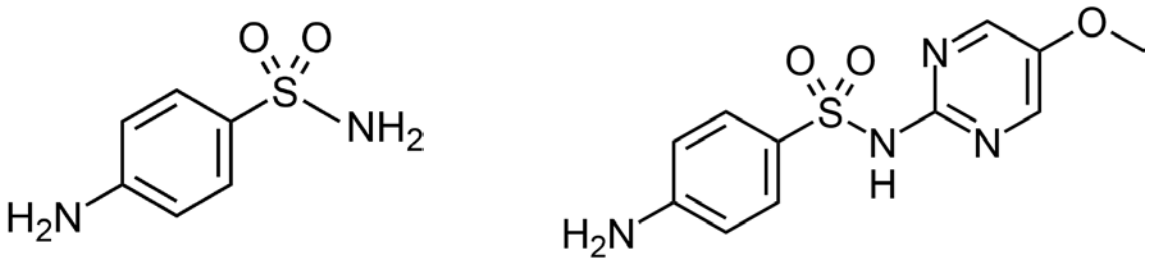


Figure 4.2: A structural representation of sulfanilamide is shown at left. A structural representation of a sulfonamide, sulfameter, is depicted at right.

Various experimental techniques have been shown to have the capability to separate different sulfonamides. Cai *et al.*[90] showed simultaneous detection of twenty-four sulfonamide residues from meat samples using ultra-performance liquid chromatography and tandem mass spectrometer. (UPLC-MS/MS). Berger and Berger[91] used ultra-high performance supercritical fluid chromatography (UHPSFC) in order to quickly separate nine sulfonamides and test the feasibility of detecting these drugs in a milk matrix. While no studies have been done explicitly on the separation of sulfonamides using IMS techniques, traveling wave ion mobility spectrometry (TWIMS) has been used to determine the collision cross section (CCS) of ten sulfonamide compounds in a larger study of human and veterinary drugs.[92] Other techniques used include HLPC[89,93], HPLC with diode ray ultraviolet detection[94], LC-MS/MS[88], and ultra-high performance liquid chromatography tandem mass spectrometry (UHPLC-MS/MS).[86]

The following chapter of this thesis will aim to expand upon the work described in Chapter 3, with experimental DMS data and computational binding energies and CCSs for the diastereomers (+)-ephedrine and (+)-pseudoephedrine, as well as three isomeric groups of sulfonamides. DMS data is

obtained alongside nine modifier gases. Through combining the results from this chapter with previous results, we obtain a greater insight into the behavior of ions within the DMS cell and further show the value of DMS as an analytical tool for the separation of isomeric compounds. Additionally, work will be conducted to follow up on the ML study of Walker *et al.*[1] In that study connections were shown between DMS behavior as well as these properties. Applying the CCS and BE models from this previous work to new compounds will provide a test for the effectiveness of the ML models. Additionally, a second test will be conducted for each property involving the introduction of several of the testing compounds into the training set. The effects of this inclusion on the efficacy of the ML model will then be observed.

4.2 Experimental Methods:

The compounds (+)-ephedrine (EPH), (+)-pseudoephedrine (PsEPH), sulfadimethoxine, sulfadimidine sulfadoxine, sulfameter, sulfamethoxypyridazine, sulfamonomethoxine, and sulfisomidine were chosen for analysis in this experiment. EPH and PsEPH are shown in Figure 4.1 and the sulfonamide structures are shown in Figure 4.3, below.

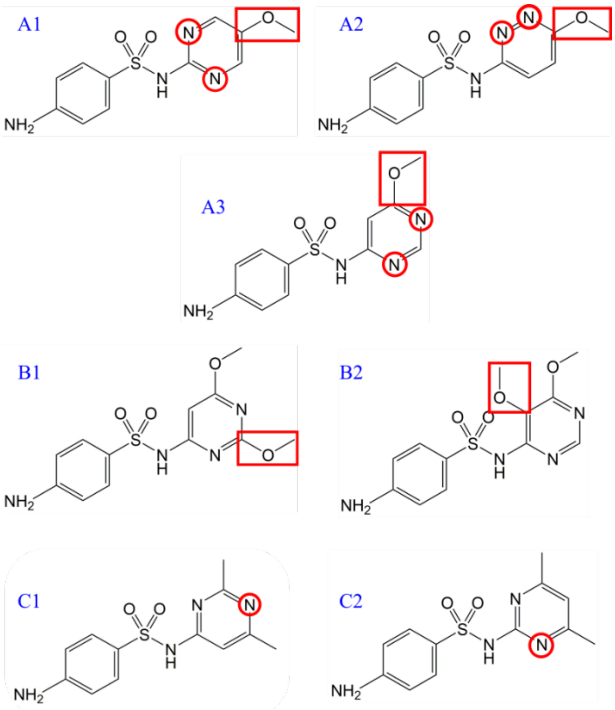


Figure 4.3: Structures of sulfonamides used in this study: A1: sulfameter A2: sulfamethoxypyridazine A3: sulfamonomethoxine B1: sulfadimidine B2: sulfadoxine C1: sulfadimethoxine C2: sulfisomidine. Structural

isomers are given the one letter designation (A,B,C) and a numerical designation. Red outlines indicate the functional groups that are being moved in each grouping.

A commercial SelexION QTRAP 5500 instrument available from SCIEX was used in order to perform the following experiments. Default DMS parameters were used unless otherwise stated. Experiments were carried out separately for each analyte using EPI mode by the Hopfgartner group at the University of Geneva. Compounds were sorted into groups by their m/z values in order to compare the behavior of the different isomers. The diastereomers (+)-ephedrine and (+)-pseudoephedrine formed one group, while the sulfonamides were broken into three groups of positional isomers: A, B, and C. Group A (see Figure 4.3) is comprised of sulfameter (A1), sulfamethoxypyridazine (A2), and sulfamonomethoxine (A3) which vary by the position of the ring nitrogens and the nearby methoxy group. Group B is comprised of sulfadimidine (B1) and sulfadoxine (B2) which differ by the position of a methoxy group which moves from the meta position to the ortho position on the pyrimidine ring relative to the sulfonamide group. Group C is comprised of sulfadimine (C1) and sulfisomidine (C2) which differ by the ortho versus para position of one ring nitrogen on the pyrimidine ring relative to the sulfonamide group. A variety of different gas modifiers were tested for their suitability in separating analytes: including H_2O, methanol (MeOH), ethanol (EtOH), propanol (PrOH), isopropyl alcohol (IPA), ethyl acetate (AcOEt), acetone (ACE) and toluene (Tol).

4.3 Computational Methods:

As in Chapter 3, candidate structures were generated by optimization of neutral guess structures at the PM7 level of theory, followed by the generation of protomers at the B3LYP/6-311++G(d,p) level of theory. Basin hopping was then employed in order to generate global minimum candidate structures using UFF for each analyte, and are reoptimized and the frequencies calculated at the B3LYP/6-311++G(d,p) level of theory. CCSs were computed using MobCal-MPI as discussed previously. See Section 3.3 for further details on this process. In this chapter the structures were identified from their Gibbs energies at 100 °C, rather than at room temperature, in order to more accurately reflect the conditions within the DMS cell.[95] For a DMS temperature of 150 °C for molecules at or near the edges of the DMS cell should

be around 150 °C, as this is where the temperature is being applied, however 100 °C has been shown to give a more accurate representation of the temperature of the along the transmission axis through the DMS cell into the MS.[95] Singly solvated structures were computed using structures for each of the modifier gases that were employed for experiments. A thorough investigation of each potential solvation site was first performed for H_2O using basin hopping. In order to reduce the number of subsequent calculations, it was assumed that the other species capable of hydrogen bonding (MeOH, EtOH, PrOH, and IPA) would share similar sites for solvation as H_2O. The solvent species which cannot undergo the same degree of hydrogen bonding (ACE, AcOEt and Tol) were individually investigated to determine their optimal solvated structures. ACE and AcOEt can each accept a hydrogen bond from a protonated group; however, neither molecule can donate a hydrogen bond like H_2O or the other OH containing modifiers; while Tol is unable to form any hydrogen bonds and relies on charge/dipole interactions. Gibbs energies for solvated structures were also calculated at 100 °C to match the bare ions.

In order to examine the general applicability of the machine learning work by Walker *et al.*[1] to other compounds outside the training set, a model is constructed from their work which includes N_2, H_2O and MeOH modifier DMS data, calculated CCS, and ion-solvent modifier binding energies. A random forest algorithm within the Orange python package[39] has been chosen for the learner in this work due to its previous successes in predicting CCS and other properties.[1] Hyperparameters were chosen to minimize fitting errors, 500 decision trees were used to generate the RF algorithm, and each node was allowed to split the database into groups as small as two entries. Other hyperparameters for the RF learner were chosen by Orange based upon the number of data entries and feature space. For CCS fits features were chosen as their m/z values along with a DMS experimental data represented by CV peak values at SV = 2000 V, 2500 V, 3000 V, 3500 V, and 4000 V, as well as the experimental temperature. In the case of BEs m/z and CCS are used as features, as well as DMS data for SV=0 V-4000 V in 500 V increments. BEs used in the ML model were calculated as zero-point corrected energies in order to match with the BEs used by Walker *et al.*[1] Each model is evaluated using 10-fold cross validation and also by applying

the model to a test set of compounds not included in model training. Specifically, the test set in this case is comprised of ephedrine, pseudoephedrine, and the sulfonamides introduced in this chapter. After the model evaluation, a few of the test set of compounds will be moved to the training set in order to expand the applicability of the ML model. This new training set will then be tested against the remaining test set of molecules in order to evaluate the effects on the ML model's predictive abilities.

4.4 Results and Discussion

4.4.1 Pure N_2 Experiments

Prior to using any of the modifier gasses, each analyte was examined in a pure N_2 environment within the DMS. Dispersion plots for EPH, PsEPH and each of the different sulfonamides (Figure 4.3) are shown below in Figure 4.4. While there are small differences in optimal CV for transmission of specific isomers at high separation voltages, separation between isomers as measured from ΔCV is always less than the error as would be expected from the half width at half maximum (HWHM) of the ionogram. Thus, the isomeric systems under study cannot be separated using a pure N_2 environment.

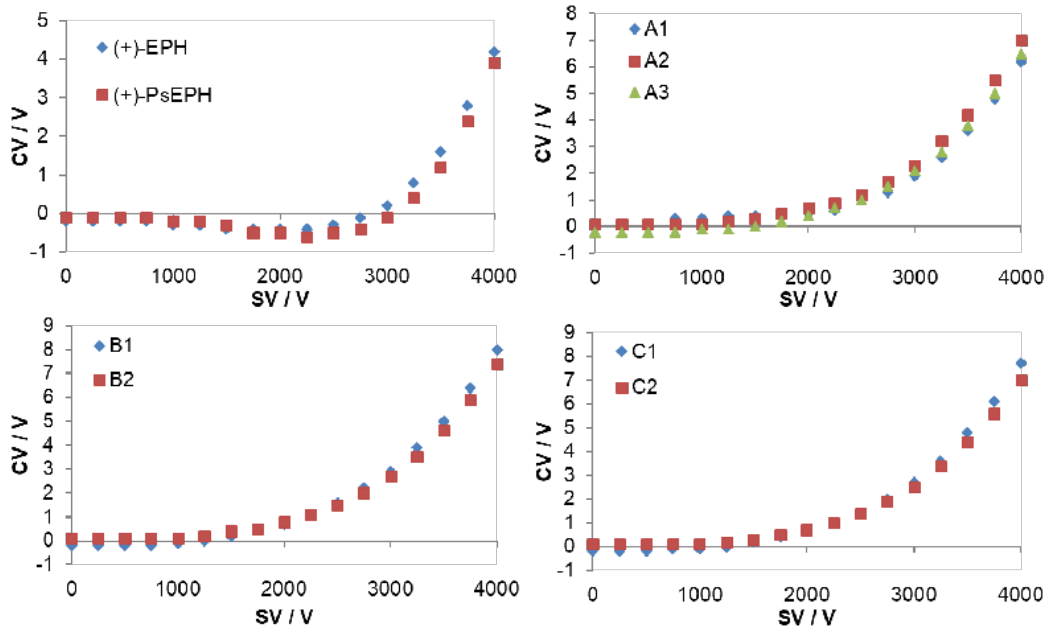


Figure 4.4: Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with no gas modifier (pure N_2). CV was obtained from a fit to the ionogram of specific SV values.

In IMS ions separate within a N_2 environment based upon their relative mobilities, which from the Mason-Schamp equation (Eq. 4.1) is inversely proportional to their CCS.

$$K(0) = \frac{3}{16}\left(\frac{2\pi}{\mu k_B T}\right)^{1/2} \frac{ze}{N\Omega(T)} \qquad \text{Eq. 4.1}$$

In Eq. 4.1, $K(0)$ is the zero-field mobility, μ is the ion-buffer reduced mass, k_B is the Boltzman constant, T is temperature, ze is the ion charge, N is the number density, and $\Omega(T)$ is the ion CCS. In DMS the CCS has been shown to behave dynamically as it moves through the cell. Experiments performed by Iertiano *et al.*[96] and Haack *et al.*[95] have shown the effect fluctuating temperature and alternating separation voltage has upon the CCS. Temperature influences the population of different isomer structures for each analyte which can allow for separation as field induced heating causes the population of ions to change, resulting in an average CCS that can differ in high versus low field. If CCS does not change appreciably between high versus mobilities, K will still differ due to changes in ion-buffer interactions under different field conditions. Change in mobility with field is commonly represented with Eq. 2.5. In a pure N_2 environment (no solvent-ion clusters) α will strongly depend on the ion CCS, with a larger CCS requiring a more positive CV value for detection to occur and therefore ions experience hard sphere (Type C) behavior. As stated previously, CCSs were computed using the MobCal-MPI software.[35] See Table 4.1 for summary comparison of DMS measure CV values at SV = 4000 V and ion CCS. In each of the isomer groups the differences separation between calculated CCSs was negligible, with all differences falling within the standard deviation in the MobCal-MPI calculations. This ultimately agrees with the experimental DMS results that showed little to no separation between the compounds within each group.

Group	Compound	CV / V	CCS / Å^2	Std Dev. / Å^2
EPH/PsEPH	EPH	4.2	133.4	1.64
	PsEPH	3.9	134.0	1.23
A	A1	6.2	161.2	1.85
	A2	7.0	159.6	2.06
	A3	6.5	161.6	2.53
B	B1	8.0	170.5	2.33
	B2	7.4	169.3	1.67
C	C1	7.7	163.9	1.85
	C2	7.0	163.9	1.78

Table 4.1: Summary of behavior in pure N_2 environment. Compound names are given in Figure 4.3. Compensation voltages are taken at a SV of 400 V as shown for each compound as well as their MobCal-MPI determined CCS values in Å^2 along with the standard deviation in Å^2.

In order to build a ML model for CCS prediction, the training set from Walker *et al.*[1] was used and tested using 10-fold cross validation method with a random forest algorithm as the learner. In Figure 4.5 the results of the RF prediction are shown plotted alongside the MobCal-MPI CCS values. A linear fit shows an R^2 value of 0.951, which indicates good performance for the model. Additionally, the root mean square error (RMSE) is 6.443 Å^2 and the mean absolute error (MAE) is 4.793 Å^2 across the complete molecular set. The rank module, available in orange, also allows us to determine which features have the largest impact on the fit. For the case of CCS fitting m/z values for each compound was the most important feature. This connection makes intuitive sense, as the larger molecules will also have larger m/z and vice versa. In fact, a preliminary investigation showed that CCS can be predicted using only m/z with an R^2 for the fit of 0.817. The values for predicted CCS were then influenced by the compounds' DMS behavior, most prominently the CV at SV = 2000 V and SV = 4000 V, and lastly by the temperature of the DMS experiment, which had the least influence on the ML fit.

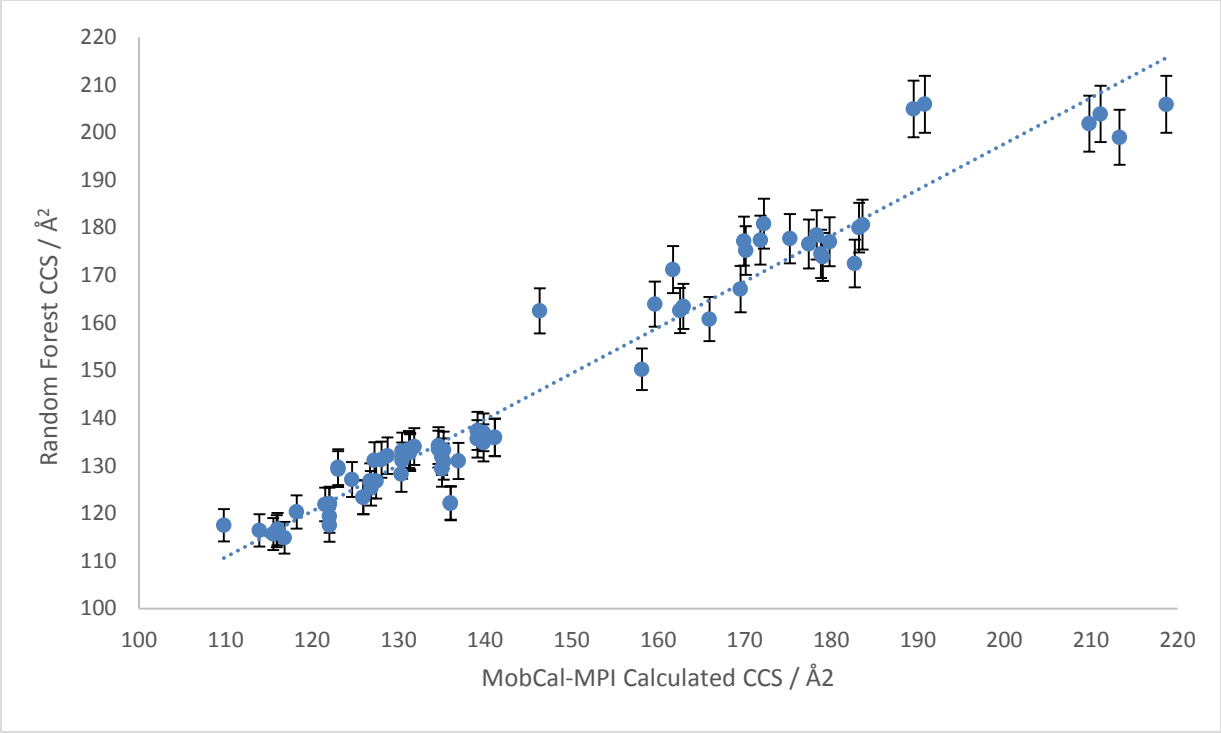


Figure 4.5: Plot of MobCal-MPI calculated CCS vs RF predicted CCS (in Å^2). The RF fit to experimental data gathered from the ML work of Walker *et al.*[1] shows strong correlation between predicted and calculated values when employing 10-fold cross validation for CCS prediction. This plot displays an R^2 = 0.951 as well as a MAE of 4.793 Å^2 and a percent error of 2.90%.

The model was then used to test on the EPH and PsEPH diastereomer pair, along with the sulfonamides examined in this chapter. (Figure 4.6a). The R^2 for prediction was 0.867 for this test set, with a RMSE of 4.801 Å^2 and a MAE of 4.203 Å^2. While this result is promising, it should be further improved by the introduction of new chemical functionalities into the training set which were not fully represented by the original Walker *et al.*[1] data set. By moving three compounds, PsEPH, A3 and B2 from the test set into the training set, the accuracy of the training set changed negligibly (R^2= 0.944, RMSE = 6.374 Å^2, MAE = 4.72 Å^2), however the prediction of the test set jumped from an R^2 of 0.867 to 0.946, approximately equal to that of the training set. (Figure 4.6b) Additionally, the RMSE of for test set prediction fell to 2.708 Å^2 and the MAE to 2.18 Å^2, close to half of the previous values. PsEPH, A3 and B2 were chosen to be moved into the training set as samples to cover the full range of CCS observed within the testing set.

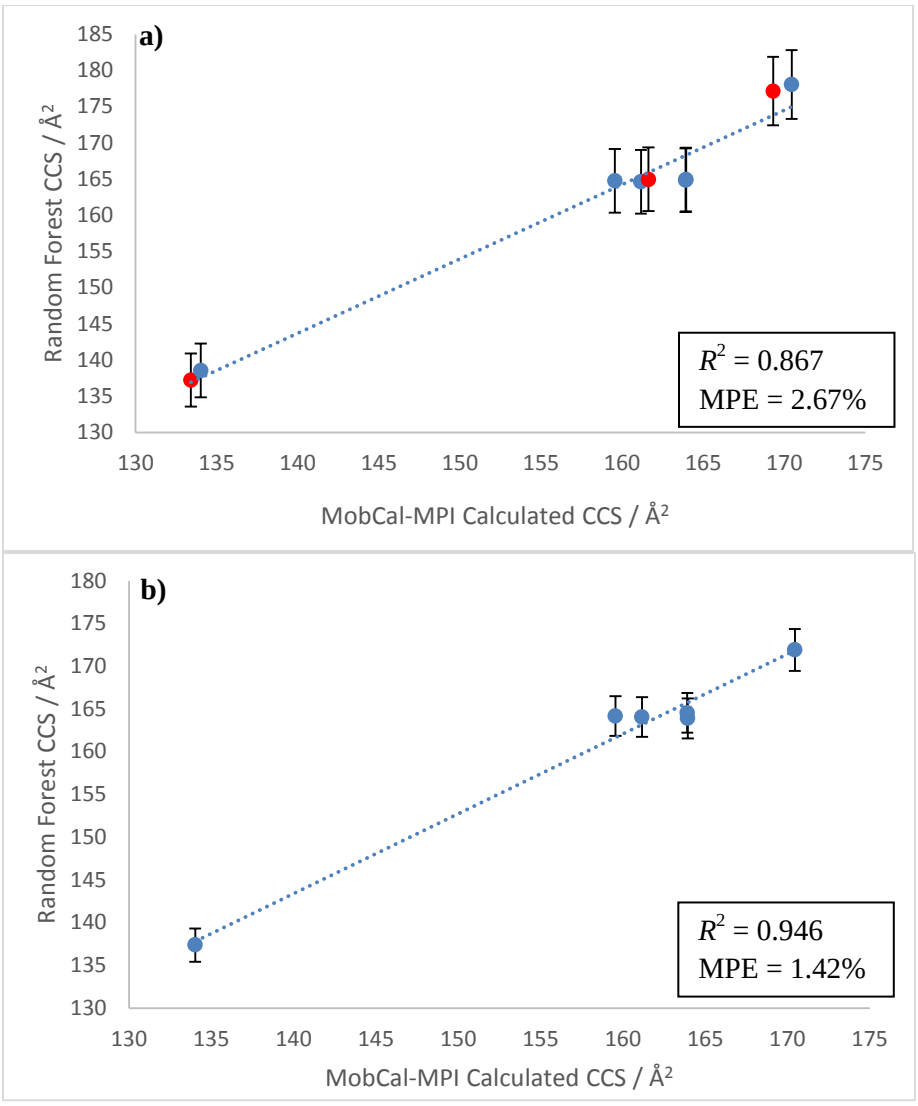


Figure 4.6: Plots comparing calculated and predicted CCS for: a) the test set containing EPH/PsEPH and the sulfonamides and b) the test set containing the remaining compounds after three compounds, EPH, A3 and B2 were moved to the training set. The compounds moved from the testing set to the training set are denoted by red markers in plot a). Notably, B1 seems to have decreased its RF CCS noticeably after the introduction of B2 to the training set. The original training set showed an R^2 of 0.867 and MAE of 4.203 Å^2 which was improved to an R^2 of 0.946 and MAE of 2.180 Å^2. This corresponds with a percent error of 2.67% for the original test set vs. 1.42% for the augmented test set.

While this result is promising for the addition of new compounds into the training set of the ML model, it must also be regarded carefully. Due to the small number of compounds being examined, removing a few data points from the test set can have strong effects on the criteria used for evaluating performance, such as the individual errors generated for the ML fit. As RMSE and MAE are used to evaluate results as

opposed to individual errors, the model training and testing is still relatively insensitive to seize but more test points are desired for confidence.

4.4.2 Experiments with Gas Modifiers

In order to achieve improved DMS resolution as compared with pure N_2 experiments, several different gas modifiers were seeded into the DMS chamber. Modifiers for which type B (weak binding) behavior was observed, such as H_2O, MeOH, and Tol, showed mixed results in separating the isomer groups. While toluene was able to resolve all isomer groups successfully, H_2O failed to resolve isomers A2 and A3 from each other as well either in group D, and MeOH modifier also failed to resolve isomers A2 and A3. The modifier gases that displayed mainly Type A (strong clustering) DMS behavior; EtOH, PrOH, IPA, ACE, and AcOEt proved to be yield larger separations. In only a few cases did the CV shifts remain similar between isomers when using these modifiers (around ~1V separation), such as in the case of EPH/PsEPH for PrOH modifier, and A2/A3 for IPA modifier. See below for examples of dispersion plots for modifier experiments using MeOH (Figure 4.7), Tol (Figure 4.8), and ACE (Figure 4.9).

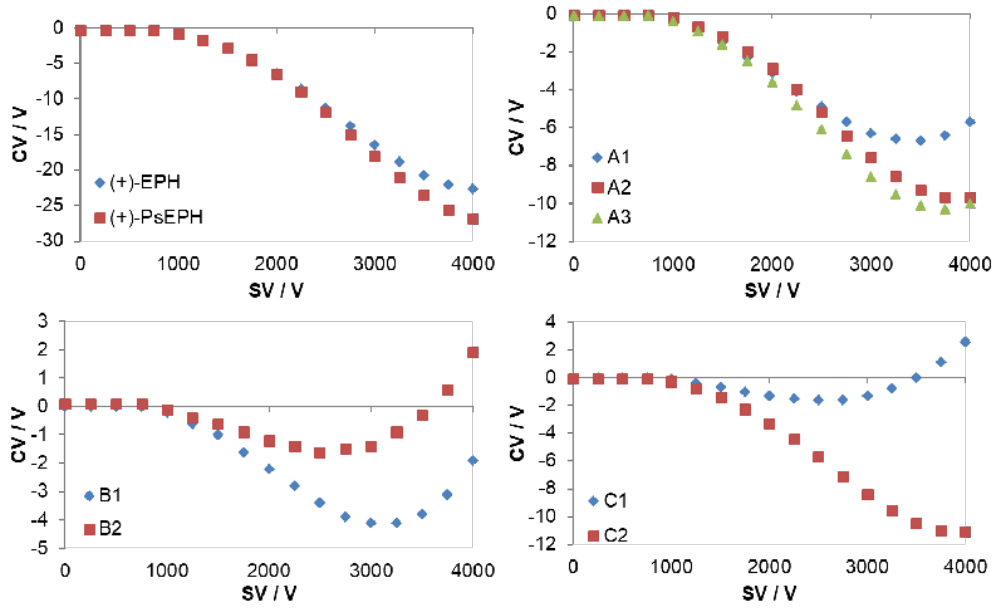


Figure 4.7: Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with MeOH gas modifier. CV was obtained from a fit to the ionogram of specific SV values.

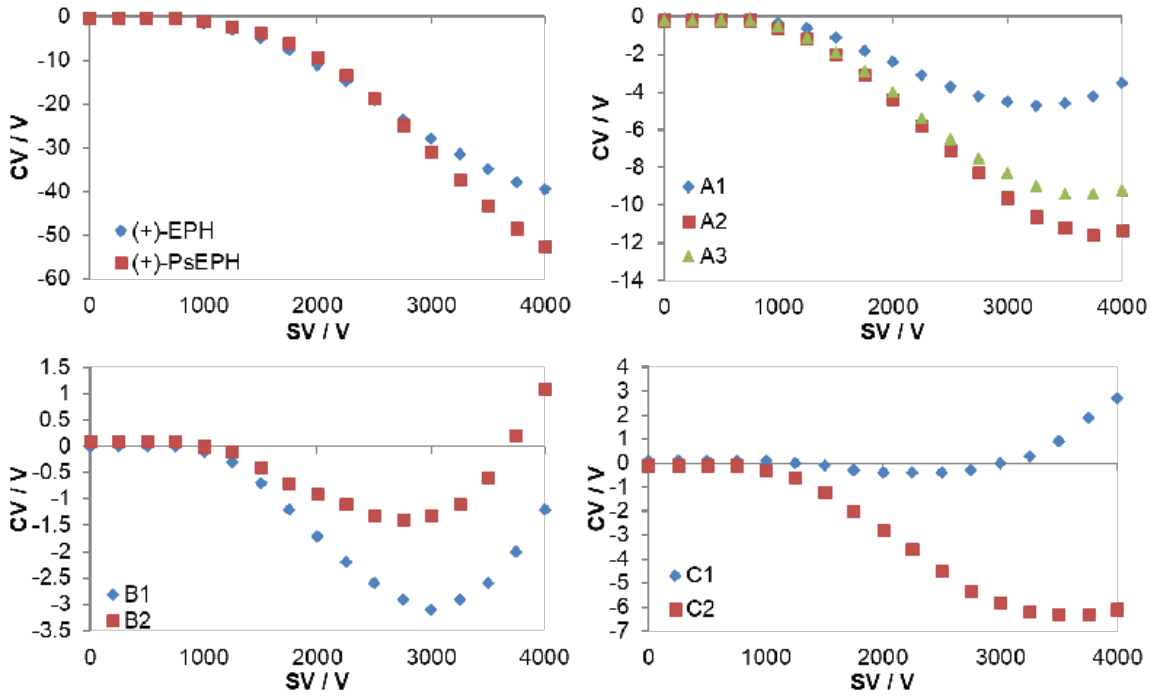


Figure 4.8: Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with Tol gas modifier. CV was obtained from a fit to the ionogram of specific SV values.

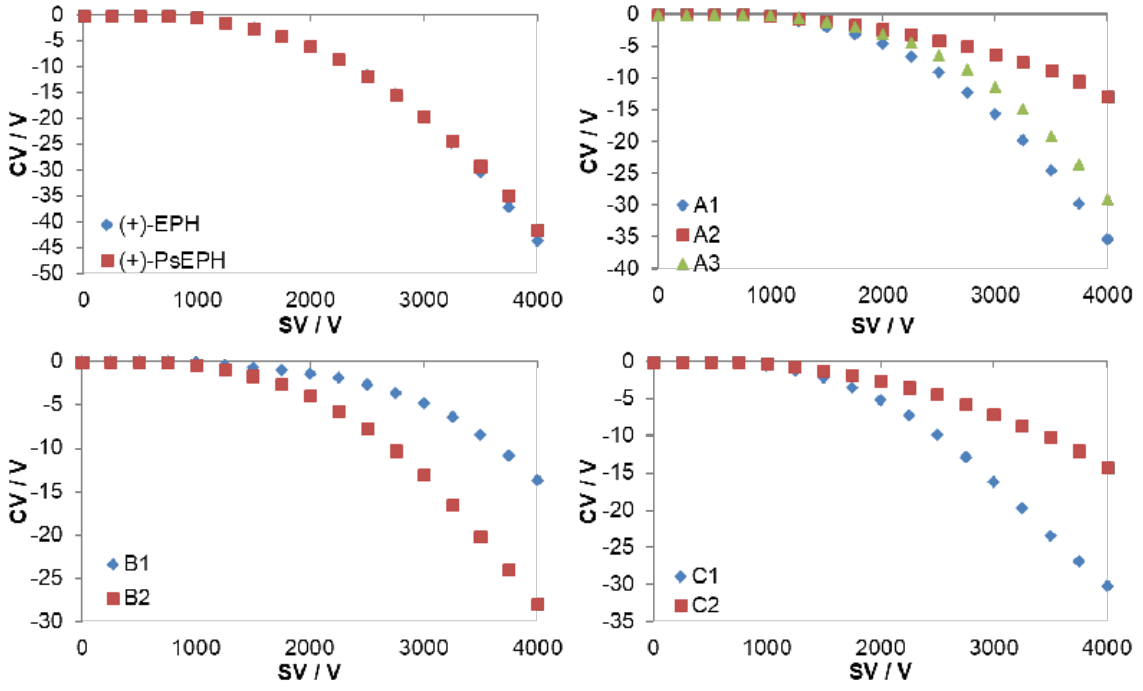


Figure 4.9: Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with ACE gas modifier. CV was obtained from a fit to the ionogram of specific SV values.

Differences in DMS behavior between pure N_2 and modifier experiments are presumed to be due to ion-solvent interactions.[95] In an effort to predict DMS behavior using computational properties, calculated single solvent binding energies were compared to experimental DMS results. In previous work the calculated binding energy has been shown to correlate with the SV value at CV_{min} for type B dispersion plots in quinoline-based drugs by Liu *et al.*[22,97] This correlation tested for some of the dispersion plots in Figure 4.7 and Figure 4.8, but is less readily employed for the Type A curves shown in Figure 4.9.

For Tol this relationship between ion-solvent BE and SV at CV_{min} is qualitatively observed. In each of the four plots shown in Figure 4.8, the ordering of the SV at CV_{min} observed in experiment matched with that predicted by the computational binding energies. EPH and PsEPH show binding energies with Tol of 1.43 kcal/mol and 4.06 kcal/mol, respectively, with local minimum around SV = 4000 V for EPH and outside the DMS range (SV > 4000 V) for PsEPH. Group A provided the most ambiguous results as each show BEs with Tol (A1: 1.10 kcal/mol, A2: 1.70 kcal/mol, A3: 1.34 kcal/mol) though they all show the expected trend with SV at CV_{min} values: SV at CV_{min} = 3250 V for A1, SV at CV_{min} = 3750 V for A2 and SV at CV_{min} = 3625 V for A3. Group B placed B1 (1.10 kcal/mol) as more strongly binding than B2 (−1.36 kcal/mol) which accorded with their minima at ~3000 V and ~2750 V. In group C, C2 was predicted to be more strongly binding than C1 (−0.29 kcal/mol vs −1.53 kcal/mol BE) which agreed with experimental minima. (C2: 3750 V C1: ~2250 V) It is intuitive that Tol would correlate well with single solvent binding energies due to the bulkiness of Tol relative to some of the other solvents used in this work. Due to its size a single molecule of Tol may prevent the interaction of other Tol molecules with suitable sites on the ion. This comes in contrast with MeOH and H_2O, which may be better described using multiple solvent-ion binding energies. The negative BEs observed in the latter two groups are indicative that ion solvent binding has become energetically unfavorable. While the associated dispersion plot traces are typically closer to Type C behavior than the traces with positive BEs, Type B DMS behavior is still observed in each case. It should be noted that although the expected SV at CV_{min} is well followed between similar compounds it is not generally followed between groups.

The other modifier gases which showed primarily Type B DMS behavior did not fare as well as Tol in these comparisons. The SV at CV_{min} for H_2O experiments successfully predicted the behavior in two groups, namely the separation of A1 (SV at CV_{min} = 2500 V) from A2 (SV at CV_{min} = 2750 V) and A3 (SV at CV_{min} = 2750 V) which corresponded with BEs of −0.07 kcal/mol, 2.82 kcal/mol and 2.10 kcal/mol, respectively. A1 versus A2/A3 shows the expected SV at CV_{min} trend but no A2/A3 separation is observed. BEs for B1 (1.55 kcal/mol) and B2 (1.27 kcal/mol) are similar and result in a small degree of separation in their minima with SV at CV_{min} of ~2000 V and ~1875 V, respectively. H_2O experiments with EPH and PsEPH did not align with the calculated H_2O ion binding energies with BEs of 4.64 kcal/mol and 4.57 kcal/mol and SV at CV_{min} of 3500 V for EPH and 3750 V for PsEPH. MeOH modifier experiments matched the expected trend with binding energies for one group in four, successfully predicting the separation of C1 (SV at CV_{min} = ~2625 V) and C2 (SV at CV_{min} = ~4000 V) with BEs of 1.43 kcal/mol and 4.16 kcal/mol. MeOH experiments with group B were less successful with B1 having a BE of 2.29 kcal/mol and B2 having a BE of 2.70 kcal/mol with corresponding SV at CV_{min} of ~3125 V and 2500 V, respectively, which run contrary to the expected trend. This demonstrates that SV at CV_{min} can be correlated with ion-solvent binding energy for cases where single solvent ion binding is predominant but does not work as well when multiple sites of solvation are available to smaller solvents that will not interfere with each other's binding.

Additionally, the SV at CV_{min} is difficult to predict for BEs of strongly binding modifiers. This is because for type A behavior the SV at CV_{min} needs to be predicted from extrapolation past the experimental limit of SV = 4000 V. Separation in CV space for SV = 4000 V may also correlate with binding energy, see Table 4.2. In the case of EPH and PsEPH, the small degree of observed separation seemed to match with calculated binding energies for each of the strongly binding modifiers except for ACE modifier, where EPH-ACE and PsEPH-ACE BEs varied significantly as 5.78 kcal/mol and 7.84 kcal/mol respectively, while no distinction was observed in ACE experiment any correlation with CV at SV = 4000 V. Results for group A did not show any correlation with CV at SV = 4000 V. Group B, similar to group A, shows

no correlation of BEs with experimental results in EtOH, PrOH and ACE. A sizable difference in binding energy was observed when clustering AcOEt modifier (6.12 kcal/mol for B1 vs 3.47 kcal/mol for B2), which matches with separation observed in experiment. Group C showed the highest degree of agreement between experimental and computational results, with largely different BEs corresponding to large differences in CV at SV = 4000 V although in ACoET and ACE the trend is reversed relative to the other modifiers.(Table 4.2)

The mixed results in the prediction of experimental DMS behavior from computational binding energies can partially be explained by several factors. Firstly, it was initially assumed that single solvent cluster binding energies would be able to describe the interactions that are occurring within the DMS cell. This is not necessarily the case, and the clustering/declustering process could be occurring with multiple solvent molecules, or from single solvent clustering/declustering to a strongly bound ion-solvent complex. This is especially possible with smaller solvent molecules that would not interfere with other points of interaction on the analyte. Further investigation should be conducted to determine if this is the case. Additionally, in the cases of MeOH, EtOH, PrOH, and IPA, not all possible sites of solvation were investigated. It was assumed that these compounds would simply solvate the same site as water due to their similar ability to form H-bonds. However, with the given results, further investigation is warranted to determine if different sites of solvation can better capture the DMS results. It is worth noting that Haack *et al.*[95] show that single solvent binding is dominant for molecules studied therein and we expect similar results in this study.

Modifier		EtOH		PrOH		IPA	
Group	Compound	CV / V	BE (kcal/mol)	CV / V	BE (kcal/mol)	CV / V	BE (kcal/mol)
EPH/PsEPH	EPH	−44.8	6.52	−60.1	6.68	−57.6	11.23
	PsEPH	−46.5	6.09	−59.1	6.96	−55.8	10.54
A	A1	−19.6	4.30	−34.7	4.77	−37.4	7.70
	A2	−22.5	4.85	−31.8	5.10	−29.5	8.96
	A3	−25.1	3.65	−37.8	4.01	−36.3	7.89
B	B1	−16.3	3.63	−26.5	3.20	−25.6	7.22
	B2	−7.1	3.58	−20.1	3.22	−20.0	6.36
C	C1	−7.5	1.22	−19.5	2.74	−19.4	5.11
	C2	−24.0	4.18	−33.0	3.77	−30.9	8.53

Modifier		ACE		AcOEt	
Group	Compound	CV / V	BE (kcal/mol)	CV / V	BE (kcal/mol)
EPH/PsEPH	EPH	−43.8	5.78	−51.7	7.72
	PsEPH	−41.7	7.84	−47.7	8.13
A	A1	−35.4	3.32	−45.6	4.78
	A2	−13.0	5.64	−14.4	7.17
	A3	−29.1	4.83	−34.5	6.69
B	B1	−13.7	4.61	−17.8	6.12
	B2	−28.0	2.17	−38.3	3.47
C	C1	−30.2	1.27	−39.3	2.84
	C2	−14.3	4.73	−17.2	7.43

Table 4.2: Experimental and computational results for strongly binding modifier gases. CV at SV = 4000 V is chosen to represent maximal ion separation and is shown alongside calculated single solvent binding energies for each compound modifier pair.

A more rigorous method of predicting BEs with DMS data was demonstrated by Walker *et al.*[1] who created ML training set fits for MeOH and H_2O BEs with a series of drug candidate compounds using the full dispersion plot data. Training sets were then created using the Walker *et al.*[1] data and the resulting ML model was tested upon the compounds examined in this chapter, as was done previously for CCS. (See Figure 4.10) The ML database made use of m/z, charge state, temperature, SV, CV and single H_2O-ion BEs in order to make predictions. The fit appears to be practically perfect in the training set, with a $R^2 = 1$ and a MAE of 0.008 kcal/mol between predicted BE and calculated BE. However, when this training set is applied to the isomer groups from this chapter the predicted BE values are unsatisfactory (Figure 4.11). A linear fit to predicted BE versus calculated BE for the test set of data achieved a ML fit with an R^2 of -31.99, a MAE of 1.00 kcal/mol. As was previously done for the CCS ML testing, several of these compounds (PsEPH, A3, and B2,) were then moved from the test set into the training set of the ML

model to examine the effects expanding the chemical space within the training set. This resulted in an increase in the R^2 of the test set to 0.596, and a reduction MAE to 0.83 kcal/mol (Figure 4.12) While there is still much improvement that can be made in the ML fit, this does demonstrate that by increasing the applicability domain of the training set, the quality of the fitting on test data sets will also be improved. In order to capitalize on this result, data should be gathered for more compounds in order to improve the efficacy of the ML model for testing upon new compounds. As the process for calculating BEs can be a computationally expensive and time-consuming process a method for generating these values from similar experiments would be a valuable tool for use in the future.

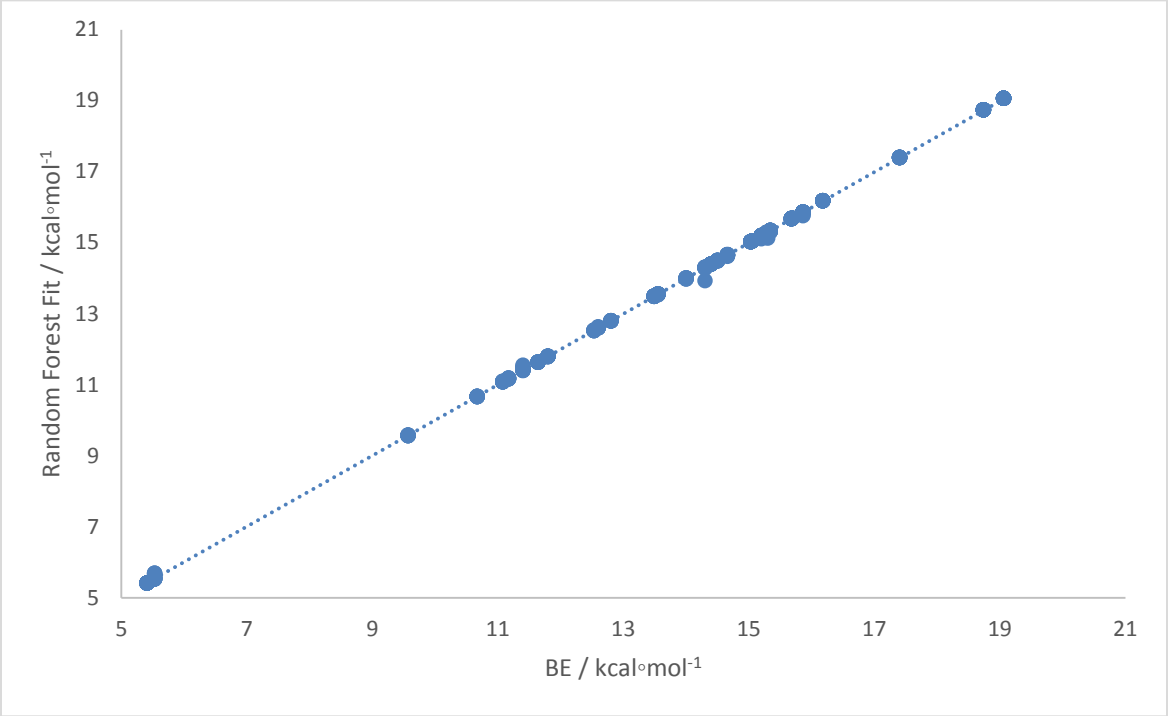


Figure 4.10: Training set results for ML fits with H_2O modifier DMS data from Walker *et al.*[1] Measured BEs are plotted against ML predicted fit. BEs in kcal/mol. The fit for the plot displays an R^2 = 1 as well as a MAE of 0.008 kcal/mol.

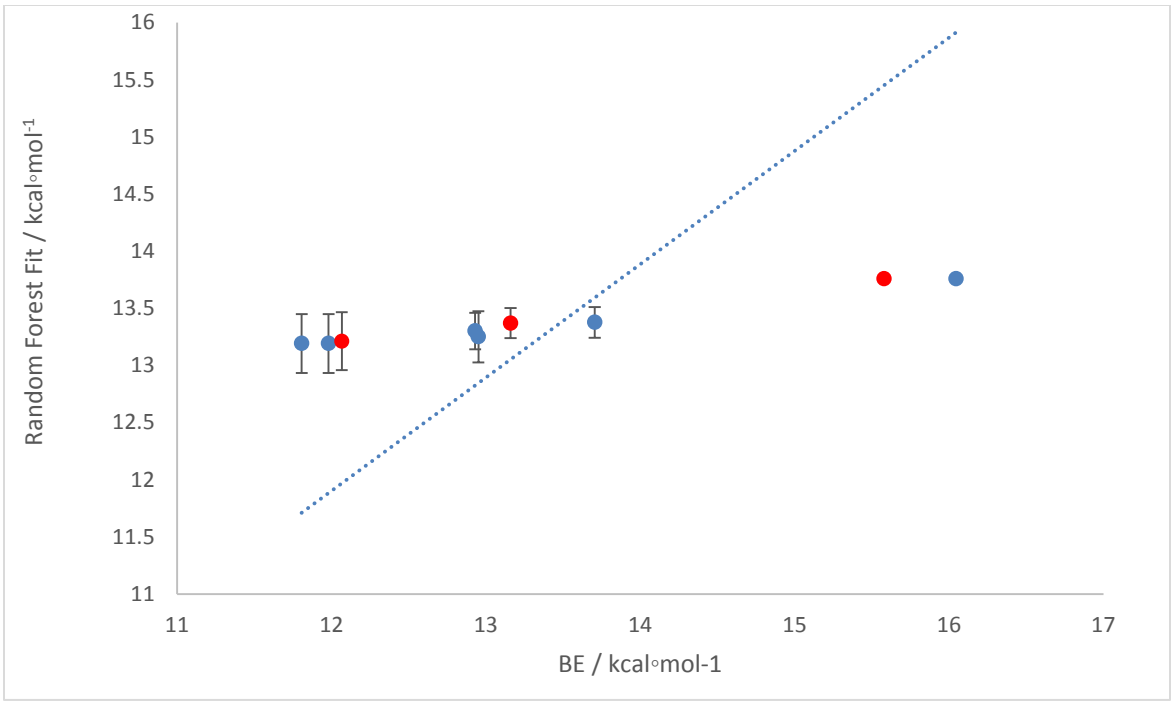


Figure 4.11: ML predicted BEs for the test set using the model depicted in Figure 4.10 The fit for this plot (set to pass through 0) displays an R^2 = -31.99 (no correlation) and a MAE of 1.00 kcal/mol. Compounds moved from the testing set to the training set are indicated as red markers.

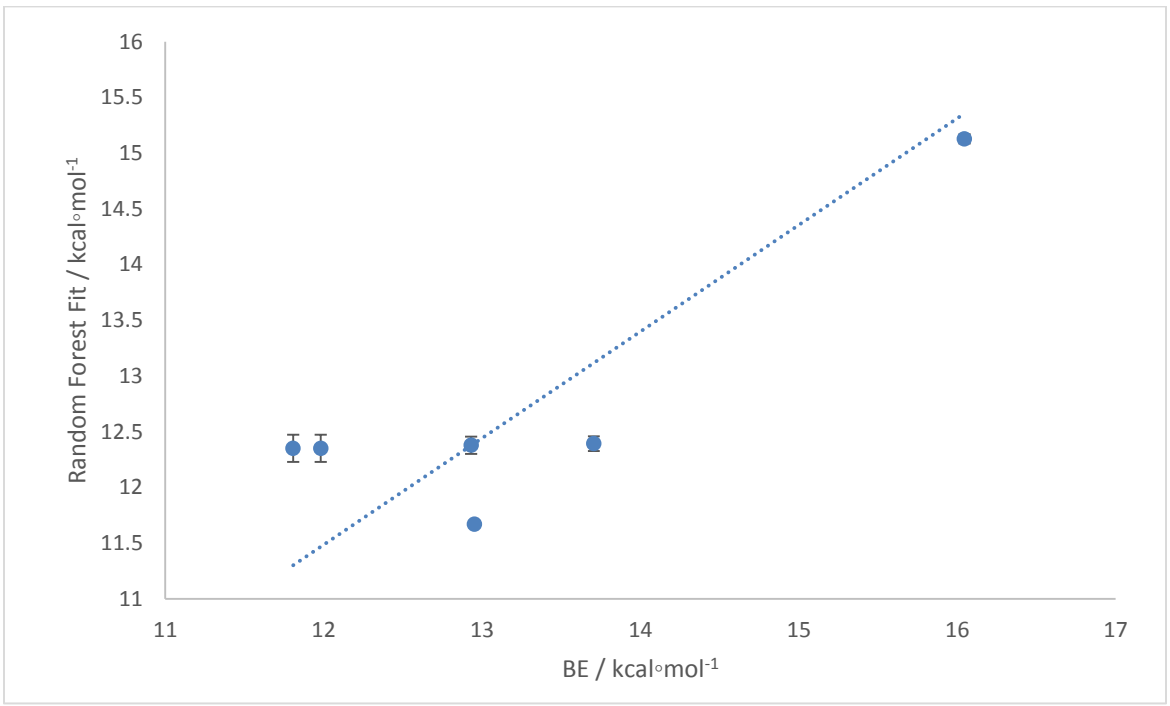


Figure 4.12: ML predicted BEs for augmented test set using the expanded training model. The fit for this plot displays an R^2 = 0.596 and a MAE of 0.83 kcal/mol, showing in increase in fit and a decrease in error compared to the original test set.

A similar train/test ML experiment was carried out using experimental and computational data for MeOH modifier from Walker *et al.*[1] As was the case with the H_2O modifier training set fit, the MeOH modifier training set fit was nearly perfect with an R^2 of 1, and MAE of 0.009 kcal/mol. However, when the training model was used upon the testing set, a negative R^2 was obtained (−1.138), indicating that no correlation exists between the test and training sets. Unfortunately, after some of the test compounds were introduced into the training set, no significant improvement was observed in the R^2 (−1.295). The MeOH data set is similar in size to the H_2O set (forty compounds examined for MeOH compounds versus twenty-eight examined for H_2O), and it is likely that ML model is overfitting on the training set, resulting in a low training error but extremely high error when applied to the molecules outside the training set. As with the H_2O model, this can be potentially resolved by the addition of more data points to the model.

4.5 Conclusions

A diastereomeric pair (EPH/PsEPH) and three groups of sulfonamide isomers were examined extensively using DMS. These experiments could be broken down into three main groups: Pure N_2 gas, weakly binding modifiers and strongly binding modifiers. Computational investigations were conducted in order to show the relationship between calculated parameters, specifically single-solvent binding energies and collision cross sections, and experimental results. DMS experiments with pure N_2 gas showed little to no separations between the compounds within each isomer group; this was expected based on the results of CCS calculations which showed very little difference between the compounds in each group. The data used by Walker *et al.*[1] was used to train a random forest ML model for the prediction of CCS using DMS behavior in N_2 and was tested upon the isomer groups examined in this chapter. Initial results were good (fit of R^2 = 0.867, compared to R^2 = 0.944 of the training set) and were further improved by the incorporation of several isomer compounds (EPH, A3, B2) into the training set. (R^2 = 0.946) In order to strengthen this result, more train/testing should be conducted on a larger data set in order to ensure that the size of the test set does not influence the result and in order to improve the ML model.

The weakly binding solvents showed DMS separation of isomers in most cases, however, computational results were mixed. Previous work has shown a relation between SV at CV_{min} and ion-solvent BE, which in general was not observed in this study. Though calculated Tol-analyte binding energies quantitatively predicted the ordering of DMS isomer separation using calculated binding energies, H_2O-analyte and MeOH-analyte binding energies did not show the expected DMS behavior with calculated BE. This could be a result of these smaller solvent molecules being able to interact with multiple sites on the analyte ions at once instead of a single solvent molecule. Strong binding solvents show good resolution between the various isomers in almost all cases, but the qualitative orderings with calculated BE were often incorrect or inconclusive. Because SV at CV_{min} is not directly measurable for Type A (strongly binding) DMS behavior, experimental conditions should be changed to move these minima into the range viewable on the DMS instrument (i.e., increasing the temperature) or by predicting the minima location by curve fitting and extrapolation. SV at 4000 V was examined as an alternative in these cases, with a more negative CV shift expected to correlate with a stronger BE. However, this was not observed to be the case. Further solvation sites and multi-solvent BEs should be investigated in both strong and weak clustering solvents in order to resolve possible discrepancies between experimental and computational results. ML fitting to full dispersion plot data was also attempted for H_2O and MeOH, with training data from Walker *et al.*[1] Training set error was negligible in both H_2O and MeOH modifiers, prediction of BEs for the test set not possible in the case of H_2O (R^2 = -31.99) and MeOH. (R^2 = −1.138). By introducing compounds from the testing into the training set the predictive ability of the H_2O BE model was increased (R^2 = 0.596), while the MeOH BE model remained unable to find any correlations between its own data and that within the training set. (R^2 = −1.295) The low error in training set fits and large error in the test set indicates overfitting which can potentially be corrected by the addition of more data or by using a simpler model. As our BE database is quite small the former should be pursued.

A variety of paths exist for moving forward with the work presented within this chapter. In order to better compare inconclusive experimental results with BEs, experiments could be rerun with higher temperature.

This would reduce the field required to disrupt ion-solvent binding and shift Type A DMS behavior towards Type B behavior to accurately measure SV at CV_{min}. Alternatively, curve fitting could be applied to the experimental data for these compounds and their trends projected to higher SV values in order to predict where the minima may lie. The expected trend of more negative CV shift at SV = 4000 V for stronger binding energies was not observed between in any of the isomer groups. Further investigation should be made in order to better determine the dynamics of DMS behavior. Another focus for further work is the expansion of the ML training sets examined in this chapter, to allow the model to be used on a wider range of chemical space and in general to improve its accuracy. In Chapter 5, the work on CCSs will be expanded with the introduction of a number of new compounds with a range of new chemical properties.

Chapter 5 : Expanding the ML CCS Database

5.1 Introduction

A recent development in chemistry is the spike in popularity of employing ML methods in data analysis.[98] This development has followed the availability of large publicly available databases such as DrugBank[99] Pubchem[100], and ChemBank[101] Much of this increase use of ML stems from the implementation of machine learning algorithms such as random forests (RF)[40], support vector machines (SVM)[102] and artificial neural networks (ANN)[103] into programs which simplify this process for the user. Orange is an example of an interface for several python ML packages.[39] One major focus of machine learning within medicinal chemistry is determining quantitative structure activity relationships (QSAR) or quantitative structure property relationships (QSPR), which predict an activity or property for a compound based upon other known properties by finding a relationship between them.[104] An example of this type of ML is shown by Cao *et al.* in examining toxicity of ionic liquids on leukemia rat cell lines.[105] Another important area is quantum machine learning (QML), which combines quantum chemical calculations with ML methods.[106] As shown by Ramakrishnan *et al.*, this can greatly reduce computational time required to predict properties such as the thermochemistry or the electron correlation of a compound by replacing difficult to compute calculations with ML learned substitutes.[107] A third area of interest is virtual screening, which involves searching a database for compounds to that will perform well at a given task based upon their properties.[108] This approach was used by Wellenzohn *et al.* to create novel agonists of the GPR119 receptor for the treatment of Type 2 diabetes.[109]

The Hopkins lab group has also begun to make use of machine learning techniques. The group's first ML published project used unsupervised learning in the examination of the potential energy surface for the phenylalanine/serine dimer in order to investigate regions of kinetic trapping.[110] The most recent project by Walker *et al.*[1] used experimental DMS data to model and predict molecular properties (e.g., the collision cross section, pK_b) for several sets of drug candidate molecules.[1] In the previous chapter ML

was explored as a method to determine the collision cross section of analytes in pure N_2 gas, as well as ion-solvent binding energies in MeOH and H_2O modifier. It was further shown that when using this technique, ML test results were improved by the introduction of similar compounds to those in the test set into the training set of compounds. The chapter looks to expand the ML database with a variety of compounds with different functionalities so that the database's predictions will be accurate across a wide range of chemical space.

5.2 Experimental Methods:

DMS experiments were carried out on a QTRAP 5500 instrument provided by SCIEX. Each component of the experimental setup was controlled using the Analyst® software (v. 1.6.2). All compounds were examined for the full range of SVs from 0-4000 V SV, in 500 V increments from 0-2500 V and 250 V increments from 2500-4000 V SV. This allows for the creation of dispersion plots showing the different types of DMS behavior and provides a finer look at the SV range where CV shifts become more pronounced. Each experiment used MRM mode in order to maximize experimental throughput. By providing a Q1 mass, the collision energy (CE) required to generate a fragment and the Q3 fragment mass, a sizable number of compounds can be run simultaneously. In this case, compounds were examined in five groups of approximately forty compounds per group. For the purpose of this section experiments were carried out in pure N_2 gas. The "Monster Mix" was provided to by SCIEX and contains 196 compounds which range between 76 - 1279.9 m/z and contain a variety of different chemical functionalities. A list of the "Monster Mix" compounds is shown in Appendix B.

5.3 Computational Methods:

Initially, neutral structures of the compounds of interest were generated based upon structures from sources such as Pubchem.[100] Each structure was preoptimized at the PM7 level of theory, and then used to generate protonated isomers by chemical intuition. Each tautomer was optimized at the B3LYP/6-31++G(d,p) level of theory and these resulting structures were ranked by their energies in order to

determine the preferred site of protonation for each compound. B3LYP/6-31++G(d,p) was chosen as the level of theory due to the large size of some of the compounds and the well documented ability of 6-31++G(d,p) to yield accurate results for similar systems. In order to ensure that this does not have an adverse effect on the prediction of CCS, a number of structures were further optimized to B3LYP/6-311++G(d,p) in order to make a comparison of CCSs at the two levels of theory. Owing to the large number of compounds, BH searches were not conducted. MobCal-MPI[35] was used for the calculation of CCSs.

ML modeling was carried out using a RF algorithm within Orange[39], which used 500 decision trees to make its predictions, and the each was allowed to split the database into groups as small as two entries as it generated these trees. The remaining hyperparameters for the learner were determined by Orange based upon the data number and feature space. The database for this chapter included m/z values, charge states for each compound, and a subset of the DMS data covering SVs of 2000 V to 4000 V in 500 V increments. All experimental data in the section were gathered at 150 °C. A learning curve will be generated from the data in order to evaluate the ability of the training data set to test on new compounds.

5.4 Results and Discussion:

DMS Experiments were carried out for each of the compounds within the Monster Mix in order to generate more data for the CCS ML model. Several example dispersion plots are shown in Figure 5.1 below. As expected from a N_2 gas environment, most compounds exhibited Type C behavior, with the exception of some of the smaller compounds, such as acetaminophen, which displayed a small amount of Type B behavior (Figure 5.1). One issue that arose in these experiments was that of multiple CV peaks occurring within a dispersion plot. This is thought to be caused by the clustering of ions with solvent molecules such as H_2O and MeOH that originate from the Monster Mix solution, as observed previously in Anwar et al.[23] These solvent molecules remain clustered together throughout the asymmetric SV waveform resulting in hard sphere type behavior. When multiple peaks are observed, the peak with

highest ion signal intensity is taken to correspond to the global minimum structure. In the case of acetaminophen, which displayed multiple CV peaks (Figure 5.1), the Type B trace was decided to be created the global minimum structure. Due to the large amount of experimental data that needed to be processed for this project, an automated method of data extraction was created by one of the Hopkins group members, Ce Zhou. This package of python scripts allowed for the generation of ionograms as well as dispersion plots directly from experimental outputs from the QTRAP 5500 instrument (.wiff files), greatly decreasing the time spent data process required for this project.

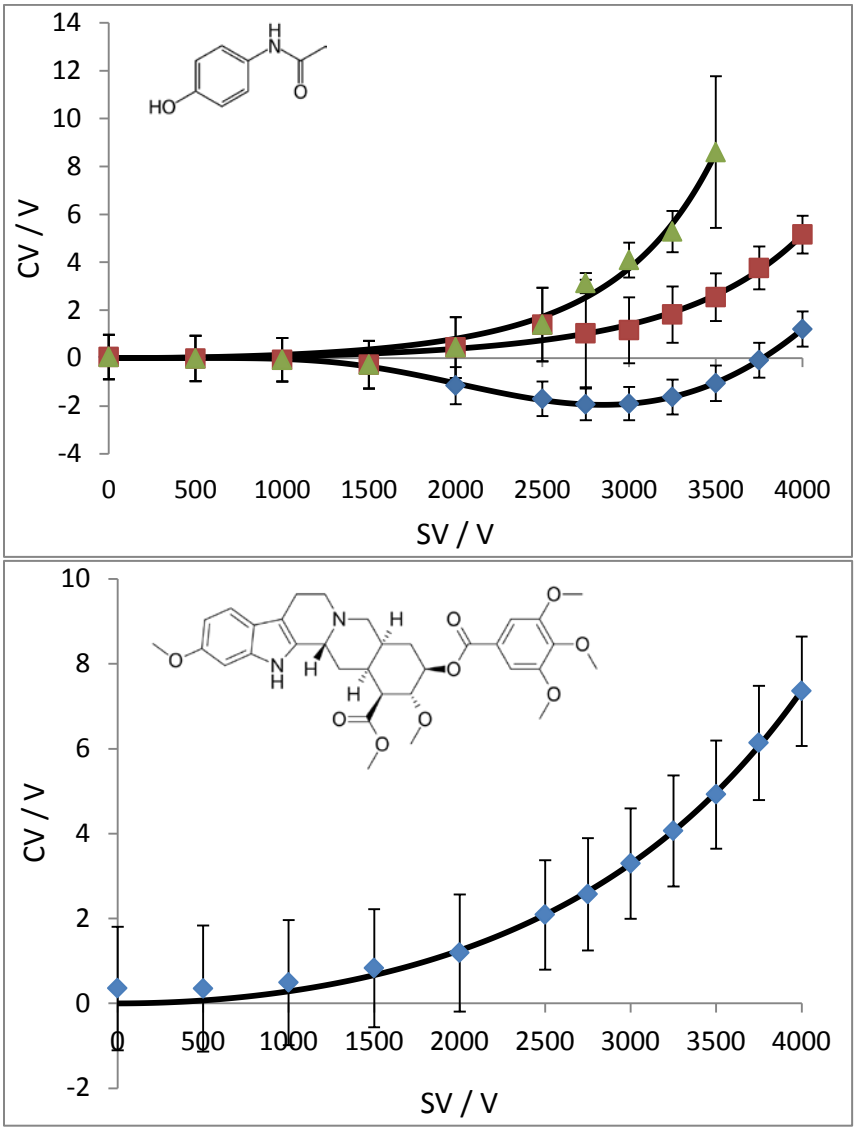


Figure 5.1: DMS Dispersion plots for acetopminophen (top) and reserpine (below). Neutral structures are provided alongsideError bars give the HWHM values of the CV peaks at a given SV value.

Of the 196 compounds investigated in the “Monster Mix” (Appendix B for list), 142 have are used in the following ML work. 44 compounds were not successfully observed by DMS for each point within the 2000-4000 V SV range and for 10 compounds computational optimization did not finish and therefore CCS were not calculated.

A subset of 47 compounds from within the “Monster Mix” were used to compute CCSs at both B3LYP/6-311++G(d,p) and B3LYP/6-31++G(d,p) levels of theory. In order to facilitate this comparison, each level of theory was used these 47 compounds to create a ML training set and were evaluated based upon their errors and fit after 10-fold cross validation. Shown in Figure 5.2b is a plot of RF predicted CCS with 6-311++G(d,p) calculated CCS as the target. Each level of theory showed similar results; treatment of the 6-31++G(d,p) data yielded an RMSE of 14.1 $Å^2$ and MAE of 9.0 $Å^2$ and 6-311++G(d,p) data yielded an RMSE of 13.6$Å^2$ and MAE of 9.4$Å^2$. This is expected as the plot of calculated CCS at each level of theory shows very little difference in the results. (Figure 5.2a)

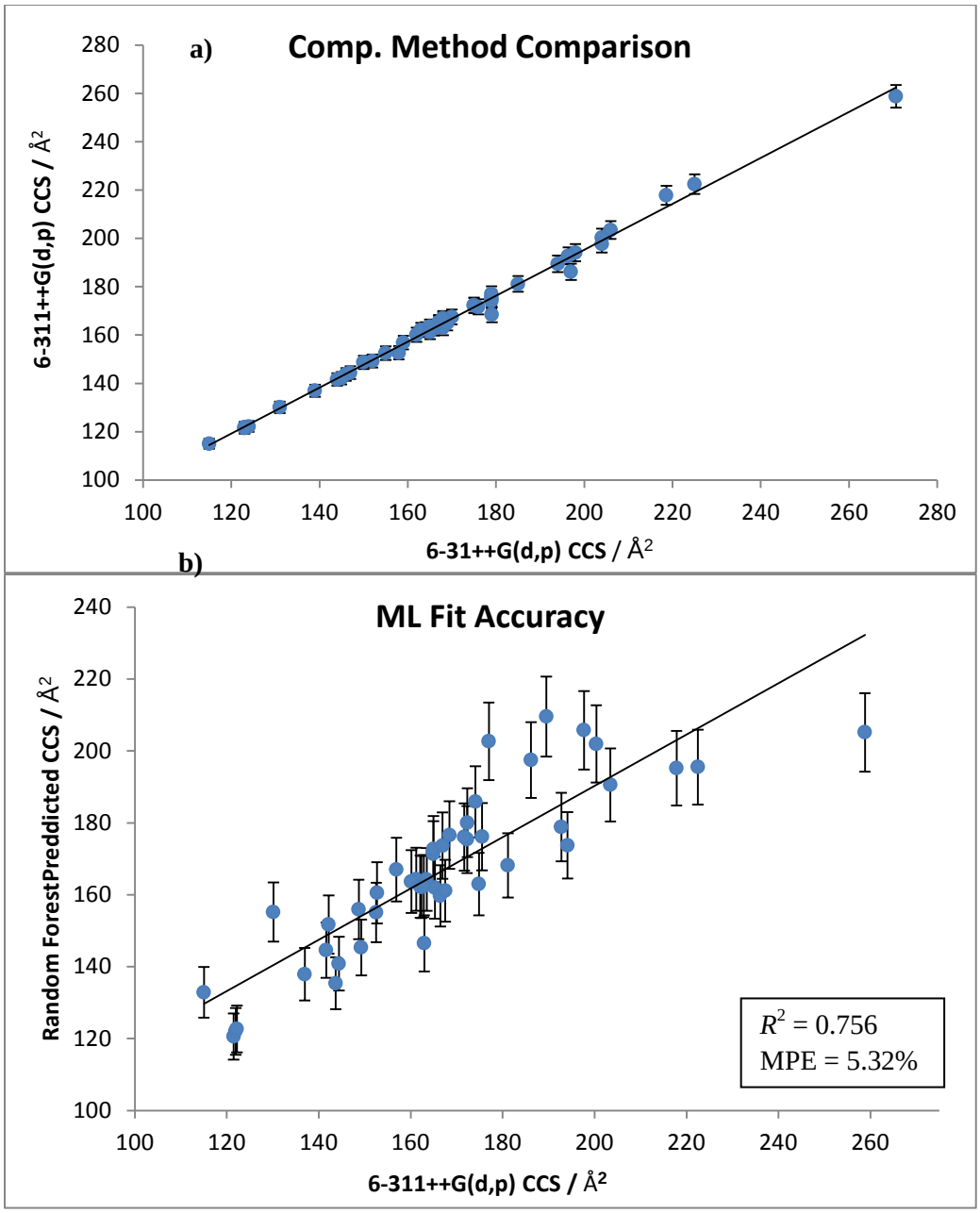


Figure 5.2 a) Plot of 6-31++G(d,p) CCS versus 6-311++G(d,p) CCS. A R^2 of 0.995 is shown for the fit, as well as a a MAE 3.23 Å^2 and MPE of 1.82%. b) Fit of calculated CCS at 6-311++G(d,p) theory for 47 "Monster Mix" compounds. A fit of R^2 = 0.756 is observed, with MAE of 9.0 Å^2 and MPE of 5.32%

To expand the set, all of the data used to predict CCSs from Chapter 4, which includes the four groups of isomers as well as the Walker *et al.*[1] ML data, were combined with the 142 new compounds in this chapter. This gave a total of 209 compounds to use for creation of the expanded ML model. The database containing this data was subsequently divided up into a series of different test set and training set

proportions; initially 10% of the data is placed into the training set and 90% is placed into the test set in order to generate a learning curve for the ML CCS model, with training set being increased in 10% increments of the total database while the test set is reduced by the same. By moving the data from the test set to the training set we show how the error in the model evolves with training size.

Figure 5.3 shows a learning curve for the full data set; 209 compounds split into training and test sets. 10 sets of training and testing pairs were generated randomly for each split of the database (i.e., 10% training, 90% testing). A clear decrease in the MAE for both sets were displayed as a greater portion of the total database is moved from the testing into the training set. The MAE for the training set plateaus to ~2.5 $Å^2$ quite quickly (at ~61 molecules, 30% of data), while the MAE for the test set follows a much slower decline reaching a value of ~8 $Å^2$ at the lowest point for the train/test sets considered (at ~105 compounds, 50% of data) . To give equal consideration to training and test sets model comparisons were carried out at the 50/50 train/test split. (Figure 5.4 and Figure 5.5) One of the 10 random database samples at this increment was then taken for further examination. The training set showed a fit of R^2 of 0.991, a RMSE of 3.54 $Å^2$ and MAE of 2.67 $Å^2$, while the test set demonstrated showed a lesser degree of fit with an R^2 of 0.905, RMSE of 11.43 $Å^2$, and a MAE of 8.13 $Å^2$. A more important figure for the large m/z range considered is the mean percent error (MPE) for the fit, 4.63%.

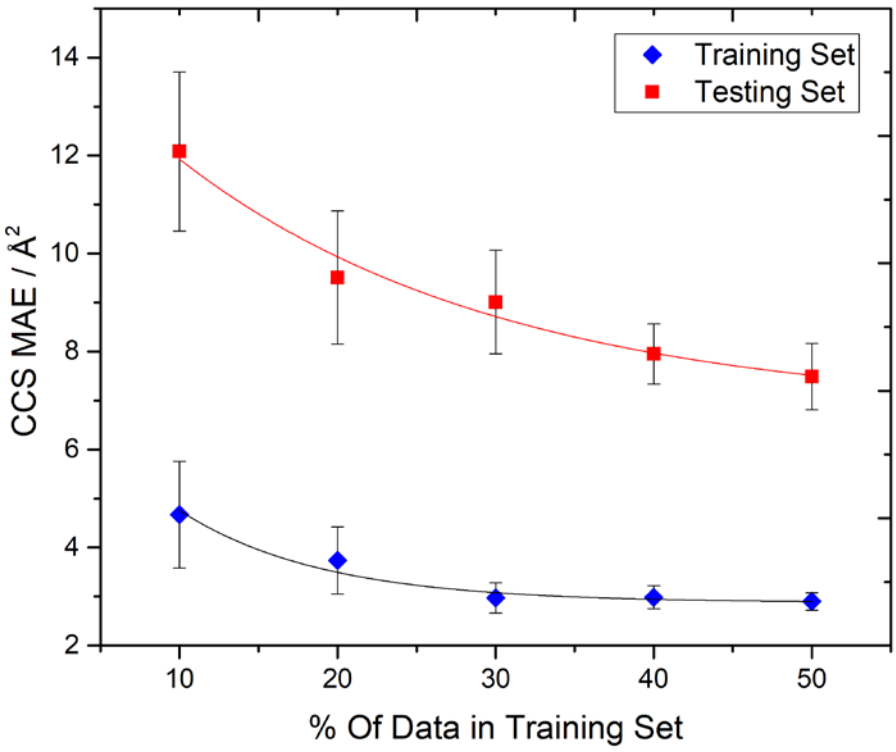


Figure 5.3: Learning curve for the combined ML database, showing the change in error as compounds are moved from the testing set into the training set. At 50% data contained in the testing set the MPE is 4.63%, as sampled from one of the random database splits.

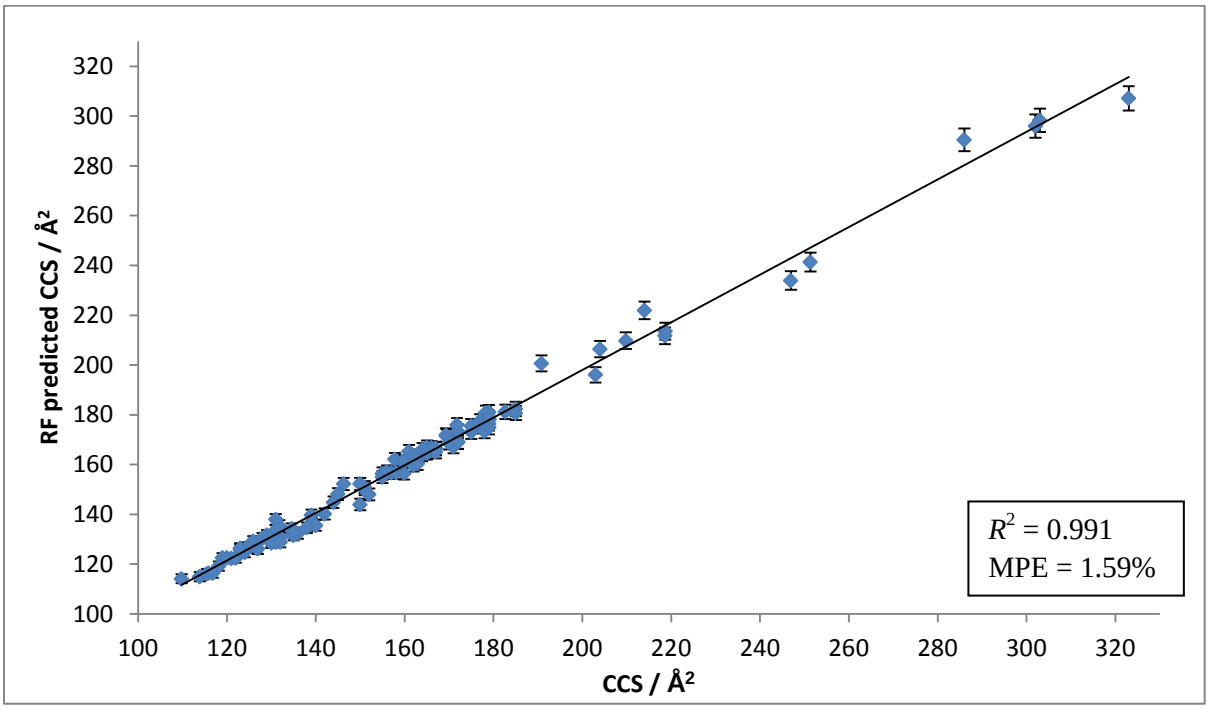


Figure 5.4: Training set ML fit for a sample split of the combined database at 50% train / 50% test split, showing calculated CCS versus RF predicted CCS. The fit gives an R^2 of 0.991 a MAE of 2.67Å^2, and a MPE of 1.59%.

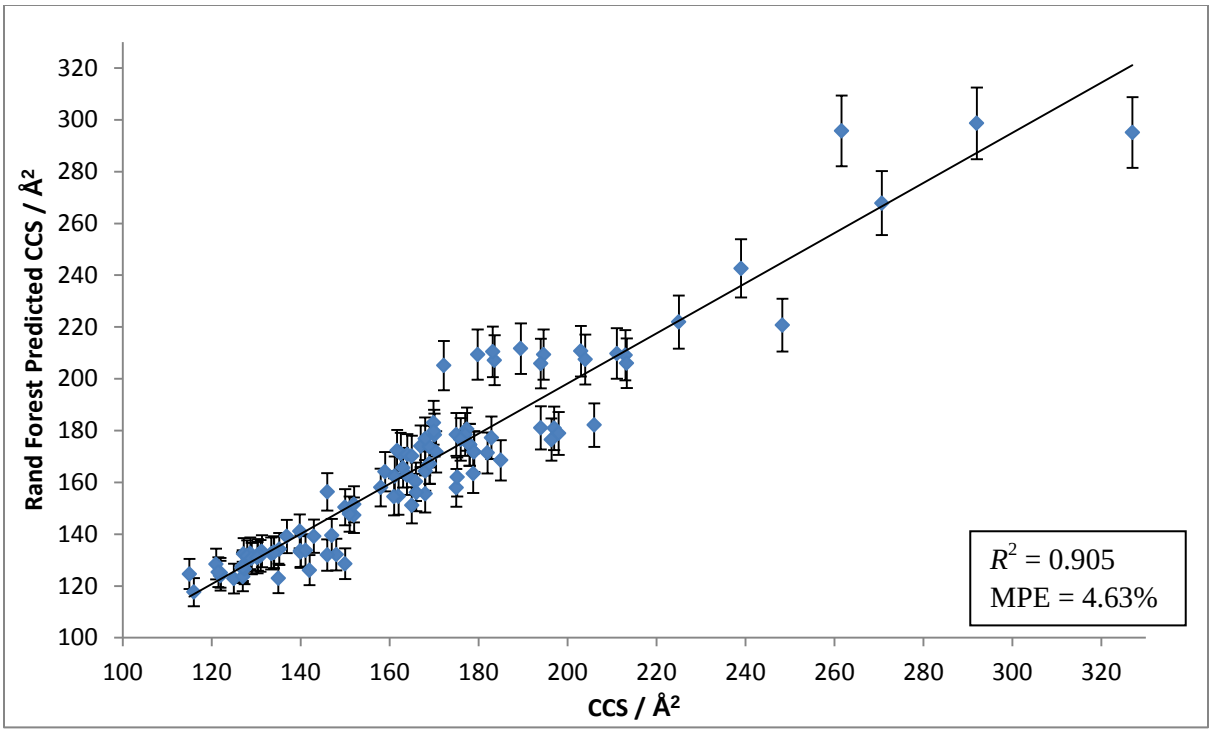


Figure 5.5: Testing set ML fit for a sample split of the combined database at 50% train / 50% test split, showing calculated CCS versus RF predicted CCS. While not as neatly fit as the training set, the overall fit is still quite good, with an R^2 of 0.905 a MAE of 8.13Å^2, and an MPE of 4.63%.

While these results are promising, further work must be completed in order to ensure the accuracy of the ML generated results. The separation between the training set and test set learning curves gives an indication of the ability for the model to make predictions about new compounds and is called the generalization gap. In an ideal case this gap should be minimized as much as possible, converging the test set error as close to the training set error as possible. The learning curve does not outright give information on how this might be accomplished, but one way that the test set error might be reduced is by the introduction of larger compounds into the ML database to fill up the 220-320 Å^2 CCS region. As shown above in Figure 5.4 and Figure 5.5, this range of CCSs is particularly sparse compared to the region below 220 Å^2. Another potential source of error the usage of only single structures to generate CCS for the Monster Mix compounds examined in this chapter. It has been shown previously in Ieritano *et al.*[96] that CCS is best represented as a Boltzmann-weighted distribution of CCS of the structures that appear at the temperatures observed within the DMS (~100 °C). By employing BH to generate additional

structures for each ion of interest and calculating CCS using a Boltzmann-weighted temperature distribution of these structures results can be further improved.

5.5 Conclusions

This chapter expands on the study described in Chapter 4 by introducing many more molecules of a variety of types (e.g., amino acids, antidepressants, stimulants, etc.) with the predictions of CCSs using a larger ML database. This chapter uses the isomer groups examined in Chapter 4 as well as the Walker *et al.*[1] data to construct a larger database with the inclusion of 142 compounds with a variety of different chemical functionalities. In order to ensure an accurate comparison of results, a comparison was made of CCSs calculated for 47 of the 142 compounds at different levels of theory: B3LYP/6-311++G(d,p) and B3LYP/6-31++G(d,p). CCS values at the two levels of theory were very similar with a mean absolute difference of ~3.2 $Å^2$. Additionally, a comparison of ML fits to CCS values for the 47 compounds were compared at the two levels of theory showed very similar model fit errors, demonstrating that using B3LYP/6-31++G(d,p) structures for CCSs is not expected to adversely affect the results.

Learning curves for training and tests sets were produced using the combined ML database. Both training and test set show a decrease in MAE sets as the portion of the database used in training was increased. Testing and training set MAE reached minimum values of ~8 $Å^2$ and ~2.5 $Å^2$, respectively. The difference between these two curves, the generalization gap, is indicative that the ML training set likely needs to be expanded further. The 50/50 train/test ML model was examined further for model predictive errors. It was observed that the fitting is particularly poor for CCS larger than ~220 $Å^2$ due to sparsity in this region. Further research should be focused on expanding the ML database with compounds that fall in the range of 220-320 $Å^2$ for better representation in the model, as well as the expansion of the model to include more compounds with different chemical functionalities.

Chapter 6 : Thesis Conclusions

In recent years isomeric structures have become increasingly important, especially within areas such as medicinal chemistry and drug development, where isomers can potentially have differing effects from each other. One of the most commonly used analytical techniques, MS, is unable to distinguish different isomeric compounds without the assistance of orthogonal techniques due to the identical masses observed. The projects in this thesis use differential mobility spectrometry to separate such compounds, along with quantum chemical calculations to reinforce obtained results. Two main properties are assumed to influence behavior within the DMS, the collision cross section of ions within the DMS, and the clustering / declustering interactions that occur with solvent modifiers that are introduced into the DMS. In Chapter 4 this assumption is tested by applying machine learning methods to predict the collision cross section of bare ions as well as ion-solvent binding energies using incorporating DMS experimental information, while in Chapter 5 the CCS ML model is further improved by the introduction of new compounds.

Chapter 3 explored the use of DMS to perform a separation of derivatized amphetamine and methamphetamine isomers. Initially, AMP and MeAMP isomers were examined in nitrogen only DMS experiments but were unable to be resolved. After a derivatization reaction with S-TPC these compounds were able to be resolved in a nitrogen only gas environment, but only after a throttle gas was applied. Unfortunately, TPC-MeAMP isomers were not distinguishable using CCS and an alternative experimental method had to be followed. Instead, DMS experiments were performed using ACN as a gas modifier. *S*-TPC-*S*-AMP/MeAMP isomers were shown in each case to experience more strongly binding interaction within the DMS and migrated to more negative CV values compared to the *S*-TPC-*R* isomers. This agreed with computational results that showed the *S*-TPC-*S* isomers experienced larger binding energies than the *S*-TPC-*R* isomers. However, this trend did not hold between AMP and MeAMP, only

between the isomer pairs. This demonstrates that DMS behavior is complex, and is dictated by not just the BEs, but also by other parameters such as ion-solvent CCS and the bare ion CCS in a nontrivial way.

Picking up where Chapter 3 left off, Chapter 4 expands upon the methodology used there with a larger number of compounds: (+)-ephedrine, (+)-pseudoephedrine, and three groups of sulfonamide structural isomers were examined. These compounds were examined with a variety of modifiers using DMS, including H_2O, MeOH, EtOH, PrOH, IPA, ACE, AcOEt, Tol and pure N_2. Firstly, experiments were performed using a pure N_2 environment. Little to no separation was observed, a result corroborated by their similar calculated CCSs. The data from these experiments was then used in order to test an ML model made with data taken from Walker *et al.*[1] The fit produced by the ML model on this new data proved be good, and it was further improved by introducing some of the studied compounds into the training set for the ML model. This demonstrated the requirement for the model to be trained upon compounds similar to the test compounds in order to ensure the best results.

Experimental results for the other modifiers were performed and then compared with ion-modifier BEs obtained from quantum chemical calculations. BEs were compared to the SV at CV_{min} values (the minima appearing within Type B dispersion plots), which had previously shown to correlate to each other by Liu *et al.*[97] This comparison showed mixed agreements between experimental and computational results. In the case of modifiers which showed mainly Type B behavior, the BEs for Tol showed excellent agreement with the ordering of SV at CV_{min}, while H_2O and MeOH were only able to predict some of the isomer separations. It was more difficult to make evaluations of the compounds that showed mainly Type A behavior, as the SV at CV_{min} values are not visible within the range of a DMS scan. Even in cases where it was possible to assume which curve would have a lower SV at CV_{min}, these orderings often did not match with the computational BE results. Further investigations should be performed on other potential solvation sites that were not investigated, as well as multi-solvent BEs in order to reconcile the differences in these results. As was performed with the CCSs, a ML model was constructed for the prediction of BEs for ion-MeOH and ion-H_2O using data from Walker *et al.*[1] and then tested upon the

isomers studied in this chapter which showed minimal fit correlation. Three of the isomers were moved from the test set into the training set for the model and the remaining compounds were evaluated again. In the case of H_2O modifier, a weak correlation was observed in testing, which was improved when some isomers were included in the training set. In the case of MeOH, no correlation was observed both before and after the introduction of isomers into the training set for the model. This is indicative of the model overfitting the training data, which has reduced its applicability in testing on molecules outside the set. Further work should focus on improving the fit of the H_2O and MeOH model by introducing new compounds to the training set for the ML model and adjusting hyperparameters for the ML model.

Chapter 5 follows up on the ML prediction of CCS shown in Chapter 4 and aims to improve the ML model by introducing 147 new compounds with a variety of chemical functionalities to its database, bringing the number up to 209. Additionally, a comparison was made between the CCSs of structures calculated at two different levels of theory: B3LYP/6-311++G(d,p) and B3LYP/6-31++G(d,p). This comparison showed little to no difference in the in the MobCal-MPI CCS values as well as the RF predicted model errors. In order to show that the training set was improved by the introduction of new compounds, a learning curve was generated using the combined ML database. A decrease in the MAE was observed in both the training and test sets as the proportion of compounds in the ML training set was increased. The minimum gap between these two plateaus gives the generalization gap, in this case around 5.5 $Å^2$. The size of this gap indicates that improvement to the model can be made by adding more compounds to the test set, particularly with CCS ranging from 220-320 $Å^2$.

Future work with these projects should be focused among three main lines: An examination of the effect of multi-solvent binding energies on DMS behaviors, the expansion of the CCS database with the addition of further compounds, and the exploration of different gas modifiers and calculated BEs using ML. While single solvent binding energies have been used in this work, this may not actually be the case due to the complex environment within the DMS. The solvation/desolvation process could be occurring from multiple solvent molecules clustering with the ion and declustering partially, or entirely. This could

explain some of the discrepancy in the ranking BEs with SV at CV_{min} values in Chapter 4. Within the Hopkins group, experiments are ongoing with a pesticide mix of similar size to the Monster Mix that may prove to be useful when incorporated into the ML database. Finally, numerous DMS experiments have been taken on the Monster Mix using different gas modifiers, such as ACN, MeOH, ACE, as well as partial completion of H_2O. These experimental results are waiting on computational BEs in order to perform ML predictions and testing with more compounds. This could be particularly useful for modifiers which display prominent type A DMS behavior and their SV at CV_{min} values can therefore not be observed.

References:

(1) Walker, S. W. C.; Anwar, A.; Psutka, J. M.; Crouse, J.; Liu, C.; Le Blanc, J. C. Y.; Montgomery, J.; Goetz, G. H.; Janiszewski, J. S.; Campbell, J. L.; et al. Determining Molecular Properties with Differential Mobility Spectrometry and Machine Learning. *Nat. Commun.* **2018**, *9* (1), 1–7. https://doi.org/10.1038/s41467-018-07616-w.

(2) Awad, H.; El-Aneed, A. Enantioselectivity of Mass Spectrometry: Challenges and Promises: ENANTIOSELECTIVITY OF MASS SPECTROMETRY. *Mass Spectrom. Rev.* **2013**, 466–483. https://doi.org/10.1002/mas.21379.

(3) Calcaterra, A.; D'Acquarica, I. The Market of Chiral Drugs: Chiral Switches versus de Novo Enantiomerically Pure Compounds. *J. Pharm. Biomed. Anal.* **2018**, *147*, 323–340. https://doi.org/10.1016/j.jpba.2017.07.008.

(4) Foster, R. H.; Markham, A. Levobupivacaine: A Review of Its Pharmacology and Use as a Local Anaesthetic. *Drugs* **2000**, *59* (3), 551–579. https://doi.org/10.2165/00003495-200059030-00013.

(5) Reading, E.; Munoz-Muriedas, J.; Roberts, A. D.; Dear, G. J.; Robinson, C. V.; Beaumont, C. Elucidation of Drug Metabolite Structural Isomers Using Molecular Modeling Coupled with Ion Mobility Mass Spectrometry. *Anal. Chem.* **2016**, *88* (4), 2273–2280. https://doi.org/10.1021/acs.analchem.5b04068.

(6) McGregor, L.; Rychkov, D. A.; Coster, P. L.; Day, S.; Drebushchak, V. A.; Achkasov, A. F.; Nichol, G. S.; Pulham, C. R.; Boldyreva, E. V. A New Polymorph of Metacetamol. *CrystEngComm* **2015**, *17* (32), 6183–6192. https://doi.org/10.1039/C5CE00910C.

(7) Modi, H. R.; Basselin, M.; Taha, A. Y.; Li, L. O.; Coleman, R. A.; Bialer, M.; Rapoport, S. I. Propylisopropylacetic Acid (PIA), a Constitutional Isomer of Valproic Acid, Uncompetitively Inhibits Arachidonic Acid Acylation by Rat Acyl-CoA Synthetase 4: A Potential Drug for Bipolar Disorder. *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **2013**, *1831* (4), 880–886. https://doi.org/10.1016/j.bbalip.2013.01.008.

(8) Fales, H. M.; Wright, G. J. Detection of Chirality with the Chemical Ionization Mass Spectrometer. "Meso" Ions in the Gas Phase. *J. Am. Chem. Soc.* **1977**, *99* (7), 2339–2340. https://doi.org/10.1021/ja00449a054.

(9) Yao, Z.-P.; Wan, T. S. M.; Kwong, K.-P.; Che, C.-T. Chiral Analysis by Electrospray Ionization Mass Spectrometry/Mass Spectrometry. 1. Chiral Recognition of 19 Common Amino Acids. *Anal. Chem.* **2000**, *72* (21), 5383–5393. https://doi.org/10.1021/ac000729q.

(10) Borges, K. B.; de Oliveira, A. R. M.; Barth, T.; Jabor, V. A. P.; Pupo, M. T.; Bonato, P. S. LC–MS–MS Determination of Ibuprofen, 2-Hydroxyibuprofen Enantiomers, and Carboxyibuprofen Stereoisomers for Application in Biotransformation Studies Employing Endophytic Fungi. *Anal. Bioanal. Chem.* **2011**, *399* (2), 915–925. https://doi.org/10.1007/s00216-010-4329-9.

(11) Kikura-Hanajiri, R.; Kawamura, M.; Miyajima, A.; Sunouchi, M.; Goda, Y. Chiral Analyses of Dextromethorphan/Levomethorphan and Their Metabolites in Rat and Human Samples Using LC-MS/MS. *Anal. Bioanal. Chem.* **2011**, *400* (1), 165–174. https://doi.org/10.1007/s00216-011-4707-y.

(12) Ruiz-Matute, A. I.; Hernández-Hernández, O.; Rodríguez-Sánchez, S.; Sanz, M. L.; Martínez-Castro, I. Derivatization of Carbohydrates for GC and GC–MS Analyses. *Enhanc. Anal. Anal. Deriv.* **2011**, *879* (17), 1226–1240. https://doi.org/10.1016/j.jchromb.2010.11.013.

(13) Waldhier, M. C.; Dettmer, K.; Gruber, M. A.; Oefner, P. J. Comparison of Derivatization and Chromatographic Methods for GC–MS Analysis of Amino Acid Enantiomers in Physiological Samples. *J. Chromatogr. B* **2010**, *878* (15), 1103–1112. https://doi.org/10.1016/j.jchromb.2010.03.021.

(14) Han, Y.; Bai, Y.; Xiao, Y.; Du, F.; Liang, Y.; Tan, Z.; Zhao, M.; Liu, H. Simultaneous Discrimination of Jasmonic Acid Stereoisomers by CE-QTOF-MS Employing the Partial Filling Technique. *ELECTROPHORESIS* **2011**, *32* (19), 2693–2699. https://doi.org/10.1002/elps.201100043.

(15) Xia, S.; Zhang, L.; Lu, M.; Qiu, B.; Chi, Y.; Chen, G. Enantiomeric Separation of Chiral Dipeptides by CE-ESI-MS Employing a Partial Filling Technique with Chiral Crown Ether. *ELECTROPHORESIS* **2009**, *30* (16), 2837–2844. https://doi.org/10.1002/elps.200800799.
(16) Zhang, Y.; Huang, L.; Chen, Q.; Chen, Z. A Silica Monolithic Column with Chemically Bonded L-Pipecolic Acid as Chiral Stationary Phase for Enantiomeric Separation of Dansyl Amino Acids by CEC–MS. *Chromatographia* **2012**, *75* (5), 289–296. https://doi.org/10.1007/s10337-012-2188-6.
(17) Bragg, W.; Shamsi, S. A. A Novel Positively Charged Achiral Co-Monomer for β-Cyclodextrin Monolithic Stationary Phase: Improved Chiral Separation of Acidic Compounds Using Capillary Electrochromatography Coupled to Mass Spectrometry. *J. Chromatogr. A* **2012**, *1267*, 144–155. https://doi.org/10.1016/j.chroma.2012.08.002.
(18) Alexander, A. J.; Staab, A. Use of Achiral/Chiral SFC/MS for the Profiling of Isomeric Cinnamonitrile/Hydrocinnamonitrile Products in Chiral Drug Synthesis. *Anal. Chem.* **2006**, *78* (11), 3835–3838. https://doi.org/10.1021/ac060326b.
(19) Zeng, L.; Xu, R.; Zhang, Y.; Kassel, D. B. Two-Dimensional Supercritical Fluid Chromatography/Mass Spectrometry for the Enantiomeric Analysis and Purification of Pharmaceutical Samples. *J. Chromatogr. A* **2011**, *1218* (20), 3080–3088. https://doi.org/10.1016/j.chroma.2011.03.041.
(20) Lapthorn, C.; Pullen, F.; Chowdhry, B. Z. Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS) of Small Molecules: Separating and Assigning Structures to Ions. *Mass Spectrom. Rev.* **2013**, *32* (1), 43–71. https://doi.org/10.1002/mas.21349.
(21) Phillips, S. T.; Dodds, J. N.; Ellis, B. M.; May, J. C.; McLean, J. A. Chiral Separation of Diastereomers of the Cyclic Nonapeptides Vasopressin and Desmopressin by Uniform Field Ion Mobility Mass Spectrometry. *Chem. Commun.* **2018**, *54* (68), 9398–9401. https://doi.org/10.1039/C8CC03790F.
(22) Liu, C.; Le Blanc, J. C. Y.; Shields, J.; Janiszewski, J. S.; Ieritano, C.; Ye, G. F.; Hawes, G. F.; Hopkins, W. S.; Campbell, J. L. Using Differential Mobility Spectrometry to Measure Ion Solvation: An Examination of the Roles of Solvents and Ionic Structures in Separating Quinoline-Based Drugs. *The Analyst* **2015**, *140* (20), 6897–6903. https://doi.org/10.1039/C5AN00842E.
(23) Anwar, A.; Psutka, J.; Walker, S. W. C.; Dieckmann, T.; Janizewski, J. S.; Larry Campbell, J.; Scott Hopkins, W. Separating and Probing Tautomers of Protonated Nucleobases Using Differential Mobility Spectrometry. *Int. J. Mass Spectrom.* **2018**, *429*, 174–181. https://doi.org/10.1016/j.ijms.2017.08.008.
(24) Campbell, J. L.; Le Blanc, J. C. Y.; Schneider, B. B. Probing Electrospray Ionization Dynamics Using Differential Mobility Spectrometry: The Curious Case of 4-Aminobenzoic Acid. *Anal. Chem.* **2012**, *84* (18), 7857–7864. https://doi.org/10.1021/ac301529w.
(25) Gehrke, R.; Reuter, K. Assessing the Efficiency of First-Principles Basin-Hopping Sampling. *Phys. Rev. B* **2009**, *79* (8). https://doi.org/10.1103/PhysRevB.79.085412.
(26) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101* (28), 5111–5116. https://doi.org/10.1021/jp970984n.
(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*; 2009.
(28) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035. https://doi.org/10.1021/ja00051a040.
(29) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91* (1), 1–41. https://doi.org/10.1016/0010-4655(95)00041-D.

(30) J P Stewart, J. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. **2012**, *19* (1), 1–32. https://doi.org/10.1007/s00894-012-1667-x.
(31) Liu, X.; Sohlberg, K. Empirical Correction for PM7 Band Gaps of Transition-Metal Oxides. *J. Mol. Model.* **2016**, *22* (1). https://doi.org/10.1007/s00894-015-2891-y.
(32) van Mourik, T.; Gdanitz, R. J. A Critical Note on Density Functional Theory Studies on Rare-Gas Dimers. *J. Chem. Phys.* **2002**, *116* (22), 9620–9623. https://doi.org/10.1063/1.1476010.
(33) Becke, A. D. Density *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913. Functional Thermochemistry
(34) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys Rev Lett* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.
(35) Ieritano, C.; Crouse, J.; Campbell, J. L.; Hopkins, W. S. A Parallelized Molecular Collision Cross Section Package with Optimized Accuracy and Efficiency. *The Analyst* **2019**, *144* (5), 1660–1670. https://doi.org/10.1039/C8AN02150C.
(36) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261* (1–2), 86–91.
(37) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100* (40), 16082–16086.
(38) Mason, E. A. Transport Properties of Gases Obeying a Modified Buckingham (Exp -Six) Potential. *J. Chem. Phys.* **1954**, *22* (2), 169–186. https://doi.org/10.1063/1.1740026.
(39) Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14* (1), 2349–2353.
(40) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32.
(41) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947–1958. https://doi.org/10.1021/ci034160g.
(42) Mitchell, J. B. O. Machine Learning Methods in Chemoinformatics: Machine Learning Methods in Chemoinformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (5), 468–481. https://doi.org/10.1002/wcms.1183.
(43) Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **2003**, *24* (1), 3.
(44) Bakhoum, S. F. W.; Agnes, G. R. Study of Chemistry in Droplets with Net Charge before and after Coulomb Explosion: Ion-Induced Nucleation *in Solution* and Implications for Ion Production in an Electrospray. *Anal. Chem.* **2005**, *77* (10), 3189–3197. https://doi.org/10.1021/ac048536g.
(45) Bruins, A. P. Mechanistic Aspects of Electrospray Ionization. *J. Chromatogr. A* **1998**, *794* (1–2), 345–357.
(46) Schneider, B. B.; Nazarov, E. G.; Londry, F.; Vouros, P.; Covey, T. R. Differential Mobility Spectrometry/Mass Spectrometry History, Theory, Design Optimization, Simulations, and Applications: DIFFERENTIAL MOBILITY SPECTROMETRY/MASS SPECTROMETRY. *Mass Spectrom. Rev.* **2016**, *35* (6), 687–737. https://doi.org/10.1002/mas.21453.
(47) Hopkins, W. S. Determining the Properties of Gas-Phase Clusters. *Mol. Phys.* **2015**, *113* (21), 3151–3158. https://doi.org/10.1080/00268976.2015.1053545.
(48) Campbell, J. L.; Zhu, M.; Hopkins, W. S. Ion-Molecule Clustering in Differential Mobility Spectrometry: Lessons Learned from Tetraalkylammonium Cations and Their Isomers. *J. Am. Soc. Mass Spectrom.* **2014**, *25* (9), 1583–1591. https://doi.org/10.1007/s13361-014-0939-3.
(49) March, R. E. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *J. Mass Spectrom.* **1997**, *32* (4), 351–369. https://doi.org/10.1002/(SICI)1096-9888(199704)32:4<351::AID-JMS512>3.0.CO;2-Y.

(50) Ohta, T.; Bouchigny, S.; Didelez, J.-P.; Fujiwara, M.; Fukuda, K.; Kohri, H.; Kunimatsu, T.; Morisaki, C.; Ono, S.; Rouillé, G.; et al. HD Gas Analysis with Gas Chromatography and Quadrupole Mass Spectrometer. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **2011**, *640* (1), 241–246. https://doi.org/10.1016/j.nima.2011.02.091.
(51) Purves, R. W.; Guevremont, R. Electrospray Ionization High-Field Asymmetric Waveform Ion Mobility Spectrometry−Mass Spectrometry. *Anal. Chem.* **1999**, *71* (13), 2346–2357. https://doi.org/10.1021/ac981380y.
(52) Heal, D. J.; Smith, S. L.; Gosden, J.; Nutt, D. J. Amphetamine, Past and Present – a Pharmacological and Clinical Perspective. *J Psychopharmacol* **2013**, *27* (6), 479–496.
(53) Yu, S.; Zhu, L.; Shen, Q.; Bai, X.; Di, X. Recent Advances in Methamphetamine Neurotoxicity Mechanisms and Its Molecular Pathophysiology. *Behav. Neurol.* **2015**, *2015*, 1–11. https://doi.org/10.1155/2015/103969.
(54) Fitzgerald, R.; Ramos, J.; Bogema, S.; Poklis, A. Resolution of Methamphetamine Stereoisomers in Urine Drug Testing: Urinary Excretion of R(-)- Methamphetamine Following Use of Nasal Inhalers. *J. Anal. Toxicol.* **1988**, *12* (5), 255–259.
(55) Jirovský, D.; Lemr, K.; Ševčík, J.; Smysl, B.; Stránský. Methamphetamine — Properties and Analytical Methods of Enantiomer Determination. *Zdeněk Forensic Sci. Int.* **1998**, *96*, 61–70.
(56) Wang, T.; Shen, B.; Shi, Y.; Xiang, P.; Yu, Z. Chiral Separation and Determination of R/S-Methamphetamine and Its Metabolite R/S-Amphetamine in Urine Using LC–MS/MS. *Forensic Sci. Int.* **2015**, *246*, 72–78. https://doi.org/10.1016/j.forsciint.2014.11.009.
(57) Wang, T.; Yu, Z.; Shi, Y.; Xiang, P. Enantiomer Profiling of Methamphetamine in White Crystal and Tablet Forms (Ma Old) Using LC–MS-MS. *J. Anal. Toxicol.* **2015**, *39* (7), 551–556. https://doi.org/10.1093/jat/bkv060.
(58) Cahn, R. S.; Ingold, C. K.; Prelog, V. The Specification of Asymmetric Configuration in Organic Chemistry. *Experientia* **1956**, *12*, 81–94.
(59) Herráez-Hernández, R.; Campíns-Falcó, P.; Verdú-Andrés, J. Strategies for the Enantiomeric Determination of Amphetamine and Related Compounds by Liquid Chromatography. *J. Biochem. Biophys. Methods* **2002**, *54* (1–3), 147–167. https://doi.org/10.1016/S0165-022X(02)00137-9.
(60) Cherkaoui, S.; Rudaz, S.; Varesio, E.; Veuthey, J.-L. On-line Capillary Electrophoresis electrospray Mass Spectrometry for the Stereoselective Analysis of Drugs and Metabolites. *Electrophoresis* **2001**, *22*, 3308–3315.
(61) Liu, J. H.; Ku, W. W. Determination of Enantiomeric N-Trifluoroacetyl-L-Prolyl Chloride Amphetamine Derivatives by Capillary Gas Chromatography/Mass Spectrometry with Chiral and Achiral Stationary Phases. *Anal. Chem.* **1981**, *53* (14), 2180–2184.
(62) Eiceman, G.; Karpas, Z.; Hill, Jr., H. H. *Ion Mobility Spectrometry; 3rd Ed.*; CRC Press: Boca Raton, FL, 2013.
(63) Dwivedi, P.; Wu, C.; Matz, L. M.; Clowers, B. H.; Siems, W. F.; Hill, H. H. Gas-Phase Chiral Separations by Ion Mobility Spectrometry. *Anal. Chem.* **2006**, *78* (24), 8200–8206. https://doi.org/10.1021/ac0608772.
(64) Holness, H. K.; Jamal, A.; Mebel, A.; Almirall, J. R. Separation Mechanism of Chiral Impurities, Ephedrine and Pseudoephedrine, Found in Amphetamine-Type Substances Using Achiral Modifiers in the Gas Phase. *Anal. Bioanal. Chem.* **2012**, *404* (8), 2407–2416. https://doi.org/10.1007/s00216-012-6365-0.
(65) Schneider, B. B.; Covey, T. R.; Coy, S. L.; Krylov, E. V.; Nazarov, E. G. Planar Differential Mobility Spectrometer as a Pre-Filter for Atmospheric Pressure Ionization Mass Spectrometry. *Int. J. Mass Spectrom.* **2010**, *298* (1–3), 45–54. https://doi.org/10.1016/j.ijms.2010.01.006.
(66) Shvartsburg, A. A.; Creese, A. J.; Smith, R. D.; Cooper, H. J. Separation of Peptide Isomers with Variant Modified Sites by High-Resolution Differential Ion Mobility Spectrometry. *Anal. Chem.* **2010**, *82* (19), 8327–8334. https://doi.org/10.1021/ac101878a.

(67) Jónasdóttir, H. S.; Papan, C.; Fabritz, S.; Balas, L.; Durand, T.; Hardardottir, I.; Freysdottir, J.; Giera, M. Differential Mobility Separation of Leukotrienes and Protectins. *Anal. Chem.* **2015**, *87* (10), 5036–5040. https://doi.org/10.1021/acs.analchem.5b00786.

(68) Psutka, J. M.; Dion-Fortier, A.; Dieckmann, T.; Campbell, J. L.; Segura, P. A.; Hopkins, W. S. Identifying Fenton-Reacted Trimethoprim Transformation Products Using Differential Mobility Spectrometry. *Anal. Chem.* **2018**, *90* (8), 5352–5357. https://doi.org/10.1021/acs.analchem.8b00484.

(69) Mie, A.; Jörntén-Karlsson, M.; Axelsson, B.-O.; Ray, A.; Reimann, C. T. Enantiomer Separation of Amino Acids by Complexation with Chiral Reference Compounds and High-Field Asymmetric Waveform Ion Mobility Spectrometry: Preliminary Results and Possible Limitations. *Anal. Chem.* **2007**, *79* (7), 2850–2858. https://doi.org/10.1021/ac0618627.

(70) Mie, A.; Ray, A.; Axelsson, B.-O.; Jörntén-Karlsson, M.; Reimann, C. T. Terbutaline Enantiomer Separation and Quantification by Complexation and Field Asymmetric Ion Mobility Spectrometry−Tandem Mass Spectrometry. *Anal. Chem.* **2008**, *80* (11), 4133–4140. https://doi.org/10.1021/ac702262k.

(71) Campbell, J. L.; Le Blanc, J. Y.; Kibbey, R. G. Differential Mobility Spectrometry: A Valuable Technology for Analyzing Challenging Biological Samples. *Bioanalysis* **2015**, *7* (7), 853–856. https://doi.org/10.4155/bio.15.14.

(72) Płotka, J. M.; Biziuk, M.; Morrison, C. Common Methods for the Chiral Determination of Amphetamine and Related Compounds I. Gas, Liquid and Thin-Layer Chromatography. *TrAC Trends Anal. Chem.* **2011**, *30* (7), 1139–1158. https://doi.org/10.1016/j.trac.2011.03.013.

(73) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100. https://doi.org/10.1103/PhysRevA.38.3098.

(74) J Lecours, M.; C Theodore Chow, W.; Scott Hopkins, W. Density Functional Theory Study of Rh n S 0,± and Rh n +1 0,± (n = 1–9). **2014**, *118*. https://doi.org/10.1021/jp412457m.

(75) Hopkins, W. S.; Marta, R. A.; McMahon, T. B. Proton-Bound 3-Cyanophenylalanine Trimethylamine Clusters: Isomer-Specific Fragmentation Pathways and Evidence of Gas-Phase Zwitterions. *J. Phys. Chem. A* **2013**, *117* (41), 10714–10718. https://doi.org/10.1021/jp407766j.

(76) Fu, W.; Xiong, J.; J. Lecours, M.; J.J. Carr, P.; A. Marta, R.; Fillion, E.; Mcmahon, T.; Steinmetz, V.; Scott Hopkins, W. The Structures of Proton-Bound Dimers of Glycine with Phenylalanine and Pentafluorophenylalanine. **2016**, *330*. https://doi.org/10.1016/j.jms.2016.07.004.

(77) Wiberg, K. B.; Rablen, P. R. Comparison of Atomic Charges Derived via Different Procedures. *J. Comput. Chem.* **1993**, *14* (12), 1504–1518. https://doi.org/10.1002/jcc.540141213.

(78) Deng, D.; Deng, H.; Zhang, L.; Su, Y. Determination of Ephedrine and Pseudoephedrine by Field-Amplified Sample Injection Capillary Electrophoresis. *J. Chromatogr. Sci.* **2014**, *52* (4), 357–362. https://doi.org/10.1093/chromsci/bmt039.

(79) Chen, H.; Chen, X.; Pu, Q.; Hu, Z.; Zhao, Z.; Hooper, M. Separation and Determination of Ephedrine and Pseudoephedrine by Combination of Flow Injection with Capillary Electrophoresis. *J. Chromatogr. Sci.* **2003**, *41* (1), 1–5. https://doi.org/10.1093/chromsci/41.1.1.

(80) Iwanicki, R. M.; Maier, K.; Zlotnick, J. A.; Liu, R. H.; Kuo, T.-L.; Tagliaro, F. Separation of Enantiomeric Ephedrine and Pseudoephedrine—High Pressure Liquid Chromatography and Capillary Electrophoresis. *J. Forensic Sci.* **1999**, *44* (3), 14496J. https://doi.org/10.1520/JFS14496J.

(81) McCooeye, M.; Ding, L.; Gardner, G. J.; Fraser, C. A.; Lam, J.; Sturgeon, R. E.; Mester, Z. Separation and Quantitation of the Stereoisomers of Ephedra Alkaloids in Natural Health Products Using Flow Injection-Electrospray Ionization-High Field Asymmetric Waveform Ion Mobility Spectrometry-Mass Spectrometry. *Anal. Chem.* **2003**, *75* (11), 2538–2542. https://doi.org/10.1021/ac0342020.

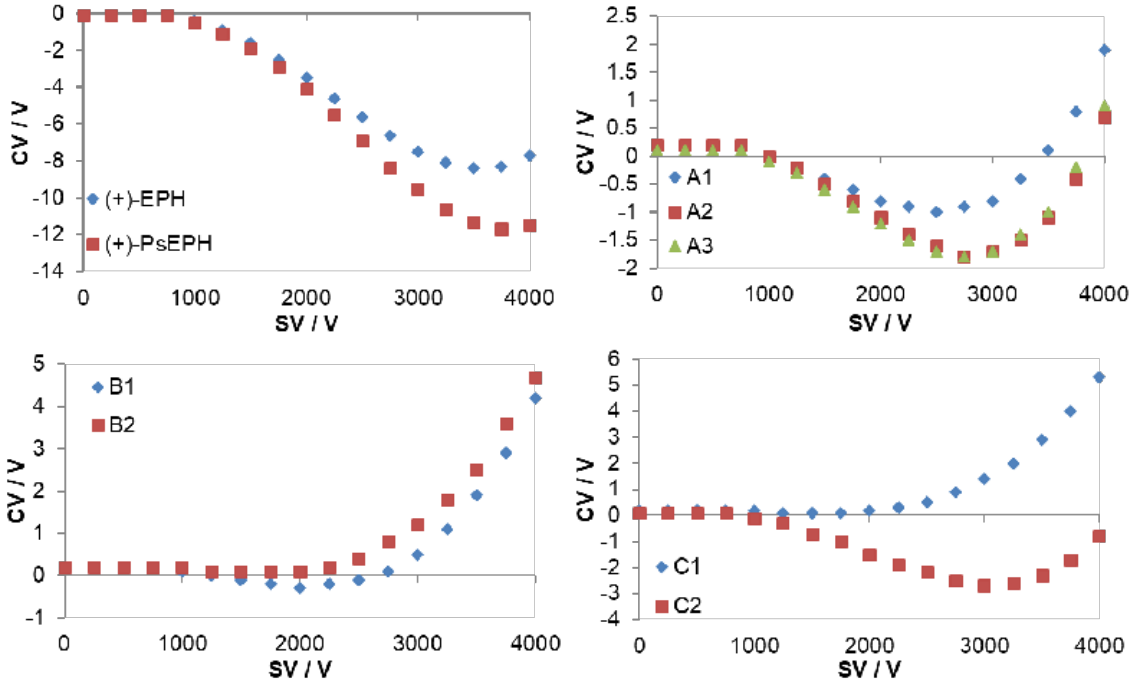
(82) Wang, S.-M.; Lewis, R. J.; Canfield, D.; Li, T.-L.; Chen, C.-Y.; Liu, R. H. Enantiomeric Determination of Ephedrines and Norephedrines by Chiral Derivatization Gas Chromatography–

Mass Spectrometry Approaches. *J. Chromatogr. B* **2005**, *825* (1), 88–95. https://doi.org/10.1016/j.jchromb.2005.01.016.
(83) Beyer, J.; Peters, F. T.; Kraemer, T.; Maurer, H. H. Detection and Validated Quantification of Nine Herbal Phenalkylamines and Methcathinone in Human Blood Plasma by LC-MS/MS with Electrospray Ionization. *J. Mass Spectrom.* **2007**, *42* (2), 150–160. https://doi.org/10.1002/jms.1132.
(84) Iwanicki, R. M.; Maier, K.; Zlotnick, J. A.; Liu, R. H.; Kuo, T.-L.; Tagliaro, F. Separation of Enantiomeric Ephedrine and Pseudoephedrine—High Pressure Liquid Chromatography and Capillary Electrophoresis. *J. Forensic Sci.* **1999**, *44* (3), 14496J. https://doi.org/10.1520/JFS14496J.
(85) Berger, T.; Berger, B. Separation of 9 Sulfonamide Drugs in ≈4 Min by Ultra-High Performance Supercritical Fluid Chromatography (UHPSFC): With a Feasibility Study for Detection in Milk. *Chromatographia* **2013**, *76* (23–24), 1631–1639. https://doi.org/10.1007/s10337-013-2539-y.
(86) Liu, R.; He, P.; Li, Z.; Li, R. Simultaneous Determination of 16 Sulfonamides in Animal Feeds by UHPLC-MS-MS. *J. Chromatogr. Sci.* **2011**, *49* (8), 640–646. https://doi.org/10.1093/chrsci/49.8.640.
(87) Perkins, J. R.; Games, D. E.; Startin, J. R.; Gilbert, J. Analysis of Sulphonamides Using Supercritical Fluid Chromatography and Supercritical Fluid Chromatography—Mass Spectrometry. *J. Chromatogr. A* **1991**, *540*, 239–256. https://doi.org/10.1016/S0021-9673(01)88813-0.
(88) Sheridan, R.; Policastro, B.; Thomas, S.; Rice, D. Analysis and Occurrence of 14 Sulfonamide Antibacterials and Chloramphenicol in Honey by Solid-Phase Extraction Followed by LC/MS/MS Analysis. *J. Agric. Food Chem.* **2008**, *56* (10), 3509–3516. https://doi.org/10.1021/jf800293m.
(89) Lin, C.-Y.; Huang, S.-D. Application of Liquid–Liquid–Liquid Microextraction and High-Performance Liquid-Chromatography for the Determination of Sulfonamides in Water. *Anal. Chim. Acta* **2008**, *612* (1), 37–43. https://doi.org/10.1016/j.aca.2008.02.008.
(90) Cai, Z.; Zhang, Y.; Pan, H.; Tie, X.; Ren, Y. Simultaneous Determination of 24 Sulfonamide Residues in Meat by Ultra-Performance Liquid Chromatography Tandem Mass Spectrometry. *J. Chromatogr. A* **2008**, *1200* (2), 144–155. https://doi.org/10.1016/j.chroma.2008.05.095.
(91) Berger, T.; Berger, B. Separation of 9 Sulfonamide Drugs in ≈4 Min by Ultra-High Performance Supercritical Fluid Chromatography (UHPSFC): With a Feasibility Study for Detection in Milk. *Chromatographia* **2013**, *76* (23–24), 1631–1639. https://doi.org/10.1007/s10337-013-2539-y.
(92) Tejada-Casado, C.; Hernández-Mesa, M.; Monteau, F.; Lara, F. J.; Olmo-Iruela, M. del; García-Campaña, A. M.; Le Bizec, B.; Dervilly-Pinel, G. Collision Cross Section (CCS) as a Complementary Parameter to Characterize Human and Veterinary Drugs. *Anal. Chim. Acta* **2018**, *1043*, 52–63. https://doi.org/10.1016/j.aca.2018.09.065.
(93) Premarathne, J. M. K. J. K.; Satharasinghe, D. A.; Gunasena, A. R. C.; Munasinghe, D. M. S.; Abeynayake, P. Establishment of a Method to Detect Sulfonamide Residues in Chicken Meat and Eggs by High-Performance Liquid Chromatography. *Food Control* **2017**, *72*, 276–282. https://doi.org/10.1016/j.foodcont.2015.12.012.
(94) Armentano, A.; Summa, S.; Lo Magro, S.; Palermo, C.; Nardiello, D.; Centonze, D.; Muscarella, M. Rapid Method for the Quantification of 13 Sulphonamides in Milk by Conventional High-Performance Liquid Chromatography with Diode Array Ultraviolet Detection Using a Column Packed with Core-Shell Particles. *J. Chromatogr. A* **2018**, *1531*, 46–52. https://doi.org/10.1016/j.chroma.2017.11.015.
(95) Haack, A.; Crouse, J.; Schlüter, F.-J.; Benter, T.; Hopkins, W. S. A First Principles Model of Differential Ion Mobility: The Effect of Ion- Solvent Clustering. *J Am Soc Mass Spectrom (In review)*, 51.
(96) Ieritano, C.; Campbell, J. L.; Hopkins, W. S. Unravelling the Factors That Drive Separation in Differential Mobility Spectrometry: A Case Study of Regioisomeric Phosphatidylcholine Adducts. *Int. J. Mass Spectrom.* **2019**, *444*, 116182. https://doi.org/10.1016/j.ijms.2019.116182.

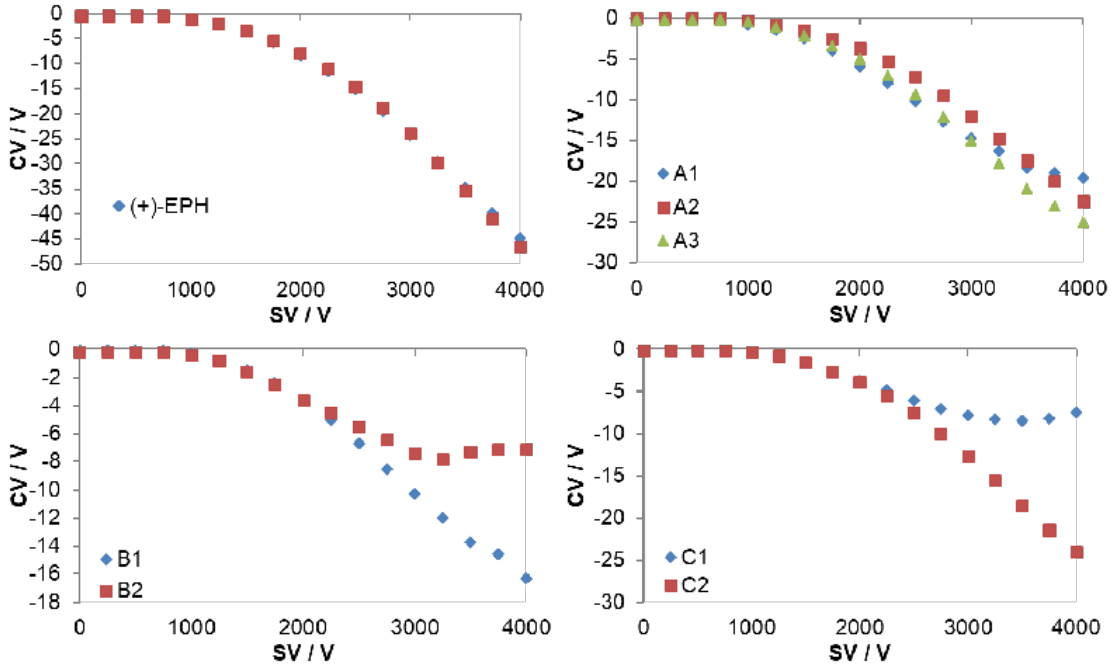
(97) Liu, C.; Le Blanc, J. C. Y.; Schneider, B. B.; Shields, J.; Federico, J. J.; Zhang, H.; Stroh, J. G.; Kauffman, G. W.; Kung, D. W.; Ieritano, C.; et al. Assessing Physicochemical Properties of Drug Molecules via Microsolvation Measurements with Differential Mobility Spectrometry. *ACS Cent. Sci.* **2017**, *3* (2), 101–109. https://doi.org/10.1021/acscentsci.6b00297.
(98) Panteleev, J.; Gao, H.; Jia, L. Recent Applications of Machine Learning in Medicinal Chemistry. *Bioorg. Med. Chem. Lett.* **2018**, *28* (17), 2807–2815. https://doi.org/10.1016/j.bmcl.2018.06.046.
(99) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2017**, *46* (D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037.
(100) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; et al. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2015**, *44* (D1), D1202–D1213. https://doi.org/10.1093/nar/gkv951.
(101) Seiler, K. P.; George, G. A.; Happ, M. P.; Bodycombe, N. E.; Carrinski, H. A.; Norton, S.; Brudz, S.; Sullivan, J. P.; Muhlich, J.; Serrano, M.; et al. ChemBank : A Small-Molecule Screening and Cheminformatics Resource Database. *Nucleic Acids Res.* **2007**, *36* (suppl_1), D351–D359. https://doi.org/10.1093/nar/gkm843.
(102) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20* (3), 273–297. https://doi.org/10.1023/A:1022627411411.
(103) Ding, S.; Li, H.; Su, C.; Yu, J.; Jin, F. Evolutionary Artificial Neural Networks: A Review. *Artif. Intell. Rev.* **2013**, *39* (3), 251–260. https://doi.org/10.1007/s10462-011-9270-6.
(104) Lee, E. Y.; Wong, G. C. L.; Ferguson, A. L. Machine Learning-Enabled Discovery and Design of Membrane-Active Peptides. *Bioorg. Med. Chem.* **2018**, *26* (10), 2708–2718. https://doi.org/10.1016/j.bmc.2017.07.012.
(105) Cao, L.; Zhu, P.; Zhao, Y.; Zhao, J. Using Machine Learning and Quantum Chemistry Descriptors to Predict the Toxicity of Ionic Liquids. *J. Hazard. Mater.* **2018**, *352*, 17–26. https://doi.org/10.1016/j.jhazmat.2018.03.025.
(106) von Lilienfeld, O. A. Quantum Machine Learning in Chemical Compound Space. *Angew. Chem. Int. Ed.* **2018**, *57* (16), 4164–4169. https://doi.org/10.1002/anie.201709686.
(107) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11* (5), 2087–2096. https://doi.org/10.1021/acs.jctc.5b00099.
(108) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432* (7019), 862–865. https://doi.org/10.1038/nature03197.
(109) Wellenzohn, B.; Lessel, U.; Beller, A.; Isambert, T.; Hoenke, C.; Nosse, B. Identification of New Potent GPR119 Agonists by Combining Virtual Screening and Combinatorial Chemistry. *J. Med. Chem.* **2012**, *55* (24), 11031–11041. https://doi.org/10.1021/jm301549a.
(110) Fu, W.; Hopkins, W. S. Applying Machine Learning to Vibrational Spectroscopy. *J. Phys. Chem. A* **2018**, *122* (1), 167–171. https://doi.org/10.1021/acs.jpca.7b10303.

Appendix A

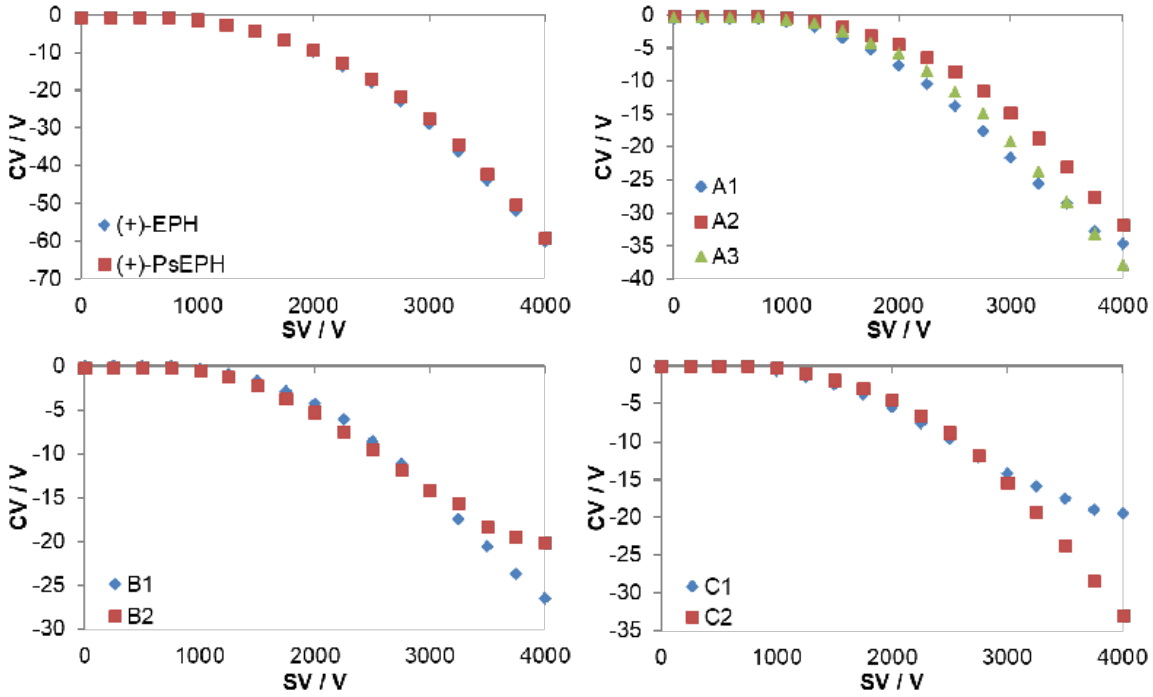
Chapter 4 Dispersion Plots:



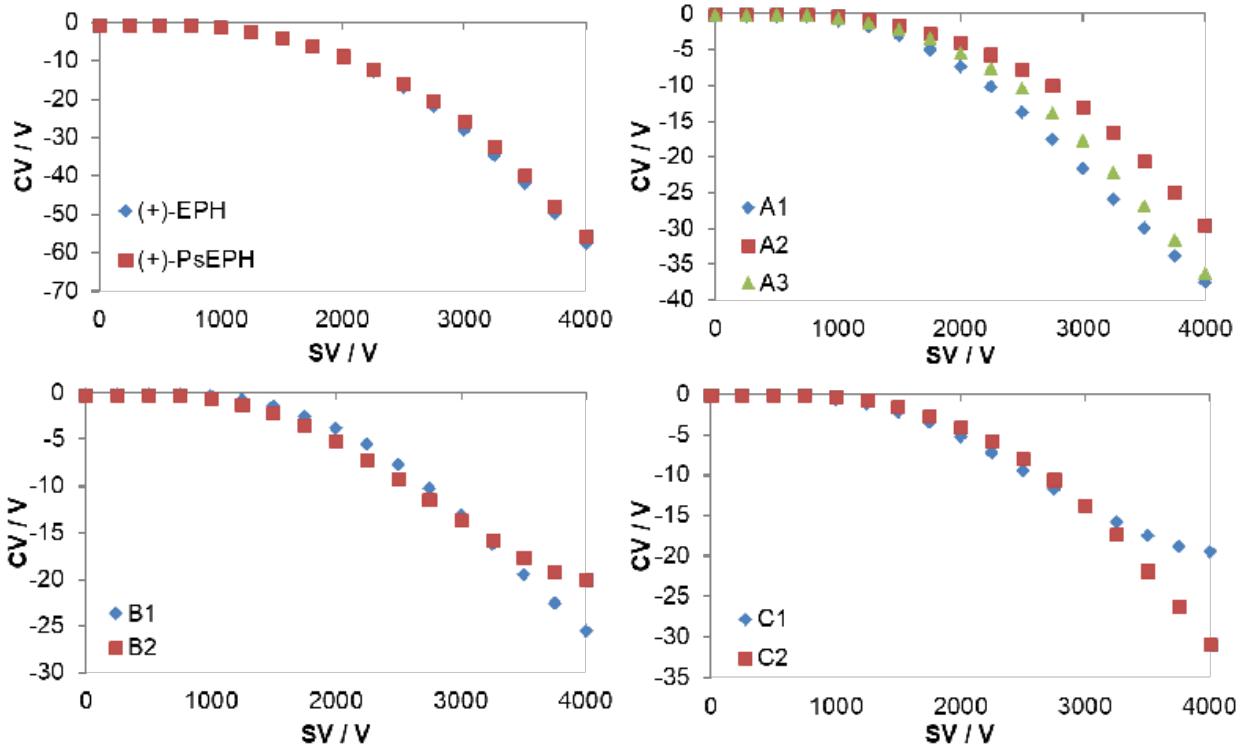
Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with H_2O gas modifier. CV was obtained from a fit to the ionogram of specific SV values.



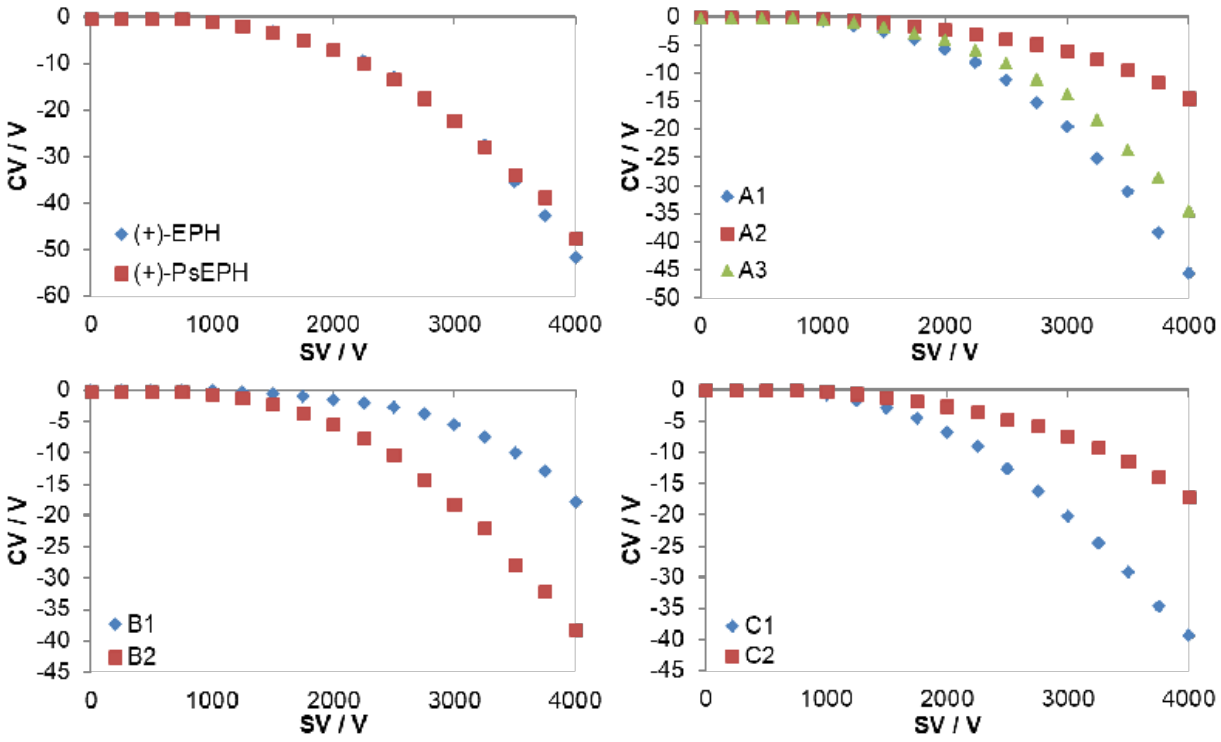
Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with EtOH gas modifier. CV was obtained from a fit to the ionogram of specific SV values.



Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with PrOH gas modifier. CV was obtained from a fit to the ionogram of specific SV values.



Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with IPA gas modifier. CV was obtained from a fit to the ionogram of specific SV values.



Dispersion plots for EPH, PsEPH, as well as the sulfonamide isomers (Figure 4.3) with AcOEt gas modifier. CV was obtained from a fit to the ionogram of specific SV values.

Appendix B

Monster Mix Compounds:

Compounds	m/z	Molecular Formula	SMILES
L-Glycine	76	C2H5NO2	C(C(=O)O)N
Sarcosine	89.8	C3H7NO2	CNCC(O)=O
L-Alanine	90.1	C3H7NO2	O=C(O)[C@@H](N)C
Choline	104.1	C5H14NO	C[N+](C)(C)CCO
γ-Aminobutyric acid (GABA)	104.1	C4H9NO2	C(CC(=O)O)CN
L-serine	106.1	C3H7NO3	C([C@@H](C(=O)O)N)O
Histamine	111.8	C5H9N3	NCCC1=C[N]C=N1
Cytosine	112	C4H5N3O	c1cnc(=O)[nH]c1N
Uracil	113	C4H4N2O2	O=C1NC=CC(=O)N1
L-proline	116.1	C5H9NO2	C1CC(NC1)C(=O)O
Fumaric Acid	117.1	C4H4O4	C(=C/C(=O)O)\C(=O)O
L-Valine	118.1	C5H11NO2	CC(C)[C@@H](C(=O)O)N
L-threonine	120	C4H9NO3	C[C@H]([C@@H](C(=O)O)N)O
L-cysteine	122.1	C6H12N2O4S2	C(C(C(=O)O)N)S
Niacinamide (Nicotinamide)	123.1	C6H6N2O	c1cc(cnc1)C(=O)N
Nicotinic acid (Niacin)	124.1	C6H5NO2	OC(=O)c1cccnc1
Methylhistamine	126	C6H11N3	CC(CC1=CN=CN1)N
Melamine	127.1	C3H6N6	c1(nc(nc(n1)N)N)N
Thymine	127.1	C5H6N2O2	Cc1c[nH]c(=O)[nH]c1=O
Vigabatrin	130	C6H11NO2	O=C(O)CCC(\C=C)N
L-Isoleucine	132.1	C6H13NO2	CC[C@H](C)[C@@H](C(=O)O)N
L-leucine	132.1	C6H13NO2	CC(C)C[C@@H](C(=O)O)N
L-asparagine	133.1	C4H8N2O3	C([C@@H](C(=O)O)N)C(=O)N

Compounds	m/z	Molecular Formula	SMILES
Malic Acid	135.1	C4H6O5	O=C(O)CC(O)C(=O)O
Adenine	136.0544	C5H5N5	NC1=NC=NC2=C1N=C[NH]2
(+-)-Amphetamine	136.1	C9H13N	NC(CC1=CC=CC=C1)C
Salicylic Acid	139	C7H6O3	c1ccc(c(c1)C(=O)O)O
Dimetridazole	141.9	C5H7N3O2	Cc1ncc([N](=O)=O)n1C
L-glutamine	147.1	C5H10N2O3	O=C(N)CCC(N)C(=O)O
L-lysine	147.2	C6H14N2O2	C(CCN)CC(C(=O)O)N
L-glutamic acid	148.1	C5H9NO4	C(CC(=O)O)C(C(=O)O)N
Methamphetamine	150	C10H15N	N(C(Cc1ccccc1)C)C
L-methionine	150.1	C5H11NO2S	CSCC[C@H](N)C(=O)O
Acetaminophen	152.2	C8H9NO2	CC(=O)Nc1ccc(O)cc1
Guanine	152.3	C5H5N5O	c1[nH]c2c(n1)c(=O)[nH]c(n2)N
Xanthine	153.1	C5H4N4O2	c1[nH]c2c(n1)nc(nc2O)O
L-histidine	156.1	C6H9N3O2	O=C([C@H](CC1=CNC=N1)N)O
Ephedrine	166.1	C10H15NO	O[C@H](c1ccccc1)[C@@H](NC)C
L-phenylalanine	166.1	C9H11NO2	C1=CC=C(C=C1)C[C@@H](C(=O)O)N
Pseudoephedrine	166.1	C10H15NO	O[C@@H](c1ccccc1)[C@@H](NC)C
Cyromazine	167.1	C6H10N6	C1CC1NC2=NC(=NC(=N2)N)N
Pyridoxal	168.1	C8H9NO3	O=Cc1c(O)c(C)ncc1CO
Gabapentin	172.1	C9H17NO2	O=C(O)CC1(CN)CCCCC1
L-arginine	175.2	C6H14N4O2	NC(CCCNC(N)=N)C(O)=O
Cis-Aconitic Acid	176.1	C6H6O6	O=C(O)CC(=CC(=O)O)C(=O)O
(-)-Cotinine	177.1	C10H12N2O	O=C2N(C)[C@H](c1cnccc1)CC2
Phenacetin	180	C10H13NO2	O=C(Nc1ccc(OCC)cc1)C
(+)-MDA (2,3-Methylenedioxyamphetamine)	180.1	C10H13NO2	CC(N)Cc2c1OCOc1ccc2
L-Tyrosine	182.1	C9H11NO3	N[C@@H](Cc1ccc(O)cc1)C(O)=O
Selegiline	188.1	C13H17N	C#CCN([C@@H](Cc1ccccc1)C)C
Caffeine	195.1	C8H10N4O2	CN1C=NC2=C1C(=O)N(C(=O)N2C)C
D-glucose	203.1	C6H12O6	C([C@@H]1[C@H]([C@@H]([C@H]([C@H](O1)O)O)O)O)O
R-(-)-Nirvanol	205.1	C11H12N2O2	CCC1(C(=O)NC(=O)N1)C2=CC=CC=C2

Compounds	m/z	Molecular Formula	SMILES
L-tryptophan	205.1	C11H12N2O2	c1ccc2c(c1)c(c[nH]2)C[C@@H](C(=O)O)N
S-(+)-Ibuprofen	207.1	C13H18O2	CC(C)Cc1ccc(cc1)[C@@H](C)C(=O)O
monoxidil	210.1	C9H15N5O	Nc1cc(nc(N)[n+]1[O-])N2CCCCC2
Butabarbital	213.1	C10H16N2O3	O=C1NC(=O)NC(=O)C1(C(C)CC)CC
Meprobamate	219.1	C9H18N2O4	O=C(OCC(COC(=O)N)(C)CCC)N
D-Pantothenic acid	220.1	C9H17NO5	CC(C)(CO)C(C(=O)NCCC(=O)O)O
Ritalinic Acid	220.1	C13H17NO2	C1CCNC(C1)C(C2=CC=CC=C2)C(=O)O
Hydroxynirvanol	221.1	C11H12N2O3	CCC1(NC(=O)NC1=O)c2ccc(O)cc2
Ethylone	222.1	C12H15NO3	CC(NCC)C(=O)c1ccc2OCOc2c1
Tapentadol	222.1	C14H23NO	Oc1cc(ccc1)[C@@H]([C@@H](C)CN(C)C)CC
Acetamiprid	223.1	C10H11ClN4	Clc1ncc(cc1)CN(\C(=N\C#N)C)C
Terbutaline	226.1	C12H19NO3	Oc1cc(cc(O)c1)C(O)CNC(C)(C)C
Amobarbital	227.1	C11H18N2O3	O=C1NC(=O)NC(=O)C1(CCC(C)C)CC
Clonidine	230	C9H9Cl2N3	Clc1c(c(Cl)ccc1)N/C2=N/CCN2
Naproxen	231.1	C14H14O3	C[C@@H](C1=CC2=C(C=C1)C=C(C=C2)OC)C(=O)O
Phenobarbital	231.1	C12H12N2O3	O=C1NC(=O)NC(=O)C1(c2ccccc2)CC
Melatonin	233.1	C13H16N2O2	COC1=CC2=C(NC=C2CCNC(C)=O)C=C1
Normeperidine	234.1	C14H19NO2	O=C(C1(CCN(CC1)C)C2=CC=CC=C2)OCC
Lidocaine	235.1	C14H22N2O	O=C(Nc1c(cccc1C)C)CN(CC)CC
Carbamazepine	237.1	C15H12N2O	c1ccc2c(c1)C=Cc3ccccc3N2C(=O)N
Secobarbital	239.1	C12H18N2O3	O=C1NC(NC(C1(CC=C)C(CCC)C)=O)=O
Salbutamol	240.1	C13H21NO3	OCc1cc(ccc1O)C(O)CNC(C)(C)C
Bentazon	241	C10H12N2O3S	O=C1N(C(C)C)S(=O)(=O)Nc2ccccc12
L-cystine	241.1	C3H7NO2S	C(C(C(=O)O)N)SSCC(C(=O)O)N
Cytidine	244.1	C9H13N3O5	O=C1/N=C(/N)\C=C/N1[C@@H]2O[C@@H]([C@@H](O)[C@H]2O)CO

Compounds	m/z	Molecular Formula	SMILES
Uridine	245.1	C9H12N2O6	O=C1NC(=O)N(C=C1)[C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O
O-Desmethyl-cis-tramadol HCL	250.1	C15H23NO2	CN(C)C[C@H]1CCCC[C@@]1(C2=CC(=CC=C2)O)O
Lacosamide	251.1	C13H18N2O3	O=C(N[C@@H](C(=O)NCc1ccccc1)COC)C
Sulfadiazine	251.1	C10H10N4O2S	C1=CN=C(N=C1)NS(=O)(=O)C2=CC=C(C=C2)N
Cimetidine	253.1	C10H16N6S	N#CN=C(NC)NCCSCc1nc[nH]c1CN#CN\C(=N/C)NCCSCc1nc[nH]c1C
Ketoprofen	255.1	C16H14O3	CC(c1cccc(c1)C(=O)c2ccccc2)C(=O)O
Lamotrigine	256	C9H7Cl2N5	NC1=NC(N)=NN=C1C2=CC=CC(Cl)=C2Cl
Ketorolac	256.1	C15H13NO3	O=C(c1ccc2n1CCC2C(=O)O)c3ccccc3
Diphenhydramine	256.2	C17H21NO	O(CCN(C)C)C(c1ccccc1)c2ccccc2
Imidacloprid	256.2	C9H10ClN5O2	[O-][N+](=O)NC/1=N/CCN\1Cc2cnc(Cl)cc2
phenyltoloxamine	256.2	C17H21NO	O(c1ccccc1Cc2ccccc2)CCN(C)C
Palmitic acid	257.2	C16H32O2	CCCCCCCCCCCCCCCC(=O)O
(+-)-Propanolol	260	C16H21NO2	CC(NCC(O)COC1=C(C=CC=C2)C2=CC=C1)C
Carisoprodol	261.1	C12H24N2O4	O=C(OCC(COC(=O)NC(C)C)(C)CCC)N
Bufuralol	262	C16H23NO2	OC(c2oc1c(cccc1c2)CC)CNC(C)(C)C
Nortriptyline	264	C19H21N	c3cc2c(/C(c1c(cccc1)CC2)=C/CCNC)cc3
Protriptyline	264	C19H21N	c3cc2c(\C=C/c1c(cccc1)C2CCCNC)cc3
Tramadol	264.2	C16H25NO2	CN(C)C[C@H]1CCCC[C@@]1(C2=CC(=CC=C2)OC)O
Sulfamerazine	265	C11H12N4O2S	O=S(=O)(Nc1nc(ccn1)C)c2ccc(N)cc2
Oxprenolol	266.2	C15H23NO3	O(c1ccccc1OC\C=C)CC(O)CNC(C)C
Atenolol	267	C14H22N2O3	O=C(N)Cc1ccc(cc1)OCC(O)CNC(C)C
Desipramine	267.2	C18H22N2	c1cc3c(cc1)CCc2c(cccc2)N3CCCNC

Compounds	m/z	Molecular Formula	SMILES
Adenosine	268.1	C10H13N5O4	n2c1c(ncnc1n(c2)[C@@H]3O[C@@H]([C@@H](O)[C@H]3O)CO)N
Tolbutamide	271.1	C12H18N2O3S	O=S(=O)(c1ccc(cc1)C)NC(=O)NCCCC
Dehydroepiandrosterone (DHEA)	271.2	C19H28O2	O=C3[C@]2(CC[C@@H]1[C@@]4(C(=C/C[C@H]1[C@@H]2CC3)\C[C@@H](O)CC4)C)C
Desomorphine	272	C17H21NO2	OC1=C2O[C@@H]3[C@@]45C2=C(C[C@H]([C@@H]5CCC3)N(C)CC4)C=C1
Sotalol	273.1	C12H20N2O3S	O=S(=O)(Nc1ccc(cc1)C(O)CNC(C)C)C
Norsertraline HCL	275.1	C16H15Cl2N	Clc1ccc(cc1Cl)[C@H]3c2c(cccc2)[C@@H](N)CC3
Clenbuterol	277.1	C12H18Cl2N2O	Clc1cc(cc(Cl)c1N)C(O)CNC(C)(C)C
Bromoxynil	277.9	C7H3Br2NO	Brc1cc(C#N)cc(Br)c1O
Amitriptyline	278.1	C20H23N	c3cc2c(/C(c1c(cccc1)CC2)=C\CCN(C)C)cc3
Sulfamethazine (Sulfadimidine)	278.9	C12H14N4O2S	O=S(=O)(Nc1nc(cc(n1)C)C)c2ccc(N)cc2
Indoprofen	282.1	C17H15NO3	O=C(O)C(c1ccc(cc1)N3C(=O)c2ccccc2C3)C
Guanosine	284.3	C10H13N5O5	c1nc2c(=O)[nH]c(nc2n1[C@H]3[C@@H]([C@@H]([C@H](O3)CO)O)O)N
Diazepam	285.1	C16H13ClN2O	CN1C2=C(C(C3=CC=CC=C3)=NCC1=O)C=C(Cl)C=C2
7-Aminoclonazepam	286.1	C15H12ClN3O	C1C(=O)NC2=C(C=C(C=C2)N)C(=N1)C3=CC=CC=C3Cl
Morphine	286.1	C17H19NO3	CN1CC[C@]23C4=C5C=CC(O)=C4O[C@H]2[C@@H](O)C=C[C@H]3[C@H]1C5
Norcodeine	286.1	C17H19NO3	COC1=C2C3=C(C[C@@H]4[C@H]5[C@]3(CCN4)[C@@H](O2)[C@H](C=C5)O)C=C1
Oxazepam	287.1	C15H11ClN2O2	OC1N=C(C2=C(NC1=O)C=CC(Cl)=C2)C3=CC=CC=C3
Etodolac	288	C17H21NO3	O=C(O)CC3(OCCc2c3nc1c(cccc12)CC)CC

Compounds	m/z	Molecular Formula	SMILES
Epitestosterone	289.1	C19H28O2	C[C@]12CC[C@H]3[C@H]([C@@H]1CC[C@H]2O)CCC4=CC(=O)CC[C@]34C
Testosterone	289.2	C19H28O2	O=C4\C=C2/[C@]([C@H]1CC[C@@]3([C@@H](O)CC[C@H]3[C@@H]1CC2)C)(C)CC4
Benzoylecgonine	290.1	C16H19NO4	CN1[C@H]2CC[C@@H]1[C@H]([C@H](C2)OC(=O)c3ccccc3)C(=O)O
Hydrocodone	300.2	C18H21NO3	O=C4[C@@H]5Oc1c2c(ccc1OC)C[C@H]3N(CC[C@]25[C@H]3CC4)C
Temazepam	301.1	C16H13ClN2O2	CN1C2=C(C(C3=CC=CC=C3)=NC(O)C1=O)C=C(Cl)C=C2
Noroxycodone HCL	302.1	C17H19NO4	COC1=C2C3=C(C[C@@H]4[C@]5([C@]3(CCN4)[C@@H](O2)C(=O)CC5)O)C=C1
Oxymorphone	302.1	C17H19NO4	O=C1[C@@H]2OC3=C(O)C=CC4=C3[C@@]2([C@]5(CC1)O)CCN(C)[C@@H]5C4
Cocaine	304.1	C17H21NO4	CN1[C@H]2CC[C@@H]1[C@@H](C(OC)=O)[C@@H](OC(C3=CC=CC=C3)=O)C2
Zolpidem	308	C19H21N3O	O=C(CC1=C(C2=CC=C(C)C=C2)N=C3C=CC(C)=CN13)N(C)C
Alprazolam	309.1	C17H13ClN4	ClC1=CC2=C(C=C1)N3C(C)=NN=C3CN=C2C4=CC=CC=C4
Warfarin	309.1	C19H16O4	CC(=O)CC(C\1=C(/O)c2ccccc2OC/1=O)c3ccccc3
(+)-Methadone	310.2	C21H27NO	CCC(C(C1=CC=CC=C1)(C2=CC=CC=C2)CC(N(C)C)C)=O
Safranin O	315.1	C20H19ClN4	n1c4c([n+](c2c1cc(c(N)c2)C)c3ccccc3)cc(c(c4)C)N
Progesterone	315.2	C21H30O2	CC(=O)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4=CC(=O)CC[C@]34C)C
ranitidine	315.2	C13H22N4O3S	[O-][N+](=O)C=C(NC)NCCSCc1ccc(o1)CN(C)C
Chlorprothixene	316.1	C18H18ClNS	Clc2cc1C(\c3c(Sc1cc2)cccc3)=C/CCN(C)C

Compounds	m/z	Molecular Formula	SMILES
Clonazepam	316.1	C15H10ClN3O3	[O-][N+](C1=CC2=C(C=C1)NC(CN=C2C3=CC=CC=C3Cl)=O)=O
Flusilazole	316.1	C16H15F2N3Si	c2ncnn2C[Si](C)(c(cc1)ccc1F)c3ccc(F)cc3
Pamaquine	316.2	C19H29N3O	O(c1cc(NC(C)CCCN(CC)CC)c2ncccc2c1)C
Fluvoxamine	319.2	C15H21F3N2O2	FC(F)(F)c1ccc(\C(=N\OCCN)CCCCOC)cc1
Chloramphenicol	323	C11H12Cl2N2O5	c1cc(ccc1[C@H]([C@@H](CO)NC(=O)C(Cl)Cl)O)[N+](=O)[O-]
Quinidine	325.2	C20H24N2O2	O(c4cc1c(nccc1[C@H](O)[C@@H]2N3CC[C@@H](C2)[C@@H](/C=C)C3)cc4)C
Midazolam	326.1	C18H13ClFN3	ClC1=CC=C2C(C(C3=CC=CC=C3F)=NCC4=CN=C(C)N42)=C1
Naloxone	328.1	C19H21NO4	O=C1[C@@H]2OC3=C(O)C=CC4=C3[C@@]2([C@]5(CC1)O)CCN(CC=C)[C@@H]5C4
Acebutolol	337	C18H28N2O4	O=C(Nc1ccc(OCC(O)CNC(C)C)c(c1)C(=O)C)CCC
Fentanyl	337.2	C22H28N2O	O=C(CC)N(C1CCN(CC1)CCc2ccccc2)c3ccccc3
Chlorthalidone	338.9	C14H11ClN2O4S	O=S(=O)(N)c1c(Cl)ccc(c1)C2(O)c3ccccc3C(=O)N2
(+)-Propoxyphene	340.1	C22H29NO2	O=C(CC)O[C@@](CC1=CC=CC=C1)([C@@H](CN(C)C)C)C2=CC=CC=C2
Sulpiride	342.1	C15H23N3O4S	O=S(=O)(N)c1cc(c(OC)cc1)C(=O)NCC2N(CC)CCC2
Propafenone	342.2	C21H27NO3	O=C(c1ccccc1OCC(O)CNCCC)CCc2ccccc2
Sucrose + H	343.3	C12H22O11	O1[C@H](CO)[C@@H](O)[C@H](O)[C@@H](O)[C@H]1O[C@@]2(O[C@@H]([C@@H](O)[C@@H]2O)CO)CO
Omeprazole	346.2	C17H19N3O3S	Cc1c(OC)c(C)cnc1CS(=O)c2nc3ccc(OC)cc3n2

Compounds	m/z	Molecular Formula	SMILES
Ampicillin	350	C16H19N3O4S	O=C(O)[C@@H]2N3C(=O)[C@@H](NC(=O)[C@@H](c1ccccc1)N)[C@H]3SC2(C)C
Buscopan	360.1	C21H30NO4	OC[C@H](c1ccccc1)C(=O)O[C@@H]2C[C@@H]3[C@H]4O[C@H]4[C@H](C2)[N+]3(C)CCCC
Prednisolone	361.1	C21H28O5	O=C\1\C=C/[C@]4(/C(=C/1)CC[C@@H]2[C@@H]4[C@@H](O)C[C@@]3([C@@](O)(C(=O)CO)CC[C@@H]23)C)C
D-trehalose	365.3	C12H22O11	OC[C@@H]1[C@@H](O)[C@H](O)[C@@H](O)[C@H](O1)O[C@@H]2[C@H](O)[C@@H](O)[C@H](O)[C@H](O2)CO
Amoxicillin	365.7	C16H19N3O5S	O=C(O)[C@@H]2N3C(=O)[C@@H](NC(=O)[C@@H](c1ccc(O)cc1)N)[C@H]3SC2(C)C
Tamoxifen	372.1	C26H29NO	CN(C)CCOc1ccc(cc1)/C(c2ccccc2)=C(/CC)c3ccccc3
Haloperidol	376.1	C21H23ClFNO2	c1cc(ccc1C(=O)CCCN2CCC(CC2)(c3ccc(cc3)Cl)O)F
25-Hydroxyvitamin D3	383.4	C27H44O2 (loses H2O easily for form C27H42O)	O[C@@H]1CC(\C(=C)CC1)=C\C=C2/CCC[C@]3([C@H]2CC[C@@H]3[C@H](C)CCCC(O)(C)C)C
Prazosin	383.8	C19H21N5O4	O=C(N3CCN(c2nc1cc(OC)c(OC)cc1c(n2)N)CC3)c4occc4
Buspirone	386.3	C21H31N5O2	O=C1N(CCCCN2CCN(CC2)C3=NC=CC=N3)C(=O)CC4(CCCC4)C1
Urapidil	388.2	C20H29N5O3	O=C1\C=C(/N(C(=O)N1C)C)NCCCN3CCN(c2ccccc2OC)CC3
Perphenazine	404.2	C21H26ClN3OS	Clc2cc1N(c3c(Sc1cc2)cccc3)CCCN4CCN(CCO)CC4
Lovastatin	405.2	C24H36O5	O=C(O[C@@H]1[C@H]3C(=C/[C@H](C)C1)\C=C/[C@@H]([C@@H]3CC[C@H]2OC(=O)C[C@H](O)C2)C)[C@@H](C)CC

Compounds	m/z	Molecular Formula	SMILES
Norbuprenorphine	414.3	C25H35NO4	C[C@]([C@H]1C[C@@]23CC[C@@]1([C@H]4[C@@]25CCN[C@@H]3Cc6c5c(c(cc6)O)O4)OC)(C(C)(C)C)O
Bendroflumethiazide	421.8	C15H14F3N3O4S2	FC(F)(F)c3c(cc1c(NC(NS1(=O)=O)Cc2ccccc2)c3)S(=O)(=O)N
Adenosine diphosphate (ADP)	428.3	C10H15N5O10P2	c1nc(c2c(n1)n(cn2)[C@H]3[C@@H]([C@@H]([C@H](O3)COP(=O)(O)OP(=O)(O)O)O)O)N
Angiotensin I 3+	433.1	C62H89N17O14	CCC(C)C(C(=O)NC(Cc1cnc[nH]1)C(=O)N2CCCC2C(=O)NC(Cc3ccccc3)C(=O)O)NC(=O)C(Cc4ccc(cc4)O)NC(=O)C(C(C)C)NC(=O)C(CCCN=C(N)N)NC(=O)C(CC(=O)O)N
Spirmycin	442.4	C43H74N2O14	O=CCC4C(OC2OC(C(OC1OC(C)C(O)C(O)(C)C1)C(N(C)C)C2O)C)C(OC)C(O)CC(=O)OC(C)C\C=C\C=C\C(OC3OC(C)C(N(C)C)CC3)C(C)C4
Verapamil	455.3	C27H38N2O4	N#CC(c1cc(OC)c(OC)cc1)(CCCN(CCc2ccc(OC)c(OC)c2)C)C(C)C
Buprenorphine	468.3	C29H41NO4	Oc7ccc5c1c7O[C@H]3[C@]6(OC)[C@H](C[C@@]2([C@H](N(CC[C@@]123)CC4CC4)C5)CC6)[C@@](O)(C)C(C)(C)C
Terfenadine	472.3	C32H41NO2	OC(c1ccccc1)(c2ccccc2)C4CCN(CCCC(O)c3ccc(cc3)C(C)(C)C)CC4
Loperamide	477.2	C29H33ClN2O2	ClC1=CC=C(C2(CCN(CC2)CCC(C3=CC=CC=C3)(C(N(C)C)=O)C4=CC=CC=C4)O)C=C1
Raffinose	505.3	C18H32O16	C([C@@H]1[C@@H]([C@@H]([C@H]([C@H](O1)OC[C@@H]2[C@H]([C@@H]([C@H]([C@H](O2)O[C@]3([C@H]([C@@H]([C@H](O3)CO)O)O)CO)O)O)O)O)O)O)O
Taurocholic Acid	516.3	C26H45NO7S	C[C@H](CCC(=O)NCCS(=O)(=O)O)[C@H]1CC[C@@H]2[C@@]1([C@H](C[C@H]3[C@H]2[C@@H](C[C@H]4[C@@]3(CC[C@H](C4)O)C)O)O)C
Angiotensin II 2+	523.9	C50H71N13O12	CCC(C)C(C(=O)NC(Cc1cnc[nH]1)C(=O)N2CCCC2C(=O)NC(Cc3ccccc3)C(=O)O)NC(=O)C(Cc4ccc(cc4)O)NC(=O)C(C(C)C)NC(=O)C(CCCN=C(N)N)NC(=O)C(CC(=O)O)N

Compounds	m/z	Molecular Formula	SMILES
Bradykinin 2+	530.8	C50H73N15O11	O=C(N[C@H](C(=O)N[C@H](C(=O)O)CCC/N=C(\N)N)Cc1ccccc1)[C@H]5N(C(=O)[C@@H](NC(=O)[C@@H](NC(=O)CNC(=O)[C@H]3N(C(=O)[C@H]2N(C(=O)[C@@H](N)CCC/N=C(\N)N)CCC2)CCC3)Cc4ccccc4)CO)CCC5
Ketoconazole	531.1	C26H28Cl2N4O4	O=C(N5CCN(c4ccc(OC[C@@H]1O[C@](OC1)(c2ccc(Cl)cc2Cl)Cn3ccnc3)cc4)CC5)C
Renin Substrate 2+	586.7	C85H123N21O20	O=C(C(NC(=O)C(CC(C)C)NC(C(NC(=O)C(Cc(c6)ncn6)NC(=O)C(NC(=O)C(C5)N(CC5)C(=O)C(Cc(n4)cnc4)NC(=O)C(NC(C(NC(=O)C(C(C)C)NC(=O)C(NC(=O)C(N)CC(O)=O)CCCNC(N)=N)Cc(c3)ccc(O)c3)=O)C(C)CC)Cc(c2)cccc2)CC(C)C)=O)C(C)C)NC(C(NC(CO)C(O)=O)=O)Cc(c1)ccc(c1)O
(Parathyroid hormone 1-34) 7+	588.8	C181H291N55O51S2	[H]N[C@@H](CO)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CO)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H]([C@@H](C)CC)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CC(C)C)C(=O)NCC(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CO)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H](CCCNC(N)=N)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H](CC1=CNC2=C1C=CC=C2)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCCNC(N)=N)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CC(O)=O)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CC1=CC=CC=C1)C(O)=O

Compounds	m/z	Molecular Formula	SMILES
Cyclosporin A (+2)	602.5	C62H111N11O12	CC[C@H]1C(=O)N(CC(=O)N([C@H](C(=O)N[C@H](C(=O)N([C@H](C(=O)N[C@H](C(=O)N[C@@H](C(=O)N([C@H](C(=O)N([C@H](C(=O)N([C@H](C(=O)N([C@H](C(=O)N1)[C@@H]([C@H](C)C/C=C/C)O)C)C(C)C)C)CC(C)C)CC(C)C)C)C)CC(C)C)C)C(C)C)CC(C)C)C
Reserpine	609.2	C33H40N2O9	O=C(OC)[C@H]6[C@H]4C[C@@H]3c2nc1cc(OC)ccc1c2CCN3C[C@H]4C[C@@H](OC(=O)c5cc(OC)c(OC)c(OC)c5)[C@@H]6OC
Angiotensin I 2+	649.1	C62H89N17O14	CCC(C)C(C(=O)NC(Cc1cnc[nH]1)C(=O)N2CCCC2C(=O)NC(Cc3ccccc3)C(=O)O)NC(=O)C(Cc4ccc(cc4)O)NC(=O)C(C(C)C)NC(=O)C(CCCN=C(N)N)NC(=O)C(CC(=O)O)N
Bromocriptine	654.1	C32H40BrN5O5	BrC1=C(C[C@H]2N(C)C3)C4=C(C=CC=C4C2=C[C@H]3C(N[C@]5(C(C)C)O[C@@]6(N([C@@H](CC(C)C)C(N7CCC[C@H]76)=O)C5=O)O)=O)N1
(Parathyroid hormone 1-34) 6+	686.8	C181H291N55O51S2	[H]N[C@@H](CO)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CO)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H]([C@@H](C)CC)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CC(C)C)C(=O)NCC(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CO)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H](CCCNC(N)=N)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H](CC1=CNC2=C1C=CC=C2)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCCNC(N)=N)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CC(O)=O)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CC1=CC=CC=C1)C(O)=O

Compounds	m/z	Molecular Formula	SMILES
Ritonavir	720.8	C37H48N6O5S2	CC(C)c4nc(CN(C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc1ccccc1)C[C@H](O)[C@H](Cc2ccccc2)NC(=O)OCc3cncs3)cs4
Vancomycin 2+	724.7	C66H75Cl2N9O24	C[C@H]1[C@H]([C@@](C[C@@H](O1)O[C@@H]2[C@H]([C@@H]([C@H](O[C@H]2Oc3c4cc5cc3Oc6ccc(cc6Cl)[C@H]([C@H](C(=O)N[C@H](C(=O)N[C@H]5C(=O)N[C@@H]7c8ccc(c(c8)-c9c(cc(cc9O)O)[C@H](NC(=O)[C@H]([C@@H](c1ccc(c(c1)Cl)O4)O)NC7=O)C(=O)O)O)CC(=O)N)NC(=O)[C@@H](CC(C)C)NC)O)CO)O)O)(C)N)O
Erythromycin	734.5	C37H67NO13	CC[C@@H]1[C@@]([C@@H]([C@H](C(=O)[C@@H](C[C@@]([C@@H]([C@H]([C@@H]([C@H](C(=O)O1)C)O[C@H]2C[C@@]([C@H]([C@@H](O2)C)O)(C)OC)C)O[C@H]3[C@@H]([C@H](C[C@H](O3)C)N(C)C)O)(C)O)C)C)O)(C)O
Nigercin	747.7	C40H68O11	OC(=O)[C@H](C)[C@@H]1O[C@H](CC[C@@H]1C)C[C@H]6O[C@]2(O[C@@](C)(C[C@H]2C)[C@H]3CC[C@](C)(O3)[C@@H]4O[C@H](C[C@@H]4C)[C@H]5O[C@@](O)(CO)[C@H](C)C[C@@H]5C)[C@H](C)[C@H](OC)C6
Azithromycin	749.5	C38H72N2O12	CN(C)[C@H]3C[C@@H](C)O[C@@H](O[C@@H]2[C@@H](C)[C@H](O[C@H]1C[C@@](C)(OC)[C@@H](O)[C@H](C)O1)[C@@H](C)C(=O)O[C@H](CC)[C@@](C)(O)[C@H](O)[C@@H](C)N(C)C[C@H](C)C[C@@]2(C)O)[C@@H]3O

Compounds	m/z	Molecular Formula	SMILES
Digitoxin	765.4	C41H64O13	O=C\1OC/C(=C/1)[C@H]2CC[C@@]8(O)[C@]2(C)CC[C@H]7[C@H]8CC[C@H]6[C@]7(C)CC[C@H](O[C@@H]5O[C@H](C)[C@@H](O[C@@H]4O[C@@H]([C@@H](O[C@@H]3O[C@@H]([C@@H](O)[C@@H](O)C3)C)[C@@H](O)C4)C)[C@@H](O)C5)C6
Digoxin	781.4	C41H64O14	O=C\1OC/C(=C/1)[C@H]2CC[C@@]8(O)[C@]2(C)[C@H](O)C[C@H]7[C@H]8CC[C@H]6[C@]7(C)CC[C@H](O[C@@H]5O[C@H](C)[C@@H](O[C@@H]4O[C@@H]([C@@H](O[C@@H]3O[C@@H]([C@@H](O)[C@@H](O)C3)C)[C@@H](O)C4)C)[C@@H](O)C5)C6
Glycyrrhizic Acid	823.4	C42H62O16	O=C(O)[C@H]7O[C@@H](O[C@@H]6[C@@H](O)[C@H](O)[C@H](O[C@@H]6O[C@@H]2C(C)(C)[C@@H]3CC[C@@]1(C)[C@]5(C(=C/C(=O)[C@@H]1[C@@]3(C)CC2)\[C@@H]4C[C@](C(=O)O)(C)CC[C@]4(C)CC5)C)C(=O)O)[C@H](O)[C@@H](O)[C@@H]7O
Rifampicin	823.4	C43H58N4O12	CN1CCN(CC1)/N=C/c2c(O)c3c5C(=O)[C@@]4(C)O/C=C/[C@H](OC)[C@@H](C)[C@@H](OC(C)=O)[C@H](C)[C@H](O)[C@H](C)[C@@H](O)[C@@H](C)\C=C\C=C(\C)C(=O)Nc2c(O)c3c(O)c(C)c5O4

Compounds	m/z	Molecular Formula	SMILES
(Parathyroid hormone 1-34) 5+	823.8	C181H291N55O51S2	[H]N[C@@H](CO)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CO)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H]([C@@H](C)CC)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CC(C)C)C(=O)NCC(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CO)C(=O)N[C@@H](CCSC)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H](CCCNC(N)=N)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CCC(O)=O)C(=O)N[C@@H](CC1=CNC2=C1C=CC=C2)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCCNC(N)=N)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](CC(C)C)C(=O)N[C@@H](CCC(N)=O)C(=O)N[C@@H](CC(O)=O)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](CC1=CNC=N1)C(=O)N[C@@H](CC(N)=O)C(=O)N[C@@H](CC1=CC=CC=C1)C(O)=O
Tylosin	916	C46H77NO17	CC[C@@H]1[C@H](/C=C(/C=C/C(=O)[C@@H](C[C@@H]([C@@H]([C@H]([C@@H](CC(=O)O1)O)C)O[C@H]2[C@@H]([C@H]([C@@H]([C@H](O2)C)O[C@H]3C[C@@]([C@H]([C@@H](O3)C)O)(C)O)N(C)C)O)CC=O)C)\C)CO[C@H]4[C@@H]([C@@H]([C@@H]([C@H](O4)C)O)OC)OC

Table of compounds in the “Monster Mix” as well as their mass to charge ratios, molecular formulas and SMILES codes.