

**Bangor University**

**DOCTOR OF PHILOSOPHY**

**MS, IMS and Computational Studies of Peptide Structure and Enzyme Activity**

Chaplin, Daniel

*Award date:*  
2021

*Awarding institution:*  
Bangor University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. Mar. 2024

**Bangor University**

**DOCTOR OF PHILOSOPHY**

**MS, IMS and Computational Studies of Peptide Structure and Enzyme Activity**

Chaplin, Daniel

*Award date:*  
2021

*Awarding institution:*  
Bangor University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 29. Jan. 2021



Bangor University

# MS, IMS and Computational Studies of Peptide Structure and Enzyme Activity

PhD Thesis



Ysgol Cemeg  
School of Chemistry  
*Rhagoriantu ers 1884 | Excellence since 1884*

Daniel Thomas Chaplin

## Abstract

Ion mobility spectrometry (IMS) is a fast growing technique which can be coupled with computational methods to elucidate properties of gas-phase ions. Proteomics, the study of proteins and their functions, is fundamentally tied to the analytical chemistry of mass spectrometry. Proteins are digested into smaller peptides before being introduced into a mass spectrometer, where the peptides are fragmented into a series of product ions. In the present thesis, IMS was used in conjunction with computational theory to investigate the structures of product ions, the smallest of these product ions being known as  $b_2$  ions. The IMS technique used here is travelling wave IMS (TWIMS), a technique that requires regular calibration to determine individual ion mobilities and a common calibrant is polyalanine. This presents a problem as this study shows that the number of chiral centres present in polyalanine can lead to multiple different ion mobilities being detected if the standard mixture is not isomerically pure. This study aims to answer the questions of how well these techniques combine, how accurate are the parameters set in Mobcal (the programme used to compute theoretical CCS values), and if polyalanine is a suitable calibrant.

Computed models of  $b_2$  ions are presented with collisional cross sections (CCS) calculated via the software Mobcal. Mobcal is found not to contain suitable parameters for nitrogen drift gas calculations and will require further development. A variety of polyalanine isomers are synthesised and their ion mobilities compared to the calibration standard used by Waters on a Synapt G2-Si and Cyclic IMS. Our data shows that polyalanine standards contain a mixture of isomers with varying mobilities, making them unsuitable for IMS calibration.

In a third project, we developed an LC-MS/MS method, using UPLC and a triple quadrupole in multi-reaction monitoring mode, to investigate the activities of nitrilase enzymes on different substrates under different reaction conditions. Known and newly identified extremophile nitrilases were incubated with a large set of substrates and under a variety of reaction conditions to study substrate promiscuity and activity in harsh reaction conditions (acidic/basic pH, organic solvents and elevated temperature). Our data show that the tested nitrilases differ not only in their substrate specificity, but also in regard to performance under adverse reaction conditions.

## Acknowledgements

First and foremost I would like to extend my sincerest thanks to my supervisor, Professor Béla Paizs, for his great support, patience and trust in my abilities. He has nurtured in me most of the skills that I consider to be my valuable, and without his continuous encouragement I may never have reached where I am today. I genuinely do not believe I could choose anyone would do better in supporting my PhD. I would also like to thank Dr Tran Hai for collaboration and support on the enzyme activity project and Keith Richardson at Waters Corporation for facilitating access to ion mobility instruments used here.

Along with my supervisor, I have been provided great support by PhD committee; Dr Andrew Davies and Dr Keith Hughes. They have continued to provide me with support and great advice throughout my studies, from both my postgraduate studies and prior to that, my undergraduate studies. I would also like to extend an additional massive thank you to Dr Lorrie Murphy and Dr Martina Lahmann, both of whom have provided invaluable support in the writing of my thesis.

I would also like to thank my colleagues in the centre for environmental biotechnology (CEB), Dr Rolf Kraehenbuehl for his help in proof reading my introduction, and Colin Potter and Dr Karin Lanthaler for their both their continues moral support throughout my writing but also all three for sharing and teaching their practical skills in the laboratory which has allowed me to expand my own skills greatly. I consider you all my closest friends.

I would like to extend big thank you to all of the staff at Bangor University, particularly those involved in the former school of chemistry for their unmatched friendliness and universal approachability.

Finally, for all the love and support provided throughout my PhD, thank you to my wife Ellie Rose Chaplin, whom I met at the start of my PhD work and married before the end. Also, thank you to my parents, my sisters, my in-laws and my dogs for their continuous support through literally everything.

## Table of Contents

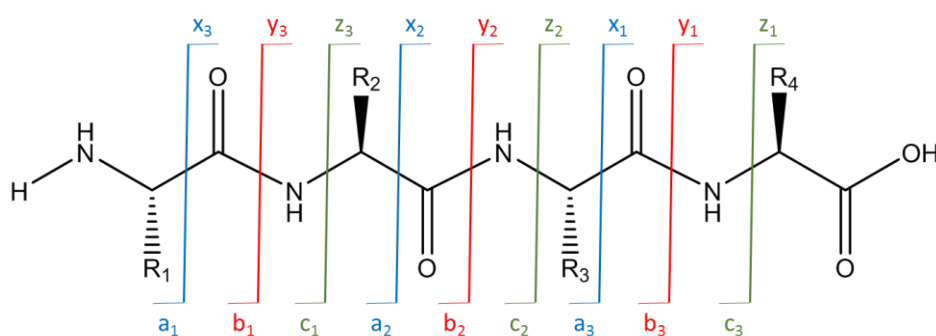
<b>Declaration and Consent</b> .....	i
<b>Abstract</b> .....	iv
<b>Acknowledgements</b> .....	v
<b>Table of Contents</b> .....	vi
<b>Chapter 1: Introduction</b> .....	1
Chapter 1.1: Peptide Fragmentation.....	1
Chapter 1.1.1: Studying Peptide Fragmentation Using Computation Theory .....	5
Chapter 1.2: Protein Analysis by Mass Spectrometry .....	9
Chapter 1.3: Mass Spectrometry .....	12
Chapter 1.4: Ion Mobility Spectrometry .....	19
Chapter 1.4.1: Calculation of CCS values for Ion Mobility Spectrometry .....	23
Chapter 1.5: Enzyme Activity .....	25
<b>Chapter 2: Suitability of Polyalanine as an IMS calibrant</b> .....	28
Chapter 2.1: Introduction .....	28
Chapter 2.2: Aims.....	30
Chapter 2.3: Experimental Methods .....	31
Chapter 2.4: Computational Methods .....	34
Chapter 2.5: Results and Discussion .....	36
Chapter 2.6: Conclusions .....	49
<b>Chapter 3: Determination of <math>b_2</math> Ion Structure Through Ion Mobility Spectrometry</b> .....	51
Chapter 3.1: Introduction .....	51
Chapter 3.2: Aims.....	53
Chapter 3.3: Experimental Methods .....	54
Chapter 3.4: Computational Methods .....	57
Chapter 3.5: Results and Discussion .....	59
Chapter 3.6: Conclusions .....	92
<b>Chapter 4: Promiscuity and Activity of Nitrilases</b> .....	94
Chapter 4.1: Introduction .....	94
Chapter 4.2: Aims.....	97
Chapter 4.3: Methods and Materials .....	98
Chapter 4.4: Results and Discussion .....	102
Chapter 4.5: Conclusions .....	114
<b>Chapter 5: Final Conclusions and Future Work</b> .....	116
<b>References</b> .....	118
<b>Appendix: <math>b_2</math> ion structures, relative energies and method development</b> .....	125

## Chapter 1: Introduction

### Chapter 1.1: Peptide Fragmentation

For as long as mass spectrometry based proteomics has been a field of research, chemists have had an interest in understanding how the peptides fragment. Peptides in a mass spectrometer undergo fragmentation to produce smaller ions, these peptides fragment in order from the terminal amino acids. This information is then used to determine the sequence of the amino acids that make up said peptide. This technique is known as De Novo Peptide sequencing.

In collision-induced dissociation (CID), protonated peptides are fragmented into various sets of ions, the majority of which take the form of  $b_n$  and  $y_n$  ions by cleavage on the amide bond between the carbonyl and amine groups. [1,2] The process follows chiefly charge-directed dissociation pathways, meaning that the cleaving amide bond is directly weakened by an ionizing proton. [3] When the amine nitrogen is protonated, the amide bond is weakened. [4] This makes the neighbouring carbonyl carbon a more suitable site for nucleophilic attack.

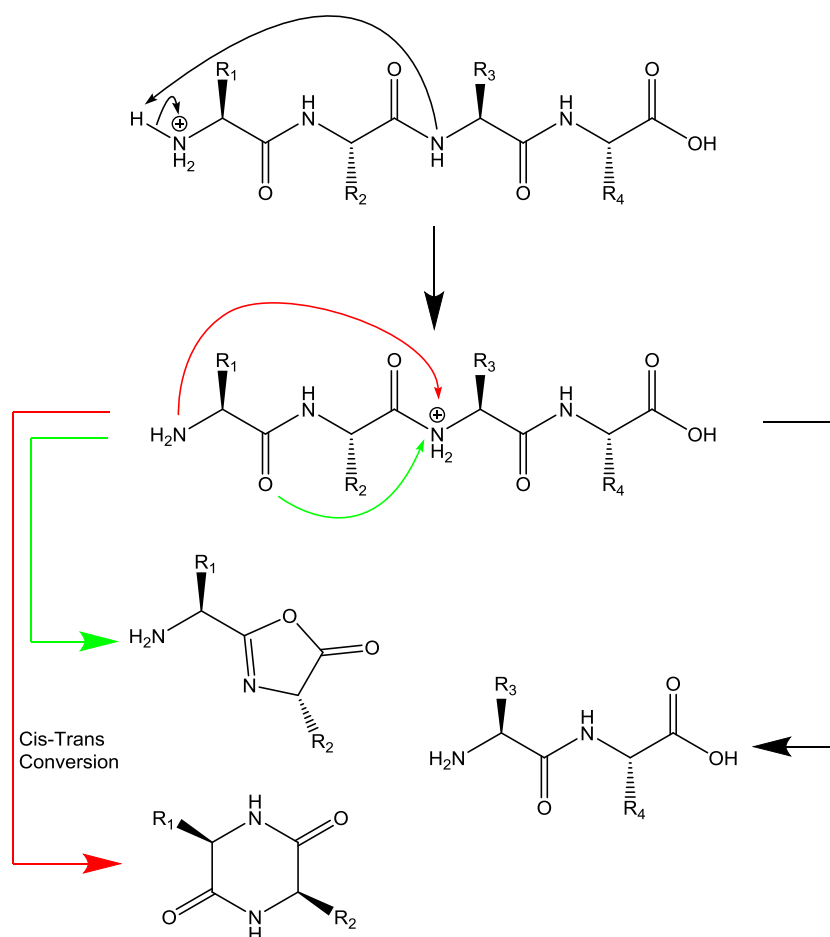


**Figure 1. Nomenclature of peptide fragmentation.** Peptide fragments are named based on whether they are derived from the C or N terminus of the parent (xyz or abc, respectively), and the number of amino acids that make up the fragment.  $b_n$  and  $y_n$  ions are the most prevalent in ESI-CID based proteomics.

Peptides are initially protonated in the sample under acidic conditions, and are protonated at the most basic sites. The most comprehensive model to describe peptide fragmentation is known as the mobile proton model. In this model, the proton initially bonds to the N-terminus or a side chain nitrogen. Within the mass spectrometer, upon excitation the proton becomes “mobile” in the sense that it is able to move between potential protonation sites, sampling

amide nitrogen atoms in the process. Upon protonation of an amide bond, the peptide is able to be cleaved at this site as described previously. [5, 6]

One of the most common forms of nucleophilic attack on the amide bond is an attack by the neighbouring amide oxygen in the direction of the N-Terminus, Scheme 1. This will produce a  $b_n$  ion with a ring derivative of an oxazolone structure (a 5-membered ring) at the C-terminus, in addition to a  $y_n$  ion consisting of the remaining C-terminus portion of the parent peptide. [7, 8] Another form of nucleophilic attack is an attack by the N-terminus nitrogen on the carbonyl carbon at the cleavage site, which produces a macrocyclic (or diketopiperazine-like) structure. This requires trans-cis isomerization of the N-terminus amide bond, which holds a significant activation energy barrier. [9]

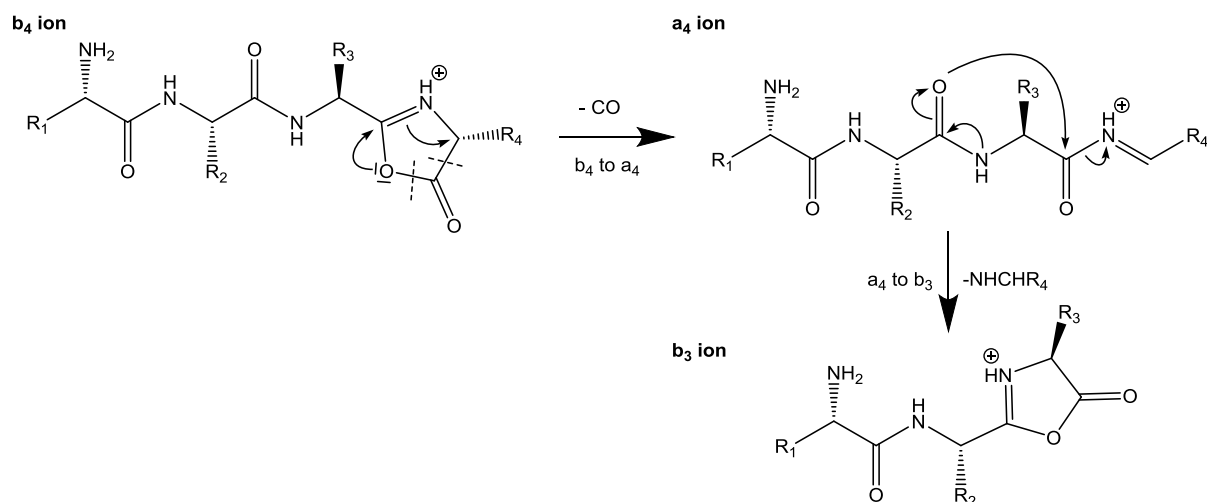


**Scheme 1. The Mobile Proton model and fragmentation pathway of peptides to form  $b_n$  and  $y_n$  ions.** The top structure shows a 4 membered peptide protonated at the N-terminus, and the proton relocating to the nitrogen of the second amide bond. The green pathway denotes formation of an oxazolone derivative  $b_2$  ion and the red pathway denotes formation of a macrocyclic  $b_2$  ion. The purple pathway shows the formation of a  $y_2$  ion.



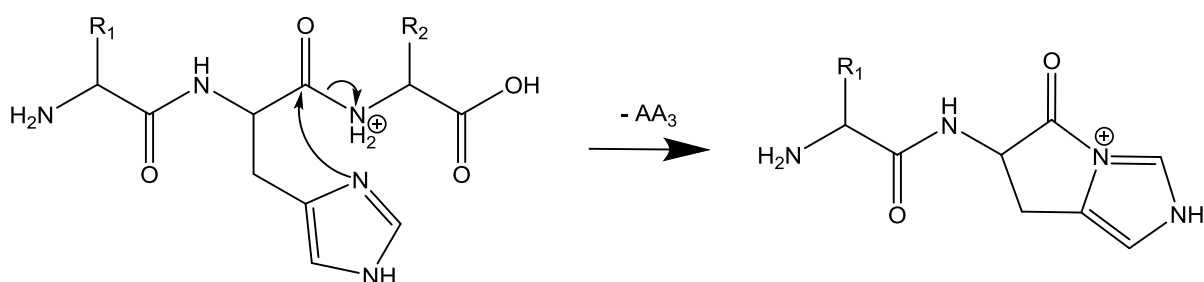
This reaction generally produces a more energetically favourable product ion than the oxazolone pathway, however, the cis-trans isomerization often has a high enough energy barrier to make the oxazolone pathway more favourable overall. The opposite side, the  $y_n$  ions, take the form of truncated peptides from the position of the amide cleavage through to the C-terminus. These ions simply take the structure of a usual protonated peptide. These final structures have been confirmed by ion IR spectroscopy, Hydrogen/Deuterium exchange (HDX) or Ion Mobility spectrometry (IMS). [<sup>9,10,11,12</sup>] Many of these studies focus on small numbers of  $b_2$  ions or larger  $b$  ions, and there continues to be disagreement as to the structure of  $b_2$  ions. In this study, these  $b_2$  ion structures will be further probed using ion mobility spectrometry and confirmed via theoretical calculations of respective geometries.

It is also possible for these  $b_n$  fragments to degrade further to produce smaller  $b_{n-1}$  ions or  $a_n$  ions. A fragmentation mechanism has been provided for the  $b_n$  to  $b_{n-1}$  degradation by Harrison, shown in scheme X. This method proposes that in the oxazolone  $b_n$  ion, the next amide oxygen along from the oxazolone ring attacks the carbon of the  $O-C=NH^+$  group in the ring, leading to the elimination of an aziridinone derivative. However, this pathway has a transition structure with an energy barrier of  $\sim 77$  kcal/mol, likely too high for many CID environments. A fragmentation pathway for the  $b_n$  to  $a_n$  degradation involves the protonation of the oxazolone nitrogen, weakening the  $O-CO$  bond, as illustrated in scheme 2, allowing for the elimination of  $C=O$ . Finally, a proposal for a combination of these fragmentation pathways has been presented for a lower transition energy pathway to reach  $b_{n-1}$ , which is referred to as  $b_n$  to  $a_n$  to  $b_{n-1}$ . In this pathway,  $C=O$  is eliminated as stated in the  $b_n$  to  $a_n$  pathway, before the  $O$  of the neighbouring carbonyl attacks the same carbon in what is now a  $O=C-NH^+$  group. This pathway requires  $\sim 40$  kcal/mol, which is significantly more favourable than the direct  $b_n$  to  $b_{n-1}$ .



**Scheme 2. A fragmentation pathway from  $b_n$  ion to form an  $a_n$  ion and then a  $b_{n-1}$  ion.** This pathway is particularly of relevance to MSn techniques, where  $b$  ions may be fragmented into smaller  $b$  ions.

In addition to the oxazolone and macrocyclic pathways, some amino acid side chains can attack the carbonyl of the weakened amide bond, producing other structures. One example of this is the histidine effect illustrated in scheme 3, in which a nitrogen of the histidine ring attacks the cleavage site carbonyl, producing a  $b_n$  ion with a double ring structure. [13] Glutamic acid and aspartic acid also produce a unique  $b_n$  structure via attacks from their side chain acids on the carbonyl carbon of the cleavage site. [14]



**Scheme 3. The histidine effect pathway forming a double ring structure.** This pathway occurs when a histidine residue is present in the amino acid closest to the amide bond being cleaved in a  $b_n$  ion.

### *Chapter 1.1.1: Studying Peptide Fragmentation Using Computation Theory.*

Theoretical tools are essential to confirm the geometries of the fragment ions. These tools have been used in the field of mass spectrometry to understand the interactions occurring within gas-phase ions, and to understand the mechanisms that lead to the fragmentation of peptides and to prove the structures of the product ions. The computational chemistry software called Gaussian is an essential tool for modelling molecules and reactions, originally developed as Gaussian 70 in 1970 at Carnegie Mellon University by John Pople. Gaussian is software which offers a variety of toolsets available, including the ability to perform molecular mechanics, semi-empirical calculations, Hartree-Fock and Density functional theory quantum calculations.

Atom Centered Density Matrix Propagation (ADMP) molecular dynamics is a powerful and computationally-efficient method of performing molecular trajectory calculations. [15] Molecular dynamics is a method of computing the movements of atoms and molecules which takes into account the forces and potentials between interacting particles. Classical trajectory methods fit one of two categories: Born-Oppenheimer methods (in which the electrical structure is calculated in each instance that the potential energy surface is probed) and the extended Lagrangian methods (in which the wave function is carried forwards with each step of the computation). ADMP is an extended Lagrangian method of performing molecular dynamics, it is well suited to gas phase reactions due to its treatment of individual atoms within a molecule separately as pseudopotentials. It also has the advantages of having a variety of optional functionals to cater for a variety of applications and allows for the inclusion of charged molecules which is of particular relevance to modelling the reactions that take place within a mass spectrometer environment. [16]

Simulated annealing (SA) can be utilized to probe all potential geometries of an ion. SA is a technique in which a large number of possible ion conformations are probed at a high temperature. These ions are then gradually cooled eliminating various high energy conformations, such as transition structures and leaving only sets of minimum energy structures, from which one can determine a global minimum. The advantage of this technique is that one produces a large number of potential structures under the high energy conditions,

reducing the probability of not finding the correct path to the global minimum structure. [17] This is relevant to the environment of the ions in IMS-MS, as the ions will generally undergo high temperature conditions in the source of the mass spectrometer of up to 600°K and high energy conditions during fragmentation. Ions are then generally under more mild, low temperature conditions during ion mobility, allowing them to reach minimum energy structures. Alternatively, this can also be performed via the use of a molecular dynamics method such as ADMP, which produces a large number of potential high energy geometries at a high temperature. All generated geometries are optimised at a low computational cost semi-empirical method such as AM1 or PM3, reducing the number of geometries down to a set of possible minimum energy structures where degenerate structures are eliminated. The level of theory and basis set can be sequentially increased to further optimise the energy and further eliminate degenerate structures. This will lead to a set of structures being produced at a density functional theory combined with a good basis, with a high confidence that potential structures will not be missed.

The Schrodinger equation describes the interactions of subatomic particles via their respective wave functions. Obtaining the wave or state function of a quantum system can be used to derive many of its electronic structural properties. [18]

$$H\Psi = E\Psi \quad (10)$$

H is the Hamiltonian of the system (the sum of kinetic and potential energies),  $\Psi$  represent the wave function, and the constant E is the energy of level of the system.

Ab initio methods use the position of nuclei and the number of electrons to solve the Schrodinger equation. [19] The words “ab initio” translate to “from first principles”, referring to the fact that only physical constants are initially included in the calculations. Hartree-Fock (HF) is an example of such a method, which is able to perform a satisfactory calculation of the geometry and vibrational frequencies of a molecule, at a computational cost that is marginally lower than a density function theory calculation. [20] To enable these calculations, HF makes a number of simplifications:

1. The greatest limitation of HF is that it neglects electron correlation. This is due to the application of mean field theory, which reduces the interactions of any single entity to an average reaction, simplifying the behaviour of a system to save on computational power. [21 p.114]
2. The Born–Oppenheimer approximation is applied, which assumes that nuclei are stationary relative to the fast interactions of the electrons. [80 p.256]
3. Relativistic effects are not accounted for. [22]
4. The stationary states of the system are each described by one Slater determinant.

Semi-empirical methods are a simplified alternative which use values derived from experimental data to solve the Schrodinger equation, therefore they assume a specific set of parameters surrounding the system. [80 p.111] Various semi-empirical methods differ based on the parameters that they assume. Due to the reduced number of parameters these methods are some of the computationally cheapest levels of theory and are suitable for operating on large systems or a large number of small systems.

Density functional theory (DFT) is a quantum chemical method, these are slightly more costly than Hartree-Fock methods, but make fewer simplifications for more accurate calculations. The most significant change between all DFT methods and HF methods is that DFT methods attempt to account for electron correlation. DFT methods split electronic energy into a series of functionals that are calculated separately: electron-nuclear interactions, coulombic repulsion, kinetic energy, electron exchange and electron correlation. [23] There are three sets of functionals that are used in DFT:

1. Local functionals for exchange and correlation that use only electron spin densities.
2. Gradient-corrected functionals that use electron spin densities with their gradients.
3. Hybrid functionals, which combine HF, local and gradient-corrected electron exchange with local and gradient-corrected electron correlation functions. [80 p.119]

An example of a DFT hybrid functional Becke-3 Lee Yang Parr (B3LYP), which is among the most popular levels of theory for DFT calculations. B3LYP gives a reasonably accurate structure, energy and frequency values for many molecular models where HF calculations are relatively poor. [24]

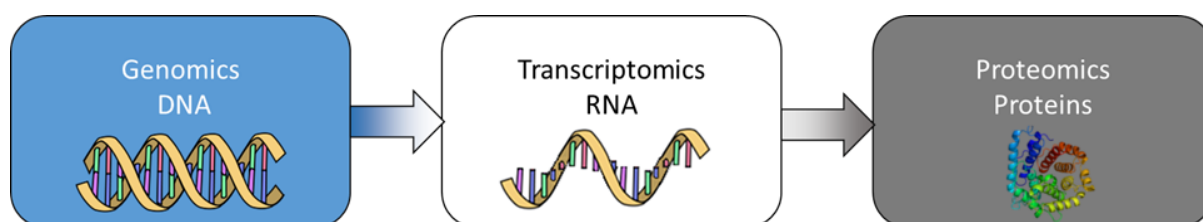
Basis sets are used in HF and DFT methods to describe the atomic orbitals of a molecule and therefore provide a solvable solution to the systems electronic wave function. A basis set is made up of a number of functions that represent different properties of the orbitals, when each is applied the respective property is computed rather than assumed or neglected. A larger basis set costs more computational power to utilise, in return it provides a more accurate representation of the atomic orbitals and a more accurate model of the system. It does this by reducing the assumptions on the positions of the electrons, which more accurately represents the physical reality of the probability of an electron existing anywhere in space. [<sup>80</sup> p94-100, <sup>25</sup>]

The basis sets can be split into categories based on the number of functions that they apply.

- Minimal basis sets (e.g. STO-3G) are the smallest basis sets required by HF or DFT methods, they describe only the inner shell orbitals and one fixed size type of valence orbital. These are cheap to compute but not accurate.
- Split valence basis sets (e.g. 6-31G) expand to include multiple sizes of valence orbital, and generally increase the number of functions that describe atoms.
- Polarised basis sets (e.g. 6-31(d)) include the ability for orbitals to change shape in addition to size, expanding the number of orbitals to include those with angular momentum (e.g. d orbitals for non-metals, p orbitals for hydrogens).
- Diffuse functions (e.g. 6-31+(d)) allow for orbitals to take a larger region of space, allowing the electrons to move further out from the nucleus. This is relevant in many systems where the electrons are repelled by a negative charge or when the system is being studied in an excited state. Generally this produces a more accurate model of the system.

## Chapter 1.2: Protein Analysis by Mass Spectrometry

Just as a genome is the entire expression of all the genes that make up an organism, a proteome is the entire expression of all proteins that make up an organism. Proteomics is the wider study of all protein properties, from their chemical structure, to their interactions on any scale and to their function as part of a greater organism. It is common knowledge that humanity, or even all life as we know it, is defined by our DNA, the main function of which is transcribed to RNA, which itself is translated into the proteins. Proteins define the structures of cells and act as functioning units of these cells. Proteins are one of the building blocks of life, but problems with proteins can cause significant problems for the life that depends upon them, and are major factors contributing to diseases. Understanding the properties and functions of proteins can lead to methods of detecting diseases and potentially targeting proteins with a variety of drugs to treat diseases. [26]



**Figure 2. The multiple layers of Omics.** Images acquired from [27] and [28].

In the earlier days of proteomics, the structure of the protein was thought to be defined by a “one gene for one enzyme” hypothesis, which proposes that each gene encodes a single protein. This theory was useful in gathering a fundamental understanding of the biochemistry that leads to protein structure and function. It led to an understanding that an organism requires amino acids to build its proteins, and if it lacks these, it lacks the fundamental components of life. Later in 1958, armed with the knowledge that peptides are made up of amino acids, Sanger was able to elucidate the structure of insulin as a string of amino acids. [29] By 1972 Stein and Moore were able to sequence an entire protein, ribonuclease S. [30] These sequences were a linear string of amino acids called the primary structure of a protein. Secondary and tertiary structures are formed by folding these strings of amino acids, held together by the amino acid intramolecular interactions. Finally, the quaternary structure is produced by multiple proteins creating a complex single structure. Quaternary structures can

be as simple as a dimer, made up of two proteins (homodimer or heterodimer). The more complex structures are very commonly being made up of several subunits of different types. Some viruses use very complex quaternary structures made up of tens of subunits to form a shell around their genome. [<sup>31</sup>]

In 1975, O'Farrell and Klose independently developed a technique known as two dimensional (2D) gel-electrophoresis. This method employs gel electrophoresis across two different dimension planes, allowing for separation of significantly more complex mixtures of proteins by their respective size/molecular weight and isoelectric point (pI). Later on this technique was used in conjunction with the ever developing techniques of mass spectrometry to analyse the composition of individual proteins on a greater scale by separating groups of proteins via gel electrophoresis before analysing them in a mass spectrometer. [<sup>32,33</sup>]

Today, the number of applications for proteomics has increased dramatically. There are wide variety of areas of research in which proteomics acts as a foundation for pharmaceutical advances, it is essential to cancer research and it is very relevant to the discovery of most diseases, as proteins are expressed differently in diseased tissue than in healthy tissue. [<sup>34</sup>]

Proteomics today is often performed on an LC-MS system, using a reverse-phase column (often one of many C<sub>18</sub> columns) and a high resolution mass spectrometer. In these cases a protein or proteome is digested via a protease enzyme such as trypsin. Trypsin hydrolyses proteins at the C-terminus side of two amino acids: lysine and arginine. The exception to this is when the next amino acid in the sequence is proline. This reduces the protein into a series of peptides, all of them with a lysine or arginine C-terminus (excluding where the parent protein C-terminus is still present). These peptides are then introduced to a reverse phase column via an LC system, starting with a mobile phase consisting primarily of water and gradually increasing the proportion of an organic mobile phase (such as methanol or acetonitrile) over a long period of time. In many systems, the LC flow is administered to a mass spectrometer via an ESI source, which uses protonation to ionise the peptides. [<sup>35</sup>]

Peptides are then analysed via a data-dependent (or in some modern systems, data-independent) mass spectrometry method. Peptides that have been separated via chromatographic techniques are then individually isolated with an accurate parent ion mass



and further fragmented produce a product ion mass spectra. This information is often enough to sequence the individual peptides, the protein and potentially a full proteome. To process this data, a user requires specialised processing software (such as Waters Progenesis QIP or Thermo Scientific Proteome Discoverer).

Proteomics software is designed to combine acquired data sets and interpret chromatographic peaks to identify potential peptide ions and their respective product ions. These data sets are then compared to reference CID databases such as UniProt and Swiss-Prot. These databases allow for searching through peptide sequences. Many researchers upload their identified peptides to these databases, whether or not they are peer reviewed (and these uploads are labelled as such). The proteomics software is able to compare the acquired data to these spectral databases to find matches between the two, assigning the peptide ions and providing users with an accuracy score based on a number of factors including a fragmentation score, mass accuracy score and isotope score. [36]

### *Chapter 1.3: Mass Spectrometry*

Tandem mass spectrometry (MS/MS) is the most commonly used technique to sequence peptides in proteomics. MS/MS allows a number of different manipulations on ions, including the isolation and fragmentation of ions. In some mass spectrometers, such as ion traps and orbitraps, those fragment ions can then be isolated and potentially fragmented further via the application of MS<sup>n</sup>. [37] MS/MS can also be used for quantitative analysis. A common method for quantification involves isotopic labelling of the samples and an analysis of the labelled fragment ion can be used to quantify the parent ion. [38, 39, 40]

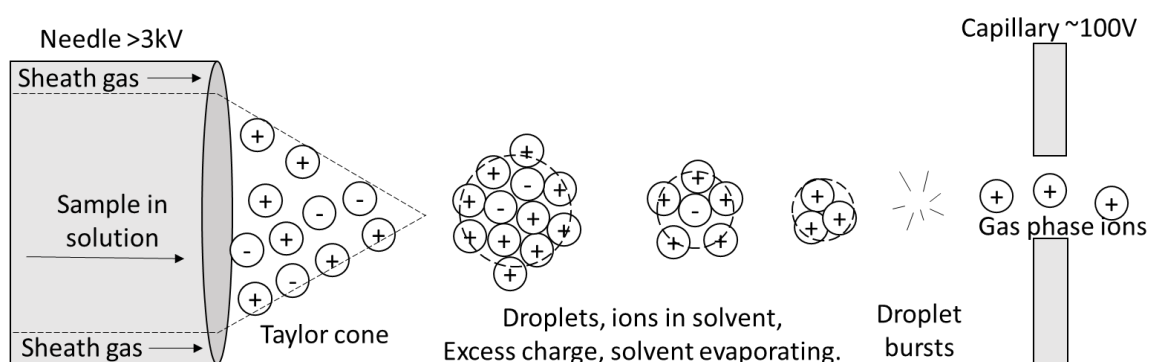
Tandem mass spectrometers are used for studying ions in the gas-phase, and are often fitted with soft ionisation sources (to avoid in-source fragmentation of large or fragile molecules and allow the determination of parent ion  $m/z$ ), such as electrospray-ionisation (ESI) and matrix assisted laser desorption ionisation (MALDI).

In MALDI, the sample is mixed and co-crystallised with matrix. A matrix consists of small molecules which act as a good proton source with high light absorption (common examples include sinapinic acid, alpha-Cyano-4-hydroxycinnamic acid and 2,5-Dihydroxybenzoic acid). [41] The matrix is then desorbed by the use of a laser, this causes the matrix to ionize the sample compounds via proton transfer. The ions are then introduced into the instrument. [42] Generally MALDI produces a majority of singly charged ions, making it less suitable to proteomic studies. [43]

Notably becoming recognised in 1988 as the work of John Bennett Fenn and earning the Nobel prize in 2002 [44, 45], ESI was demonstrated as a technique that was capable of producing multiply charged protein/peptide ions, allowing for high mass molecules to be analysed on instruments with an upper mass limit below 2000 Daltons.

ESI, as illustrated in figure 1, is a technique that applies a voltage to the solution to ionize and sprays these ions into a transfer capillary. [46] Upon the initial exit from the needle the charged solution forms a cone, held together by surface tension, this is called the Taylor cone, named after Sir Gregory Taylor. [47] At higher voltages, the cone becomes more pronounced. A stable Taylor cone is important in the formation of the electrospray, as the tip of the cone is where

the jet of charged droplets originates. [48] At higher fields the cone can become less stable. At high field energies it is possible to split the cone to produce multiple jets from multiple cone tips. These jets may produce droplets of different properties. Ultimately a single stable jet is considered the most useful mode for ESI-MS. In positive ionisation mode the charged droplets contain an excess of positively charged ions contained in a “ball” of solvent. Under the effect of the high charge the solvent in the droplet evaporates. This causes the overall size of droplets to decrease, whilst not reducing the charge. As the droplets charge to size ratio increases, the concentration of charge increases until the droplet undergoes what is called “Coulomb fission” (Figure 3). This Coulomb fission refers to the charge forcing the ions apart rapidly, essentially causing a small explosion of droplets, separating them from one another and the remaining solvent in the droplet. These separated ions are now considered to be in the gas phase and enter into the mass spectrometer via the transfer column. [49,50]



**Figure 3. Formation of gas phase ions by Coulomb fission in an ESI source.** Mobile phase is injected into the source via a narrow needle at a consistent voltage, this voltage causes the liquid to form a Taylor cone, the tip of which releases droplets of mobile phase. Solvent evaporates from the droplet, leaving only charged ions in a high concentration, the electrostatic repulsion between ions causes the droplet to undergo Coulomb vision.

In ESI, due to the efficiency of the gas phase ion production (as high as 50% with nanospray), very low concentrations of analyte can be used. Depending on the mass spectrometer used, concentrations can be in the range of pg/ml. [51] Typically, analytes are dissolved in polar solvents, such as methanol and acetonitrile. Water is a suitable solvent for ESI, but due to its relatively low vapor pressure it can be detrimental to sensitivity when used alone, therefore it is recommend that it is incorporated as a mixture with a second solvent.

ESI can be suitably coupled to liquid chromatography (LC), which has allowed for greater separation and a higher throughput for protein/peptide/metabolite analysis. The combination of LC with mass spectrometry has lead proteomics to become considerably more commercially viable, with continuing developments bringing down the cost of older instruments. LC is generally performed at the peptide level, as opposed to the protein level, due to the ease of passing through the smaller molecules through a column and for easier MS analysis and fragmentation. Proteins that are to undergo LC-MS analysis are treated with trypsin (an enzyme that primarily cleaves amide bonds at the carboxyl side of arginine and lysine) to convert them into a complex sample of peptides. [52] Peptides are easily ionisable due to their variety of ionisable functional groups and hence they are very suitable for LC-ESI-MS. HPLC allows for great separating power with appropriate chromatography columns, in the case of peptides that would be reverse-phase columns. Reverse phase columns use a non-polar hydrophobic stationary phase, which has strong affinity to other hydrophobic compounds, a category which many peptides fall into. Due to the popular commercial application of proteomics, there are a wide variety of reverse-phase columns on the market, many of which use the C18, C8 and phenyl materials as the stationary phase. The Mobile phase will consist of a polar solvent, such as methanol or acetonitrile. [53]

Tandem MS is an essential tool in the structural identification of peptide ions as it allows the parent ions to be isolated and fragmented to provide structural information about the peptides. [54] Four of the most common mass spectrometers capable of performing these techniques are the triple quadrupole mass spectrometer (QQQ), quadrupole time-of-flight (QTOF), the quadrupole ion traps and the Orbitrap. Of these instruments, two are primarily used for discovery (QTOF, Orbitrap and ion traps) and the QQQ is used for quantification. [55]

The quadrupole time-of-flight mass spectrometer consists of a quadrupole, a collision cell and a time of flight mass analyser. The primary function of the Q-TOF is qualitative mass spectrometry and discovery workflows. The quadrupole allows the instrument to isolate parent ions of specific  $m/z$  and the collision cell is used to fragment these parent ions. The time-of-flight analyser is a high resolution mass analyser which determines mass by the velocity of ions. Ions are accelerated under an electric field and released into a field free drift

region within which ions advance at a velocity determined by their kinetic energy (determined by charge) and the mass of the ion. [56]

$$E_K = \frac{1}{2}mv^2 \quad (1)$$

Under these conditions potential energy is transformed into kinetic energy ( $E_K$ ), and potential energy is equal to the charge of the ion ( $q$ ) multiplied by the potential difference ( $U$ ), we get the equation:

$$qU = \frac{1}{2}mv^2 \quad (2)$$

Velocity is equal to the length of the drift region ( $d$ ) divided by flight time ( $t$ ). So we can rearrange it to make  $t$  the subject:

$$t = \frac{d}{\sqrt{2U}} \sqrt{\frac{m}{q}} \quad (3)$$

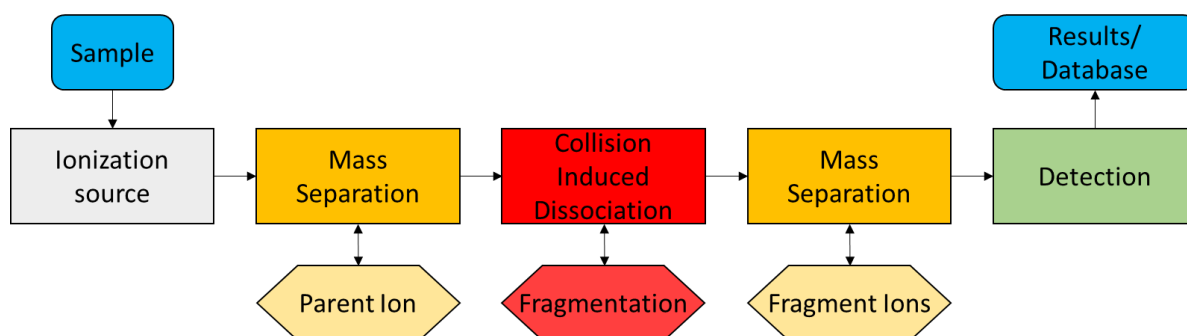
As distance and potential difference are constants defined by the instrument, the relation between mass, charge and time are clear.

Q-TOF instruments have a variety of operation modes available for different types of analyses, dependent of the manufacturer. A common discovery technique is data-dependent analysis (DDA). In DDA a range of  $m/z$  is acquired, software is then used to select ions of interest (often the highest intensity) in the quadrupole, these ions are fragmented and the product ion pattern is recorded. Data acquired via this method contains structural information about the most abundant or most well ionised compounds in a sample, and complemented by a database searching algorithm, is used to identify molecules of a complex sample. An alternative to DDA is data-independent analysis (DIA). In DIA a full range of  $m/z$  are acquired and then all (or in some cases a smaller range) of these ions are fragmented and a mixture of product ions from the various parents are acquired. Software specific to this method is used to deconvolute the data and assign product ions to their respective parents. The advantage of this method is that it allows more parent ions to be analysed, at the cost of a more complex data set. [57, 58]

The triple quadrupole is a set of three quadrupoles in series, as shown in Figure 4. Each of the quadrupoles consists of four metal “poles” parallel to one another in a square formation, with one pair of rods that are opposite to one another share the same radio frequency (RF) voltage, whilst the other pair have a separate RF voltage applied, a direct voltage is then applied in the along the poles. Ions are directed between and along the poles, whilst only ions of a specific  $m/z$  ratios are transferred from the quadrupole entrance to the exit at any specific ratio of voltages. By varying the ratio of voltages on the quadrupoles one can transfer specific  $m/z$  ratios to the next stage or mass detector of the mass spectrometer. [59]

The three quadrupoles each have a specific purpose depending on the scanning mode of the instrument. In peptide analysis, the most common mode used would be the product ion scan, single reaction monitoring (SRM) or  $ms^2$ . In this mode, the first quadrupole is used to isolate parent ions with an  $m/z$  selected by the user; this sets the DC and RF voltages to a specific ratio, letting only that specific  $m/z$ . The second quadrupole is set to have an RF voltage only, with no DC voltage, allowing all ions to pass through regardless of  $m/z$ . This second quadrupole is filled with a collision gas, a neutral molecule such as  $N_2$  or Ar. The collision gas collides with sample ions, causing them to fragment. The fragments then pass into the third quadrupole, which is set to scan a predefined  $m/z$  or range of  $m/z$  ratios by varying the voltage ratio. [60]

A triple quadrupole is mostly used in proteomics for quantitation. It requires the user to provide a list of masses of interest, derived prior to the experiment. In multi-reaction monitoring mode (MRM) the mass spectrometer requires instructions for each quadrupole. For the first quadrupole, it requires a parent ion  $m/z$  to isolate. For the second quadrupole, a collision energy is required, this is a known energy that produces optimal fragmentation pattern. In the third quadrupole, a product ion  $m/z$  is required, this product ion is unique to the parent of interest. These steps allow a user to reduce the noise of mass spectrometer to the most substantial extent and reach a higher sensitivity than other mass spectrometers, allowing the triple quadrupoles to detect and quantify much lower concentrations of analytes. [61]

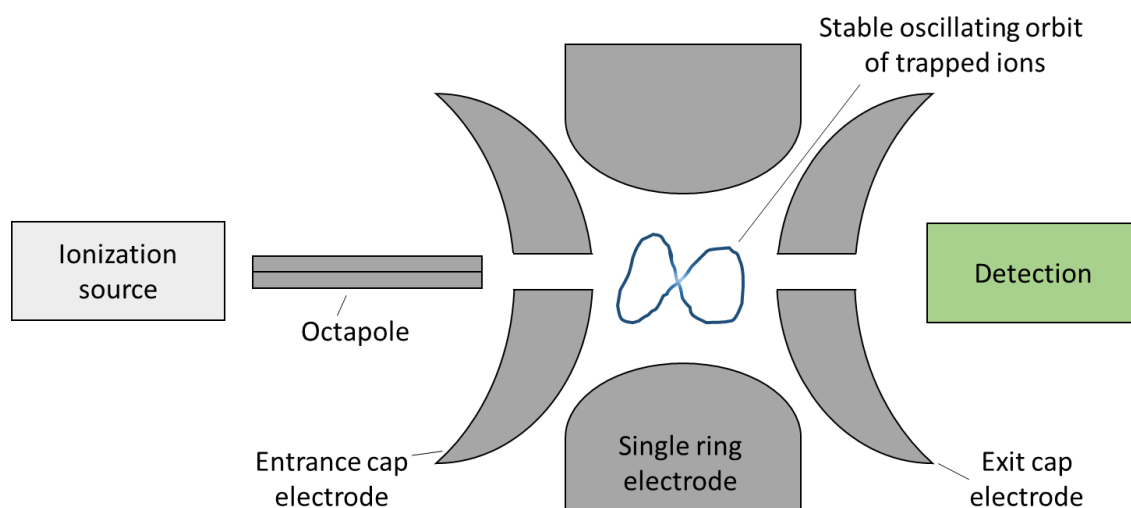


**Figure 4. A simplified schematic of a triple quadrupole mass spectrometer.** Showing the order in which ions interact with the instrument. Mass separation steps and CID each occur within individual quadrupoles.

The quadrupole ion trap, shown in Figure 5, is an instrument capable of storing and manipulating ions in the same space within the device. This is achieved by the creation of electrical fields that switch in real-time to move or oscillate the ions between the electrodes faster than the ions can reach the boundaries of the trap. The 3D ion trap is made up of a ring electrode along with a cap electrode at either end to the hollow portion of the ring electrode. The entrance cap electrode contains an opening for injecting ion into the trap and the exit cap electrode contains an opening for ejecting selected ions from the trap into the detector. [62,63] 2D ion traps have some advantages over their 3D counterparts, namely in the increase ion storage volume and higher trapping efficiencies, which can lead to a higher sensitivity. Under some conditions, 2D ion traps may also have a comparable mass accuracy to 3D ion traps. [64]

It is not possible to produce a static three dimensional field due to the restrictions of Earnshaw's theorem, which states that a stable equilibrium cannot exist in a three dimensional field without a local minima or maxima. This can be explained that for a field to have a local minima it must have a non-zero divergence, however, using Laplace's equation (equation x) that shows in a three dimensional field the divergence must always be zero in free space. [65] The solution is to create an average three dimensional field instead, this is done by alternating the direction of the field faster than the ion can reach the boundaries of the trap. The ring electrode is responsible for maintaining the oscillation of the ions, whilst the end cap electrodes remain at a constant potential. The rate of field alternation is specific to  $m/z$  ratios, and therefore the ions with only the  $m/z$  required are maintained in the trap, and other unwanted ions are allowed to collide with the trap walls. [66]

$$\nabla F = \nabla(-\nabla U) = -\nabla^2 U = 0 \quad (4)$$



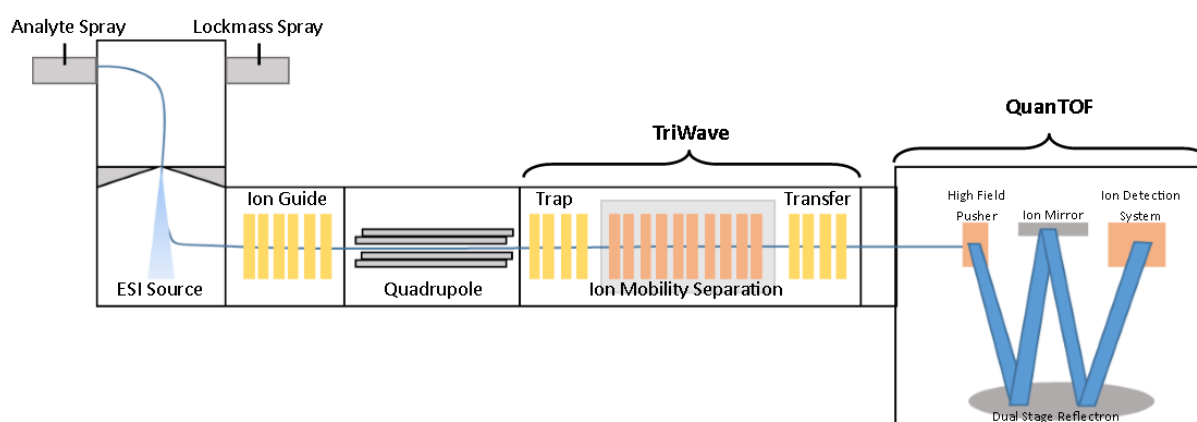
**Figure 5. A depiction of an ion trap and the ion motion within.** Ions maintain a pseudo-stable orbit within the ion trap due the oscillation of RF voltages in the ring entrance cap electrode. Varying the amplitude of RF voltages changes the size of the ion path depending on  $m/z$ , allowing specific  $m/z$  ions to be discarded, isolated or released to the detector.

The greatest advantage of the ion trap is that it is not limited to  $MS^2$ , but is capable of fragmenting product ions further, and if desired is then capable of fragmenting these second generation product ions again. This is done by trapping the parent ions in the trap, a collision gas present in the trap then interacts with excited analyte ions to cause collision induced dissociation (CID). At this point, either all ions are ejected to the detector ( $MS^2$ ) or a specific  $m/z$  product ion is now isolated in the trap while the remaining parent ions and any other fragment ions are allowed to hit the trap walls, this is called  $MS^3$ . Further, this can be repeated with as many steps and fragmentations as is desired, this is referred to as  $MS^n$ .



## Chapter 1.4: Ion Mobility Spectrometry

Ion-Mobility Spectrometry (IMS), due to the continuous development of the technique and commercial competition, is accessible to researchers and for use in industry. IMS instruments, more recently, have been coupled in to mass spectrometers (IMS-MS). IMS-MS is capable of providing mobility data regarding ions of identical  $m/z$  ratios that have (even minutely) different structures. The application of IMS is suitable to elucidate the structures of a variety of molecules, including peptides, lipids and carbohydrates. Mobility data provides information on the ability of an ion to traverse through counter-flowing drift gas whilst compelled by an electrical field. Many traditional chromatography techniques are known for having long retention times. In contrast, IMS has a relatively small “drift” time, allowing for drift time chromatograms to be produced within milliseconds. [67,68,69]



**Figure 6.** A schematic of a Waters Synapt G2-Si. Showing the path of an ion through the instrument and the position of ion mobility. The ion mobility separator in this instrument is travelling wave IMS.

The Waters Synapt G2-Si is of particular relevance to this study as the most recent model of travelling wave IMS (as of the time of the study), and because its drift region is located after the quadrupole for  $m/z$  isolation of parent ions and also after a trap cell that allows for fragmentation of the isolated ions. In the case of the Waters Synapt G2 Si (shown in Figure 6), the ions are isolated and/or fragmented in a trap cell (as in a typical mass spectrometer) and are then transferred to the drift tube. The ions are separated in the drift tube by a drift gas flowing in the opposing direction to the sample ions. Ions are separated by their relative size (or collisional cross section (CCS) also denoted as  $\Omega$ ) to charge ratio as the ions are compelled

forward by an electrical field. [70,71] Drift gases are inert atoms or molecules, the two most frequently used are helium and nitrogen, each producing different drift times due to their respective sizes.

The CCS is defined as the area around an ion of which the surface must encounter the centre of another particle (such as a drift gas) for a collision to occur, therefore larger drift gas particles produce larger CCS's as the two centres of the colliding particles are further apart. A greater CCS implies that the ion will collide with the drift gas more frequently, slowing down their progression, and producing a longer drift time.

The ion mobility separation methods do vary between different instruments. The Waters Synapt instruments utilize travelling-wave (T-Wave) separation. In this method, the ions are separated by a series of pulses of a travelling electrical field across the separation tube. Ions that are smaller (lower CCS) are carried by the wave, whilst larger ions (higher CCS) are pushed over the wave by the counter flow of the drift gas (therefore reducing mobility). [30] The complexity of the motion of ions and their interactions with the field pulses are poorly understood, therefore the CCS cannot be directly derived from the drift time. [72,73] However, CCS can be determined via the use of calibration standards. The calibration standards (a currently used example is polyalanine) are injected before samples, and an additional standard is sometimes used between samples to ensure the consistency of the instrument over the period of use. [74]

Another popular method is drift time ion mobility spectrometry (DTIMS), a widely used IMS technique. In DTIMS the drift tube has a static low voltage field compelling the ions forward, which are separated based on CCS by the drift gas. The ions motion and interaction with the field is understood and the mobility can be calculated directly from the drift time using the Mason-Schamp equations: [75]

$$V_d = KE \quad (5)$$

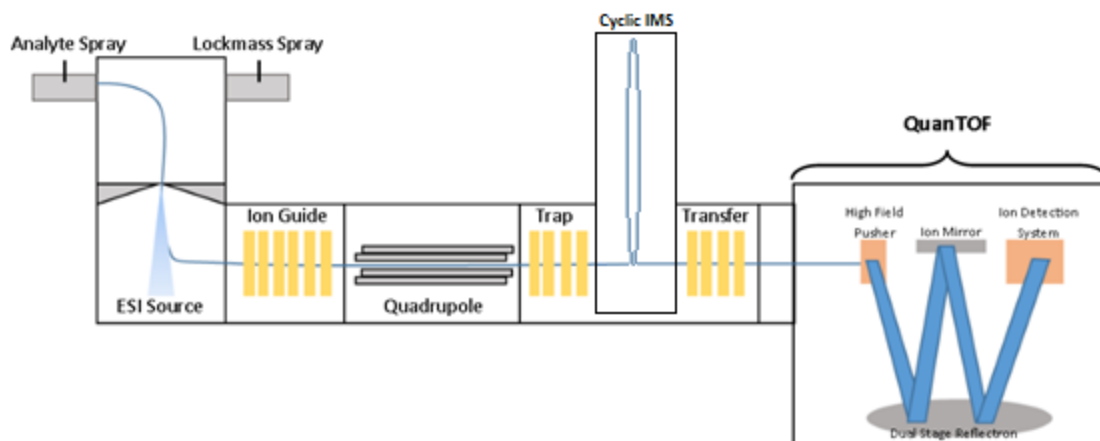
$$\Omega_T = \left( \frac{3ze}{16N} \right) \left( \frac{2\pi}{\mu kT} \right)^{\frac{1}{2}} \left( \frac{1}{K} \right) \quad (6)$$

The ion mobility (K) is derived from the drift velocity (V<sub>d</sub>) of the ions and the electric field (E). In the equation 5, the CCS (Ω<sub>T</sub>) is calculated from the numerical (z) and elementary (e) charges,

density of the buffer gas ( $N$ ), reduced mass of the ion+drift gas ( $\mu$ ), temperature ( $T$ ) and the ion mobility  $K$ . However, DTIMS has an inherent disadvantage in that its pulsed injection of ions reduces the overall sensitivity of the technique.

Alternative methods include differential mobility IMS (DMS) and field asymmetrical IMS (FAIMS), of which FAIMS uses alternating voltages to compel the sample ions forward.<sup>[76]</sup> These methods also provide quick access ion mobility information without the need for additional calibration. Unlike the previous methods, these methods do not have the same sensitivity issues due to its continuous processing of samples. However, due to clustering of ions, these methods can have issues in separating the analyte ions of similar ion mobilities.<sup>[77]</sup>

A limiting factor of IMS resolution is the limited length of the drift tube. To counter-act this, Waters Corporation have developed a cyclic-IMS (cIM). (Figure 7) This technology replaces the linear drift-tube with a circular drift-tube, in which ions can pass through the entirety of the cycle multiple times before being released in to the mass analyser. With each additional cycle through the cIM the ions are further separated by their ion mobility. The cIM is comprised of a series of drift electrodes, which consists of two parallel electrodes 5mm apart. These drift electrodes are arranged in a circle with ions passing through the gap, the gap is capped on either end with repulsive electrodes to keep ions in the centre of the path. At one point in the cycle, the electrodes are set up as an electrode array that allows for switching between facilitating the introduction of ions to the cIM and the ejection of those ions after a set number of passes. Both the separation and ejection of ions uses the T-Wave to compel the ions forwards. <sup>[78]</sup>



**Figure 7. The configuration of a Waters Cyclic IMS instrument.** The Cyclic IMS drift region takes the form a circular series of electrodes, which allows for ions to pass through the entirety of the drift region multiple times, increasing resolution with each consecutive pass.

### Chapter 1.4.1: Calculation of CCS values for Ion Mobility Spectrometry

Mass spectrometry allows molecules of different masses to be distinguished without regard for the structure the ion takes within the instrument. [79] The addition of ion mobility adds an additional degree of separation, which is determined by the ions collisional cross section, a property which can be computationally determined. Experimental IMS-MS collisional cross sections are often compared to computed counterparts to properly assign drift time chromatograms to ions. Accurate methods for computing these CCS values are being continuously developed. [80]

An approach to calculating CCS values is to compute the lowest energy geometries of the ions and calculating their respective theoretical CCSs. This is often performed by software known as mobcal, which uses atom coordinates as input to define a structure and models the interaction between this structure and drift gases. Mobcal offers a number of different methods by which to calculate ion mobility. These are the projection approximation, exact hard sphere scattering and the trajectory method.

$$\Omega_{avg}^{(1,1)} \approx \frac{1}{8\pi^2} \int_0^{2\pi} d\theta \int_0^\pi d\phi \sin\phi \int_0^{2\pi} d\gamma \pi b_{min}^2 \quad (7)$$

The projection approximation (PA) method models an ion as a collection of hard spheres, each hard sphere has a radius that represents its collisional shell. The geometric cross section is determined by averaging the radius of the whole collection from every available orientation. This method does not allow for multiple scattering, only taking into account a single encounter between a drift gas molecule and the ion. [81]

The exact hard sphere scattering method (EHSS) models the ions in the same hard sphere approximation as the PA method, however it calculates a momentum transfer cross section by orientationally averaging the ions geometry and determining the scattering angle between the incoming drift gas trajectories versus the outgoing drift gas trajectories. This method does allow for multiple scattering, but is considered to overestimate its effects. [82]

Finally there is the trajectory method (TM), where every atom that makes up the ion is represented individually by a Lennard-Jones potential, the effective and electrostatic

potential of the ion is a sum of the individual potentials. This potential is used to obtain the scattering angle (the angle of incoming drift gas trajectories and their respective departing trajectories). An orientationally averaged CCS can then be calculated. This method also allows for multiple scattering but due to the individual atoms being represent tends to avoid the error of overestimating this effect. [83] The calculation interprets the potentials using the following equation:

$$V(\theta, \phi, \gamma, R) = V(r_1, r_2, r_3, \dots, r_n) \quad (8)$$

$$= \sum_{i=1}^n 4\epsilon_1 \left[ \left( \frac{\sigma_i}{r_i} \right)^{12} - \left( \frac{\sigma_i}{r_i} \right)^6 \right] - \frac{1}{4\pi\epsilon_0} \frac{\alpha e^2}{2} \quad (9)$$

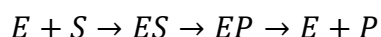
$$* \left[ \left( \sum_{i=1}^n \frac{q_i x_i}{r_i^3} \right)^2 + \left( \sum_{i=1}^n \frac{q_i y_i}{r_i^3} \right)^2 + \left( \sum_{i=1}^n \frac{q_i z_i}{r_i^3} \right)^2 \right]$$

The  $\theta$ ,  $\phi$  and  $\gamma$  define the angle of collision between the drift gas molecule and the ion,  $R$  is the distance between the centre of the ions averaged geometry and the drift gas,  $r$  is the distance between individual atoms and the drift gas molecule,  $n$  is the number of atoms within the ion.  $\epsilon$  and  $\sigma$  are Lennard-Jones potential parameters for specific atom types. [84]

Of these three methods, the TM model produces the most accurate collisional cross sections, whereas the PA model tends to underestimate the CCS and the EHSS model tends to overestimate the CCS. This is attributed to variations primarily to how these model account for the multiple scattering effect. Mobcal has been optimised for helium drift gas calculations and specific atoms including: H, C, O, N and  $\text{Na}^+$ . Performing calculations using other parameters such as  $\text{N}_2$  as drift gas or Cl as a constituent atom of the ion, requires significant modification of the Mobcal program. [43]

## Chapter 1.5: Enzyme Activity

Enzymes are macromolecules consisting of one or more proteins. The primary function of an enzyme is to act as a catalyst in reacting with substrate molecules to convert them to their respective product molecules. An enzyme possesses an active site to which substrates bind, here substrates are converted to products via an enzymatic mechanism. Enzymes vary in specificity and may have a single purpose, or may perform a more general reaction. [85] Enzymes that are highly substrate specific react with only one (or a small number) of substrates. An example of such an enzyme is lactase-phlorizin hydrolase, which only breaks down lactose into galactose and glucose by hydrolysing the  $\beta$ -1-4 glycoside bond. [86] Enzymes that are highly promiscuous, or have low substrate specificity, react with a broad range of substrates. An example of a promiscuous enzyme is decarboxylase which is able to remove COOH groups as CO<sub>2</sub> from a variety of substrates. [87]



Enzyme activity is influenced by a number of factors. Activity can be regulated via activation and inhibition. Enzyme activators can bind to enzymes in some cases to increase or start enzyme activity, these are common features of biological pathways to regulate enzymes. More frequently, activity is regulated by enzyme inhibition, in which a molecule binds to the enzyme to reduce activity via one of three methods: competitive, non-competitive and uncompetitive. Competitive inhibitors compete with the substrates to bind to the active site; binding to this site in place of the substrate prevents the reaction from taking place. A competitive inhibitor might also bind to a site other than the active site (known as allosteric site), providing that it both: Stops the substrate from binding to the enzyme and cannot bind to the enzyme if a substrate is already bound. A non-competitive inhibitor binds to a site other than an active site and does not affect the enzymes ability to bind to substrates. Instead it reduces the efficiency of the enzymes catalytic mechanisms. Uncompetitive inhibitors are a rare type of inhibitor that only bind to the enzyme-substrate complex, rendering the complex inert. [88]

Enzymes can be modified to improve or reduce activity. Post-translational modification (PTM) refers to a modification to a protein after it has been synthesised. This can occur via a reaction

at either the N or C terminus or on an amino acid side chain, modifying the functional groups at these locations to change the folding of the protein and its overall chemistry. This process allows enzymes to be synthesised in one location within an organism, where it should be inactive (for example, to avoid digesting undesirable targets, such as organ tissue), then be transported to another location where it undergoes PTM to become active and perform its primary function. [89] This is the case with the protease chymotrypsin, which is produced in the pancreas in an inactive form known as chymotrypsinogen; once transported to the gut it is activated via cleavage of the amide bond between arginine-15 and isoleucine-16. This cleavage is performed by the enzyme trypsin. [90]

Many industrial syntheses are inefficient and expensive, increasing the production costs in the food, plastic, pharmaceutical and other industries, with the addition costs being passed on to the consumer. Enzymes are a potential source of cost reduction in these processes. However, mesophilic enzymes generally are not suitable for adverse process purpose due to the reaction conditions like heat, which can lead to enzyme deactivation. [91] In contrast, extremophile micro-organisms possess enzymes that are active at much higher (thermophilic) or lower (psychrophilic) temperatures, and in harsher conditions. These enzymes can therefore be used in some of these industrial syntheses. For example, psychrophilic enzymes can increase the efficiency of the synthesis of heat sensitive products; conversely, thermophilic enzymes can catalyse reactions that require high temperatures but also have the property of being more resistant to proteases. Other useful properties of some extremophiles include: halophilicity, which enables organisms thrive in high salt concentrations; [92] acidophilicity, which enables growth in a low pH environments; [93] Alkaliphilicity, which enables growth in high pH environments; [94] radiophilicity, which enables resistance to higher levels of radiation; and heavy metal resistances. [95]

Some industries have particular interests for extremophile enzymes, four of which are discussed here. Nitrilases are enzymes that can catalyse the conversion of a nitrile substrate to non-nitrile products, such as carboxylic acids and amides. Many common nitriles are highly toxic compounds. Many plants and micro-organisms use nitrilases as a defence against these dangerous compounds converting them to less harmful or beneficial acids and amides. In industry, the desire to convert nitriles to less toxic compounds, particularly in the



pharmaceutical industry, is relevant because nitriles are side products in the synthesis of hormones. [96]

Proteases are enzymes that break down proteins by catalysing the hydrolysis of amide bonds between amino acid units. In nature these enzymes play a variety of vital roles, including digestion of proteins, performing critical roles in the immune system and other controls over metabolism. Proteases are used in multiple fields: in research they are used to break down proteins into their constituent peptides for proteomics; in industry they are often part of the product, such as influencing blood-clotting in medical applications and in detergents to break down protein on the surface of other materials. In the case of laundry detergent, a desire exists to have proteases which are active in cold water, therefore a psychrophilic derived enzyme that can be active at these temperatures would be highly valuable to this industry. [97]

Lipases are enzymes that catalyse the hydrolysis of lipids. They are a subclass of esterases, and they specifically target the ester bonds at the connection between a hydrocarbon chain and the head group of a lipid. In humans they are used in the digestion of consumed fats, in both the gut and within the blood. They are popular within the food industry and are used in the production of cheeses, yoghurts and in baking. They are also used in laundry detergents in a similar role to proteases, targeting lipids instead of proteins. Thermophile lipases are of particular interest to industry: because lipids are often treated in high temperature environments, this creates the requirement for polluting reagents to compensate for enzyme inefficiency. [98]

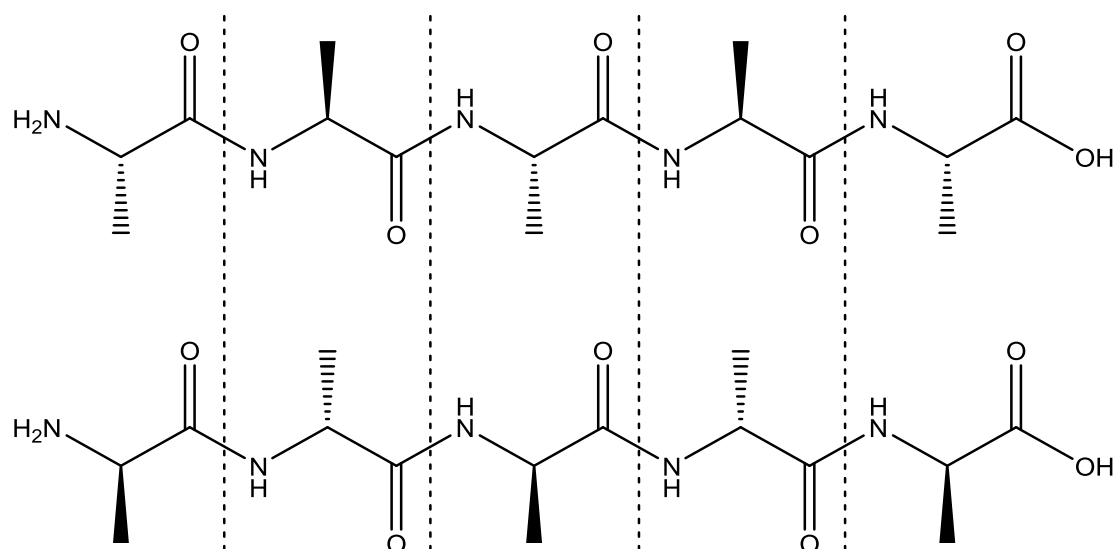
Glycosidases are enzymes that catalyse the breakdown of carbohydrates via hydrolysis of glycosidic bonds, converting complex sugars to smaller carbohydrate products. In animals they are found in saliva and digestive fluids where they digest complex sugars. In the food industry, glycosidases are used to increase the concentration of specific sugars by converting more complex sugars. Additionally, invertases are used to convert sugars from one structure to another. Halophilicity is a valuable property found in extremophiles that can help the conversion of the aforementioned sugars. [99]

## Chapter 2: Suitability of Polyalanine as an IMS calibrant

### *Chapter 2.1: Introduction*

IM-MS instruments equipped with T-wave devices show superb sensitivity and robustness in practical applications, [100] however these need to be calibrated to provide CCS values. Calibration is usually achieved using drift-tube derived measurements of poly-DL-Alanine as a reference. The polyalanine calibrant used by Waters contains a mixture of lengths of polyalanine, produced at low cost and is considered to be stable at 4°C. This calibrant is useful for TWIMS as it is readily available, contains an array of molecule sizes at equal mass differences to one another and is readily ionised in both negative and positive ionisation modes. The potential problem with this is that the standard is assumed to be stereospecific, however peptides do have chiral centres, and have isomers known L (the most common) and D. Whilst L-alanine is readily available, some less expensive peptide synthesis methods can lead to racimisation, and therefore a solution cannot be assumed to be isomerically pure unless tested.

If the polyalanine isomers are found to have a different CCS, and the polyalanine standard is found to contain a mixture of isomers, then the standard could produce inconsistent results on higher resolution instruments. Specific isomers can be produced in the lab via solid-phase synthesis, then these isomers can be tested in a Synapt G2 to test their separation. If it is confirmed that this standard is not suitable for calibration, and alternative standard will need to be proposed.



**Figure 8. LLLLL (Top) and DDDDD (Bottom) enantiomers of polyalanine.** *L and D refer to the enantiomers of the alanine units that make up the peptide.*

It is possible that should the resolving power of the IMS instruments further improve, that the varying quantities of peptide isomers in standards may make polyalanine unsuitable for IMS calibration. In this study we will display the difference in drift time for different penta-alanine isomers using the relatively new technology of cyclic IMS. Cyclic IMS is an IMS instrument in which the drift tube is replaced with a cyclic-separator that allows ions to pass through the “loop” multiple times before being ejected to the transfer cell, allowing for higher resolution separation. <sup>[41]</sup>

## *Chapter 2.2: Aims*

The hypothesis of this study is that peptides, in particular polyalanine, present stereochemistries with different ion mobilities. Without ensuring standards are enantiomerically pure, the CCS will vary depending on the mixtures of isomers. To achieve this, the study will compute CCS values for various enantiomers of polyalanine and compare these to experimentally obtained values. This will also provide the opportunity to examine the accuracy of mobcal for nitrogen drift gas calculations and assess the application of mobcal for determining the collisional cross section of isomers. Utilizing cyclic IMS provides a unique opportunity to further separate these enantiomers where they may not be distinguishable in a linear IMS cell.

### *Chapter 2.3: Experimental Methods*

All peptides (LLLLD-A5, LLLDD-A5, LLDLL-A5, DDL LL-A5, D LLL-A5, LLLLLLLL-A8, D LLLLLLL-A8, D D LLLLLLL-A8, LLLLLLLLLL-A10, D LLLLLLLLL-A10, and D D LLLLLLLLL-A10) were manually synthesized using standard Fmoc solid-phase synthesis [101]. Fmoc-Ala-OH and Fmoc-D-Ala-OH were purchased from ChemPep. HOBt, HBTU, Piperidine, DIPEA and TFA were purchased from Sigma-Aldrich. All peptides were synthesized and then diluted in 50:50 H<sub>2</sub>O:Methanol solution with 0.1% formic acid to concentrations of approximately 10–50  $\mu$ M. Poly-DL-Alanine standard was purchased from Sigma-Aldrich.

The Synapt G2-Si was used to obtain ion mobilities of polyalanine. The solutions were directly infused into the mass spectrometer at a rate of 5ml/min. Peptides were ionized with an ESI source in positive ion mode under the following source conditions: Desolvation gas 600 L/hr; Cone gas 0 L/hr; Collision gas 2 ml/min; Drift gas flow 90 ml/min; Ion spray nebulizer gas 6.5 bar; Capillary voltage 3 kV; Cone voltage 40 V. The resolution was set to 20,000 to balance the sensitivity and mass accuracy of the instrument. CCS values were obtained by acquiring repeated calibrations of Waters Major Mix standard and utilising Driftscope to match the peptide mobility to the calibrated CCS values.

The in-development Cyclic IMS mass spectrometer was used to collect higher resolution polyalanine ion mobilities. The peptides were ionized with an ESI source in positive ion mode under the following source conditions: Desolvation gas 800 L/hr; Cone gas 0 L/hr; Source temperature 70°C; Collision gas 4.7 ml/min; Drift gas flow 40 ml/min; Ion spray nebulizer gas 6 bar; Capillary voltage 2 kV; Cone voltage 20. The resolution was set to 10,000, however this is relatively arbitrary for the in development instrument with incomplete software. CCS values cannot be derived from this data as the instrument could not yet be accurately calibrated.

Polypropylene syringes with polypropylene filters were utilised as reaction vessels for SPPS reactions. For each synthesis 0.1 mmol of Wang resin was functionalised by swelling for one hour in 2ml of DMF. The first amino acid (The C-terminal unit of the desired peptide) was introduced as 0.5mmol of an Fmoc protected amino acid, and was dissolved in 2ml of DMF with 0.49 mmol HBTU. HBTU is used as the coupling reagent bind the carboxylic acid of the amino acid to the last unprotected unit bound to the Wang resin. 1 mmol of DIPEA is added

to the resin solution and shaken for two minutes before adding the fmoc solution, to ensure appropriately basic conditions for coupling. The full solution in the reaction vial is left to shake for at 150rpm for three hours. The reaction vessel and solution should be thoroughly washed with 4 washes of 2ml DMF to remove any unreacted FmocAA and HOBt, and again with four washes of 2ml DCM to remove the DMF.

Successfully coupled and washed FmocAA on the resin is deprotected with 2ml of 2:8 piperidine:DMF over 5 minutes, as piperidine will remove Fmoc protecting groups. The solution should then be washed twice with 2ml DMF and then the deprotection should be repeated with another 2ml of 2:8 piperidine:DMF over 5 minutes to ensure complete deprotection. It is then important that all piperidine is completely removed from the solution to prevent premature deprotection of the next FmocAA added. This is done by washing with four times 2ml of DMF and followed by four times DCM. The last 2ml of DCM should be retained in the syringe until the next FmocAA solution is added to prevent the resin drying out.

Additional amino acids beyond the first should be added via slightly different method. Alongside the 0.5mmol FmocAA and 0.49 mmol HBTU in the 2ml DMF, 0.5 mmol HoBT is added to ensure there is enough benzotiazole tetramethyl units to sustain the coupling reactions. 1 mmol of dipea should still be added to ensure the correct pH for the reaction vessel two minutes prior to adding the Fmoc solution. The coupling reaction should be shaken at 150rpm for only two hours. The same 4x2ml DMF and 4x2ml DCM wash should take place after each coupling. After each coupling stage, the previously reported deprotection method should be completed, ensuring proper washing takes place each time. The reaction mixture can be stored at 4°C in 2ml DCM after the coupling steps and washing, but prior to deprotection steps, for multiple days allowing a partially completed reaction to be stored across the weekend.

After the final step of deprotection was complete on the desired N-terminal FmocAA, the peptide then needed to be cleaved from the resin. Before commencing with the cleavage, wash the solution again with two lots 2ml methanol and another four times with 2ml DCM to ensure all piperidine and any other unreacted components are completely removed. The

peptide was cleaved by adding 1.5ml 95:5 TFA:H<sub>2</sub>O to the solution. The water is present as the cleavage is hydrolytic. This solution was then shaken for two hours. After shaking, the solution was washed into a pre-weighed flask with 1ml of 95:5 TFA:H<sub>2</sub>O and dried under vacuum. The product was then stored in 5ml of MeOH at -20°C.

The mass spectra were collected on a standard and a modified Waters Synapt G2i in positive ion mode using electrospray ionization. The source voltage was set between 3 and 4 kV with cone and extractor voltages set to ~30V and ~5V, respectively. Ions were selected in the initial quadrupole and fragmentation performed with argon gas in the trap stacked ring ion guide located immediately before the IMS cell of the instrument. The TWIMS drift region was filled with Nitrogen as a drift gas.

## *Chapter 2.4: Computational Methods*

The potential energy surfaces of the N-terminal amine protonated polyalanine isomers (LLLLL, DLLLL, DDLLL, LLDLL, LLLDD and LLLLD) were scanned using molecular dynamics simulations (AMBER force field) and quantum chemical calculations performed at the PM6, HF-3-21G, B3LYP/6-31G(p), and B3LYP/6-31+G(d,p) levels of theory using Gaussian [102]. Early steps of the aforementioned molecular dynamics was done in conjunction with Professor Bela Paizs. Theoretical CCSs were computed using the Mobcal function for the Trajectory Method (TM) for helium and nitrogen background gases adapting TM parameters from Ref [43].

The  $b_2$  ions were initially modelled using Tinker as uncharged molecules. Each modelled molecule was then protonated at each heteroatom to simulate proton mobility. A combination of Tinker, GaussView and Molden were used to monitor and document each ion throughout the set of calculations.

Initial calculations for each polyalanine ion involved an Amber forcefield simulated annealing (SA) calculation to generate a large variance of confirmations. [103] The Amber14 package was used to perform these calculations and 20 SA calculations were performed for each ion. In each SA calculation the temperature was scaled between 300°K and 900°K over 30,000 steps. The step size was 0.001 ps. 150 geometries were saved from each of the 20 SA calculations, for a total of 3,000 geometries for each ion to be processed with quantum mechanics. Amber was chosen over ADMP for the geometry generation due to it allowing flexible bonds whilst not allowing individual bonds to break, whereas ADMP leads to dissociation at high temperatures for larger molecules if bonds are allowed to be flexible.

Each of the 1000 geometries were inputted into a pm6 geometry optimisation calculation, run with a single core and 1gb of memory each, individual calculations took approximately 20 seconds to run, depending on the structure of the ion and the similarity between the starting position and end position. Degenerate structures were then removed, significantly reducing the number of geometries, the exact quantity varying by structure. Unique ion geometries were then optimised at HF/3-21G, as a reasonably quick and inexpensive method. Final geometries were again investigated for degenerate structures, and any ions that were degenerate were removed. Any unique ion geometries were again optimised at B3LYP/6-



31G(d), a more expensive but energetically accurate method than used in the prior step. Geometries found to be degenerate at this level were again removed, but these were uncommon at this stage. Geometries found at this stage were considered to be of a high accuracy, and energies are also considered to be sufficiently competently calculated. However, this level of theory is computationally expensive. Due to the many previous steps at PM6, HF and B3LYP/6-31G(d), the initial geometries input into the B3LYP/6-31+G(d,p) optimisations were reasonably close to the local minimum, therefore reducing the cost of these final calculations significantly. At this level of theory, partial charges for atoms were also computed.

Pople basis sets were chosen as they are well established in the field peptide fragmentation, and can be observed in the majority of literature referenced in this chapter. The most common alternative at this level of theory for these basis sets are correlation consistent Dunning basis sets. For gas phase small molecules, these two types of basis sets produce very similar results, with Pople basis sets being arguably the computationally cheaper of the two.

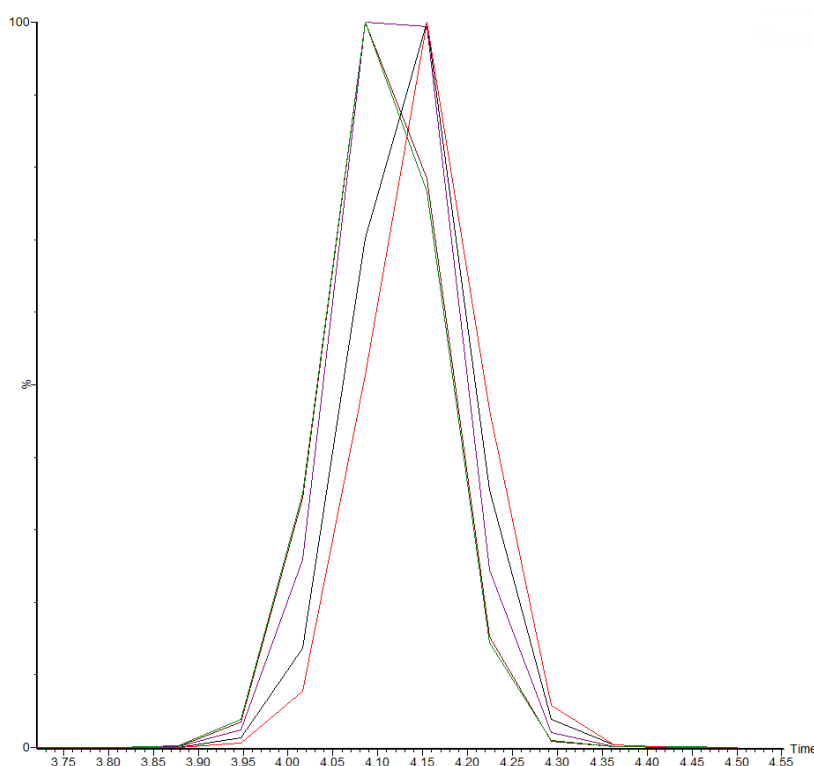
All geometries produced at the B3LYP/6-31+G(d,p) level were then used to generate relevant collisional cross section (CCS) values. The programme mobcal, produced by the MFJ Research Group at Indiana University, was used to compute these CCS values. The trajectory model was used for these theoretical values, this is the most expensive of the methods available in mobcal, but is considered to be the most accurate. At the time of computing the CCS values of the structures produced here, mobcal was the only established programme freely available capable of completing this task, since this time collidoscope and mobcal-MPI have become accessible. The programme can be used to generate CCS values with a He or N<sub>2</sub> drift gas molecule, and touts an accuracy of within 3% of experimentally obtained values. The He parameters supplied in the mobcal programme are well established to be competent within this margin of error, however the N<sub>2</sub> parameters are less well known and are considered to be experimental.

The theoretical CCS values are produced to be compared to experimentally observed CCS values produced on a Waters Synapt G2-Si ion mobility spectrometer filled with both helium and nitrogen drift gases at different points throughout the study.

## Chapter 2.5: Results and Discussion

### 2.5.1: Ion mobility of 5A-Polyalanine Isomers

The five isomers split into three distinct peaks, separated by only a single “bin” (a unit that describes a single capture of ions as a measure of drift time).

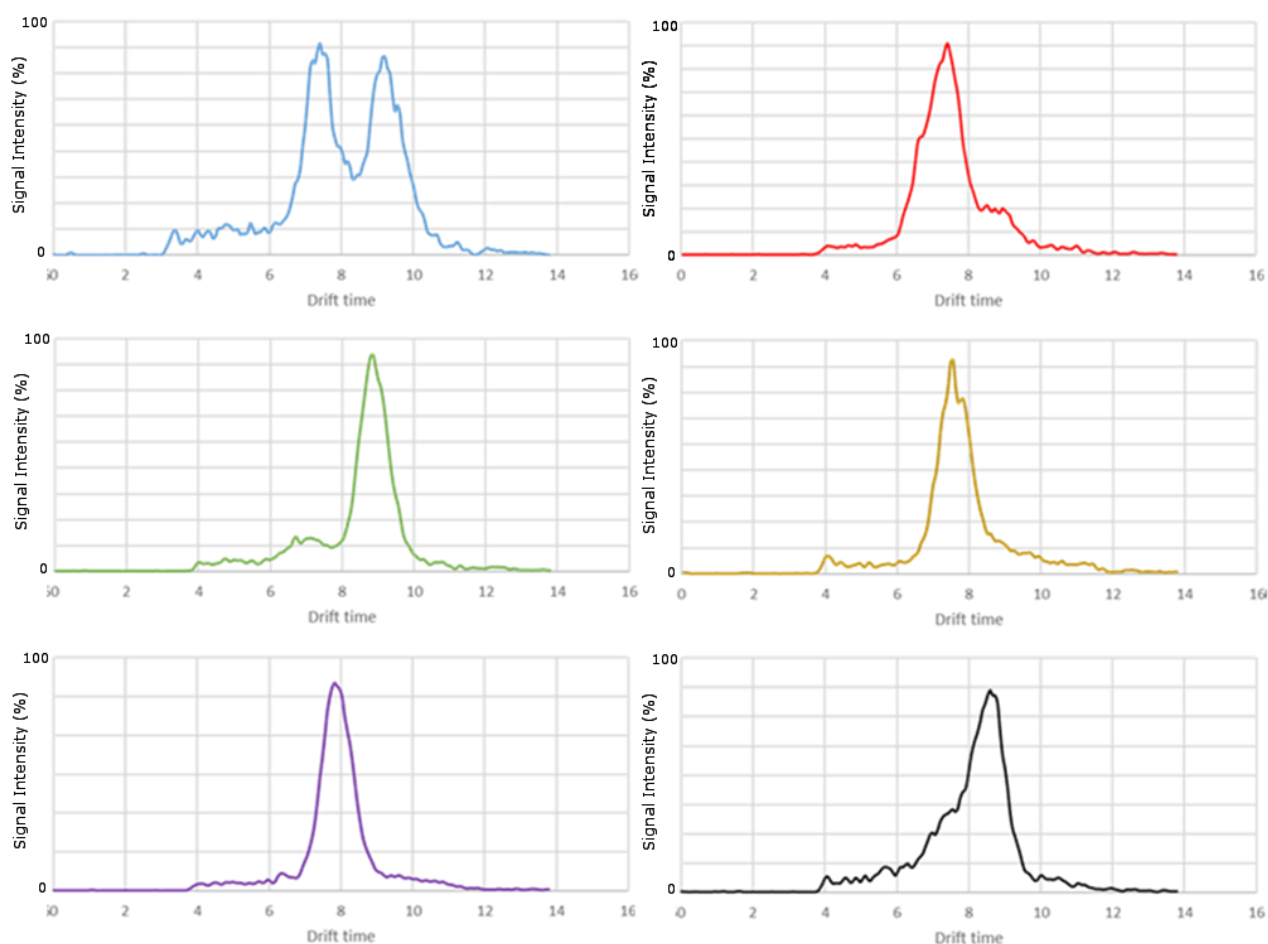


**Figure 9.** Five isomers of penta-alanine separated by TWIMS. (DDLLE (red), DLELE (green), LLELE (purple), LLELD (black), LLELD (brown)) were injected into a Waters Synapt G2-Si individually, with a TWIMS cell filled with N<sub>2</sub> drift gas. Peptide isomers were separated by a single measurement, this repeatable over multiple injections, months apart.

DLELE and LLELD form the lower drift time peak, DDLLE and LLELD form the higher drift time peak, whilst LLELE sits between the two peaks with a flat top suggesting the tip of the peak falls between the two bins. The resolution of the Synapt G2 leads to these peaks being difficult to objectively distinguish. (Figure 9). To clearly distinguish the mobilities of the five isomers, the new cyclic-IMS was used to gradually increase the time the ions spent separating.

As the number of passes through the loop of the cyclic-IMS is increased, the five peptide enantiomers become more clearly separated from one another, which allows them to be clearly distinguished. With 10 Passes giving the best separation whilst retaining enough ions,

the same settings were used to examine the five individual isomers and the 5-membered peptide from a polyalanine mixture standard. (Figure 10)



**Figure 10. Drift time chromatograms of 5-A polyalanine peptides with 10 passes through a cyclic IMS.** The injected polyalanine solutions were the standard mixture used for calibration (top left), DLLLL (top right), DDLLL (centre left), LLLLD (centre right), LLDLL (bottom left) and LLLDD (bottom right). The positions of the individual peptide peaks correspond to the relative positions of peaks observed in a Synapt G2-Si.

The five isomers retain their positions relative to Figure 9, with DLLLL and LLLLD both being detected around the 7.5ms window, DDLLL and LLLDD having a higher drift time, being detected around the 8.5ms window and LLDLL having a drift time between the others at about 7.9ms. It is important to note that the prototype cyclic IMS was not able to provide accurate drift times, therefore these times are arbitrary and are for the use of referencing the relative drift times of these ions. Figure 10 clearly displays the effect that isomers can have on a peptide drift time at high resolutions. These can be compared to the Sigma Aldrich polyalanine that is used as a calibrant for IMS, at low resolutions it shows only one distinct

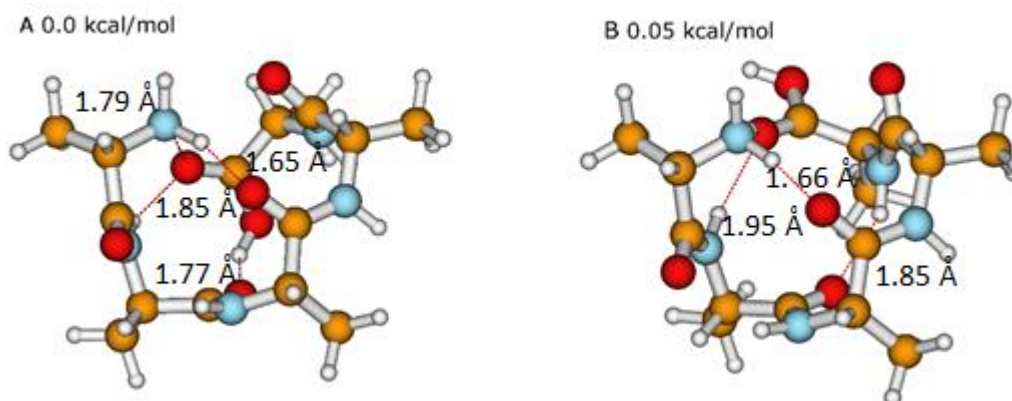
peak as all isomers merge into one peak. However, as shown in Figure 10, the polyalanine standard isolated on the cyclic IMS clearly has two major peaks, potentially made of more than two isomers. The apparent effect of isomers on the polyalanine drift times suggests that this is not a suitable standard for the high resolution IMS instruments.

*Table 1. Relative energies (kcal/mol) and He and N<sub>2</sub> theoretical CCS values (Å<sup>2</sup>) for selected N-terminal amine protonated species. Energies and optimised geometries were computed using the B3LYP/6-31+G(d,p) level of theory and basis set, whilst the CCS values ( $\Omega$ ) were computed using these geometries in mobcal with the trajectory model.*

Species	$\Delta E$	$\Omega_{\text{He}}$	$\Omega_{\text{N}_2}$	Exp. $\Omega_{\text{N}_2}$
LLLLL_A5_A	0	112.8	178.7	180.8 [ <sup>104</sup> ]
LLLLL_A5_B	0.05	114.2	182.7	
LLLLD_A5_A	0	113.9	182.9	180.2
LLLLD_A5_B	1.6	114.9	183.2	
DLLLL_A5_A	0	111.9	177.8	180.3
DLLLL_A5_B	0.5	113.1	178.4	
DLLLL_A5_C	0.8	112.6	178.3	
DDLLE_A5_A	0	113.2	179.9	181.8
DDLLE_A5_B	0.2	115.5	181.1	
LLLDD_A5_A	0	114.1	178.6	181.4
LLLDD_A5_B	0.7	112.6	179.3	
LLDLL_A5_A	0	112.6	179.1	180.7
LLDLL_A5_B	0.7	116.9	186.1	

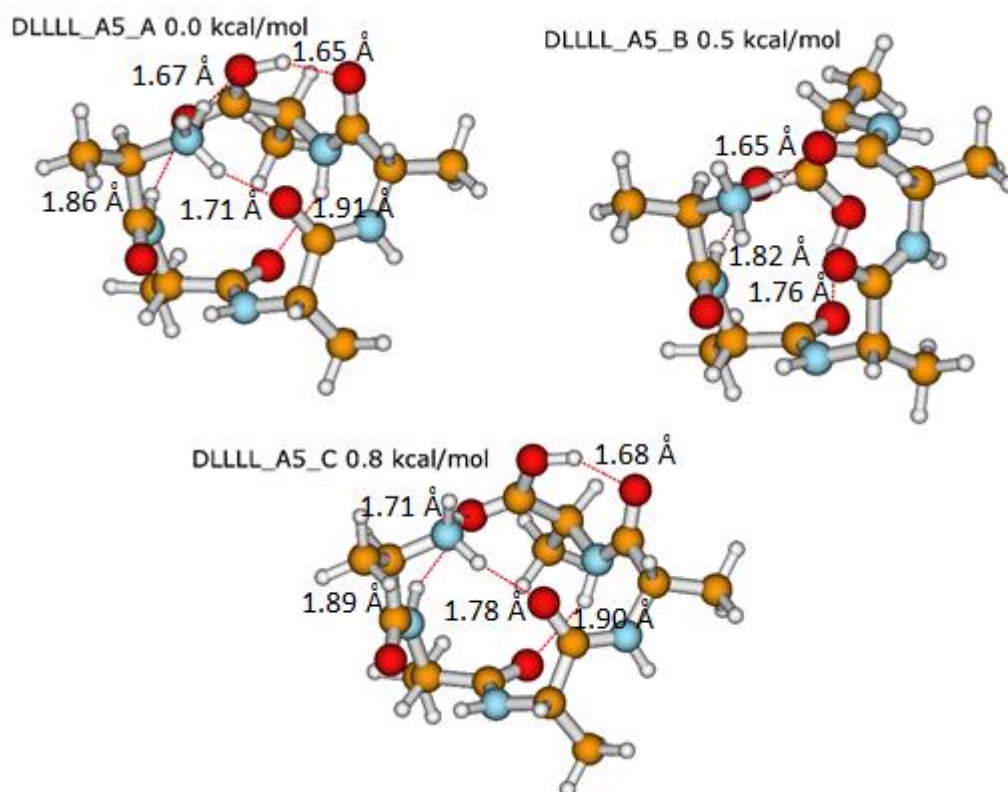
### 2.5.2: Computed structures of 5A-Polyalanine Isomers

The structures for the DLLLL, DDLLE, LLDLE, LLLLD and LLLDD been computed for comparison to the experimental data, the relative energies and computed CCS values for these structures are listed in table 1. Notably, the structures of DLLLL and LLLLD or DDLLE and LLLDD are similar but not identical to one another, the reasons for this can be readily seen in the interaction between the heteroatoms and neighbouring groups (which differ between the C and N terminus). Therefore a change from L to D next to a carboxylic acid will have different effect to a change from L to D next to a terminal amine.



**Figure 11. Low energy structures of N-terminal amine protonated LLLLL-A5.** Energies and Optimised structures computed at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. A) Mobcal was used to generate a CCS value of  $112.8 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $178.7 \text{ \AA}^2$  for Nitrogen drift gas. B) Mobcal was used to generate a CCS value of  $114.2 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $182.7 \text{ \AA}^2$  for Nitrogen drift gas. The respective experimentally determined CCS value for a LLLLL-A5 ion was not obtained.

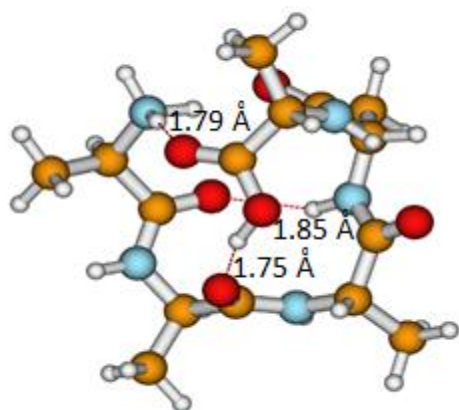
Structures A and B, illustrated in in figure 11, are similar of LLLLL are similar to one another, but have one main difference in the rotation of the carboxylic acid group. In structure A the carboxylic acid is hydrogen bonded to two other groups, whilst in structure B it is only hydrogen bonded to the N-terminus. Energetically, the two structures are very close, with only 0.05 kcal/mol between them, determined with B3LYP/6-31+G(d,p). It is likely that A and B interconvert, the backbones are quite similar so averaging the two CCS values is appropriate



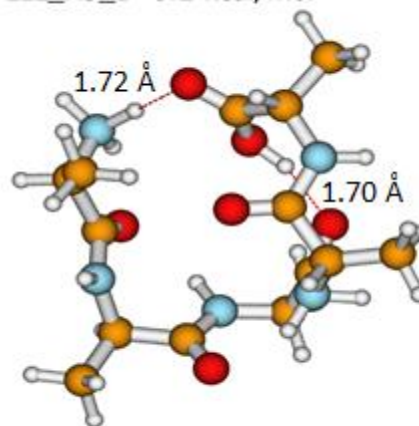
**Figure 12. Low energy structures of N-terminal amine protonated DLLL-A5.** Energies and Optimised structures computed at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. A) Mobcal was used to generate a CCS value of  $111.9 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $177.8 \text{ \AA}^2$  for Nitrogen drift gas. B) Mobcal was used to generate a CCS value of  $113.1 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $178.4 \text{ \AA}^2$  for Nitrogen drift gas. C) Mobcal was used to generate a CCS value of  $112.6 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $178.3 \text{ \AA}^2$  for Nitrogen drift gas. The respective experimentally determined CCS value for a DLLL-A5 ion was  $180.3 \text{ \AA}^2$  with a nitrogen drift gas.

DLLL computations produced three similar structures, illustrated in in figure 12, with relative energies less than 1.0 kcal/mol, determined with B3LYP/6-31+G(d,p). Structures A and C are very similar, the only difference between them is around the orientation of the terminal COOH group and its role in solvating the charge. The main difference between structures A and B is in the position of the A4-A5 amide bond and the carboxylic acid H-bonding pattern. It is likely that A, B and C interconvert, the backbones are quite similar so averaging the three CCS values is reasonable.

DDLLL\_A5\_A 0.0 kcal/mol

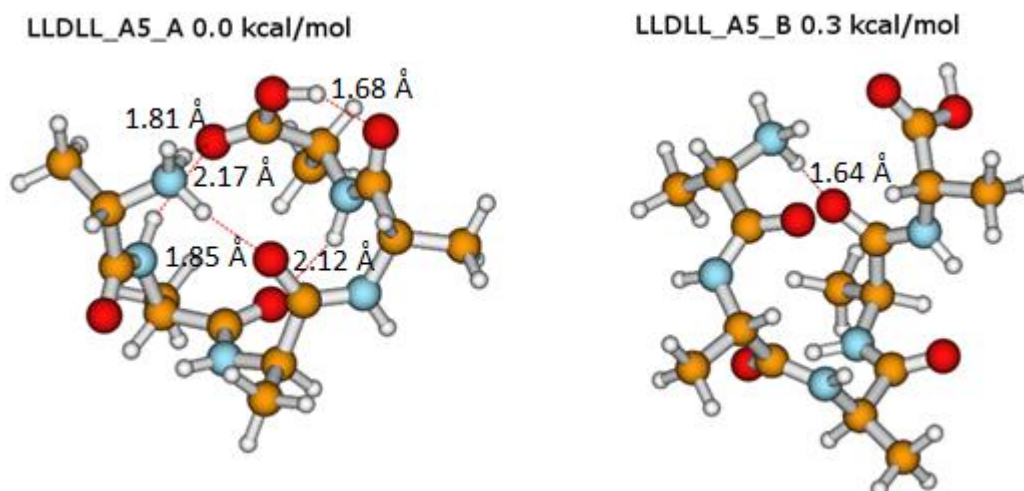


DDLLL\_A5\_B 0.2 kcal/mol



**Figure 13. Low energy structures of N-terminal amine protonated DDLLL-A5.** Energies and Optimised structures computed at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. A) Mobcal was used to generate a CCS value of  $113.2 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $179.9 \text{ \AA}^2$  for Nitrogen drift gas. B) Mobcal was used to generate a CCS value of  $115.5 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $181.1 \text{ \AA}^2$  for Nitrogen drift gas. The respective experimentally determined CCS value for a DDLLL-A5 ion was  $181.8 \text{ \AA}^2$  with a nitrogen drift gas.

From the geometries computed for DDLLL, illustrated in in figure 13, the two lowest energy structures had  $\Delta E$ 's of 0 kcal/mol (A) and 0.2 kcal/mol (B), determined with B3LYP/6-31+G(d,p). Structures A and B have some significant differences, the main differences are A4-A5, A3-A4 amide bonds, which significantly changes H-bonding pattern. For DDLLL it is not likely that A and B interconvert, the backbones are quite different so averaging the two CCS values is not appropriate.

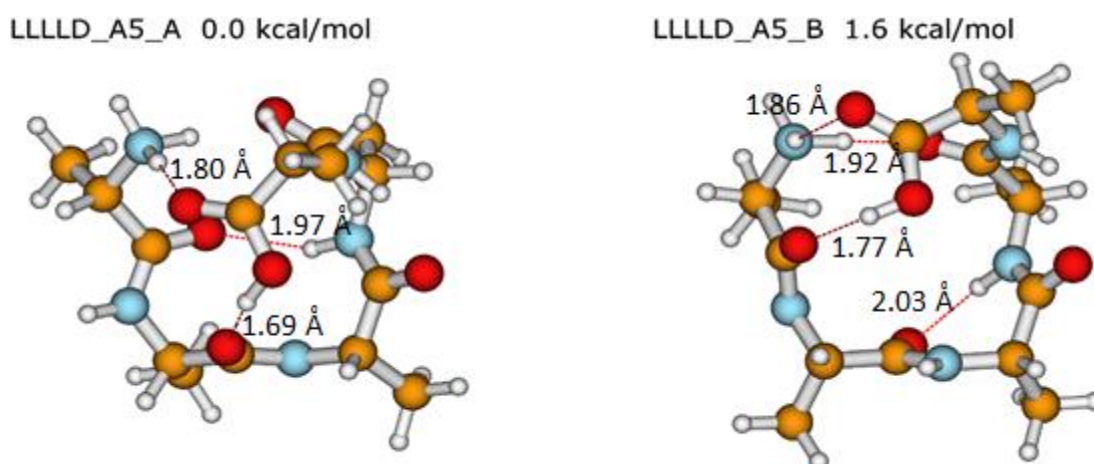


**Figure 14. Low energy structures of N-terminal amine protonated LLDLL-A5.** Energies and Optimised structures computed at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. A) Mobcal was used to generate a CCS value of  $112.6 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $179.1 \text{ \AA}^2$  for Nitrogen drift gas. B) Mobcal was used to generate a CCS value of  $116.9 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $186.1 \text{ \AA}^2$  for Nitrogen drift gas. The respective experimentally determined CCS value for a LLDLL-A5 ion was  $180.7 \text{ \AA}^2$  with a nitrogen drift gas.

The LLDLL two minimum energy structures, illustrated in in figure 14, with a difference 0.3 kcal/mol (determined with B3LYP/6-31+G(d,p)), although the energy is similar between the two structures, there are large differences in the intramolecular interactions. The primary difference between the structures is that A contains 4 clear hydrogen bonds, whilst B contains only one hydrogen bond. The largest backbone change between these structures is between A3-A4 and A4-A5 which moves the carboxylic acid significantly further from the  $\text{NH}_3$ . The

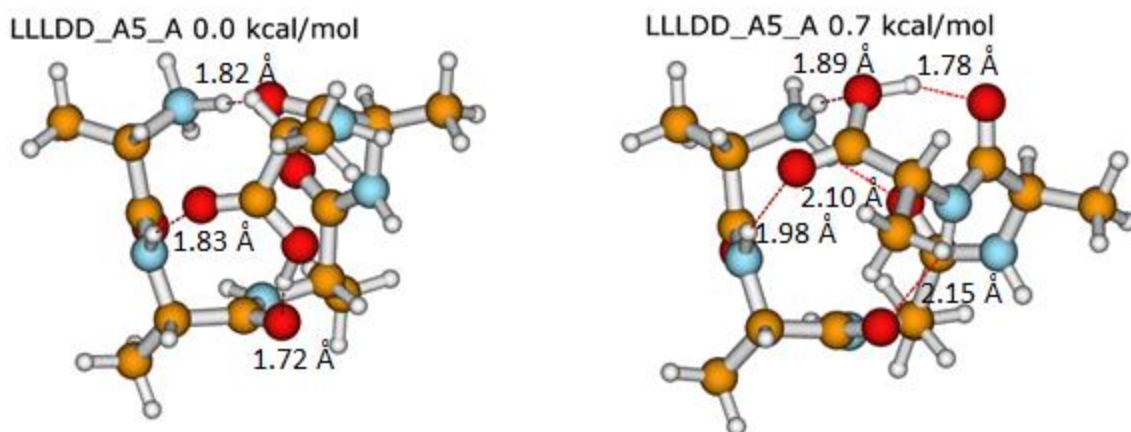


major differences between these structures means that it is not appropriate to average their collisional cross sections.



**Figure 15. Low energy structures of N-terminal amine protonated LLLD-A5.** Energies and Optimised structures computed at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. A) Mobcal was used to generate a CCS value of  $113.9 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $182.9 \text{ \AA}^2$  for Nitrogen drift gas. B) Mobcal was used to generate a CCS value of  $114.9 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas and  $183.2 \text{ \AA}^2$  for Nitrogen drift gas. The respective experimentally determined CCS value for a LLLD-A5 ion was  $180.2 \text{ \AA}^2$  with a nitrogen drift gas.

LLLLD's lowest energy geometries, displayed in in figure 15, have  $\Delta E$ 's of 0 kcal/mol (A) and 1.6 kcal/mol (B), determined with B3LYP/6-31+G(d,p). Structures A and B are quite different, with the main differences being orientation of A1-A2 and A2-A3 amide bonds and the COOH...O H-bond. It is not likely that A and B interconvert, the backbones are again quite different so averaging the two CCS values is not appropriate.



**Figure 16. Low energy structures of N-terminal amine protonated LLLDD-A5.** Energies and Optimised structures computed at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. A) Mobcal was used to generate a CCS value of 114.1 Å<sup>2</sup> for this geometry using the trajectory model with a Helium drift gas and 179.9 Å<sup>2</sup> for Nitrogen drift gas. B) Mobcal was used to generate a CCS value of 115.5 Å<sup>2</sup> for this geometry using the trajectory model with a Helium drift gas and 181.1 Å<sup>2</sup> for Nitrogen drift gas. The respective experimentally determined CCS value for a LLLDD-A5 ion was 181.4 Å<sup>2</sup> with a nitrogen drift gas.

LLLDD's structures were computed with 5 structures, illustrated in in figure 16, being similar in energy with  $\Delta E$ 's of 0 kcal/mol (A), 0.7 kcal/mol (B), 1.0 kcal/mol (C), 1.4 kcal/mol (D) and 1.5 kcal/mol (E), determined with B3LYP/6-31+G(d,p). Structures A and B differ only in A4-A5 amide bond and some hydrogen bonding, C can be generated from B by flipping around the A4-A5-COOH group, D can be generated from C just by rotating the COOH group, E is very similar to D, the only difference is in the C-terminus COOH group. Interconversion is likely.

*Table 2. A comparison of computed CCS's obtained from the lowest energy geometry of each 5-Alanine isomer and the experimental value obtain on a synapt waters G2 with a nitrogen drift gas. Energies and optimised geometries were computed using the B3LYP/6-31+G(d,p) level of theory and basis set, whilst the theoretical CCS values were computed using these geometries in mobcal with the trajectory model.*

Peptide	Experimental $\Omega_{N_2}$	Computational $\Omega_{N_2}$
DDL <sup>LL</sup>	181.8	179.9
D <sup>LL</sup> LL	180.3	178.2
LLD <sup>LL</sup>	180.7	179.1
LL <sup>LL</sup> D	180.2	182.9
LLLDD	181.4	179.0

The collisional cross sections produced from the computational data were compared to the CCS's produced from the experimental data, as displayed in table 2. The theoretical values

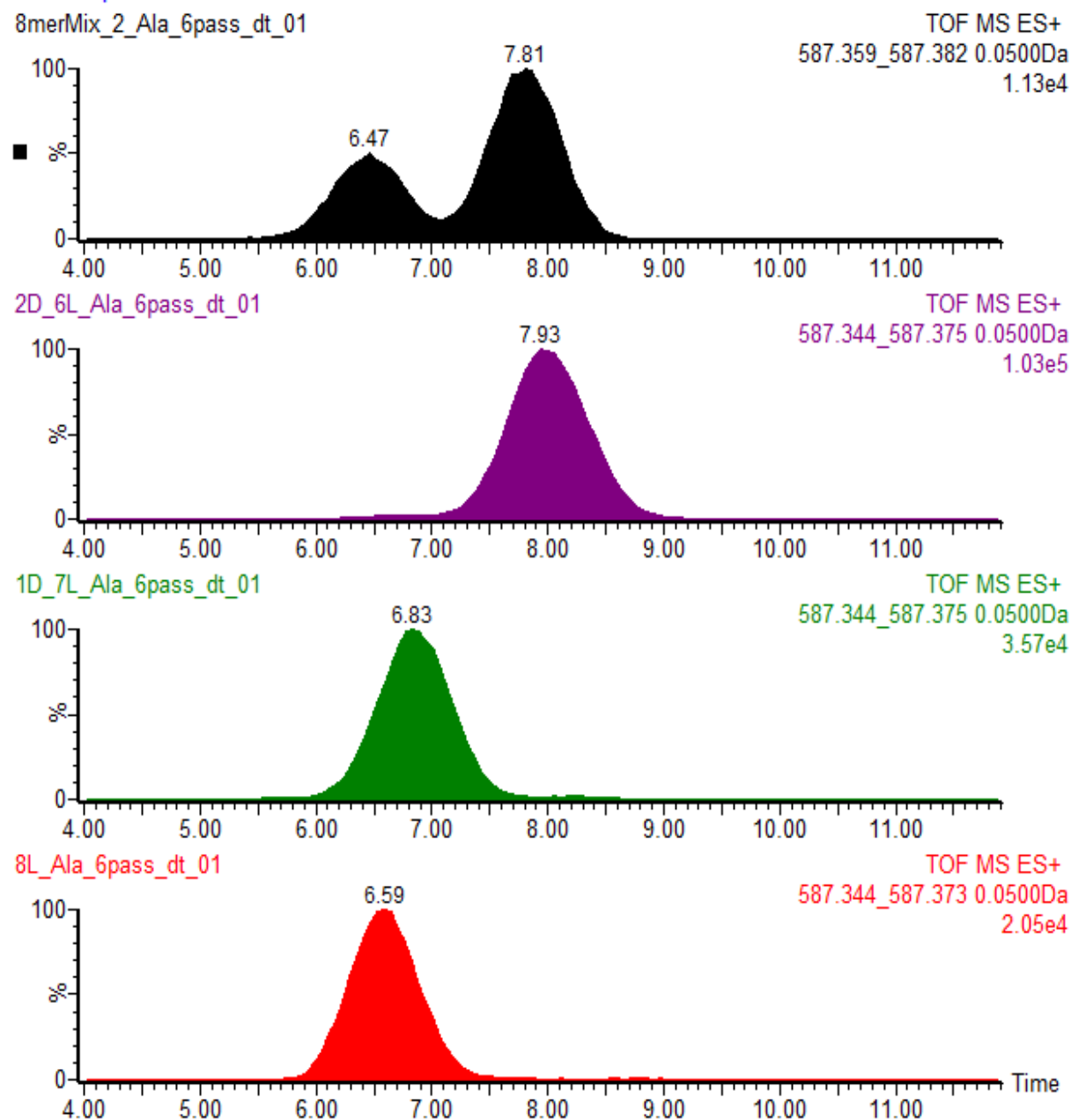
and the experimental values have a poor overall match, with the lowest experimental values have the greatest theoretical values and vice versa. This shows clearly that the mobcal program is not yet accurate enough to distinguish isomers from one another, and the generally accepted 3% error of margin is too wide for such similar ions.

### *2.5.3: Ion Mobility Investigation of 8A- and 10A- Polyalanine Isomers*

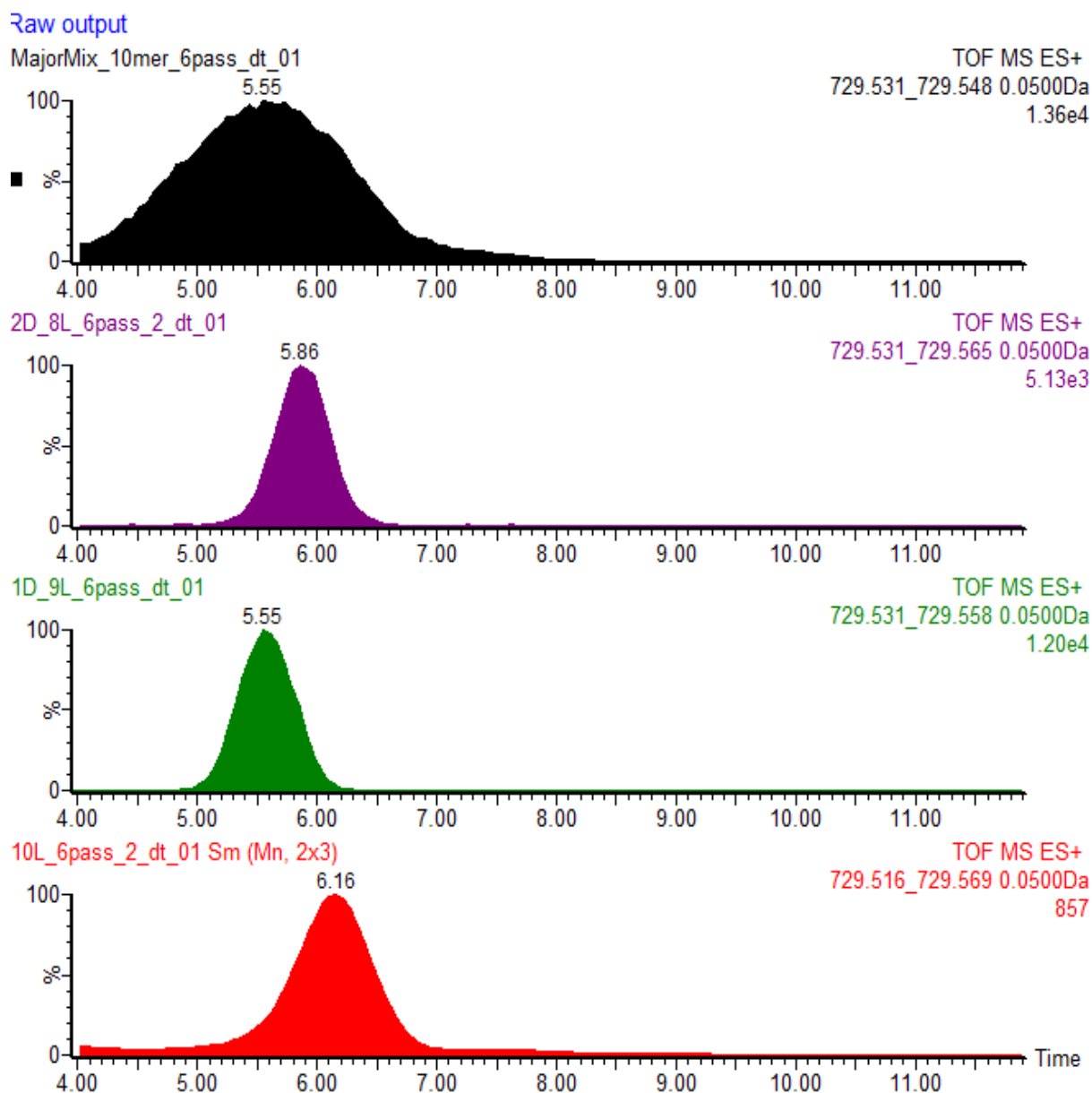
5-Alanine is only present in the polyalanine standard as a fragment/degradation of larger polyalanine peptides. Two peptides that are present from the synthesis include 8-Alanine and 10-Alanine, both of which were synthesised by SPPS. The isomers of 8-Alanine synthesised were LLLLLLLL, DLLLLLLL and DDLLLLLL. The isomers of 10-alanine synthesised were LLLLLLLLLL, DLLLLLLLLL and DDLLLLLLLL. Each of these peptides were run on the cyclic IMS and compared to the polyalanine standard equivalents.

Whilst the 5-Alanine results show clearly that isomers of polyalanine do have different ion mobilities (and CCS's), 5-Alanine is not intentionally synthesized into the standard mixture, but is a product of degradation of larger polyalanine peptides. Therefore 8-Alanine and 10-Alanine were synthesized, as these are present in the standard mixture.

# Raw output



**Figure 17. Separation of three 8-alanine isomers.** Three isomers of 8-alanine, LLLLLLLL (Bottom), DLLLLLLL (Centre-bottom) DDLLLLLL (Centre-top) and the polyalanine standards 8-alanine (Top) acquired after 6 passes on the prototype cyclic IMS, shown separately.



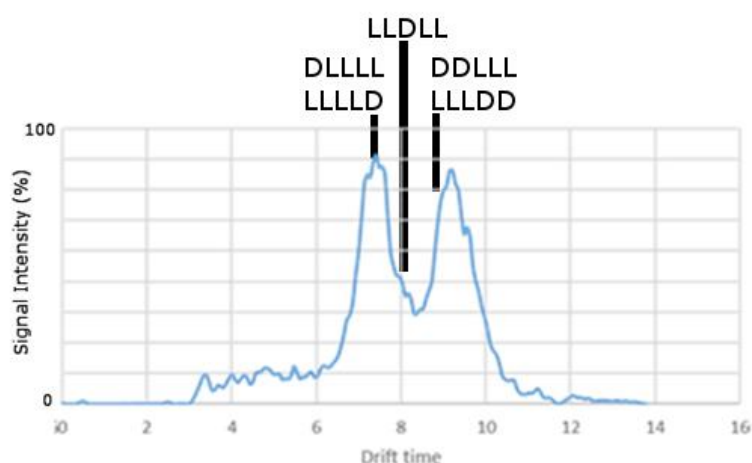
**Figure 18. Separation of three 10-alanine isomers.** Three isomers of 10-alanine, LLLLLLLLL (Bottom), DLLLLLLLL (Centre-bottom) DLLLLLLLL (Centre-top) and the polyaniline standards 10-alanine (Top) acquired after 6 passes on the prototype cyclic IMS, shown separately.

The results of 8-Alanine experiment show that the 3 tested peptides have different ion mobilities in the cyclic-IMS, displayed in figure 17. LLLLLLLL has the lowest drift time, DLLLLLLL has a slightly higher drift time (easily distinguishable on the cyclic-IMS) and DLLLLLLL has a significantly higher drift time with an almost completely separated peak. This reinforces the evidence that polyaniline isomers have significantly distinct ion mobilities, and thus the polyaniline standard was also tested for its 8-Alanine component. The ion mobility spectra for the standards 8-Alanine peptide shows two distinct peaks, aligning approximately with the known isomers.

The 10-Alanine experiment produced similar results, shown in figure 18. Of the three 10-alanine isomers run the LLLLLLLLLL polyalanine has the longest drift time, DLLLLLLLLL has the shortest drift time and the DDLLLLLLLL's drift time sits between the other two. Each of these peaks are clearly distinct. The 10-alanine peptide present in the polyalanine standard is made up of a single peak which is very broad, encompassing the drift times of each of the known isomers and potentially others. Due to the similarities found within the CCS values for 5-A polyalanine, it was not considered a reasonable use of resources to similarly compute theoretical structures for 8-A and 10-A as they would be unable to be differentiated based on their theoretical CCS values generated by mobcal.

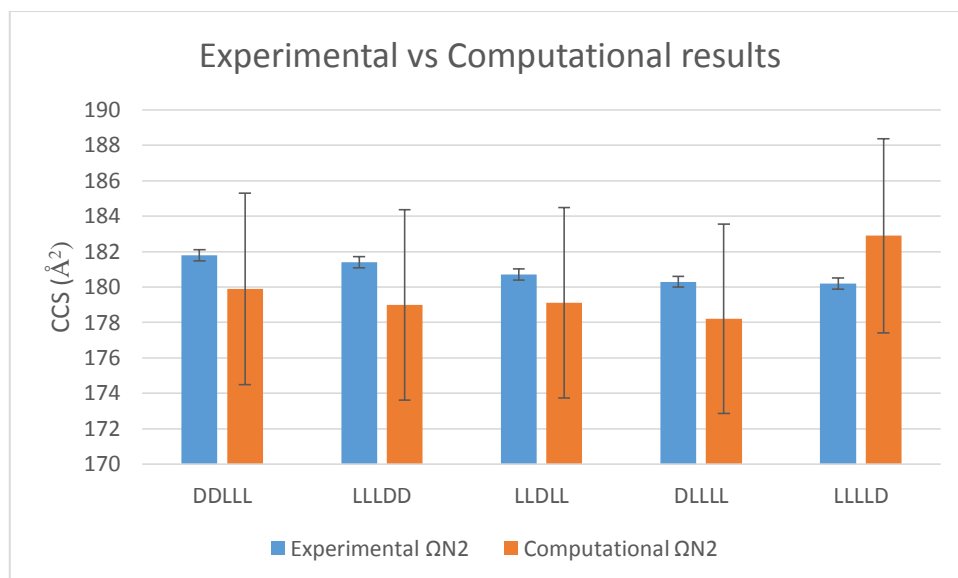
## Chapter 2.6: Conclusion

The present results strongly suggest that the polyalanine standard is not monoisomeric, and the enantiomers that are present are not identical in ion mobility. Each length of polyalanine has displayed different mobilities for different isomers, and most importantly, the polyalanine standard use for calibration presents a broad spectrum of ion mobilities. With no evidence that the quantity of each isomer is controlled, the polyalanine mixture is not an appropriate calibrant for IMS. Alternatives to polyalanine have already been proposed for other reasons, such as polymers; polyethylene glycol (PEG) for lower drift times and polylactide (PLA) for higher drift times. [105] Although these options are restricted to positive ionisation mode, whereas polyalanine is suitable for both positive and negative.



**Figure 19.** *The relative mobilities of polyalanine isomers in a polyalanine standard. The displayed polyalanine mobility was of the calibrant used by Waters. The positions of the individual polyalanine isomers correspond to the relative positions of peaks observed in a Synapt G2-Si.*

Additionally, the application of mobcal calculation of CCS values was found to be too inaccurate for identifying individual isomers of polyalanine. All computed CCS values fell within the 3% error of the experimental values that is expected with mobcal calculations. However, all experimental CCS values for 5A-polyalanine were also within 3% of each other and the computed values did not match the trend displayed by experimental value. Mobcal is therefore not a suitable software for distinguishing the mobilities of structures with such similar compositions and is likely not suitable to distinguish any set molecule of molecules within the 3% error threshold.



**Figure 20.** The difference seen between the computational CCS values versus the experimentally obtained CCS values. The experimental values were obtained on a synapt waters G2 with a nitrogen drift gas. Energies and optimised geometries were computed using the B3LYP/6-31+G(d,p) level of theory and basis set, whilst the theoretical CCS values were computed using these geometries in mobcal with the trajectory model.



## Chapter 3: Determination of $b_2$ Ion Structure through Ion Mobility Spectrometry

### *Chapter 3.1: Introduction*

Whereas  $y_n$  ions have predictable truncated peptide structures [3], the structure of  $b_n$  ions can take various forms, and this form is influenced by the size of the  $b_n$  ion. In the past, all  $b_n$  ions were assumed to take an acylium ion structure [106], however the  $b_1$  ions (which would be expected to take an acylium structure) are very unstable, which suggests that larger ions that take other forms where possible [107], which has been explained with the oxazolone and cyclic structures. It has been observed that in larger  $b_n$  ions (where  $n > 5$ ) that the ions tend to take the form of macrocyclic structures due to the reduction in steric hindrance, whilst smaller  $b_n$  ions ( $n = 3$  or  $4$ ) tend to take the structure of oxazolone derivatives due to the higher energy requirements of interacting the C and N terminus of the product ion. However, in  $b_2$  ions specifically a mixture of oxazolone and macrocyclic structures have been observed. Macrocyclic structures often have a deeper minimum energy but have an energy barrier of a cis to trans conformational change. This energy barrier may be overcome in the environment of the mass spectrometer, but more frequently  $b_2$  ions are observed to take the structure of an oxazolone derivative.

The structure of peptide fragments can be confirmed via a few methods, one of which is addition fragmentation and another is infrared multiple photon dissociation (IRMPD). In ion trap and orbitrap mass spectrometers, ions can be fragmented more than once, therefore a  $b_2$  ion can be isolated and fragmented which can give structural information about the  $b_2$  ion. Oxazolone ions can be fragmented to  $a_2$  ions more readily than macrocyclic  $b_2$  ions, however due to the influence of side chain interactions, not all oxazolone ions will readily fragment to  $a_2$  ions and this method requires a specific mass spectrometer. IRMPD produces vibrational spectral data, which can be combined with theoretically computed vibrational spectral data to determine the structure of an ion. IRMPD is a mass spectrometry technique that uses an infrared laser on the ions in a mass spectrometer, most often used in conjunction with FT-ICR. These instruments are expensive and there are very few of them available, an example of which is the CLIO Free Electron Laser at Paris-Sud University in Orsay, France. This is a

powerful strategy to elucidate a structure for a specific ion but is not a readily available technique. [108]

An alternative to the previously discussed methods is to use ion mobility spectrometry. Ion mobility spectrometry allows separation of fragment ions based on their respective collisional cross sections, an attribute that can be computationally calculated. Oxazolone and diketopiperazine derivative ions will likely have different ion mobilities and with a comprehensive computational analysis of these ions show be able to prove or disprove a trend towards either one type of structure. Advantages of this technique include the commercial availability of IMS instruments, and that IMS devices are often coupled to TOF detectors. This allows for the on-line determination of a structural property that with a sufficiently accurate algorithm to determine a structures CCS will allow ions to be ions to be proven in an accurate, efficient and cost effective manner. As these instruments can be used for routine proteomics, the computed CCSs can potentially be used in a database to more accurately determine the sequence peptides and proteins.

### *Chapter 3.2: Aims*

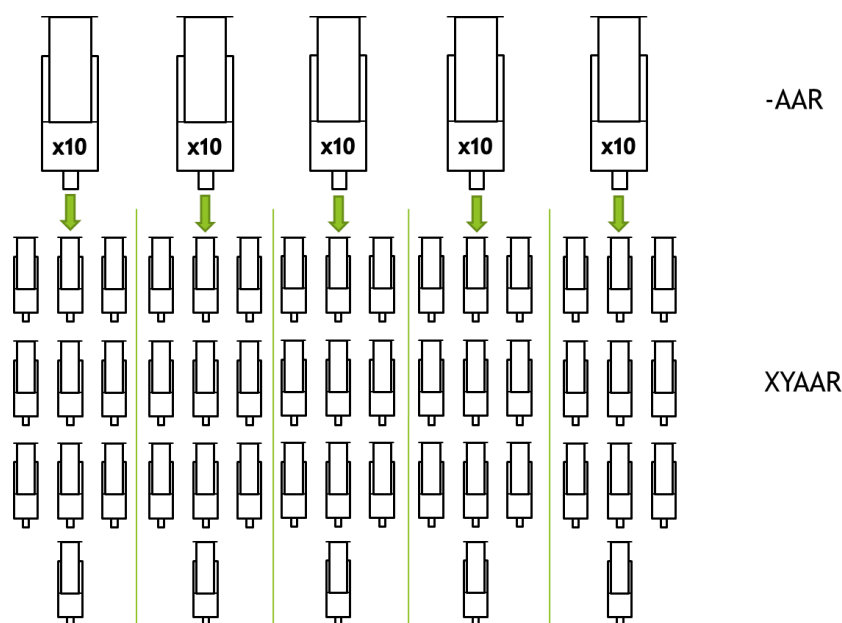
The aim of this chapter is to provide an insight into the structure of peptide  $b_2$  ions via computational and experimental means. The approach is to compute the CCS of an array of  $b_2$  ion structures in the forms of oxazolone and macrocyclic diketopiperazine derivatives and then compare these results to experimentally obtained CCS values for the respective  $b_2$  ions. Additionally, the study aims to examine and compare the competency of the helium and nitrogen parameters of the mobcal programme. Helium drift gasses are known to have the most accurate parameters for computing trajectory model mobilities, however helium is the drift gas (of those commonly used) that produces the lowest resolution ion mobility spectra. Nitrogen is a commonly used drift gas with a good resolutions, therefore it would prove valuable if the  $N_2$  drift gas function of mobcal was able to produce CCS values of a similar accuracy to the helium function.

### *Chapter 3.3: Experimental Methods*

Eleven peptides used in this analysis were obtained from Heidelberg University, this includes: CTVVVR, ETFANR, TCLIIR, FCLSLR, GCAQALR, YGGFL, ACSFIIR, FLVGSTNK, FLLDTNK, FLEQQNK and FLDLWNK. These peptides were chosen for preliminary analysis due to their ionisation potential, presence in literature and ready availability. The peptide MRFA was purchased from Sigma Aldrich. All other XYAAR peptides were synthesized by solid phase peptide synthesis (SPPS). All peptides were dissolved in MeOH and spiked with formic acid. The IMS analysis of the peptides was performed on a Waters Synapt G2 Si by direct infusion.

The peptides were ionized with an ESI source in positive ion mode under the following source conditions: Desolvation gas 800 L/hr; Cone gas 30 L/hr; Collision gas 0.4 ml/min; Drift gas flow 100 ml/min; Ion spray nebulizer gas 6.5 bar; Capillary voltage 3 kV; Cone voltage 50 V. The resolution was set to 20,000 to balance the sensitivity and mass accuracy of the instrument. CCS values were obtained by acquiring repeated calibrations of Waters Major Mix standard and utilising Driftscope to match the peptide mobility to the calibrated CCS values.

Polypropylene syringes with polypropylene filters were utilised as reaction vessels for SPPS reactions. For each synthesis 0.1 mmol of Wang resin was functionalised by swelling for one hour in 2ml of DMF. The first amino acid (The C-terminal unit of the desired peptide) was introduced as 0.5mmol of an Fmoc protected amino acid, and was dissolved in 2ml of DMF with 0.49 mmol HBTU. HBTU is used as the coupling reagent bind the carboxylic acid of the amino acid to the last unprotected unit bound to the Wang resin. 1 mmol of DIPEA is added to the resin solution and shaken for two minutes before adding the fmoc solution, to ensure appropriately basic conditions for coupling. The full solution in the reaction vial is left to shake for at 150rpm for three hours. The reaction vessel and solution should be thoroughly washed with 4 washes of 2ml DMF to remove any unreacted FmocAA and HOBt, and again with four washes of 2ml DCM to remove the DMF.



**Figure 21. The method of efficiently synthesising a large number of peptides with the same C-Terminal amino acids.** In the initial steps of the reaction, amino acids are added in sequence from the C-terminus in a large reaction vessel. Once the sequence begins to diverge at the AA<sub>2</sub> position, the reaction solution is divided into multiple smaller reaction vessels and the reaction is continued with separate additions of different amino acids.

A single peptide is synthesised in a 6ml syringe. If multiple peptides are being synthesised, a larger syringe is used and the quantity of each material is multiplied by the number of peptide samples being synthesised. Peptides are synthesised from the C-terminus, so a peptide such as VAAAR would first have the R group added. When multiple peptides are being synthesised and the amino acid added next is no longer the same for each peptide, the material can be split into different syringes after coupling reactions and subsequent washing is complete.

Successfully coupled and washed FmocAA on the resin is deprotected with 2ml of 2:8 piperidine:DMF over 5 minutes, as piperidine will remove Fmoc protecting groups. The solution should then be washed twice with 2ml DMF and then the deprotection should be repeated with another 2ml of 2:8 piperidine:DMF over 5 minutes to ensure complete deprotection. It is then important that all piperidine is completely removed from the solution to prevent premature deprotection of the next FmocAA added. This is done by washing with four times 2ml of DMF and followed by four times DCM. The last 2ml of DCM should be retained in the syringe until the next FmocAA solution is added to prevent the resin drying out.

Additional amino acids beyond the first should be added via slightly different method. Alongside the 0.5mmol FmocAA and 0.49 mmol HBTU in the 2ml DMF, 0.5 mmol HoBT is added to ensure there is enough benzotiazole tetramethyl units to sustain the coupling reactions. 1 mmol of dipea should still be added to ensure the correct pH for the reaction vessel two minutes prior to adding the Fmoc solution. The coupling reaction should be shaken at 150rpm for only two hours. The same 4x2ml DMF and 4x2ml DCM wash should take place after each coupling. After each coupling stage, the previously reported deprotection method should be completed, ensuring proper washing takes place each time. The reaction mixture can be stored at 4°C in 2ml DCM after the coupling steps and washing, but prior to deprotection steps, for multiple days allowing a partially completed reaction to be stored across the weekend.

After the final step of deprotection was complete on the desired N-terminal FmocAA, the peptide then needed to be cleaved from the resin. Before commencing with the cleavage, wash the solution again with two lots 2ml methanol and another four times with 2ml DCM to ensure all piperidine and any other unreacted components are completely removed. The peptide was cleaved by adding 1.5ml 95:5 TFA:H<sub>2</sub>O to the solution. The water is present as the cleavage is hydrolytic. This solution was then shaken for two hours. After shaking, the solution was washed into a pre-weighed flask with 1ml of 95:5 TFA:H<sub>2</sub>O and dried under vacuum. The product was then stored in 5ml of MeOH at -20°C.

The mass spectra and ion mobility spectra were collected on a standard and a modified Waters Synapt G2i in positive ion mode using electrospray ionization by direct infusion at a flow rate of 0.5 ml/min. The source voltage was set between 3 and 4 kV with cone and extractor voltages set to ~30V and ~5V, respectively. Ions were selected in the initial quadrupole and fragmentation performed with argon gas in the trap stacked ring ion guide located immediately before the IMS cell of the instrument. The TWIMS drift region was filled with Nitrogen as a drift gas.

### *Chapter 3.4: Computational Methods*

All computational work was performed SCW servers, formerly HPCW, including molecular modelling and Gaussian calculations. All calculations from ADMP to the final geometry optimisation calculations were performed using the Gaussian09 suite of programmes.

The  $b_2$  ions were initially modelled using Gaussview as uncharged molecules. Each modelled molecule was then protonated at each heteroatom to simulate proton mobility. A combination of GaussView and Molden were used to monitor and document each ion throughout the set of calculations.

Initial calculations for each  $b_2$  ion, protonated at each heteroatom, involved an ADMP calculation to generate a large variance of conformations. The initial kinetic energy was set to 0.36 Hartrees, with 20,000 steps (of which each step was 1.5 femtoseconds), SCF was converged at each point to within  $10^{-5}$ . Additionally, the simulation was temperature controlled at 800°K, velocity was scaled for the first 500 steps and checked/maintained every 10 steps. Distance matrices were not printed. After ADMP calculations were complete, the geometry at every 20 steps of the trajectory were saved, for a total of 1000 geometries for every  $b_2$  ion.

Each of the 1000 geometries were inputted into a pm6 geometry optimisation calculation, run with a single core and 1gb of memory each, individual calculations took approximately 10 seconds to run, depending on the structure of the ion and the similarity between the starting position and end position. Degenerate structures were then removed, significantly reducing the number of geometries, the exact quantity varying by structure. Unique ion geometries were then optimised at HF/3-21G, as a reasonably quick and inexpensive method. Final geometries were again investigated for degenerate structures, and any ions that were degenerate were removed. Any unique ion geometries were again optimised at B3LYP/6-31G(d), a more expensive but energetically accurate method than used in the prior step. Geometries found to be degenerate at this level were again removed, but these were uncommon at this stage. Of the remaining geometries only the 30 lowest in energy (as determined in the B3LYP/6-31G(d) step) were selected for optimisation at the final stage of calculations. In most cases, this included all geometries as less than 30 unique minimums had

been found, however this limit of 30 was used to restrict the potential computational cost of the following steps. Ions were then optimised at the B3LYP/6-31+G(d,p) level of theory. Geometries found at this stage were considered to be of a high accuracy, and energies are also considered to be sufficiently competently calculated. However, this level of theory is computationally expensive. Due to the many previous steps at PM6, HF and B3LYP/6-31G(d), the initial geometries input into the B3LYP/6-31+G(d,p) optimisations were reasonably close to the local minimum, therefore reducing the cost of these final calculations significantly. At this level of theory, partial charges for atoms were also computed.

All geometries produced at the B3LYP/6-31+G(d,p) level were then used to generate relevant collisional cross section (CCS) values. The programme mobcal, produced by the MFJ Research Group at Indiana University, was used to compute these CCS values. The trajectory model was used for these theoretical values, this is the most expensive of the methods available in mobcal, but is considered to be the most accurate. The programme can be used to generate CCS values with a He or N<sub>2</sub> drift gas molecule, and touts an accuracy of within 3% of experimentally obtained values. The He parameters supplied in the mobcal programme are well established to be competent within this margin of error, however the N<sub>2</sub> parameters are less well known and are considered to be experimental.

The theoretical CCS values are produced to be compared to experimentally observed CCS values produced on a Waters Synapt G2-Si ion mobility spectrometer filled with both helium and nitrogen drift gases at different points throughout the study.



## Chapter 3.5: Results and Discussion

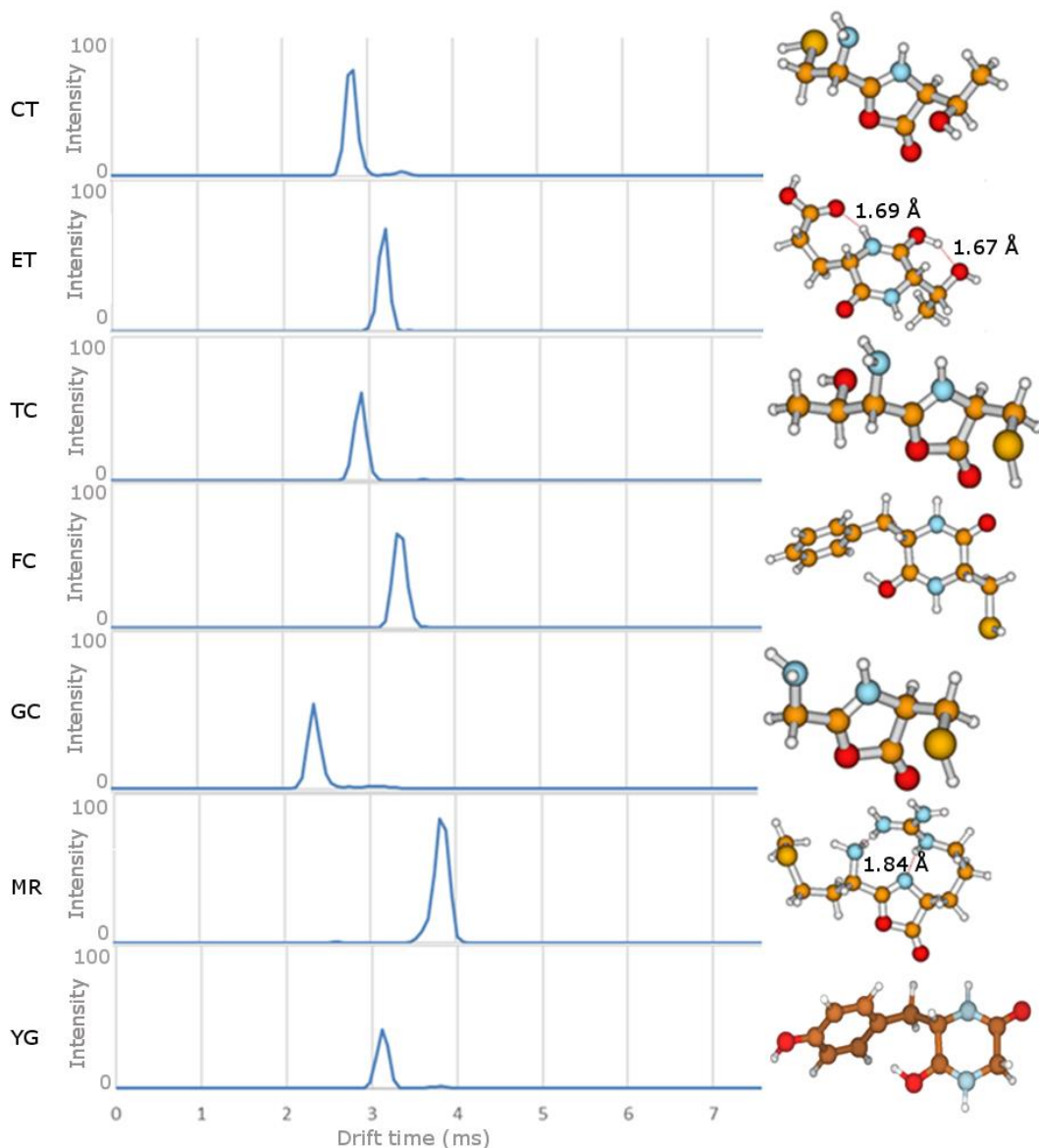
### 3.5.1: Preliminary investigation.

Seven peptides (CTVVVR, ETFANR, TCLIIR, FCLSLR, GCAQALR, MRFA and YGGFL) were chosen for comparison between the experimentally and computationally obtained collisional cross sections. Those peptides with an arginine (R) in the C-terminal position are similar to those that would be observed in a peptide produced via trypsin digestion in proteomics. MRFA, is a common calibrant for mass spectrometry. YGGFL has been used in prior studies into peptide fragment cyclisation. [110] These peptides were positively charged and fragmented in the Synapt trap cell prior to ion mobility spectrometry and the  $b_2$  ions were isolated post-analysis using Waters Mass Lynx software. The drift gas used was nitrogen, however the instrument was also calibrated to produce helium drift gas values. The helium values should be considered to be approximations as this method does not account for addition drift gas-ion interactions. The same  $b_2$  ion structures were computed and each of the geometries had a respective CCS calculated via the Mobcal programme. (Table 3 and Figures 22 & 23)

**Table 3. Results from initial trials of experimental and computation  $b_2$  ion CCS values.** Structures for each  $b_2$  ion were optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal produced the respective theoretical CCS values using the trajectory model with helium and nitrogen drift gases. The experimental values were obtained with a nitrogen drift gas on a Synapt G2-Si, helium values were approximately derived with a nitrogen drift gas calibrated for helium CCS values.

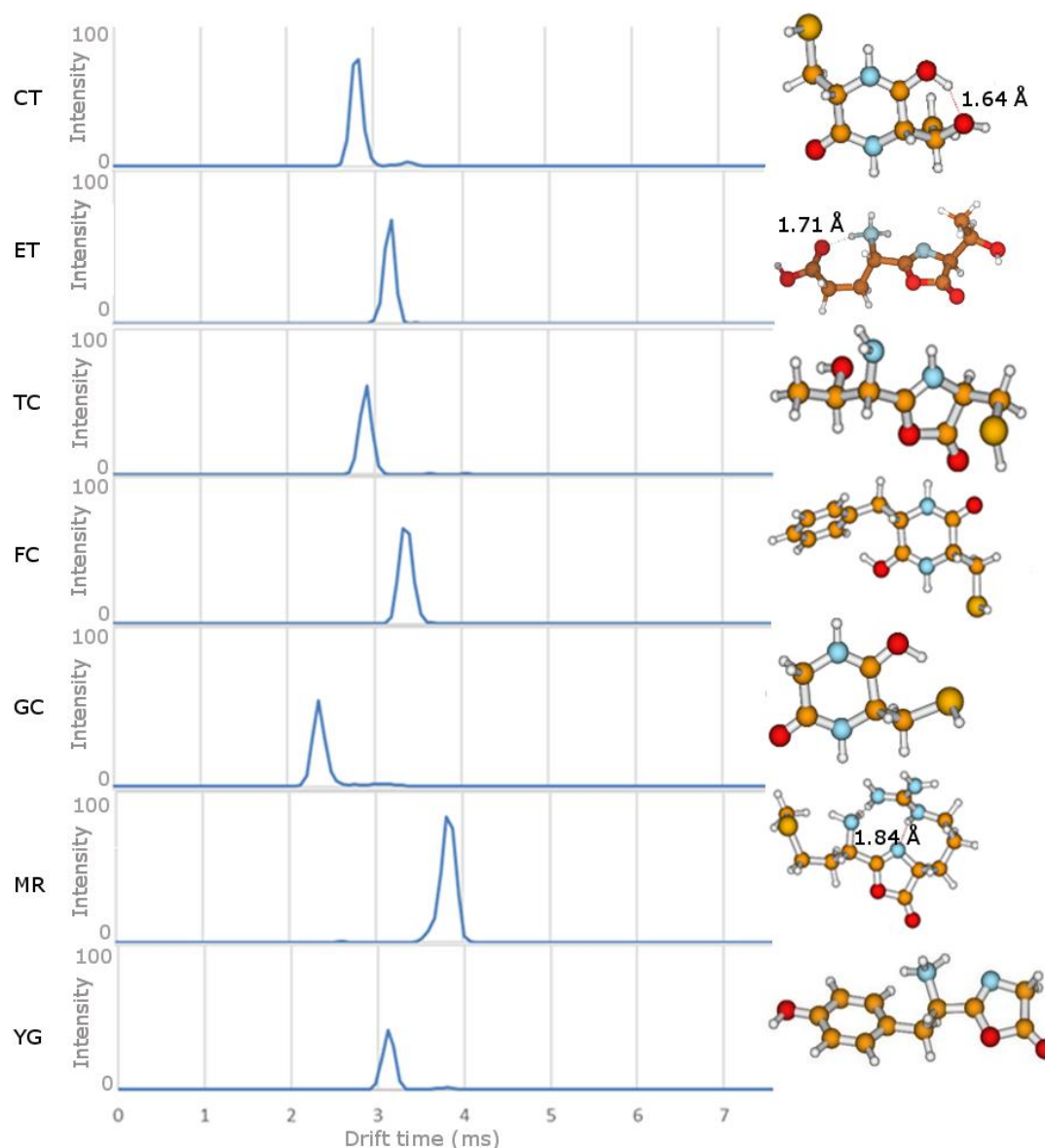
$b_2$ ion	Helium Drift Gas			Nitrogen Drift Gas		
	Experimental CCS ( $\text{\AA}^2$ )	Theoretical Oxazolone CCS ( $\text{\AA}^2$ )	Theoretical Macrocycle CCS ( $\text{\AA}^2$ )	Experimental CCS ( $\text{\AA}^2$ )	Theoretical Oxazolone CCS ( $\text{\AA}^2$ )	Theoretical Macrocycle CCS ( $\text{\AA}^2$ )
CT	76.2	78.6	75.9	139.6	142.7	143.4
ET	85.6	84.3	80.2	149.4	152.5	148.7
TC	78.7	78.9	75.8	142.3	142.8	140
FC	90.6	95.2	92.6	154.7	164.7	158.7
GC	64	66.7	63.9	127.4	130.6	130.8
MR	102.8	105.2	107.8	167.9	172.9	177.8
YG	85	84.6	81.8	148.7	155.7	151.1

The N<sub>2</sub> computed b<sub>2</sub> ions have a varied accuracy, with the TC oxazolone ion being only 0.5 Å<sup>2</sup> off the experimental value and the ET macrocycle ion being 0.7 Å<sup>2</sup> off. Other theoretical b<sub>2</sub> ions were generally about 3 Å<sup>2</sup> above the experimental values. The theoretical He values are generally more in-line with their respective experimental values, however they often indicate different structures than the N<sub>2</sub> values. The TC macrocyclic structure is within 0.3 Å<sup>2</sup> of the experimental value, and generally ions are within 2 Å<sup>2</sup> of the experimental values.



**Figure 22.** A plot of the helium drift time (x-axis) for seven b<sub>2</sub> ions, each accompanied by a computed structure most closely matching the experimentally obtained collisional cross section. Structures for each b<sub>2</sub> ion were optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal produced the respective theoretical CCS values using the trajectory model with a helium drift gas. The helium values were approximately derived with a nitrogen drift gas in a Synapt G2-Si calibrated for helium CCS values.

The results indicate a difference in structure for various samples. The CT  $b_2$  ion is suggested to take an oxazolone structure, however the  $N_2$  CCS difference between the minimum energy structures is only  $0.7 \text{ \AA}^2$  which may be due to error. The ET  $b_2$  ion is strongly suggested by the  $N_2$  data to take the macrocyclic structure, with a significant difference in the relevant CCSs. However, the He data disagrees and suggests that oxazolone is the dominant structure.



**Figure 23.** A plot of the nitrogen drift time (x-axis) for seven  $b_2$  ions, each accompanied by a computed structure most closely matching the experimentally obtained collisional cross section. Structures for each  $b_2$  ion were optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal produced the respective theoretical CCS values using the trajectory model with a nitrogen drift gases. The experimental ion mobilities were obtained with a nitrogen drift gas on a Synapt G2-Si.

From the N<sub>2</sub> data the TC b<sub>2</sub> ion is strongly suggested to take the form of an oxazolone, with a 0.5 Å<sup>2</sup> difference versus the 2.3 Å<sup>2</sup> difference of the macrocyclic structure. The He data disagrees on the structure, with the macrocycle having a difference of 0.3 as opposed to the oxazolone that has a 2.4 Å<sup>2</sup> difference.

From the N<sub>2</sub> value, the FC b<sub>2</sub> ion is strongly suggested to take a macrocyclic structure, as the theoretical CCS value of the macrocycle is 4 Å<sup>2</sup> above the experimental value, compared to the 10 Å<sup>2</sup> of the oxazolone structure. However, neither of these values are acceptably close to the experimental value. The He data does agree with the macrocyclic structure, with the macrocycle He CCS being 2 Å<sup>2</sup> above the experimental value, and the oxazolone being 4.6 Å<sup>2</sup> above the experimental value.

The GC b<sub>2</sub> ion is suggested by the N<sub>2</sub> data to take the form of an oxazolone, however the difference between this CCS and that of the macrocycle is only 0.2 Å<sup>2</sup>, which is too similar to prove that these values are not due to error. This is the structure that is expected. However, the He data disagrees, with the macrocycle being within 0.1 Å<sup>2</sup> of the experimental value, versus 2.7 Å<sup>2</sup> of the oxazolone structure.

The MR b<sub>2</sub> ion is strongly suggested by N<sub>2</sub> values to take the oxazolone structure, with the theoretical value being 5 Å<sup>2</sup> above the experimental CCS, and the theoretical macrocycle is 10.1 Å<sup>2</sup> above the experimental value. The He data agrees with an oxazolone CCS 2.4 Å<sup>2</sup> above the experimental value, opposed to the 5 Å<sup>2</sup> of a macrocyclic structure.

The YG b<sub>2</sub> ion is strongly suggested to take the macrocycle structure, being 4.6 Å<sup>2</sup> closer to the experimental value of than the macrocyclic structure, however previous works have shown this ion to take the structure of the oxazolone, with the parent ion YGGFL. [109] The He data, however, matches previous works, being 0.4 Å<sup>2</sup> off the experimental value with the oxazolone structure, and 3.2 Å<sup>2</sup> off with the macrocycle.

A peptide, ACSFIR, with a similar b<sub>2</sub> (175.05) mass and y<sub>1</sub> (175.11) mass was analysed to observe the difference between ions of very similar mass. The computed structures for the b<sub>2</sub> ion have a CCS of 134.7 Å<sup>2</sup> (oxazolone) and 133.9 Å<sup>2</sup> (macrocyclic), the experimental CCS value is 137.1 Å<sup>2</sup>. This suggests the AC b<sub>2</sub> ion takes the structure of an oxazolone. The R y<sub>1</sub> ion takes

the form of a protonated arginine, the theoretical CCS was computed as 132.6 Å<sup>2</sup>, whilst the experimental CCS obtained was 135.2 Å<sup>2</sup>. The comparison between these values shows that the two ions are clearly distinguishable in the drift time chromatogram. The values of the theoretical CCS values are both approximately 2.5 Å<sup>2</sup> below the respective experimental values, they still correspond to the correct peak positions.

Four peptides, with an FL b<sub>2</sub> ion and NK y<sub>2</sub> ion, were analysed as the b<sub>2</sub> ions and y<sub>2</sub> ions have similar masses. The singly charged peptides were fragmented, with an expectation of producing a majority of y<sub>2</sub> ions, and the doubly charged peptides were fragmented with the intention of producing more b<sub>2</sub> ions. The computed structures for the b<sub>2</sub> ion have a CCS of 168.6 Å<sup>2</sup> (oxazolone) and 167.4 Å<sup>2</sup> (macrocycle), the experimental CCS value produced from the doubly charged peptide is 159.2 Å<sup>2</sup>. These values are too far apart to assign the AC b<sub>2</sub> ion to either structure. The NK y<sub>2</sub> ion takes the form of a truncated NK, the theoretical CCS was computed as 158.7 Å<sup>2</sup>, whilst the experimental CCS obtained from the singly charged peptide was 156.3 Å<sup>2</sup>. All four FL\*NK peptides produce consistent b<sub>2</sub> ions and y<sub>2</sub> ions with identical drift times. The y<sub>2</sub> theoretical CCS is quite similar to the CCS obtained from the doubly charged fragmentation, and the b<sub>2</sub> theoretical CCS is significantly different from both peaks. Whilst it is possible that the mobcal software is not accurate enough, it also suggests that the two peaks may be both due to the y<sub>2</sub> ion, possibly taking different structures.

### *3.5.2: Large scale study of peptide b<sub>2</sub> ion structures.*

These initial results prompted theoretical studies and exploration of another 68 b<sub>2</sub> ions, 20 of which had been previously synthesised and another 48 which were synthesised for the purpose of this study. All of these peptides had the sequence X<sup>1</sup>X<sup>2</sup>AAR, where X<sup>1</sup> and X<sup>2</sup> amino acids were substituted for V, A, G, F, D, Q, P, S, T, H and L. Of these 68 b<sub>2</sub> ions, all have been computed. This format was chosen as penta-peptides are readily observed in proteomics, they are able to reliably hold multiple charges (2+), arginine is commonly observed as a terminal amino acid in proteomics due to its interaction with trypsin. Alanine has a small residue and does not interfere with the formation of b<sub>2</sub> ions. The specific amino acids for the X positions contain only hetero atoms and generally form classical oxazolone and diketopiperazine ring structure b<sub>2</sub> ions.

**Table 4. Theoretical collisional cross sections of all minimum energy computed b<sub>2</sub> ions, including helium and nitrogen drift gas CCS's of both macrocyclic and oxazolone structures.** Structures for each b<sub>2</sub> ion were optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal produced the respective theoretical CCS values using the trajectory model with helium and nitrogen drift gases.

ION	MACROCYCLIC CCS HE	MACROCYCLIC CCS N2	OXAZOLONE CCS HE	OXAZOLONE CCS N2
VA	70.5	136.3	73.8	134.3
VF	93.6	159.9	97.1	162.7
VG	65.7	132.7	68.3	131.6
VH	90.0	159.9	89.6	157.4
VL	85.9	152.6	90.0	150.6
VP	77.0	141.0	78.8	139.9
VQ	85.6	151.2	90.2	150.8
VS	72.5	138.6	75.8	138.3
VT	76.0	141.3	81.1	143.0
VY	95.9	165.4	97.0	166.1
DA	69.0	136.3	71.7	140.4
PA	67.6	132.9	70.4	136.5
QA	77.4	143.6	80.9	145.5
SA	63.5	129.3	66.3	128.0
CT	75.9	143.4	78.6	142.7
ET	80.2	148.7	84.3	152.5
TC	75.8	140.0	78.9	142.8
FC	92.6	158.7	95.2	164.7
GC	63.9	130.8	66.7	130.6
MR	107.8	177.8	105.2	172.9
YG	81.8	151.1	84.6	155.7
LA	77.7	145.9	80.6	141.7
FA	84.6	149.4	85.7	151.7
FL	100.7	167.4	101.9	168.6
AC	69.1	133.9	71.7	134.7

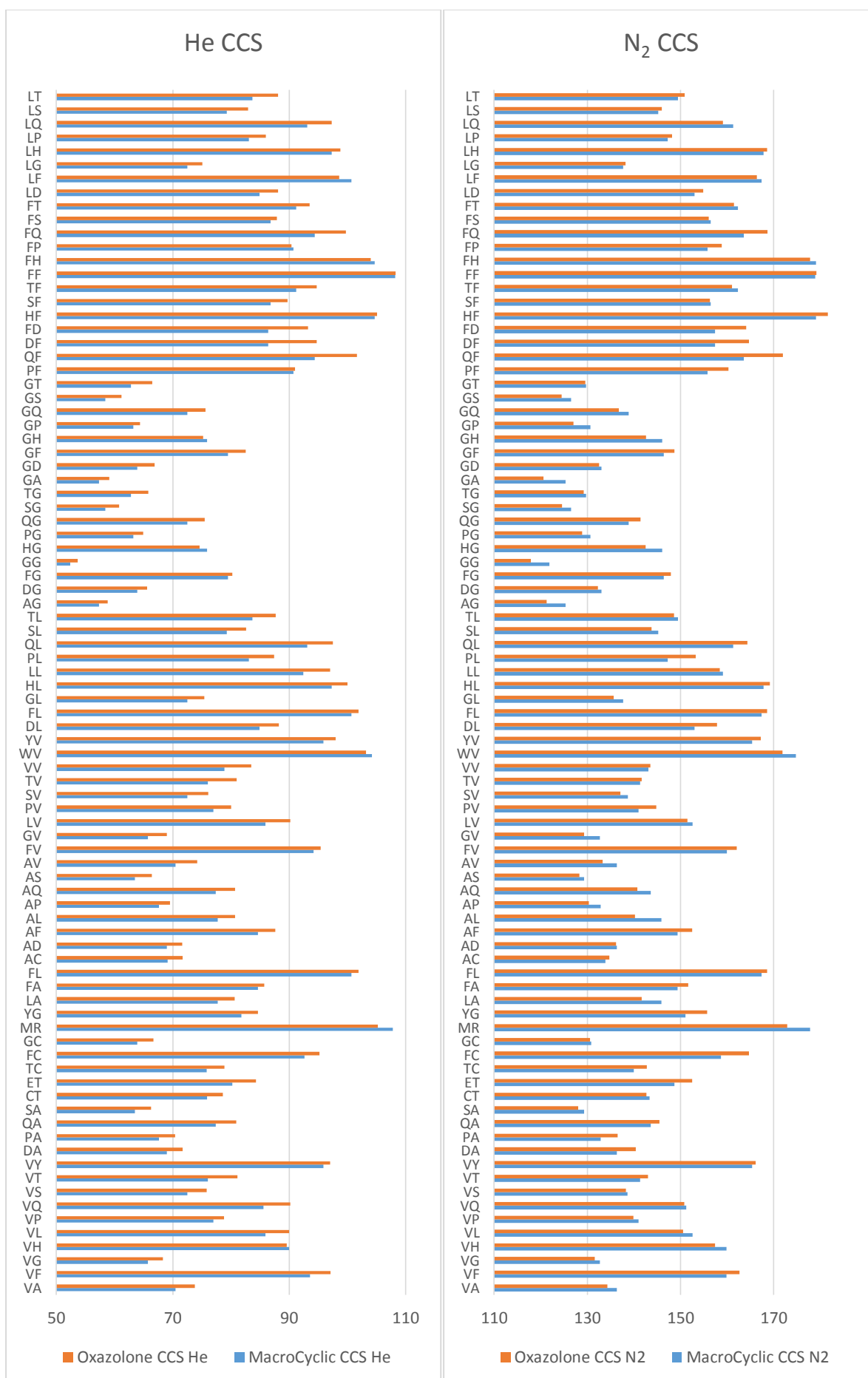
AD	69.0	136.3	71.6	136.2
AF	84.6	149.4	87.6	152.5
AL	77.7	145.9	80.7	140.2
AP	67.6	132.9	69.5	130.3
AQ	77.4	143.6	80.7	140.7
AS	63.5	129.3	66.4	128.3
AV	70.5	136.3	74.2	133.3
FV	94.2	160.0	95.4	162.1
GV	65.7	132.7	69.0	129.3
LV	85.9	152.6	90.2	151.5
PV	77.0	141.0	80.0	144.8
SV	72.5	138.7	76.1	137.1
TV	76.0	141.3	81.0	141.7
VV	78.9	143.1	83.5	143.5
WV	104.2	174.8	103.2	171.9
YV	95.9	165.4	98.0	167.2
DL	84.9	153.0	88.2	157.8
FL	100.7	167.4	101.9	168.6
GL	72.5	137.7	75.4	135.7
HL	97.3	167.8	100.0	169.2
LL	92.4	159.1	97.0	158.4
PL	83.1	147.3	87.4	153.3
QL	93.1	161.3	97.5	164.4
SL	79.3	145.2	82.6	143.8
TL	83.7	149.5	87.7	148.6
AG	57.3	125.3	58.8	121.3
DG	63.9	133.0	65.6	132.3
FG	79.5	146.4	80.2	147.9
GG	52.4	121.9	53.7	117.9
HG	75.9	146.1	74.6	142.5

PG	63.2	130.7	64.9	128.9
QG	72.5	138.9	75.5	141.4
SG	58.4	126.5	60.8	124.6
TG	62.8	129.7	65.8	129.2
GA	57.3	125.3	59.1	120.6
GD	63.9	133.0	66.9	132.5
GF	79.5	146.4	82.5	148.7
GH	75.9	146.1	75.2	142.6
GP	63.2	130.7	64.4	127.0
GQ	72.5	138.9	75.6	136.8
GS	58.4	126.5	61.2	124.5
GT	62.8	129.7	66.5	129.6
PF	90.7	155.8	91.0	160.3
QF	94.4	163.6	101.6	172.0
DF	86.4	157.4	94.7	164.7
FD	86.4	157.4	93.2	164.1
HF	104.7	179.1	105.1	181.6
SF	86.8	156.5	89.7	156.3
TF	91.2	162.3	94.7	161.1
FF	108.2	178.9	108.3	179.2
FH	104.7	179.1	104.0	177.8
FP	90.7	155.8	90.4	158.9
FQ	94.4	163.6	99.7	168.7
FS	86.8	156.5	87.9	156.1
FT	91.2	162.3	93.5	161.5
LD	84.9	153.0	88.1	154.9
LF	100.7	167.4	98.6	166.4
LG	72.5	137.7	75.1	138.2
LH	97.3	167.8	98.8	168.6
LP	83.1	147.3	86.0	148.2



LQ	93.1	161.3	97.3	159.1
LS	79.3	145.2	82.9	146.0
LT	83.7	149.5	88.1	150.9

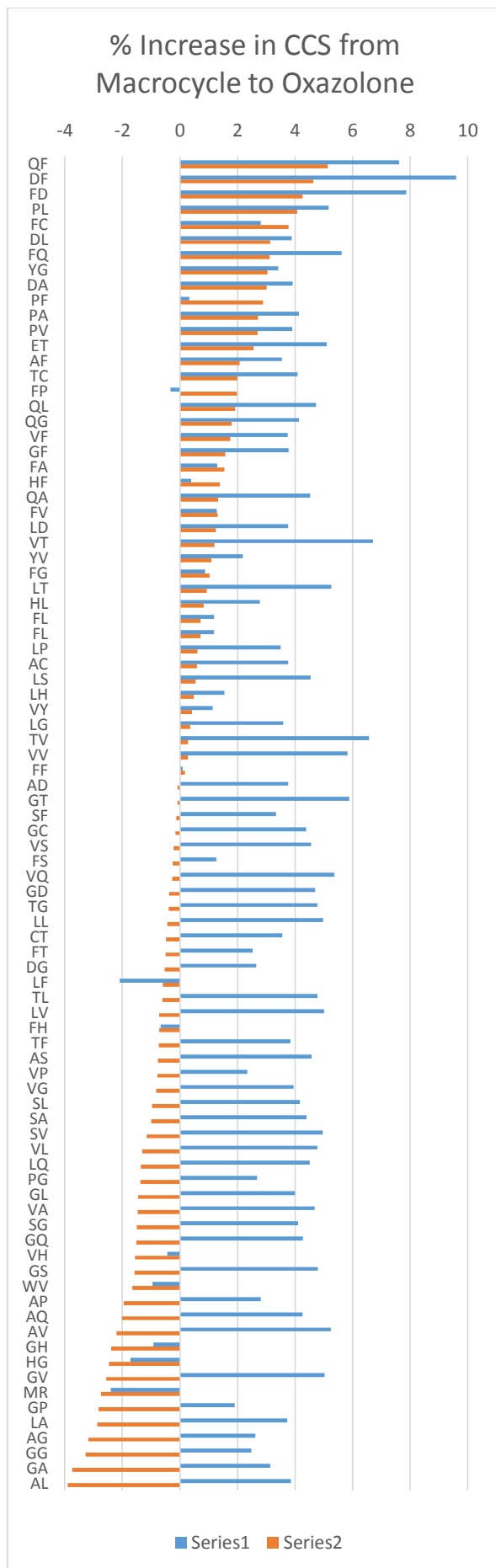
The oxazolone and macrocyclic computed structures shown in Table 4 are generally within 3% of one another, which is often used as the error in which computed CCS' are expected to fall. If the computed CCS for many structures falls within the 3% error for both the oxazolone and macrocyclic structures and are statistically imprecise, then the computational method may be proven not be accurate enough to predict the structure of  $b_2$  ions. These samples have been analysed on the Water Synapt G2 Si Instrument available at Bangor University. The experimental data is compared to the theoretical data to provide an insight into the structural tendencies of  $b_2$  ions and the accuracy of our computational methods.



*Figure 24. A comparison of computed macrocyclic and oxazolone CCS, with both helium and nitrogen as the drift gas. The CCS values were computed using the trajectory method of the Mobcal programme. It is readily apparent from these values that the relative CCS values between the oxazolone and helium structures is not consistent between nitrogen and helium drift gas calculations for individual ions.*

Graphically displaying the theoretical collisional cross sections of all computed ions shows that there is a significant difference in the results when switching between a helium and nitrogen drift gas. It is expected that the collisional cross section scales proportionally to the size of a drift gas molecule/atom, and that the collisional cross section for two structures should remain relatively proportional to one another when a large or heavier drift gas is used. In figure 24 it is clear that the relative CCS values between the two structures changes with different drift gases. To clarify this phenomenon, figure 25 shows that whilst the helium calculations show that the oxazolone structures are most often between 3 to 6 % larger than the macrocyclic structure. The nitrogen calculations almost seem to have a random difference between the two structures, with about half of the oxazolone structures being smaller than the macrocycles, and the other half being larger.

The conclusion of this result is that the helium and nitrogen calculations are not of equivalent accuracy. Due to the consistency of the helium calculation with the relative size of the oxazolone versus the macrocycles, it would appear that the helium programme is the more accurate method of computing CCS. In addition, literature has previously shown the helium calculations to be sufficiently competent, therefore suggesting that the less well developed nitrogen mobcal programme is the less accurate. This is a matter of concern as helium is the lowest resolution of the common drift gases, with nitrogen being the most commonly used with a good resolution and CO<sub>2</sub> being higher resolution but less commonly used. The conclusion is that the nitrogen drift gas calculations are not yet ready for deployment and need further development.



**Figure 25. A representation of the difference between macrocyclic and oxazolone CCS for nitrogen (Series 2) versus helium (Series 1) drift gases.** The CCS values were computed using the trajectory method of the Mobcal programme. From this ordered data, it is visible that using He drift gas in these computations produce oxazolone values that are slightly greater than macrocyclic structures, whilst N<sub>2</sub> drift gas produces a random array of differences CCS value differences between structures with no observable consistency.

All synthesised peptides have been stored in a -20°C freezer until they could be analysed on the Water Synapt G2 Si Instrument. The IMS cell was filled with helium drift gas and was calibrated using Waters Major Mix solution. The experimental data is compared to the theoretical data to provide an insight into the structural tendencies of b<sub>2</sub> ions and the accuracy of our computational methods.

**Table 5. Experimental and theoretical CCS values with a helium drift gas for 58 b<sub>2</sub> ions.** Structures for each b<sub>2</sub> ion were optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal produced the respective theoretical CCS values using the trajectory model with a helium drift gas. The experimental values were obtained with a helium drift gas on a Synapt G2-Si.

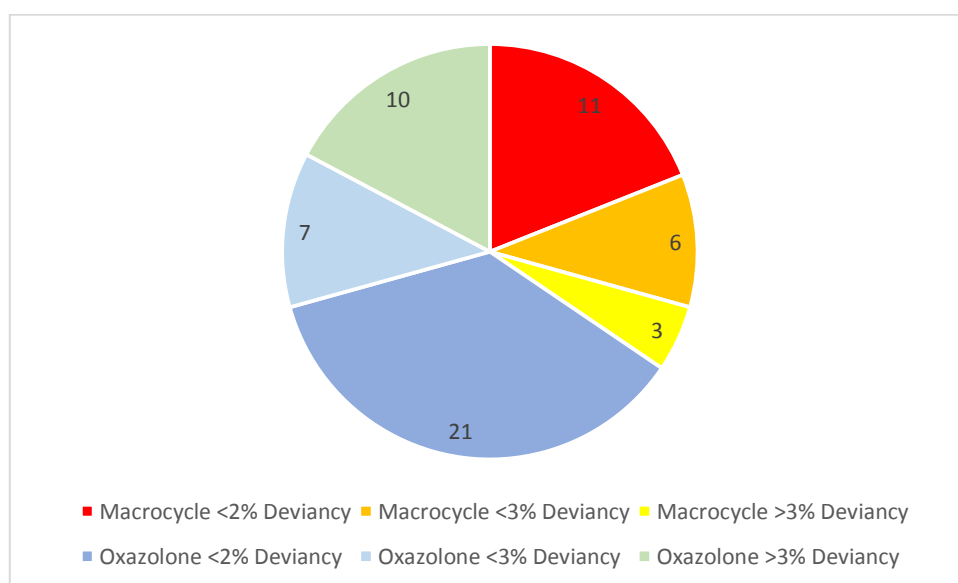
b <sub>2</sub> ion	[M+H] <sup>+</sup>	[M+2H] <sup>+</sup>	M/Z b <sub>2</sub> ion	Experimental CCS	Theoretical Macrocyclic CCS	Theoretical Oxazolone CCS
VG	544.3202	272.6637	157.09716	68.01	65.7	68.3
VA	558.3359	279.6716	171.11281	71.93	70.5	73.8
VP	584.3515	292.6794	197.12846	80.74	77	78.8
VS	574.3308	287.669	187.10772	80.05	72.5	75.8
VT	588.3464	294.6768	201.12337	81.96	76	81.1
VL	600.3828	300.695	213.15976	87.93	85.9	90
VQ	615.3573	308.1823	228.13427	88.8	85.6	90.2
VH	624.3576	312.6825	237.1346	90.9	90	89.6
VF	634.3672	317.6872	247.14411	95.2	93.6	97.1
VY	650.3621	325.6847	263.13902	97.96	95.9	97
GV	544.3202	272.6637	157.09716	65.8	65.7	69
AV	558.3359	279.6716	171.11281	72.15	70.5	74.2
VV	586.3672	293.6872	199.14411	80.4	78.9	83.5
LV	600.3828	300.695	213.15976	88.17	85.9	90.2
SV	574.3308	287.669	187.10772	80.93	72.5	76.1
TV	588.3464	294.6768	201.12337	81.56	76	81
PV	584.3515	292.6794	197.12846	81.64	77	80
FV	634.3672	317.6872	247.14411	94.92	94.2	95.4
YV	650.3621	325.6847	263.13902	99.21	95.9	98
WV	673.3781	337.1927	286.15501	104.33	104.2	103.2
SA	537.2668	269.137	159.07642	68.9	63.5	66.3
LS	517.3093	259.1583	201.12337	82.56	79.3	82.9
LP	527.33	264.1687	211.14411	86.05	83.1	86

LF	577.3457	289.1765	261.15976	98.87	100.7	98.6
LG	487.2987	244.153	171.11281	76.98	72.5	75.1
LQ	558.3358	279.6716	242.14992	93.58	93.1	97.3
LD	545.3042	273.1557	229.11829	88.85	84.9	88.1
LT	531.3249	266.1661	215.13902	85.78	83.7	88.1
LH	567.3362	284.1717	251.15025	96.28	97.3	98.8
SG	461.2467	231.127	145.06077	56.19	58.4	60.8
PG	471.2674	236.1374	155.08151	64.61	63.2	64.9
FG	521.2831	261.1452	205.09716	82.84	79.5	80.2
GG	431.2361	216.1217	115.05021	54.43	52.4	53.7
QG	502.2732	251.6403	186.08732	82.93	72.5	75.5
DG	489.2416	245.1244	173.05569	71.59	63.9	65.6
TG	475.2623	238.1348	159.07642	65.11	62.8	65.8
HG	511.2736	256.1404	195.08765	80.86	75.9	74.6
AG	445.2518	223.1295	129.06586	53.76	57.3	58.8
GS	461.2467	231.127	145.06077	58.79	58.4	61.2
GP	471.2674	236.1374	155.08151	62.47	63.2	64.4
GF	521.2831	261.1452	205.09716	81.52	79.5	82.5
GQ	502.2732	251.6403	186.08732	82.66	72.5	75.6
GD	489.2416	245.1244	173.05569	71.1	63.9	66.9
GT	475.2623	238.1348	159.07642	63.55	62.8	66.5
GH	511.2736	256.1404	195.08765	77.21	75.9	75.2
PF	561.3144	281.1608	245.12846	92.45	90.7	91
QF	592.3202	296.6637	276.13427	101.22	94.4	101.6
FF	611.33	306.1687	295.14411	105.64	108.2	108.3
DF	579.2886	290.1479	263.10264	97.26	86.4	94.7
TF	565.3093	283.1583	249.12337	94.6	91.2	94.7
HF	601.3205	301.1639	285.1346	101.72	104.7	105.1
SF	551.2936	276.1505	235.10772	90.61	86.8	89.7
FP	561.3144	281.1608	245.12846	92.52	90.7	90.4
FQ	592.3202	296.6637	276.13427	101.3	94.4	99.7
FT	565.3093	283.1583	249.12337	94.89	91.2	93.5
FD	579.2886	290.1479	263.10264	100.57	86.4	93.2
FH	601.3205	301.1639	285.1346	101.88	104.7	104
FS	551.2936	276.1505	235.10772	92.1	86.8	87.9

All peptides were injected directly into the ESI source of the Synapt G2-Si. Each doubly charged ion was isolated in the quadrupole before being fragmented in the trap cell prior to entering the ion mobility chamber filled with helium gas. Of the product ions detected, the specific  $b_2$  ions were selected in the driftscope software to determine their ion mobility and

a CCS value via a calibration. The experimental values are displayed table 5, alongside the theoretical values computed for both the oxazolone and macrocyclic structure.

This data was compared to understand which structure each  $b_2$  ion had taken in the mass spectrometer. Matching the closest theoretical value to the ions experimental value, the likely structure of the ion was assigned. 38 of the  $b_2$  ions take an oxazolone structure, whilst 20 of the ions take a macrocyclic structure. The deviance from the experimental value varied between ions, and each structure has been separated into 3 levels of confidence: Very High Confidence (<2% deviance), Good Confidence (<3% deviance) and Poor Confidence (>3% deviance). These values are displayed in figure 26. Of those that meet the criteria of being at least of good confidence, 28 are oxazolone and 17 are macrocyclic.



**Figure 26.** The deviation of oxazolone and macrocycles  $b_2$  ion computed CCS values from experimentally obtained values. The experimental values were obtained via IMS on a Synapt G2-Si with a helium drift gas. Computed CCS values were calculated using the trajectory model of Mobcal with a helium drift gas.

Whilst taking the closest value gives us a “best match”, there is the problem that the acceptable error between an experimental and theoretical value is 3% and it is possible for both the oxazolone and macrocycle to fall within this range. If both values are within this range, it is not possible to assume that the best matching structure is the structure taken in reality. The percentage difference between the macrocyclic or oxazolone theoretical values

versus the experimental value is tabulated in table 6. Here the table is split between those that more closely match the oxazolone and macrocycle for clarity.

Of those 28 values oxazolone theoretical CCS that fall within a 3% error, 18 experimental values clearly match oxazolone, whilst 10 fall within 3% for both structures. Of the 17 macrocyclic theoretical CCS that fall within a 3% error, only 6 experimental values clearly imply a macrocyclic structure and the other 11 fall within 3% error for both structures. This shows 64% of those structures that best match oxazolone cannot be macrocyclic, however only 35% of those structures that best match macrocycles cannot be oxazolone. The implication of this is that the majority of those ions assigned “Oxazolone” have a high confidence and are unlikely to take another structure, whilst a majority of those assigned “Macrocyclic” have a low confidence and for many of these there is no certainty that they take the macrocyclic structure.

**Table 6. The percentage difference between the calculated CCS values and the experimentally obtained values for each  $b_2$  ion. The experimental values were obtained via IMS on a Synapt G2-Si with a helium drift gas. Computed CCS values were calculated using the trajectory model of Mobcal with a helium drift gas.**

$b_2$ ion	% $\Delta$ Macrocyclic	% $\Delta$ Oxazolone	Closest match	$b_2$ ion	% $\Delta$ Macrocyclic	% $\Delta$ Oxazolone	Closest match
VG	3.5	0.4	Oxazolone	VA	2.0	2.5	Macrocyclic
VP	4.9	2.5	Oxazolone	VH	1.0	1.5	Macrocyclic
VS	10.4	5.6	Oxazolone	VF	1.7	2.0	Macrocyclic
VT	7.8	1.1	Oxazolone	GV	0.2	4.6	Macrocyclic
VL	2.4	2.3	Oxazolone	AV	2.3	2.8	Macrocyclic
VQ	3.7	1.6	Oxazolone	VV	1.9	3.7	Macrocyclic
VY	2.1	1.0	Oxazolone	WV	0.1	1.1	Macrocyclic
LV	2.6	2.3	Oxazolone	LQ	0.5	3.8	Macrocyclic
SV	11.6	6.3	Oxazolone	LT	2.5	2.6	Macrocyclic
TV	7.3	0.7	Oxazolone	LH	1.0	2.6	Macrocyclic
PV	6.0	2.1	Oxazolone	SG	3.8	7.6	Macrocyclic
FV	0.8	0.5	Oxazolone	HG	6.5	8.4	Macrocyclic
YV	3.5	1.2	Oxazolone	AG	6.2	8.6	Macrocyclic
SA	8.5	3.9	Oxazolone	GS	0.7	3.9	Macrocyclic
LS	4.1	0.4	Oxazolone	GP	1.2	3.0	Macrocyclic
LP	3.5	0.1	Oxazolone	GT	1.2	4.4	Macrocyclic
LF	1.8	0.3	Oxazolone	GH	1.7	2.7	Macrocyclic
LG	6.2	2.5	Oxazolone	FF	2.4	2.5	Macrocyclic

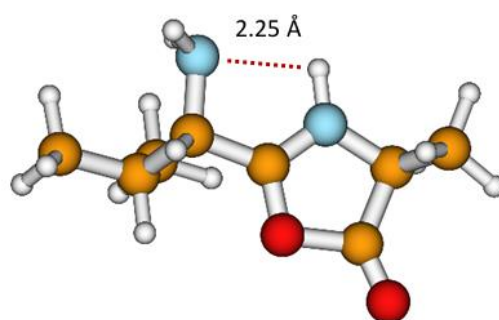


LD	4.7	0.9	Oxazolone	HF	2.8	3.2	Macrocycle
PG	2.2	0.4	Oxazolone	FP	2.0	2.3	Macrocycle
FG	4.2	3.3	Oxazolone				
GG	3.9	1.4	Oxazolone				
QG	14.4	9.8	Oxazolone				
DG	12.0	9.1	Oxazolone				
TG	3.7	1.0	Oxazolone				
GF	2.5	1.2	Oxazolone				
GQ	14.0	9.3	Oxazolone				
GD	11.3	6.3	Oxazolone				
PF	1.9	1.6	Oxazolone				
QF	7.2	0.4	Oxazolone				
DF	12.6	2.7	Oxazolone				
TF	3.7	0.1	Oxazolone				
SF	4.4	1.0	Oxazolone				
FQ	7.3	1.6	Oxazolone				
FT	4.0	1.5	Oxazolone				
FD	16.4	7.9	Oxazolone				
FH	2.7	2.0	Oxazolone				
FS	6.1	4.8	Oxazolone				

### 3.5.3: $b_2$ Ion Computed Structures.

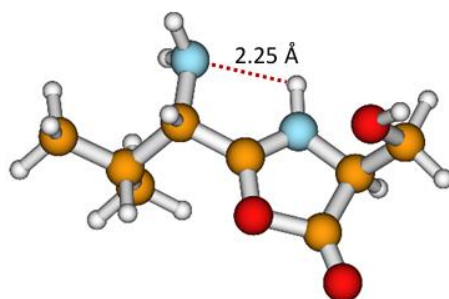
Whilst 136 structures were computed, all of which are displayed in appendix 1, there are a number of shown structural features shown within the minimum energy structures. These features are unique to either oxazolone derivatives or macrocyclic structures. Among the oxazolone structures, the ionising proton most commonly sits on the oxazolone ring nitrogen.

Oxazolone derivative structures:



**Structure 1. The minimum energy structure of a valine-alanine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $73.8 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VA  $b_2$  ion was  $71.93 \text{ \AA}^2$ .

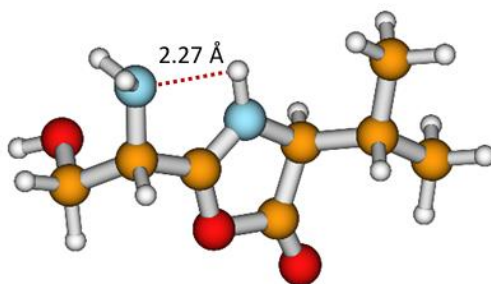
In structure 1, the VA  $b_2$  ion shows the most common feature seen within almost all oxazolone structures, this is the hydrogen bond between the N-H of the oxazolone ring and the lone pair of the neighbouring N-Terminus nitrogen. The prevalence of this trait among oxazolone structures suggests that this is a highly stabilizing interaction.  $b_2$  ions without side chains containing heteroatoms consistently take this configuration.



**Structure 2. The minimum energy structure of a valine-serine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a

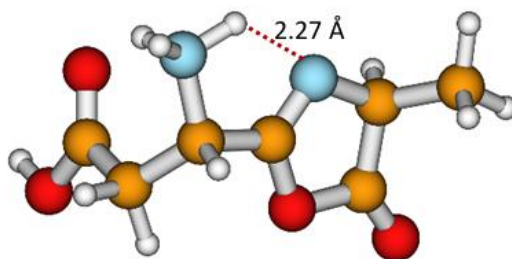
CCS value of  $75.8 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VS  $b_2$  ion was  $80.74 \text{ \AA}^2$ .

In structure 2, the VS  $b_2$  ion has a hydroxyl group on the second side chain. The hydroxyl group in these structures takes a position to the side of the ring N-H group, with its own hydrogen pointing away from the ring. This is counter to what might be expected, in which the hydroxyl would interact with the ring carbonyl oxygen to form a hydrogen bond, however the interaction with the ionising proton produces a lower energy structure.



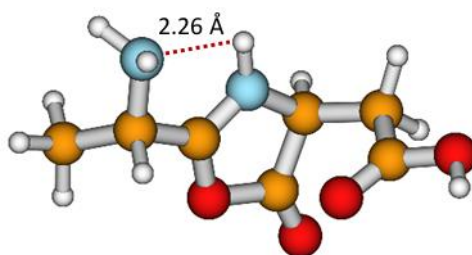
**Structure 3.** *The minimum energy structure of a serine-valine oxazolone  $b_2$  ion.* Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $76.1 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a SV  $b_2$  ion was  $80.93 \text{ \AA}^2$ .

The SV  $b_2$  ion shown in structure 3 has a hydroxyl group on the first side chain. In this case the interaction between the N-terminus nitrogen and its own hydrogens has been weakened due to the strong interaction with the ionising proton. One of these hydrogens are stabilised with a favourable interaction with the oxygen of the hydroxyl group, and this is shown to be more energetically favourable than the hydroxyl group forming a hydrogen bond with the ring oxygen. This N-H---N-H---O interaction is prevalent throughout the oxazolone structures where it is possible to form.



**Structure 4. The minimum energy structure of an aspartic acid-alanine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $71.7 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. An experimental CCS value for the DA  $b_2$  ion was not obtained.

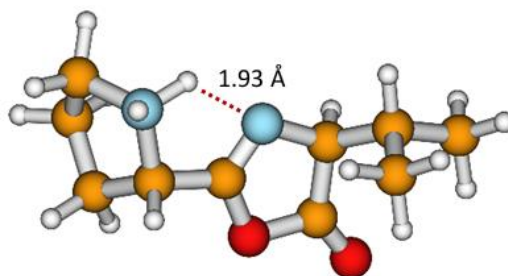
DA  $b_2$  ions, shown in structure 4, has a carboxylic acid group on the first side chain. In the case of DA and DL, the ionising proton moves across to the N-terminus nitrogen, still interacting favourably with the ring nitrogen. The other hydrogens on the N-terminus nitrogen interact with the carbonyl oxygen of the carboxylic acid, whilst the hydroxyl hydrogen of the acid interacts with the other oxygen of the acid. This interaction is seen throughout the carboxylic acid side chains of  $b_2$  ions, these interactions are consistently more energetically favourable than the carboxylic acid forming hydrogen bonds with ring oxygens.



**Structure 5. The minimum energy structure of an alanine-aspartic acid oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $71.6 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. An experimental CCS value for the AD  $b_2$  ion was not obtained.

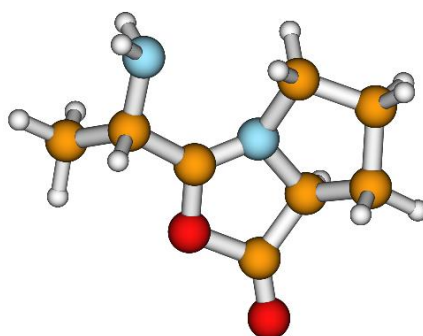
The AD  $b_2$  ion, shown in structure 5, has a more unexpected conformation in that the carbonyl oxygen of the carboxylic acid when situated on the second side chain consistently takes a position to the side of the ionising proton. It may be expected that the hydrogen of the acid would form a hydrogen bond with the ring carbonyl oxygen, or perhaps take a position closer to ionising proton. This structure has been investigated to see if a more favourable position can be found, however the structure displayed is the most favourable found. Other

protonation positions have also been probed and not found to be more favourable than this structure. Further work could be done to discover why this position specifically is so energetically favourable.



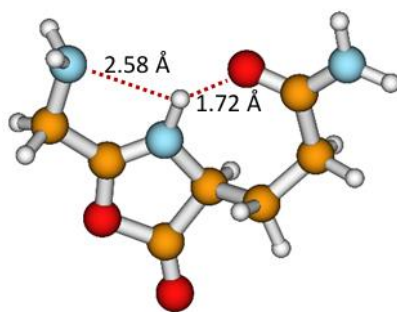
**Structure 6. The minimum energy structure of a proline-valine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $80.0 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a PV  $b_2$  ion was  $81.64 \text{ \AA}^2$ .

In structure 6, we see that the proline N-terminus nitrogen is a more energetically favourable position to which the ionising proton can be bound than the oxazolone nitrogen. This feature is seen in most structures with a proline N-terminus, with an exception of PG.



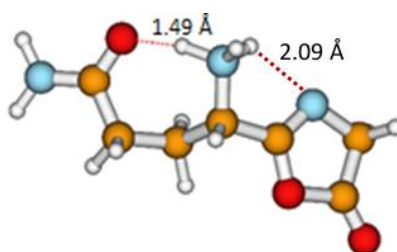
**Structure 7. The minimum energy structure of a valine-proline oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $78.8 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VP  $b_2$  ion was  $80.74 \text{ \AA}^2$ .

When proline is the second amino acid to make up a  $b_2$  ion, as seen in structure 7 for VP, the ionising proton is no longer present as it is not required to maintain a positive charge. This is due to the nitrogen in the double ring having an additional bond to the proline ring, causing a dissociated positive charge along the double bond between the nitrogen and the neighbouring carbon that is bound to three atoms.



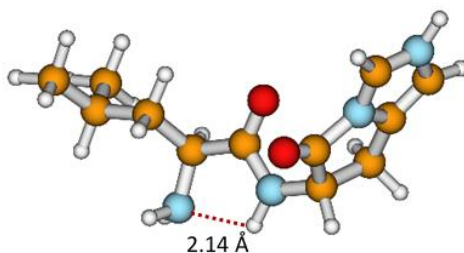
**Structure 8. The minimum energy structure of a glycine-glutamine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $75.6 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GQ  $b_2$  ion was  $82.66 \text{ \AA}^2$ .

In structure 8, a GQ  $b_2$  ion, we see that the oxygen of the amide group on the second side chain interacts with the ionising proton. In this case it is observed that the ionising proton sits between the oxygen of the amide and the N-terminus nitrogen. This structural features is observed across structures that contain a glutamine as the second amino acid.



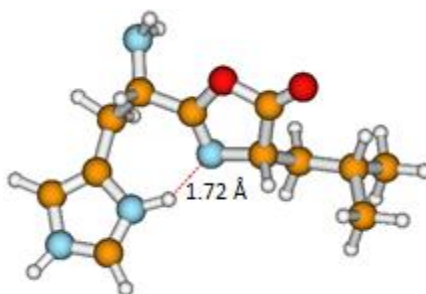
**Structure 9. The minimum energy structure of a glutamine-glycine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $75.5 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a QG  $b_2$  ion was  $82.93 \text{ \AA}^2$ .

With a glutamine as the N-terminal amino acid, as seen in structure 9, we see a feature consistent with having a carboxylic acid in the same position as the amide, as seen in structure 4. With the ionising proton moving to the N-terminus nitrogen (a feature prevalent across QX  $b_2$  ions), we see a strong hydrogen bond between oxygen of the amide and loosened hydrogens on the N-terminus nitrogen. There continues to be a favourable interaction between the N-terminus nitrogen and the ring nitrogen, but it is less pronounced than in other structures.



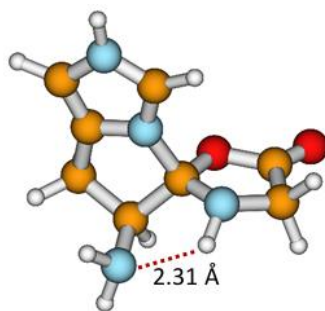
**Structure 10. The minimum energy structure of a leucine-histidine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $98.8 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LH  $b_2$  ion was  $96.28 \text{ \AA}^2$ .

In the case of structure 10, a LH  $b_2$  ion, we see the histidine effect in action. In this case the histidine effect produces a double ring structure that is more energetically favourable than the oxazolone structures. The ionising proton is lost in the formation of this structure and the ion carries a positive charge across a double bond between the double ring nitrogen and the neighbouring C-H group.



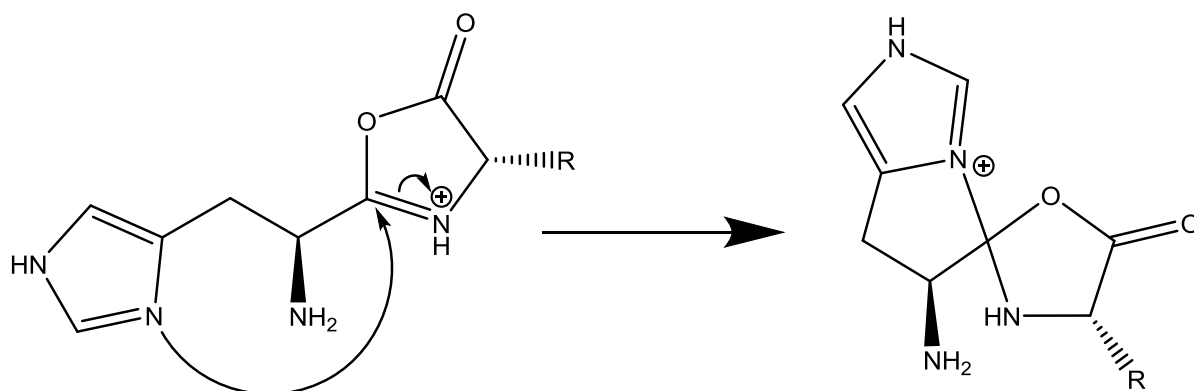
**Structure 11. The minimum energy structure of a histidine-leucine oxazolone  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $75.5 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. An experimental CCS value for the HL  $b_2$  ion was not obtained.

It is possible, when a histidine is in the N-terminal position, for the histidine to hold the ionising proton, in this case we see a hydrogen bond between the ionising proton on the histidine and the oxazolone ring nitrogen. This is one of the few cases in which the N-terminus nitrogen is not interacting with the oxazolone ring nitrogen, instead it forms a favourable interaction between its hydrogens and oxygen in the oxazolone ring. This is shown in structure 11.



**Structure 12.** *The minimum energy structure of a histidine-glycine oxazolone  $b_2$  ion.* Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $74.6 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a HG  $b_2$  ion was  $80.86 \text{ \AA}^2$ .

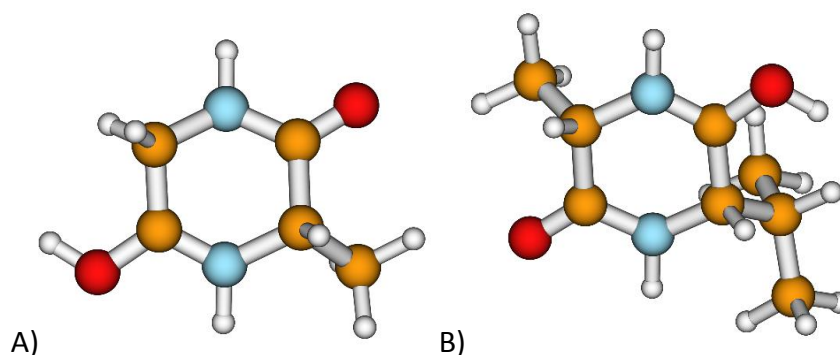
Alternatively, the nitrogen of the oxazolone ring may retain the ionising proton and continue to interact with the N-terminus nitrogen. In this case, we see the double bond of the oxazolone ring be broken by the histidine nitrogen, forming a bond between the carbon that previously made up the double bond, and relocating the positive charge to the nitrogen of the histidine, illustrated in scheme 4. This creates a new triple ring structure.



*Scheme 4. A visual description of how a histidine in the N-terminal position can form a triple ring structure.*

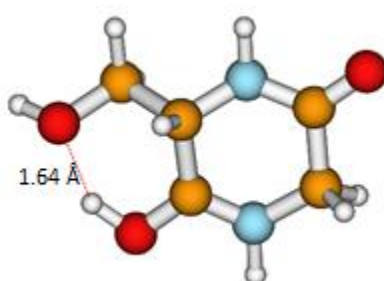


Macrocyclic/diketopiperazine derivative structures:



**Structure 13.** *The minimum energy structures of A) glycine-alanine and B) alanine-valine macrocyclic  $b_2$  ions.* Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used these geometries to generate a CCS value of 57.3 Å<sup>2</sup> for GA and 70.5 Å<sup>2</sup> for AV using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a AV  $b_2$  ion was 72.15 Å<sup>2</sup>. An experimental CCS value was not obtained for a GA  $b_2$  ion.

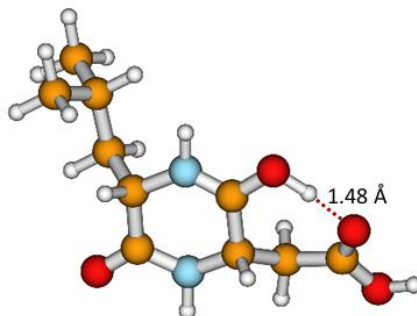
Macrocyclic  $b_2$  ion structures are degenerate to their respective amino acids having switched position, therefore a cyclic GA  $b_2$  ion would be identical to a cyclic AG  $b_2$  ion. The ionising proton most commonly sits on what of the two carbonyl oxygens in the ring, and this is always the case when no heteroatoms are present in side chains, as shown in structures 13 A and B. The proton always points away from the nearest N-H group, as the neighbouring hydrogen has a favourable interaction with the ionised oxygen. These structures are following similar trends to those reported in literature for tri-alanine fragments. [<sup>110</sup>]



**Structure 14.** *The minimum energy structure of a glycine-serine macrocyclic  $b_2$  ion.* Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 58.4 Å<sup>2</sup> for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GS  $b_2$  ion was 58.79 Å<sup>2</sup>.

As seen in structure 14, a GS cyclic  $b_2$  ion, when a hydroxyl group is present in a side chain, the ionising proton is able to form a hydrogen bond from the nearby carbonyl oxygen. This is

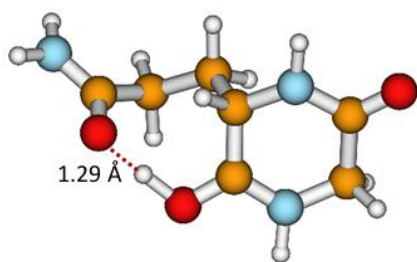
feature is prevalent in all cyclic  $b_2$  ions that contained a serine side chain, although there were few examples in which competition with other side chains containing heteroatoms was available.



**Structure 15. The minimum energy structure of a leucine-aspartic acid macrocyclic  $b_2$  ion.**

Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $84.9 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LD  $b_2$  ion was  $88.85 \text{ \AA}^2$ .

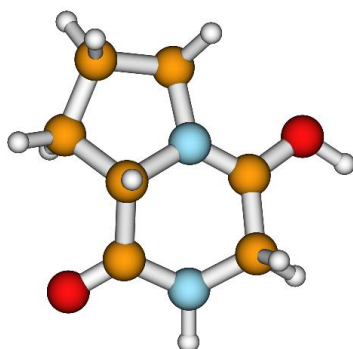
When an acid makes up a side chain in a cyclic  $b_2$  ion, as seen in structure 15, there are favourable interactions between the ionising proton and the oxygen of the acid. This stabilising interaction also includes the hydrogen of the acid interacting with the same oxygen that ionising proton is. This feature is present in all cyclic ions with an acid present in the side chains.



**Structure 16. The minimum energy structure of a glutamine-glycine macrocyclic  $b_2$  ion.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $72.5 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a QG  $b_2$  ion was  $82.93 \text{ \AA}^2$ .

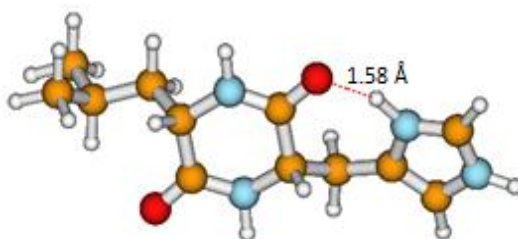
Having an amide group in the side chain of a cyclic  $b_2$  ion forms a very similar feature to that observed in structure 15. Displayed in structure 16, a favourable interaction is formed with between the ionising proton and the side chain amide oxygen, which also has an interaction

with the amide nitrogens hydrogen atoms. Like with the acid interaction, all side chain amides produce this feature.



**Structure 17.** *The minimum energy structure of a proline-glycine macrocyclic  $b_2$  ion.* Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $63.2 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a PG  $b_2$  ion was  $64.61 \text{ \AA}^2$ .

A proline present in the parent peptide leads to a double ring structure (shown in structure 17). Notably in all structures containing proline, the ionising proton takes a position on the oxygen closest to the nitrogen of the proline ring, with the proton facing away from the proline ring. Due to not interacting with more than one proton, each oxygen is more stable than if they were to split electron density between both the ionising proton and the neighbouring N-H. These structures have been proposed by RN Grewal, et al and LL Smith, et al, in the context of discussing the proline effect and how the fragment ion further breaks down by the loss of 71 m/z. [<sup>111</sup>,<sup>112</sup>]

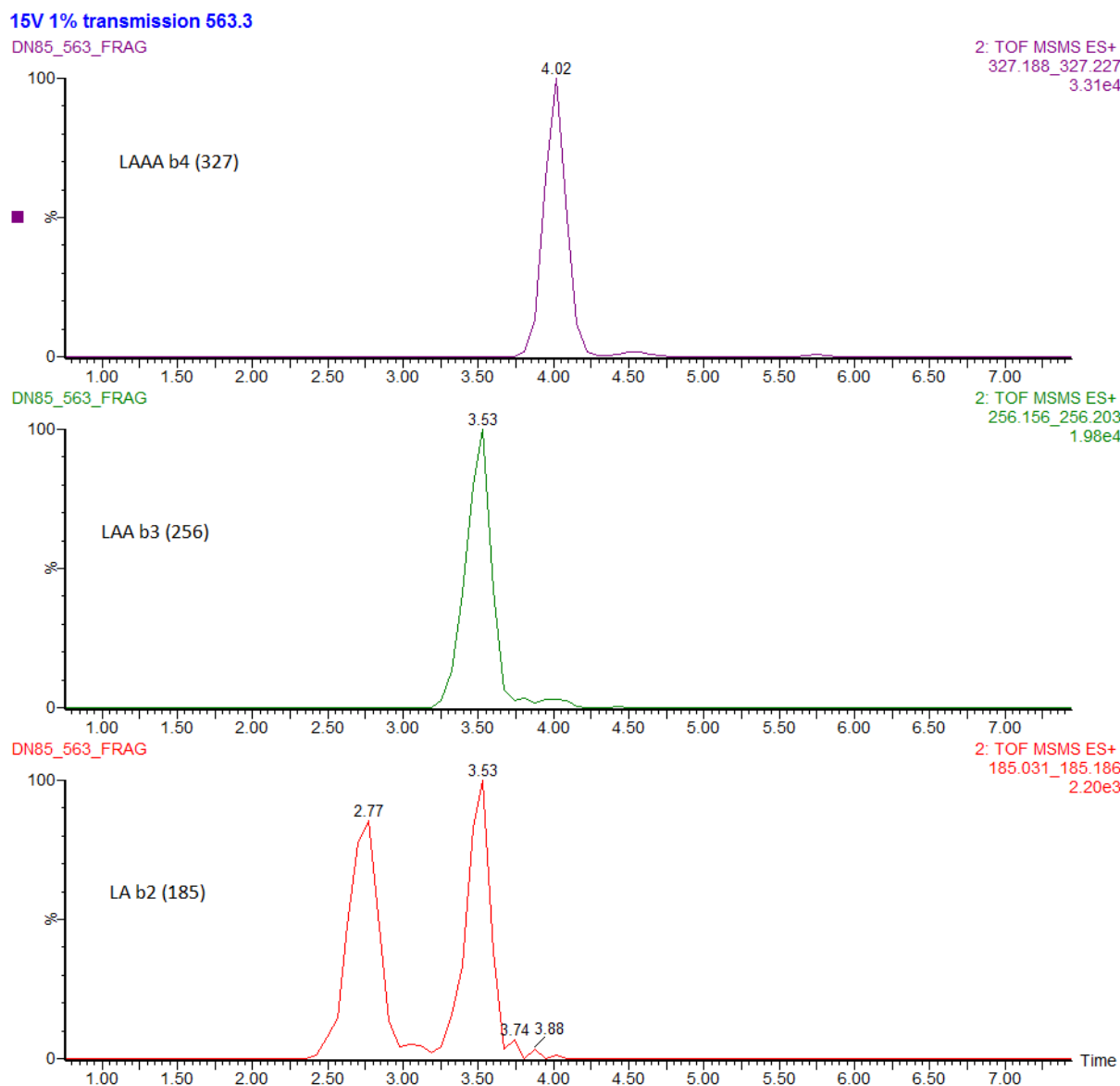


**Structure 18.** *The minimum energy structure of a leucine-histidine macrocyclic  $b_2$  ion.* Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $97.3 \text{ \AA}^2$  for this geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LH  $b_2$  ion was  $96.28 \text{ \AA}^2$ .

When a histidine group is present, such as in structure 18, we see the histidine side chain holds the ionising proton, forming a hydrogen bond with the nearest oxygen. This is the only

case of those cyclic  $b_2$  ions calculated in which the ionising proton does not sit on one of the ring carbonyl oxygens. Due to the rigid ring structure of cyclic  $b_2$  ions, there is considerably less variance between ions. The fact that the position of the two amino acids that make up a cyclic ion does not change the structure also reduces variance, whereas an oxazolone structure will be differently influenced by a side chain on the N-terminus versus in on the oxazolone ring. These structures are consistent with those reported in 2009 by Brittany R. Perkins, et al. [109]

### 3.5.4: Post-separation fragmentation



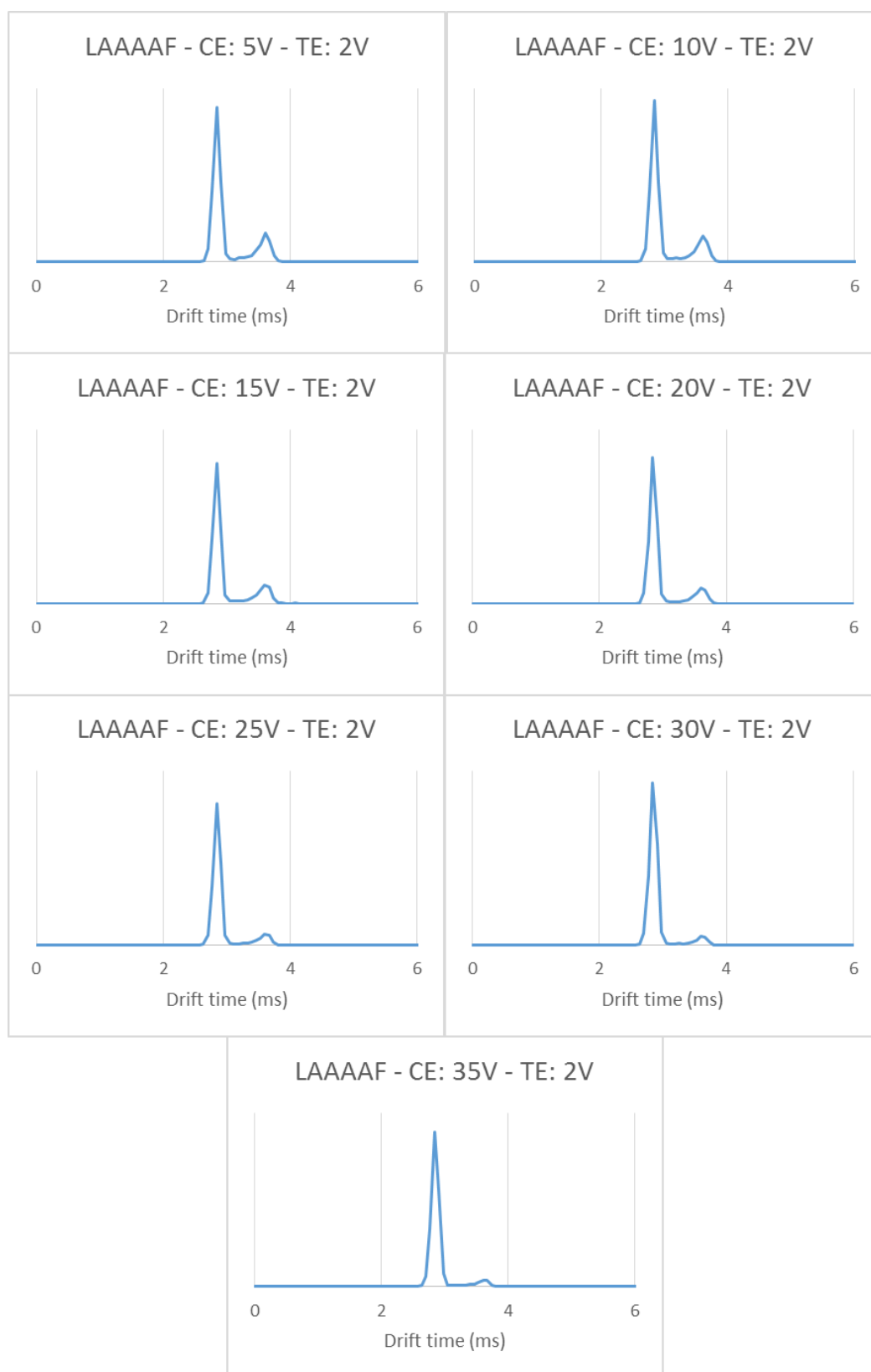
**Figure 27. Three drift time chromatograms of LAAAF fragment mass to charge ratios.** The b2 ion  $m/z$  (185) is shown at the bottom. The b3 ion  $m/z$  (256) is shown centre. The b4 ion  $m/z$  (327) is shown at the top. These mobility chromatograms were obtained on a Synapt G2-Si with a Nitrogen drift gas.

Internally synthesised peptides have been analysed, an example of these is LAAAF. In the case of LAAAF, the intensity of the b2 peak was low in the first attempt at isolation (Figure 27). When isolated, the b2 ion was represented by two almost equal intensity peaks with CCSs of 141.2 Å<sup>2</sup> and 166.6 Å<sup>2</sup>. An experimental CCS was obtained for the LA b2 ion of 141.7 Å<sup>2</sup>, and for the LAA b3 ion of 167.1 Å<sup>2</sup>. This suggests that the second peak in the drift time chromatogram is produced by the b3 ion, with this ion fragmenting to a b2 ion post

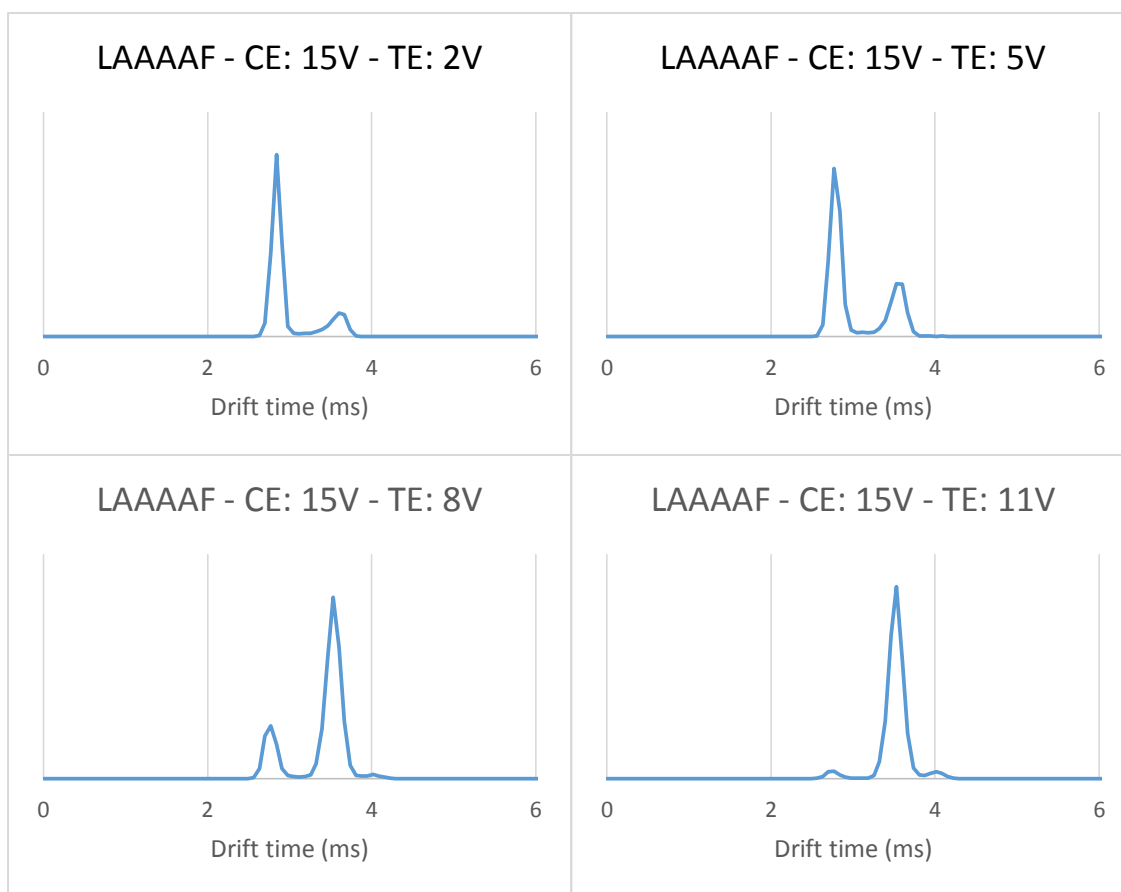
separation. This could potentially cause problems with the analysis and assignment of samples. This is caused by the parent peptide ion fragmenting in trap cell before separation, forming both b2 and b3 ions (and, of course, other ions). The b2 and b3 ions separate in the separation cells. A portion of the b3 ions then fragment in the transfer cell (which is a collision cell used mainly to transfer ions) to b2 ions, which are then detected by the mass spectrometer, but come through later than those produced in the trap cell.

It is important to understand that in the Synapt G2-Si ion mobility mass spectrometers, the TWIMS drift cell sits in between a trap cell and a transfer cell. These two components are essentially the same component with two different names based on their intended function (To trap/fragment ions and to transfer them from the drift region to the TOF region in pulses, respectively). Whilst their primary function is different, these two devices can be controlled by the user of the instrument in the same manner as one another via the Mass-Lynx mass spec tune window. Generally, to help with the maintaining the path of the ions, a small voltage of approximately 2V is applied to the transfer cell. This voltage is the same value that is altered in the trap cell to fragment ions. We observed that this voltage, even as low as 2V, was enough to cause fragmentation of some fragments, producing additional ion mobility signals of the smaller  $b_n$  ions that it may fragment into, misrepresenting the CCS of these smaller fragments.

To study and prove the cause of this phenomenon, the same LAAAAF sample was run and the collision voltage (trap cell voltage) was altered from 5V to 35V in steps of 5, whilst the transfer energy remained constant at 2V (Figure 28). With each increase to the collision voltage, the relative intensity of the 141.2 Å<sup>2</sup> peak increased and the 166.6 Å<sup>2</sup> peak decreased. This is due to the number of b2 ions being produced in the trap cell increasing, and outnumbering the smaller portion of b2 ions being produced post-separation.



**Figure 28. Varying the trap cell voltage of the Synapt G2-Si whilst maintaining the transfer cell energy.** The TWIMS drift region was filled with a nitrogen drift gas. Collision and transfer energies were controlled via Waters Mass Lynx software.



**Figure 29. Varying the transfer cell voltage of the Synapt G2-Si whilst maintaining the trap cell energy.** The TWIMS drift region was filled with a nitrogen drift gas. Collision and transfer energies were controlled via Waters Mass Lynx software.

The opposite was also done, with transfer energy being increased in steps of 3V, from 2V to 11V, the collision energy remained constant at 15V (Figure 29). In this case the difference is far more dramatic, with the 2V 166.6 Å<sup>2</sup> peak being dwarfed by the 141.2 Å<sup>2</sup> peak and in the 11V chromatogram the 166.6 Å<sup>2</sup> peak is almost completely dwarfed by the 141.2 Å<sup>2</sup> peak. This is due to the quantity of b<sub>3</sub> ions being produced in the trap cell outnumbering the b<sub>2</sub> ions being produced in the trap cell significantly, when the b<sub>3</sub> ions reach the transfer cell, they fragment to produce a larger number of post separation b<sub>2</sub> ions.

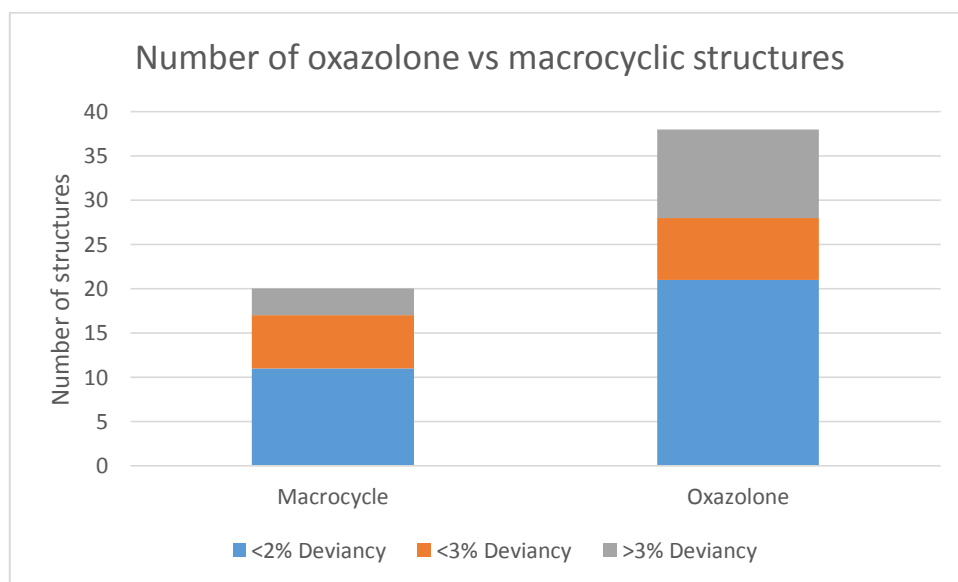
A consequence of this observation is that users of this instrument should be aware of the potential for a second fragmentation of b<sub>n>3</sub> ions, producing incorrect mobility times for these third generation fragments. This issue will be more pronounced in systems in which the larger second generation fragment is much more prevalent than the smaller second generation fragment and may produce third generation fragments with an intensity of magnitudes of



order greater than the otherwise identical second generation fragment. This can causes issues in the growing use of ion mobility for proteomics, in which fragments may be verified by their mobility measurements to improve accuracy. In these cases, caution should be taken to avoid recording CCS values incorrectly into CCS databases. This phenomenon may not be exclusive to proteomics, and could present issues in metabolomic fields such as glycomics.

### Chapter 3.6: Conclusion

The results of the experiment provide an insight into the structures of  $b_2$  ions, strongly suggesting (but does not conclusively prove) the hypothesis that  $b_2$  ions have a tendency to take an oxazolone derivative structure. Very few of the macrocyclic structures correlate well to their respective experimental CCS value, or can be clearly differentiate from the oxazolone structure based on CCS alone. A fair number of oxazolone structures, however, both correlate well to experimentally obtained data and also can be differentiated from macrocycles by the Mobcal margin of error. A large proportion of structures cannot be proven to taken either form, due to either both theoretical values being outside of the 3% error, or both being within. The experiment could be continued with an improved algorithm for computing the nitrogen collisional cross sections and measuring  $b_2$  ions using a nitrogen drift gas. This could also be further expanded to use  $\text{CO}_2$  as a drift gas both experimentally and computationally, however the algorithms used in mobcal to calculate the CCS values would require an extensive rework, as  $\text{CO}_2$  has two polar bonds with different electrostatic charges on the carbon and oxygen, in addition to being a more complex molecule than  $\text{N}_2$  with an additional atom to account for.



**Figure 30.** The number of oxazolone and macrocycles  $b_2$  ions that match each of the proposed structures by CCS values. The experimental values were obtained via IMS on a Synapt G2-Si with a helium drift gas. Computed CCS values were calculated using the trajectory model of Mobcal with a helium drift gas. Deviancy is the difference percentage between the experimentally obtained value and theoretical value, 3% is considered an acceptable deviancy in the IMS community.

With the structures already computed, the mobilities and collisional cross sections may be recalculated using a newer software such as ISiCLE. ISiCLE, In Silico Chemical Library Engine, is a new (as of 2018) freely available software for computing CCS values. ISiCLE refactors Mobcal code to produce a significantly more efficient method of computing trajectory model values. ISiCLE claims to achieve within 3.2% accuracy when calculating CCS values, which is not significantly improved versus the results detailed here, however it may prove to be more accurate for specific sets of small molecules such as these  $b_2$  ions. <sup>[113]</sup> Other new programmes, as of 2019, include Collidoscope and Mobcal-MPI, both of which claim to improve the accuracy of nitrogen drift gas collision parameters. Whilst the structures are already computed for this study, it would take a substantial amount of time and computational power to recalculate the theoretical CCS values with alternative software. This would be a suitable avenue for future work, utilising data generated from this study.

## Chapter 4: Promiscuity and Activity of Nitrilases

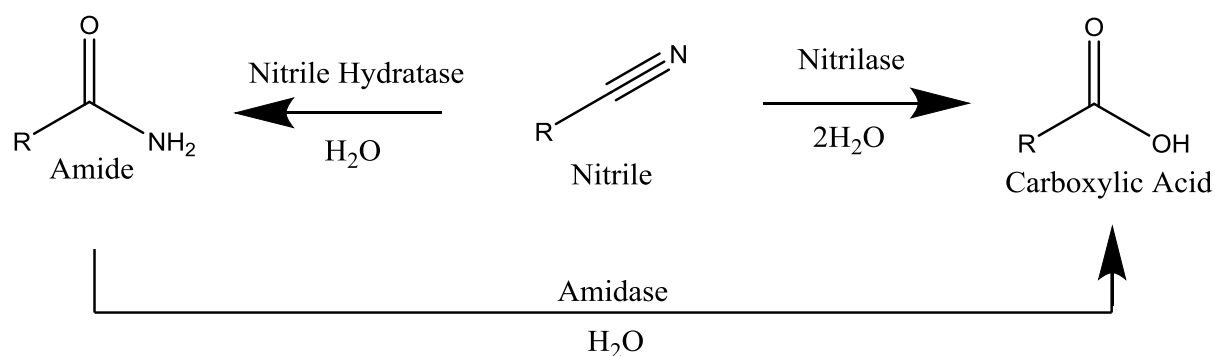
### Chapter 4.1: Introduction

Proteomics and the understanding of the techniques used in the field exist for the purpose of a greater understanding of the application and use of proteins. These fields include an understanding of the proteome available in organisms, the mechanisms of reactions undertaken by proteins and the uses of enzyme catalysis. Proteomics links into a greater understanding of enzyme activity with the metabolome, the full set of small molecules found within an organism. Enzymes are essential to the development of the metabolome, and mass spectrometry is equally useful for investigating the activity of enzymes as it is for understanding their structure. For the following chapter, we will focus on a small subsection of enzymes, nitrilases, and a study of their activity by mass spectrometry.

Most nitriles are highly toxic due to the cyano functional group, a popular example of which is cyanide, known for its effectiveness as a chemical weapon, but it was also an essential component in the early production of amino acids, proteins and lipids. Some plants produce cyanide and other nitriles as part of their metabolism while others survive in environments where cyanide is prevalent. These plants are often sources of nitrilase enzymes [<sup>114</sup>]. Fungi in particular are a source of cyanide hydratase, which converts cyanide into formamide, which further degrades into CO<sub>2</sub> and NH<sub>3</sub> [<sup>115</sup>].

Nitrilases can be divided into two classes; bifunctional nitrilases and real nitrilases. Bifunctional nitrilases are defined by their ability to produce two alternative products, carboxylic acids and/or amides. This allows a bifunctional nitrilase to simultaneously act as a nitrile hydratase and as a nitrilase, illustrated in scheme 5. This bifunctional property of some enzymes has previously been ascribed to a contaminating nitrile hydratase but has more recently been shown to be a dual property of individual enzymes with the *Pseudomonas*

*fluorescens* nitrilase. This property has also been shown to be influenced by the pH and temperature of the reaction environment. [116]



*Scheme 5. The metabolomic pathway of nitriles under bio-catalysis to produce amides and carboxylic acids.*

Real nitrilases convert nitriles into only carboxylic acids without the formation of amides. These nitrilases can be further divided into three categories: aromatic, aliphatic and arylaliphatic nitrilases. These terms denote the type of nitrile substrate to which the nitrilase is most specific [117]. In addition to these three categories of nitriles with defined substrate specificities, there's another class of nitrilases termed promiscuous nitrilases. These nitrilases react with a broader range of nitriles and are not particularly specific to only one group [118]. Whilst the exact mechanism of nitrilase catalysis is not well known, this family of nitrilases are distinguishable by a highly conserved catalytic triad of the amino acids Glu54, Lys127 and Cys169. [119] Via protein modelling, the cysteine residue has been shown to act as the nucleophile, activated by the nearby presence of the glutamic acid, whilst the lysine catalyses the protonation of the functional groups being converted. [120]

Nitriles are often produced as by-products in industrial processes and are introduced into waste water systems, ultimately creating a risk for the health of humans, animals and the environment. A potential solution to reduce the threat of high nitrile concentration is nitrile bio-catalysis, using enzymes produced by organisms to convert nitriles present in the environment to less harmful compounds, particularly carboxylic acids and amides. Nitrilases are also used in industry to create valuable products from waste materials. Some examples include the production of nicotinic acid and mandelic acid [121]. Interest in the application of

nitrilases for environmental and industrial purposes has accelerated rapidly over recent years, and is a highly active research topic. [115]

In this chapter, the quantification of products of various nitrilase reactions will be presented. The chromatography of each product will be displayed, substrate promiscuity experiments will be analysed and discussed, and finally, data obtained from testing a subset of nitrilases under harsh reaction conditions will be shown.

## *Chapter 4.2: Aims*

The aim of this study is to investigate the promiscuity of extremophile nitrilases with different substrates and to probe their functionality under different harsh conditions, like acidic and basic pH, or in organic solvents. To this end, we developed UPLC and MRM methods to quantify numerous carboxylic acid and amide products of enzyme reactions. The methods were to be developed on a UPLC system with a C18 column and a Xevo TQ-XS triple quadrupole mass spectrometer. UPLC-TQMS is a suitable technique for the quantification of enzyme products due to the rapid development time, competent sensitivity and high throughput. Samples have been taken from the Menai strait in North Wales to investigate what valuable enzymes may be available from this local resource. This project is also partially a collaboration with Bayer AG, in which they have provided a number of chemical sewage waste water samples that are to be probed for nitrilases that may be promiscuous or active in harsh reaction conditions.

## Chapter 4.3: Methods and Materials

### 4.3.1: Sources of Recombinant Nitrilases.

Samples analysed in this study were provided by Tran Hai of Bangor University and the Centre for Environmental Biotechnology. Enzymes involved in this study are a mixture of extremophile nitrilases and new nitrilases obtained from A) waste water from a chemical sewage plant of Bayer AG in Dormagen, Germany and B) marine sediment from the Menai Strait. DNA encoding the nitrilases under investigation was done via sequencing on the Illumina MiSeq Platform and cloning into *E. coli* via Polymerase chain reaction (PCR) techniques by Tran Hai. After Expression, the nitrilases were purified via affinity chromatography by Tran Hai.

Nitriles labelled NIT-0 to NIT-20 were obtained from waste water and marine sediment, nitriles labelled NIT-21 to NIT-35 are previously known annotated extremophile nitriles. The source of individual nitriles was not made available to this study. All work prior to mass spectrometry analysis was performed by Tran Hai at the Centre for Environmental Biology. Work performed as part of this study is the method development and quantification of enzyme activity via mass spectrometry. Details of individual experiments were not made fully available to this study.

### 4.3.2: Substrate Sources.

The individual substrates were sourced as displayed in table 7.

*Table 7. Enzyme substrates and their respective sources.*

No.	Nitriles	Source	MW
1	2-(4-fluorophenyl)-2-phenylacetonitrile	Fluorochem	211.239
2	(4-dimethylaminophenyl)piperidin-1yl)-acetonitrile.	ZYLAXA/Pharma	243.35
3	2-amino-2-methylpropanenitrile	Fluorochem	84.069
4	$\alpha$ -phenylglycinonitrile hydrochloride	Fluorochem	168.581
5	R-(+)-mandelonitrile	Aldrich	133.15
6	4-Chlorophenylacetonitrile	Sigma	151.6



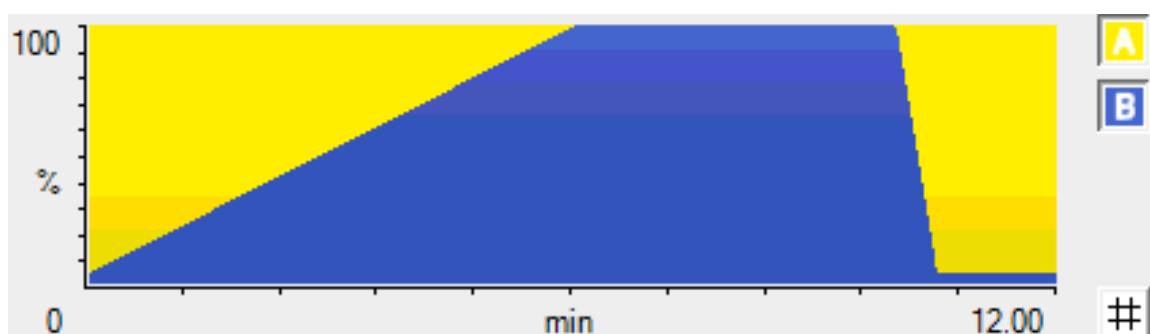
7	2-Fluoro-(4-trifluoromethyl benzonitrile	abcr	189.4
8	(S)-(+)-2-methylbutyronitrile	Sigma	83.13
9	3-Pyridinocarbonitrile	Sigma	104.11
10	4-Amino-2-chlorobenzonitrile	Sigma	152.58
11	Acrylonitrile	Sigma	53.06
12	Glycolic acid nitrile	Sigma	57.06
13	4-Bromomethyl benzonitrile	Sigma	196.04
14	2,2,3,3-Tetramethylsuccinonitrile	Sigma	136.19
15	Adiponitrile	Sigma	108.14
16	Benzonitrile	Sigma	103.12
17	2-amino-5-chlorobenzonitrile	Sigma	152.58
18	Bromoxynile	Sigma	276.91
19	Mandelonitrile	Sigma	133.15
20	Succinonitrile	Sigma	80.09
21	Succinamide	Sigma	116.12
22	Sulfamide	Sigma	96.11
23	Malonamide	Sigma	102.09
24	Nicotinamide	Sigma	122.12
25	Hexanoamide	Sigma	115.17
26	NaCN	Sigma	49.12
27	Acetonitrile	Sigma	41.05

#### 4.3.3: Liquid Chromatography – Tandem Mass Spectrometry (LC-MS/MS)

UPLC-MS/MS experiments were performed on an Acquity I Class UPLC coupled to a Xevo TQ-XS triple quadrupole mass spectrometer. Fisher Scientific Optima grade methanol, water and formic acid were purchased from Fisher Scientific UK. Standards used for method development and calibration were all purchased from Sigma Aldrich.

An Ascentis Express C18 column (150 mm × 2.1 mm, 2.7 µm) from SUPELCO (Bellefonte, PA, USA) was used in the UPLC separation. The column oven was kept at a constant 40°C throughout the experiment. The mobile phase consisted of water with 0.1% formic acid (A) and methanol with 0.1% formic acid (B). The gradient program was as follows: 95% A at 0 min,

0% A at 6 min, 0% A at 10 min, 95% A at 10.5 min, 95% A at 12 min (Figure 31). The flow rate was set at 0.2 mL min<sup>-1</sup>. The injection volume was 1 µL. The analytes were ionized with an ESI source in positive and negative ion modes (utilizing fast switching) under the following source conditions: Desolvation gas 500 L/hr; Cone gas 150 L/hr; Collision gas 0.14 ml/min; Ion spray nebulizer gas 6.8 bar; Capillary voltage 3 kV; Detector voltage 0.56 kV; Cone voltage 53 V; Entrance Potential -2.1 V; Exit Potential -2.1 V. Analysis was in a polarity specific to each compound.



**Figure 31. UPLC gradient applied in the method.** Developed on the Waters Acquity I-Class with the Mass-Lynx software.

Ionisation of individual products was optimised by direct infusion of standards, and MRM methods were developed during the direct infusion of standards and then validated by UPLC-MS/MS. The MRM transitions and collision energies used to identify the compounds are given in table 8:

**Table 8. MRM precursor/product ion transitions.** CE: Collision energy in eV. MRM Methods were developed for the Waters Xevo TQ-XS using the Mass-Lynx software.

Substrate	Product	MRM Method	CE
<u>Positive Ion Mode</u>			
4	2-amino-2-phenylacetic acid	152.0 -> 135.0	12
8	2-methylbutyric acid	103.2 -> 75.1	20
20	Succinamide	117.1 -> 100.0	20
<u>Negative ion Mode</u>			
2	2-(4-(dimethylamino)phenyl)-2-(piperidin-1-yl)acetic acid	263.2 -> 178.1	18
2	2-(4-(dimethylamino)phenyl)-2-(piperidin-1-yl)acetamide	262.2 -> 177.1	18
3	2-aminoisobutyric acid	104.0 -> 58.0	10
5	R-(-)-Mandelic acid	151.0 -> 125.0	15
6	4-Chlorophenyl acetic acid	169.0 -> 125.0	12

7	2-fluoro-4-(trifluoromethyl)benzoic acid	207.0 -> 163.0	10
11	Acrylic acid	No good MRM	N/A
13	4-(bromomethyl)benzoic acid	214.0 -> 153.0	20
14	2,2,3,3-tetramethylsuccinic acid	173.0 -> 129.0	10
15	Adipic acid	145.0 -> 83.0	10
16	Benzoic acid	121.0 -> 77.0	15
18	3,5-dibromo-4-hydroxybenzoic acid	292.0 -> 249.0	20
20,21	Succinamic acid	115.8 -> 98.0	10
20,21	Succinic acid	117.3 -> 73.0	10
23	Malonic acid	103.0 -> 59.0	10
24	Nicotinic acid	122.0 -> 78.0	12

## Chapter 4.4: Results and Discussion

### 4.4.1: Substrate promiscuity of nitrilases

Enzyme activity is reported as  $\mu\text{g}$  of product produced per ml per hour ( $\mu\text{g}/\text{ml}/\text{h}$ , represented by U) per mg of protein (U/mg). This unit is calculated directly from the measurements obtained via mass spectrometry, reported in appendix 2. Quantitation was achieved by producing a concentration curve of commercially available standards. Each enzyme reaction was measured twice. The data in the results tables represent the average of the two LC-MS/MS injections.

Samples were submitted in three lots of experiments and organised in a table format, in which rows denote the enzyme used and the columns denote the substrate used. Each reaction in experiment 1 was performed as a 1ml solution made up of 900 $\mu\text{L}$  of 20 mM Triethylamine (TEA).HCl buffer, 50 $\mu\text{L}$  of enzyme and 50 $\mu\text{L}$  of 100mM substrate (final concentration of 5mM). The reactions were incubated at 37°C for 15 hours. The reactions were stopped with the addition of 1 volume of methanol and placing the reaction mixtures in a freezer at -80°C. Each sample was diluted 50 times for analysis.

*Table 9. Products of experiment 1 quantified in U/mg. O represents the acid product and N represents the amide product. OO represents the succinic acid product, NN represents the succinamide product and ON represents the succinamic acid product. Concentrations were calculated from standard calibrations on a Xevo TQ-XS.*

ENZYMES	2-O	2-N	4	5	18	20-OO	20-ON	20-NN	24
NIT-0	1	1	2	3	13	0	0	0	2528
NIT-7	1	2	2	2	12	0	0	0	3043
NIT-8	1	1	2	3	32	0	0	0	2697
NIT-16	2	2	2	2	370	8	0	0	2743
NIT-24	1	1	2	3	26	0	0	0	2612
NIT-25	2	2	2	16	0	0	0	0	2519
NIT-26	1	1	2	277	17113	1069	493	178	2703
NIT-28	2	2	2	1160	29767	2004	604	35	2709

Each reaction in experiment 2 was performed as a 0.5ml solution made up of 450 $\mu\text{L}$  of 20 mM of a TEA.HCl buffer, 20 $\mu\text{L}$  of enzyme and 30 $\mu\text{L}$  of 100mM substrate (final concentration of 6mM). The reactions were incubated at 37°C for 5 hours. The reactions were stopped with

the addition of 1 volume of methanol and placing the reaction mixtures in a freezer at -80°C. Each sample was diluted 50 times for analysis.

The benzoic acid standard available for calibration was not of a sufficient quality for the quantification of the respective products, therefore all reactions involving substrate 16 use only peak area to represent activity and cannot be compared to other substrates directly.

**Table 10. Products of experiment 2 quantified in U/mg.** Concentrations were calculated from standard calibrations on a Xevo TQ-XS.

CE or Pure	3	6	7	8	14	15	16
NIT-0	2	348	5	0	4936	0	2000
NIT-1	2	46	1	0	5107	0	0
NIT-3	2	252	1	0	5107	0	0
NIT-7	2	466	214	0	4446	0	4000
NIT-7	2	456	4	0	4920	0	0
NIT-8	2	404	300	0	4789	0	8000
NIT-15	2	373	9	0	5196	0	2000
NIT-16	2	3732	1157	0	3436	0	94000

Each reaction in experiment 3 was performed as a 0.5ml solution made up of 450µL of 20 mM of a TEA.HCl buffer, 20µL of enzyme and 30µL of 100mM substrate (final concentration of 6mM). The reactions were incubated at 37°C for 14 hours. The reactions were stopped with the addition of 1 volume of methanol and placing the reaction mixtures in a freezer at -80°C. Each sample was diluted 50 times for analysis.

**Table 11. Products of experiment 3 quantified in U/mg.** O represents the acid product and N represents the amide product. OO represents the succinic acid product, NN represents the succinamide product and ON represents the succinamic acid product. Concentrations were calculated from standard calibrations on a Xevo TQ-XS.

CE or Pure	2-O	2-N	3	5	6	7	8	13	14	15	16	20-OO	20-ON	20-NN
NIT-17	3	2	2	24	13	0	0	0	311	4	1400	0	0	0
NIT-19	2	2	2	1604	4550	2	0	0	284	491	2300	1538	230	11
NIT-20	3	2	2	2	0	0	0	0	295	5	400	0	0	0
NIT-24	2	1	2	32	93	0	0	0	295	1	400	0	0	0
NIT-25	2	2	3	3	0	0	0	0	280	2	0	0	0	0
NIT-26	1	1	2	162	149	1	0	0	287	296	1300	547	131	53
NIT-28	1	1	1	758	6573	2	0	0	289	3211	2900	1152	309	23
NIT-35	2	1	2	548	5148	1065	15	0	294	4079	132000	49	9	0

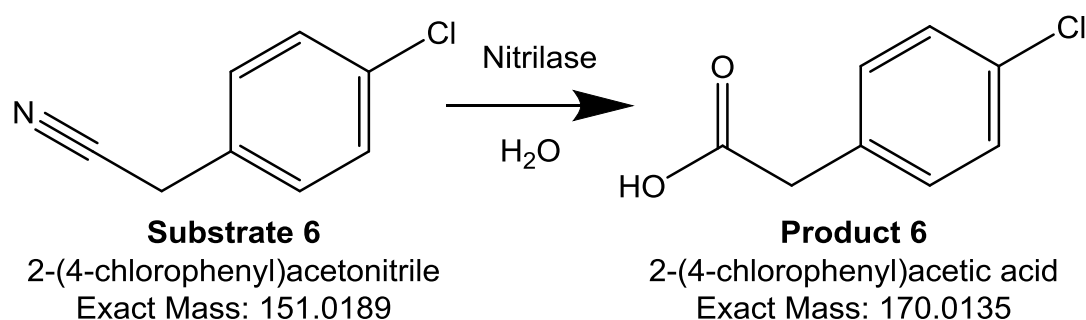
Promiscuity experiments on the various enzymes revealed that most examined nitrilases are not highly specific, with some even being highly promiscuous. Notably NIT-35, NIT-28, NIT-26, NIT-19 and NIT-16 show a very high level of promiscuity, each being active on most substrates they are tested with. NIT-35 has not only very high promiscuity, but also has a high activity for most substrates, often greater than any of the other nitrilases that also show activity with the same substrates. NIT-35 was active with all substrates that showed activity with any other enzymes. NIT-0, NIT-1, NIT 17 and NIT-20 are all highly substrate specific enzymes, each active with only one substrate excluding those substrates (such as 2, 4, 14 and 24) which react with all tested enzymes. Enzymes that are active with substrate 20, succinonitrile, produce primarily the carboxylic acid product, succinic acid, whilst succinamic acid is always the second most abundant product and succinamide is the least abundant product. This may be due to enzymes that produce amides also act as amidases and converting amides to carboxylic acids.

The practical outcomes of this set of experiments is an insight into which novel nitrilases have a high substrate promiscuity. Primarily NIT-35, NIT-28, NIT-26 and NIT-16 are very interesting enzymes as they all catalyse a large number of nitrile hydrolysis reactions. These nitriles could be suitable for reducing nitrile contamination in waste systems, with the adaptability to be used in complex mixtures and potentially remove contaminants that may not be known.

#### *4.4.2: Influence of reaction conditions (pH, solvent) on product generation with 2-(4-chlorophenyl)acetonitrile*

In a further experiment, enzyme activity under varying reaction conditions was examined with two substrates, 2-(4-chlorophenyl)acetonitrile (substrate 6) and 2-fluoro-4-(trifluoromethyl)benzonitrile. These two substrates were chosen for this experiment due to their varying (and often high) activity with multiple nitriles in previous experiments. These variable conditions include the pH of the reaction, the solvents that make up the reaction medium and the incubation temperature. The pH conditions were controlled via the addition of HCl and NaOH to reduce and increase the reaction pH, respectively. The target pH values for the reaction conditions are pH 5, 7.5 and 9. This allows us to monitor activity of enzymes under moderately acidic and basic conditions that may be present in waste water from

industrial chemical processes, and allow us to probe optimal conditions for the respective enzymes in future experiments. The solvents were added as a 10% volume/volume and were either methanol (MeOH), ethanol (EtOH) or isopropanol (Iso-P). These organic solvents were selected as they are common alcohols in many industrial synthesis, such as plastic production and are present often in waste systems that contain nitrilase waste by-products. The presence of organic solvents can also increase the solubility of hydrophobic materials, which would help to facilitate the enzymatic conversion of these materials were the enzymes able to remain active in the presence of these solvents. [122] Experiment 4 focuses on the enzymes activities with substrate 6 and how each harsh reaction conditions affect ability of the enzymes to hydrolyse this substrate as displayed in scheme 6.



*Scheme 6. The conversion of 2-(4-chlorophenyl)acetonitrile to 2-(4-chlorophenyl)acetic acid by a nitrilase.*

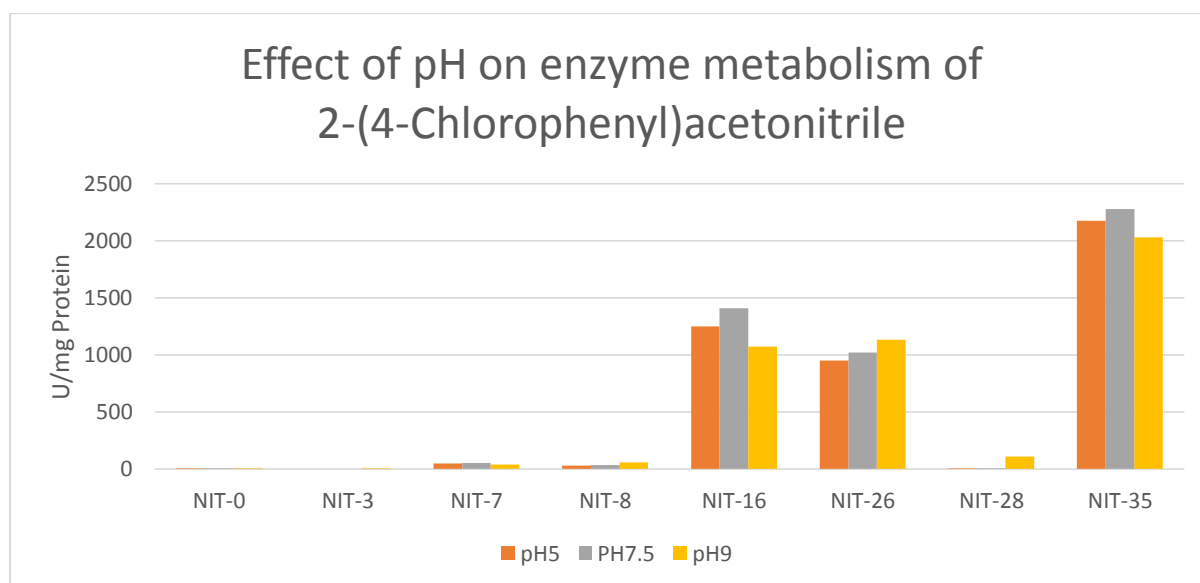
All experiment 4 reactions were performed as a 0.5ml solution made up of 450μL of 20 mM of a TEA.HCl buffer, 20μL of enzyme and 30μL of 100mM substrate (final concentration of 6mM). The reactions were incubated at 37°C for 15 hours. The T37 samples were positive control samples with a pH of 7. The reactions were stopped with the addition of 1 volume of methanol and placing the reaction mixtures in a freezer at -80°C. Each sample was diluted 50 times for analysis.

**Table 12. Products of experiment 4 quantified in U/mg.** “T37 –S” represents the negative control without substrate, “T37 E\*+ S” represents the negative control with enzymes boiled at 90°C. Concentrations were calculated from standard calibrations on a Xevo TQ-XS.

Experiment 4	SUBSTRATE 6								
ENZYMES	T37 -S	T37 E*+ S	T37	pH5	pH7.5	pH9	MeOH	EtOH	Iso-P
NIT-0	13	0	0	7	3	4	4	4	2

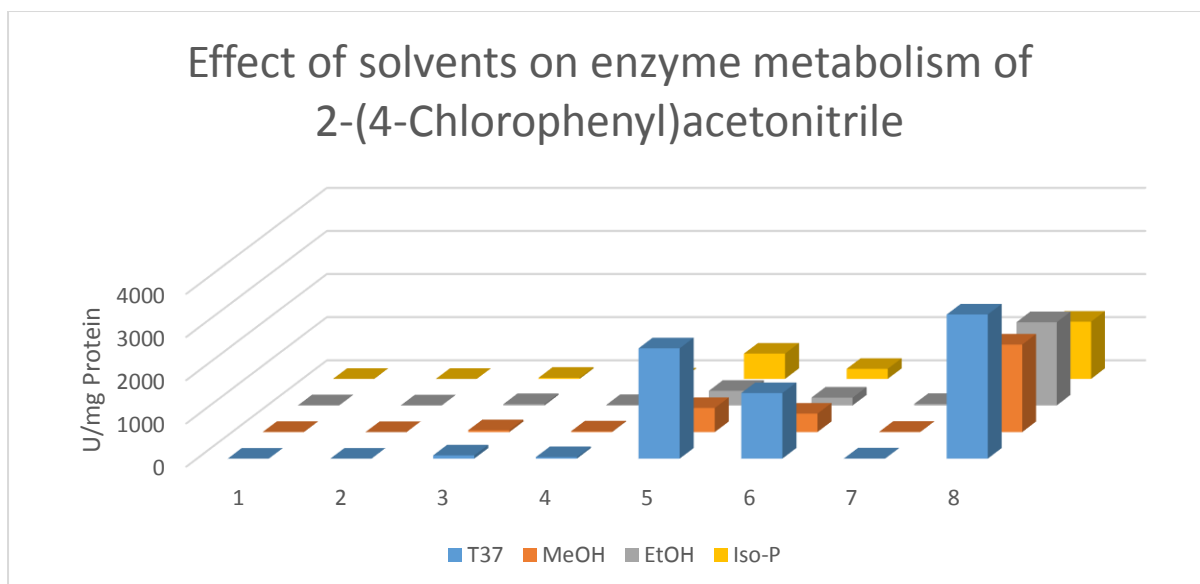
<b>NIT-3</b>	0	0	0	0	0	5	0	2	0
<b>NIT-7</b>	0	0	74	50	55	40	42	32	26
<b>NIT-8</b>	0	0	35	32	36	56	16	9	8
<b>NIT-16</b>	0	0	2550	1249	1407	1072	556	339	585
<b>NIT-26</b>	0	0	1516	951	1019	1132	426	182	233
<b>NIT-28</b>	0	0	4	9	8	108	5	36	30
<b>NIT-35</b>	0	0	3335	2173	2278	2031	2023	1926	1325

The data in experiment 4 (Table 12 and Figure 32) show that reducing or increasing the pH does not significantly reduce the activity for most enzymes, with the exception of NIT-28. In the case of NIT-35, compared to a neutral pH of 7.5, pH5 and pH 9 only reduce the efficiency of the enzyme by 4.7% and 11%, respectively. NIT-7 shows a small increase (5%) in activity for pH7.5 versus pH5 and a 33% increase versus pH9. NIT-28 shows a massive improvement in activity when exposed to the more basic conditions, with pH9 producing ten times the concentration of product compared to pH5. NIT-26 shows a smaller leaning towards higher pH conditions with 11.1% increase in product versus pH7.5 and a 19.3% increase versus pH5. NIT-8 also shows an affinity towards higher pH with a 70% increase in product from pH5 to pH9. NIT-3 only showed activity under basic conditions, and this activity was also very low.



**Figure 32.** The enzyme activity for each nitrilase at different pH levels for the substrate 2-(4-Chlorophenyl)acetonitrile. Quantification was performed via a standard calibration on a Xevo TQ-XS.





**Figure 33.** The enzyme activity for each nitrilase in different solvents for the substrate 2-(4-Chlorophenyl)acetonitrile. Quantification was performed via a standard calibration on a Xevo TQ-XS.

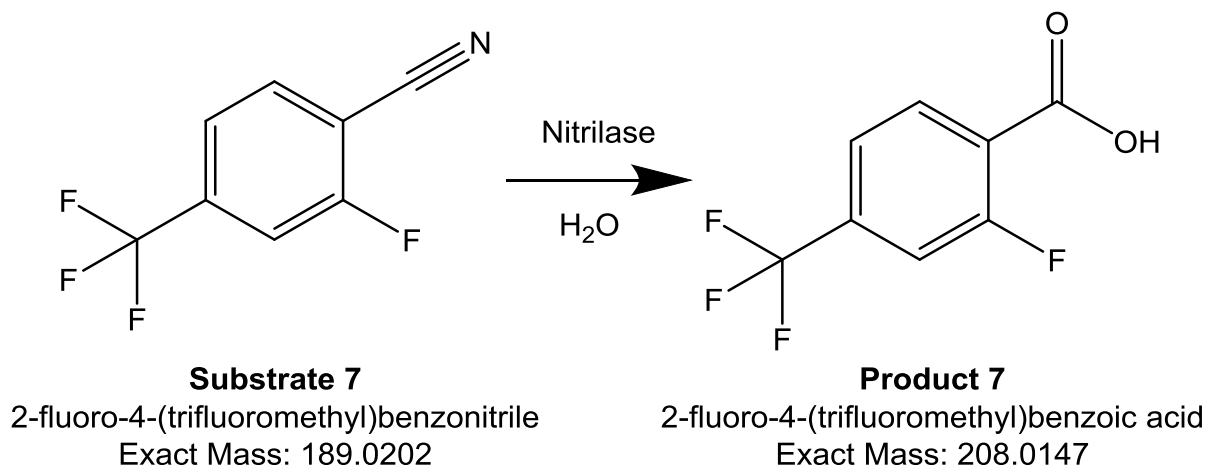
Solvent conditions are also shown to have a significant impact on the activity of these enzymes. All enzymes are less active in these solvents versus the positive control, as shown in Table 12 and Figure 33. NIT-0 and NIT-3 both display no or hardly any activity under these conditions. NIT-7 shows greater activity in MeOH versus other solvents, with EtOH causing a 23.1% reduction in activity and isopropanol (Iso-P) causing a 38.5% reduction in activity. NIT-8 shows low activity for all conditions, but notably better activity in MeOH, with a 66% higher concentration of product versus EtOH and Iso-P. NIT-16 has similar levels of activity for MeOH and Iso-P with a 4.8% difference in favour of Iso-P, however, incubation in EtOH produces 39% less product. NIT-26 shows greater activity in MeOH than in EtOH and isopropanol, which leads to a decrease in activity levels of 57% and 45.3%, respectively. NIT-28 is the only enzyme to show improved activity in EtOH and also has improved activity against the T37 conditions, yielding an 11 times higher product concentration. Iso-P shows an 18.1% reduction in activity versus EtOH and MeOH shows an 82.8% reduction in activity, however both of these produce more activity than the standard conditions. NIT-35 shows good activity in MeOH and EtOH, with only a 5% difference between the two, whilst Iso-P falls short with a 34.6% reduction in activity versus MeOH.

These results all follow the expected trend of enzyme activity in organic solvents. Typically, organic solvents reduce the activity of enzymes due to the reduction in access to

water molecules at the surface of the enzyme. [123] Reducing the water access increases the rigidity of enzyme, hindering interactions with substrates and reducing the ability of an enzyme to facilitate catalysis of a substrate at its active site. The lack of water and interaction of the solvent molecules themselves can interfere with hydrogen bonding between protein subunits, as well as active site unit interactions with substrates. [124] It's been reported that enzyme activity in organic solvents may be improved by the presence of disulfide bonds on the surface of the protein. [125]

#### 4.4.3: Influence of reaction conditions (pH, solvent) on product generation with 2-fluoro-4-(trifluoromethyl)benzonitrile.

Experiment 5 focuses on the enzyme activities with substrate 7, 2-fluoro-4-(trifluoromethyl)benzonitrile, and how each harsh reaction condition affects the enzyme's abilities to hydrolyse this substrate as displayed in scheme 7. 2-fluoro-4-(trifluoromethyl)benzonitrile was chosen due it's high activity with a wide range of nitrilase enzymes under the previously reported standard reaction conditions.



*Scheme 7. The conversion of 2-fluoro-4-(trifluoromethyl)benzonitrile to 2-fluoro-4-(trifluoromethyl)benzoic acid by a nitrilase.*

All reactions in this experiment were performed as a 0.5ml solution made up of 450μL of 20 mM of a TEA.HCl buffer, 20μL of enzyme and 30μL of 100mM substrate (final concentration of 6mM). The reactions were incubated at 37°C for 15 hours. The T37 samples were positive control samples with a pH of 7. The reactions were stopped with the addition of 1 volume of

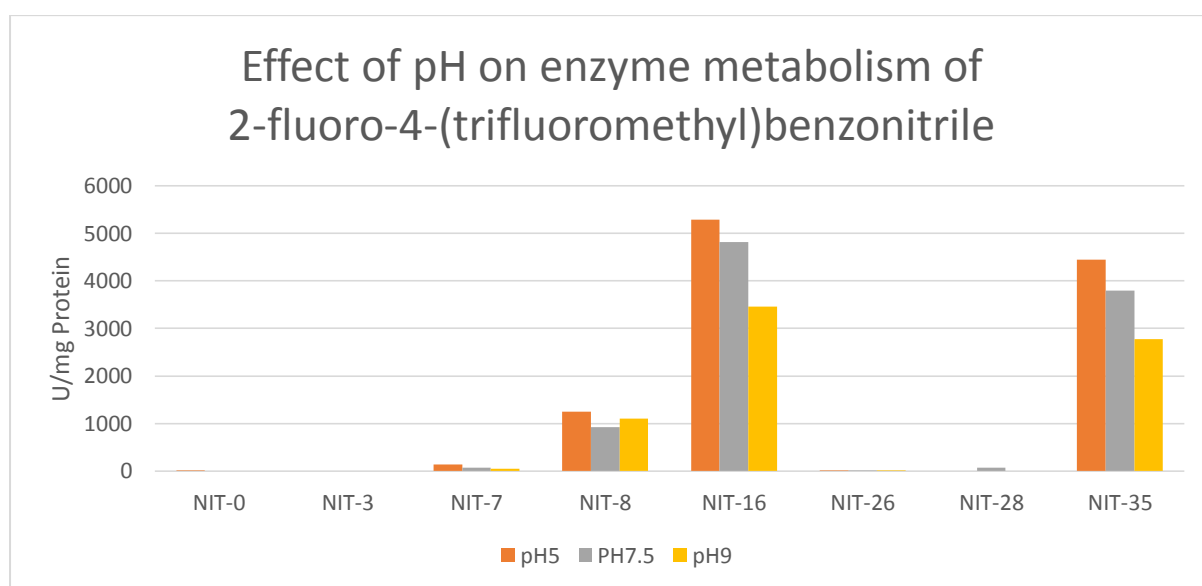
methanol and placing the reaction mixtures in a freezer at -80°C. Each sample was diluted 50 times for analysis.

**Table 13. Products of experiment 5 quantified in U/mg.** “T37 –S” represents the negative control without substrate, “T37 E\*+ S” represents the negative control with enzymes boiled at 90°C. Concentrations were calculated from standard calibrations on a Xevo TQ-XS.

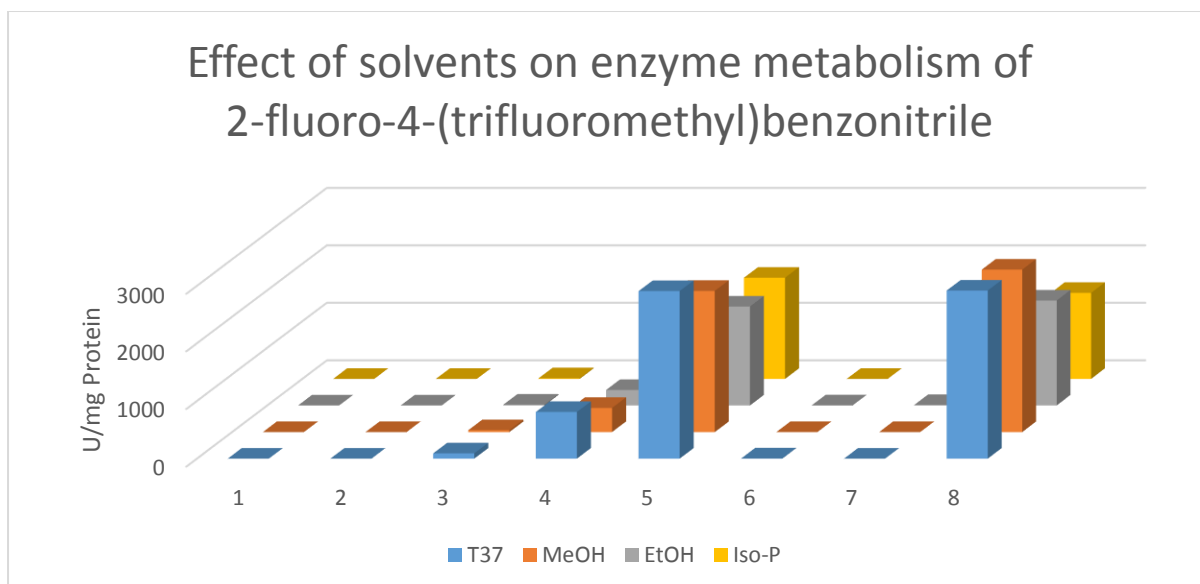
Experiment 5	SUBSTRATE 7								
ENZYMES	T37 S	T37 E*+ S	T37	pH5	pH7.5	pH9	MeOH	EtOH	Iso-P
NIT-0	0	0	0	3	0	0	0	0	0
NIT-3	0	0	0	0	0	0	0	0	0
NIT-7	0	0	91	142	70	54	33	15	11
NIT-8	0	0	810	1247	922	1101	419	269	220
NIT-16	0	0	2908	5285	4812	3460	2451	1718	1758
NIT-26	0	0	3	11	9	2	1	1	0
NIT-28	0	0	0	0	70	0	0	9	23
NIT-35	0	2	2918	4447	3796	2774	2822	1822	1494

From the data for experiment 5 for substrate 7 the activity of the nitrilases can be compared with the activities determined in experiment 4 with substrate 6, showing whether the environmental effects on the enzyme are due to the nature of the enzyme alone or whether enzymes may be more active under some conditions for specific substrates. For substrate 7, the varying pH conditions generally show greater activity versus the standard conditions at pH7 (T37). However this is the inverse of what is shown for substrate 6 where a more acidic or more basic pH generally decreases the activity of the enzyme compared to the standard conditions (T37). NIT-0 and NIT-3 both show little to no activity on both substrates under any of the tested conditions. On substrate 7, NIT-7 shows greater activity in more acidic conditions, with pH7.5 showing a 50.5% reduction in activity versus pH5 and a 61.9% reduction for pH9 versus pH5. With substrate 6, NIT-7 shows a 61.4% reduction in activity for its best pH condition versus the optimal pH condition for substrate 7. With substrate 7, NIT-8 again also shows its highest activity in acidic conditions, pH7.5 results in 26.1% lower activity than pH5, and pH9 results in 11.7% lower activity than pH5. NIT-8 is overall much more active on substrate 7 than on substrate 6, as shown for all conditions, in which the concentration of product was never observed to exceed 56 U/mg for substrate 6, whilst substrate 7 reaches a

minimum of 922 U/mg for the pH conditions and 220 U/mg for the solvent conditions. NIT-16 shows a massive increase in activity under the varying pH conditions for substrate 7, with the greatest activity at pH5, while incubation at pH7.5 leads to an 8.9% reduction in activity and pH9 reduces activity by 34.5%. Whilst under standard conditions the two substrates show similar levels of activity, the pH conditions for substrate 6 generally show half as much activity versus the standard conditions and substrate 7 shows about twice the activity for pH conditions versus the standard conditions. NIT-26 has minimal activity towards substrate 7, with a maximum of 11 U/mg of product across all conditions; it has some activity at pH5 and pH7.5 but hardly any activity under all other tested conditions. In contrast, NIT-26 provides good activity with substrate 6, producing as much as 1516 U/mg of product under standard conditions. NIT-28 shows low, but specific activity with substrate 7, limited to the pH7.5 condition where it produced 70 U/mg of product, while incubation at more acidic or basic pH yielded no product. A preference for a particular pH could also be observed on substrate 6, albeit in this case maximum activity was observed at pH9. NIT-35 shows improved activity for substrate 6 under acid conditions, pH7.5 has 14.6% less activity than pH5 and pH9 shows 37.6% less activity. More product is recorded for each condition for the substrate 7 experiments of NIT-35, excluding the T37 condition which showed 14% more activity for substrate 6.



**Figure 34.** The enzyme activity for each nitrilase with 2-fluoro-4-(trifluoromethyl)benzoic acid at different pH levels. Quantification was performed via a standard calibration on a Xevo TQ-XS.



**Figure 35.** The enzyme activity for each nitrilase with 2-fluoro-4-(trifluoromethyl)benzonitrile in different solvents. Quantification was performed via a standard calibration on a Xevo TQ-XS.

The data for enzyme activities when incubated with different solvents are similar for both substrates, with the two notable differences of NIT-8 and NIT-26 (see below). On substrate 7, NIT-7 shows greater activity in MeOH, with a 55.4% reduction in EtOH and a 67.5% reduction in Iso-P. This follows a similar trend to what is observed for substrate 6. NIT-8 shows a greater activity in MeOH versus a 35.8% reduction in activity for EtOH and a 47.4% reduction for Iso-P. While this is a similar trend as shown for substrate 6, NIT-8 is much less active for this substrate with a 96.2% reduction in activity versus substrate 7. NIT-16 again shows greater activity in MeOH for substrate 7, with a 29.9% reduction in activity for EtOH and a 28.3% reduction in activity for Iso-P. In contrast, NIT-16 was most active in Iso-P on substrate 6, although the enzyme activity was only slightly increased compared to incubation with MeOH. However, the product generation in EtOH and Iso-P differed more for substrate 6 than for substrate 7, where almost equal amounts of product for both reaction conditions were measured. NIT-26 shows no activity under any solvent conditions on substrate 7, which is consistent with its general lack of affinity for this substrate, and in contrast to its activity on substrate 6. NIT-28 also displays limited activity for substrate 7, showing no activity in MeOH, little activity in EtOH, and most activity in Iso-P. Similar levels of activity are observed for substrate 6, but with a higher activity in EtOH, but still with a similar level of activity for Iso-P. NIT-35 shows its greatest activity in MeOH, a similar level of activity to the standard conditions, EtOH shows a reduction of 35.4% versus MeOH and Iso-P shows a reduction of

47% versus MeOH. Notably MeOH was significantly better for the activity of NIT-35 with substrate 7 than 6, further exaggerating the difference between MeOH and EtOH, whereas for substrate 6 the enzyme activity was similar in both conditions.

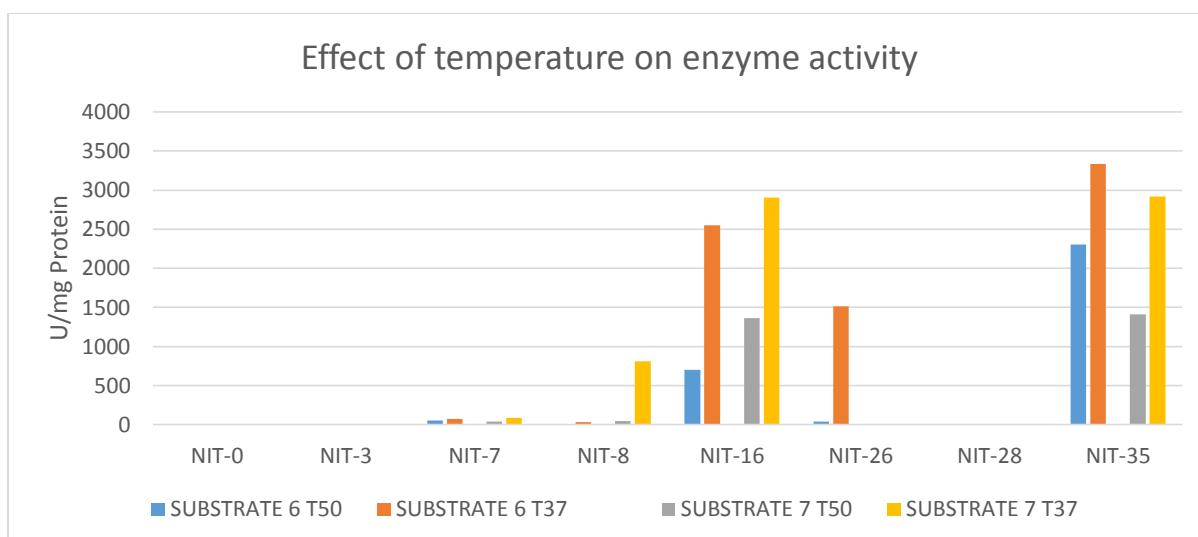
The results of experiments 4 and 5 show that whilst the pH affects the activity of the enzymes, it is also dependent on the substrate being metabolised, with some enzymes performing better at lower pH for one substrate and better at a higher pH for another substrate. However, the solvent environment experiments show that the enzyme activity is influenced differently depending primarily on the solvent present but not significantly affected by the substrate being metabolised.

#### 4.4.4: Influence of incubation temperature on product generation.

In addition to the experiments on pH and solvent conditions, which were carried out at 37°C, an additional reaction for each substrate-enzyme mixture was performed at a higher temperature of 50°C.

*Table 14. Products of experiment 6 quantified in U/mg. Concentrations were calculated from standard calibrations on a Xevo TQ-XS.*

Experiment 6	SUBSTRATE 6		SUBSTRATE 7	
ENZYMES	T50	T37	T50	T37
NIT-0	0	0	0	0
NIT-3	1	0	0	0
NIT-7	56	74	43	91
NIT-8	2	35	47	810
NIT-16	699	2550	1362	2908
NIT-26	38	1516	0	3
NIT-28	9	4	0	0
NIT-35	2303	3335	1414	2918



**Figure 36.** The enzyme activity for each nitrilase with substrates 6 and 7 under different temperatures. Quantification was performed via a standard calibration on a Xevo TQ-XS.

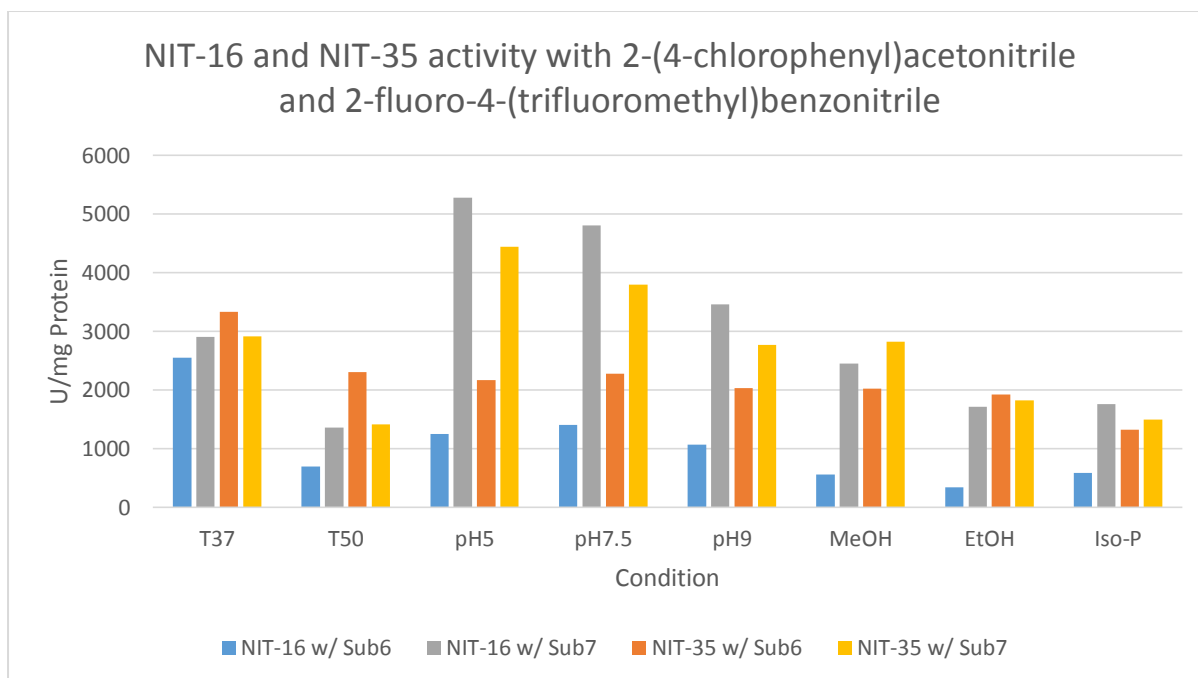
With one exception, NIT-28, the 50°C reduced the enzyme activity significantly, although the degree to which the activity was reduced varied from enzyme to enzyme. NIT-0 and NIT-3 continue to show no activity for either substrate at either temperature. NIT-7 shows a moderate decrease in activity at 50°C, with activity with substrate 6 being reduced by 24% and substrate 7 by 52.5%. NIT-8 shows a great reduction in activity at the increased temperature, dropping by approximately 95% for both substrates. NIT-16 shows a substantial decrease in activity at 50°C, losing 72.6% of its activity with substrate 6 and 53.2% of activity with substrate 7. NIT-26 does not show significant activity with substrate 7 at any temperature, but shows a great reduction in activity (97.5%) for substrate 6 at 50°C vs 37°C, making high temperatures unsuitable for this enzyme despite still producing 38 U/mg of product. Whilst NIT-28 does not produce any product from substrate 7, it produces a small quantity from substrate 6 at 37°C and a 124% increased quantity at 50°C. Although these quantities are very small, it may be a sign of increased activity at higher temperatures and could provide an avenue for further investigation. NIT-32 shows a similar trend to NIT-7, showing a moderate decrease in activity of 30.9% and 51.5% for substrates 6 and 7, respectively, at 50°C.

#### *Chapter 4.5: Conclusion*

A suitable LC-MS/MS method, using UPLC and a triple quadrupole mass spectrometer, has been developed to analyse a range of carboxylic acids with individual MRMs. UPLC-Mass spectrometry has generally proven to be highly adaptable to monitoring enzyme activity with carboxylic acid products. Despite the fact that carboxylic acids are acidic, they tend to easily deprotonate in the presence of formic acid, this is a common feature of formic acid, making it suitable for both positive and negative ionisation modes. [126] This is observable in the significant signals observed for all products quantified here. Additionally, C18 chromatography separates the various carboxylic acids easily due to the structures of the acids. The COOH functional group adds a strong polar component to the molecule, which binds poorly and thus is eluted reasonably easily. However, the various bodies of the carboxylic acids reduce or enhance the polarity of the overall molecule substantially, generating a variation in the acids affinity for the stationary phase, producing a good spread of retention times. The methods explored here are both suitable and efficient for the rapid separation of carboxylic acids.

Of the analysed nitrilases, NIT-35, NIT-28, NIT-26 and NIT-16 are highly promiscuous, with variable degrees of activity. Under the various abnormal reaction conditions that were trialled, NIT-16 and NIT-35 were found to be consistently active. Both of these nitrilases show some of the most promising results with high activity with a wide range of substrates, but also strong activity under a wide array of adverse conditions. However, NIT-16 is a new nitrilase discovered by the Centre for Environmental Biotechnology at Bangor University, whilst NIT-35 is a known extremophile enzyme synthesised from COG and Pfam databases. Further experiments should be performed on NIT-16 to further understand the limits of its promiscuity and ability to withstand harsh reaction conditions. NIT-8 shows a lower promiscuity than the aforementioned enzymes, however it shows a reasonable independence to pH levels and therefore might be a useful enzyme for reacting with specific substrates. NIT-28, whilst not particularly active in most conditions, displayed a significant increase in activity when exposed to physiological or basic pH conditions (substrate dependent); therefore it may be a reasonable avenue of future research to investigate the activity of this enzyme on different substrates in suitable pH conditions.





**Figure 37.** The enzyme activity for nitrilases NIT-16 and NIT-35 with 2-(4-chlorophenyl)acetonitrile (Substrate 6) and 2-fluoro-4-(trifluoromethyl)benzoic acid (substrate 7) under different conditions. Quantification was performed via a standard calibration on a Xevo TQ-XS. U represents  $\mu\text{g}$  of product produced per ml per hour.

The primary route of future work is to take a subset of nitrilases studied here and further probe their abilities with additional conditions and substrates. NIT-32 and NIT-16 are good candidates for monitoring further promiscuity due to their aforementioned resilience to a broad range of conditions. 4-Chlorophenyl acetic acid and 2-fluoro-4-(trifluoromethyl)benzoic acid are both good acids for monitoring the enzyme activity of the nitrilases studied here, due to the high ionisation potential and generally high conversion from their respective nitrile substrates with a broad range of enzymes. Therefore, if we were to expand our selection of nitrilases, these acids should be included to study the condition promiscuity of further enzymes. One route available in the near future is to probe the activity of enzymes in the presence of heavy metal ions, such as Cu or Fe, or other common inhibitors such as EDTA. These types of promiscuity can be useful in processes in which you want to cease the activity of enzymes that are readily inhibited by them (eg, during production reactions), whilst sustaining the activity of enzymes that remove waste materials such as nitriles (eg, in waste water run off systems).

## Chapter 5: Conclusion and Future Work.

In this thesis we have explored the world of proteomics using ion mobility practically and theoretically. We have shown conclusively that polyalanine is not a suitable calibrant for high resolution ion mobility by examining enantiomers using the new cyclic ion mobility spectrometer from Waters. We observed the standards used for calibration contain a racemic mixture of polyalanines that separate into distinct peaks under high resolution. We compared this to theoretical values of 5-A polyalanine molecules and found that mobcal is not accurate enough to distinguish enantiomers by generating theoretical CCS values. Future work on this project could potentially involve exploring alternative options for IMS calibration, ideally cheap, non-chiral molecules that can produce a regular range of mobilities and masses. Whilst the mobcal is not yet suitable for distinguishing these enantiomers, the workflow for generating structures could be improved in future by including FEP calculations to generate free energy maps to better confirm whether these structures might interconvert. Additional structures could also be generated for 8-A and 10-A polyalanine, although it would be best to ensure that a more accurate method of computing CCS values was available before spending the resources on these calculations.

By exploring the gas phase ion mobilities of peptide  $b_2$  fragments we have attempted to gain a greater understanding of the structures that these ions take in a mass spectrometer. We've observed, using a theoretical approach to determining collisional cross sections, that a majority of the ions take an oxazolone structure, which matches well with what previous literature has predicted. In performing this investigation, we have found that the primary programme used to compute these values, mobcal, is not sufficiently accurate to generate nitrogen drift gas CCS values. For this reason, the majority of the work was performed using the lower resolution helium drift gas. Future work in this field would benefit well from improving the trajectory model approximations for nitrogen drift gases, utilising some more newly developed programmes that were not available at the time of this study. Some examples of this include collidascope, Isicle and mobcal-MPI. Additionally, more potential structures could be investigated for non-classical  $b_2$  structures, in addition to including amino acid residues that form such structures, including aspartic acid, asparagine, glutamic acid and glutamine.

We finally investigated the promiscuity of a set of nitrilases from various sources including industrial waste water and the local Menai strait in north Wales. We created an efficient and repeatable UPLC-MRM method for monitoring the activity of 20 amides and carboxylic acids. In conjunction with Tran Hai at Bangor University, we were able to monitor the activity of novel nitrilases with 20 substrates under a range of conditions including pH, temperature and solven environment. A small subset of analysed enzymes produced promising activity with a range of substrates, showing good substrate promiscuity, whilst also showing potentially promising activity in a range of reaction conditions, presenting good condition promiscuity. In future work, we could examine these enzymes in greater detail, testing the exact limits of their individual promiscuity and expanding the criteria for promiscuity to look at metals, other buffers and solvents. Additionally, the workflow explored he could be utilised and modified to suit other types of enzymes, in particular plastic degrading enzymes could be studied as they are a growing interest in protecting the environment. Such enzymes, if found to be suitably stable in varying conditions could potentially be used to clean up environmentally hazardous waste before it reaches our oceans.

Overall, the work presented here provides an insight into a number of realms surrounding proteomics and mass spectrometry. We have presented some new findings in the world of ion mobility spectrometry and found an avenue of interest in nitrilase enzyme promiscuity. There are a variety of new topics of research that can be followed up from what is presented here, some examples include: Proposing new calibrants for IMS, investigating the structures of peptide fragments utilising ion mobility, improving parameters for mobcal and studying the promiscuity further of enzymes discussed here or new enzymes using the techniques shown here. The author hopes that there other areas of research could be prompted that have not been predicted and that their work inspires others in new ways.

## References

- 
- <sup>1</sup> Roepstorff, P.; Fohlmann, J. Proposals for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**, *11*, 601.
- <sup>2</sup> Biemann, K. Contributions of Mass Spectrometry to Peptide and Protein Structure. *Biomed. Env. Mass Spectrom.* **1988**, *16*, 99-111.
- <sup>3</sup> Paizs, B.; Suhai, S. Fragmentation Pathways of Protonated Peptides. *Mass Spectrom. Rev.* **2005**, *24*, 508-548.
- <sup>4</sup> Somogyi, Á.; Wysocki, V.H.; Mayer, I. The effect of pretonation site on bond strengths in simple peptides: application of ab initio and modified neglect of differential overlap bond orders and modified neglect of differential overlap energy partitioning. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 704-717.
- <sup>5</sup> Csonka, I. P., Paizs, B., Lendvay, G. and Suhai, S., Proton mobility in protonated peptides: a joint molecular orbital and RRKM study. *Rapid Communications in Mass Spectrometry*, **2000**, *14*, 417-431
- <sup>6</sup> Boyd, R. and Somogyi, Á., The mobile proton hypothesis in fragmentation of protonated peptides: A perspective, *J. Am. Soc. Mass Spectrom.*, **2010**, *21*, 1275–1278.
- <sup>7</sup> Yalcin, T.; Khouw, C.; Csizmadia, I.G.; Peterson, M.R.; Harrison, A.G., Why are B Ions Stable Species in Peptide Mass Spectra? *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1165-1174.
- <sup>8</sup> Yalcin, T.; Csizmadia, I.G.; Peterson, M.R.; Harrison, A.G., The Structures and Fragmentation of B<sub>n</sub> (n>=3) Ions in Peptide Mass Spectra. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 233-242.
- <sup>9</sup> Bythell, B. J., Somogyi, Á. and Paizs, B., What is the Structure of b<sub>2</sub> Ions Generated from Doubly Protonated Tryptic Peptides? *J. Am. Soc. Mass Spectrom.*, **2009**, *20*, 618–624.
- <sup>10</sup> Harrison, A. G. To b or not to b: The Ongoing Saga of Peptide b Ions. *Mass Spectrom. Rev.* **2009**, *28*, 640-654.
- <sup>11</sup> Polfer, N.C.; Oomens, J.; Suhai, S.; Paizs, B. Spectroscopic and Theoretical Evidence for Oxazolone Ring Formation in Collision-Induced Dissociation of Peptides. *J. Am. Chem. Soc.* **2005**, *127*, 17154-17155.
- <sup>12</sup> Riba Garcia, I., Giles, K., Bateman, R. H. and Gaskell, S. J., Studies of Peptide a- and b-Type Fragment Ions Using Stable Isotope Labeling and Integrated Ion Mobility/Tandem Mass Spectrometry, *J. Am. Soc. Mass Spectrom.*, **2008**, *19*, 1781–1787.
- <sup>13</sup> Nelson C. R., Abutokaikah M. T., Harrison A. G. and Bythell B. J., Proton Mobility in b<sub>2</sub> Ion Formation and Fragmentation Reactions of Histidine-Containing Peptides, **2015**, *27*, 487-497.
- <sup>14</sup> Men, L. and Wang, Y., Further studies on the fragmentation of protonated ions of peptides containing aspartic acid, glutamic acid, cysteine sulfinic acid, and cysteine sulfonic acid, *Rapid Commun. Mass Spectrom.*, **2005**, *19*, 23–30.
- <sup>15</sup> Iyengar, S. S., Schlegel, H. B. and Voth, G. A., Atom-centered density matrix propagation (ADMP): Generalizations using Bohmian mechanics, *J. Phys. Chem. A*, **2003**, *107*, 7269–7277.
- <sup>16</sup> Schlegel, H. B., Iyengar, S. S., Li, X., Millam, J. M., Voth, G. A., Scuseria, G. E. and Frisch, M. J., Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. III. Comparison with Born–Oppenheimer dynamics, *J. Chem. Phys.*, **2002**, *117*, 8694–8704.
- <sup>17</sup> Khachaturyan, A., Semenovsovskaia, S. and Vainshtein, B., The thermodynamic approach to the structure analysis of crystals, *Acta Crystallogr. Sect. A*, **1981**, *37*, 742–754.

- 
- <sup>18</sup> Schrödinger, E., An Undulatory Theory of the Mechanics of Atoms and Molecules, *Phys. Rev.*, **1926**, 28, 1049–1070.
- <sup>19</sup> Attig, N., Binder, K., Grubmueller, H. and Kremer, K., *Ab Initio Molecular Dynamics: Theory and Implementation*, **2004**, vol. 22.
- <sup>20</sup> Foresman, J. B., *Ab Initio Techniques in Chemistry: Interpretation and Visualization*, ACS BOOKS, Washington, DC, **1996**.
- <sup>21</sup> James, B. and Frisch, A., *Exploring chemistry with electronic structure methods*, Gaussian, Incorporated, **1996**.
- <sup>22</sup> Chang, T. C., Habitz, P., Pittel, B. and Schwarz, W. H. E., Accuracy and limitations of the pseudopotential method, *Theor. Chim. Acta*, **1974**, 34, 263–275.
- <sup>23</sup> Dobson, J. F., Vignale, G. and Das, M. P., *Electronic density functional theory: recent progress and new directions*, Springer Science & Business Media, **2013**.
- <sup>24</sup> Mardirossian, N. and Head-Gordon, M., Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals, *Mol. Phys.*, **2017**, 115, 2315–2372.
- <sup>25</sup> Davidson, E. R. and Feller, D., Basis Set Selection for Molecular Calculations, *Chem. Rev.*, **1986**, 86, 681–696.
- <sup>26</sup> Tyers, M. and Mann, M., From genomics to proteomics, *Nature*, **2006**, 422, 193–197.
- <sup>27</sup> Forluvoft and Leyo, Simple diagram of double-stranded DNA (without text), *Wikipedia*, **2009**, [https://en.wikipedia.org/wiki/File:DNA\\_simple.svg](https://en.wikipedia.org/wiki/File:DNA_simple.svg)
- <sup>28</sup> Emw, Structure of the C3 protein. Based on PyMOL rendering of PDB 1c3d, *Wikimedia*, **2010**, [https://commons.wikimedia.org/wiki/File:Protein\\_C3\\_PDB\\_1c3d.png](https://commons.wikimedia.org/wiki/File:Protein_C3_PDB_1c3d.png)  
Note: changes made to this image are: reducing size and imbedding into another image.
- <sup>29</sup> Stretton, A. O. W., The first sequence: Fred Sanger and insulin, *Genetics*, **2002**, 162, 527–532.
- <sup>30</sup> NobelPrize.org, The Nobel Prize in Chemistry, *Nobel Media AB*, **1972**
- <sup>31</sup> Janin, J., Bahadur, R. P. and Chakrabarti, P., Protein-protein interaction and quaternary structure, *Q. Rev. Biophys.*, **2008**, 41, 133–180.
- <sup>32</sup> O’Farrell, P. H., High resolution two-dimensional electrophoresis of proteins., *J. Biol. Chem.*, **1975**, 250, 4007–4021.
- <sup>33</sup> Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues, *Humangenetik*, **1975**, 26, 231–243.
- <sup>34</sup> Lottspeich, F., in *Proteomics: Methods and Protocols*, eds. J. Reinders and A. Sickmann, Humana Press, Totowa, NJ, **2009**, pp. 3–10.
- <sup>35</sup> Aebersold, R. and Mann, M., Mass-spec proteomics, *Nature*.
- <sup>36</sup> Kuharev, J., Navarro, P., Distler, U., Jahn, O. and Tenzer, S., In-depth evaluation of software tools for data-independent acquisition based label-free quantification, *Proteomics*, **2015**, 15, 3140–3151.
- <sup>37</sup> J. Mitchell Wells; Scott A. McLuckey. Collision-Induced Dissociation (CID) of Peptides and Proteins. *Academic Press*. **2005**, 402, 148–185.
- <sup>38</sup> Zieske, L. R., A perspective on the use of iTRAQ<sup>TM</sup> reagent technology for protein complex and profiling studies, *J. Exp. Bot.*, **2006**, 57, 1501–1508.

- 
- <sup>39</sup> Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B., Quantitative mass spectrometry in proteomics: a critical review, *Anal. Bioanal. Chem.*, **2007**, 389, 1017–1031.
- <sup>40</sup> Nikolov, M., Schmidt, C., & Urlaub, H. Quantitative mass spectrometry-based proteomics: an overview. *Quantitative methods in proteomics*, **2012**, 85-100.
- <sup>41</sup> Fitzgerald, M. C., Parr, G. R. and Smith, L. M., Basic matrixes for the matrix-assisted laser desorption/ionization mass spectrometry of proteins and oligonucleotides, *Anal. Chem.*, **1993**, 65, 3204-3211
- <sup>42</sup> Knochenmuss, R., Ion formation mechanisms in UV-MALDI, *Analyst*, **2006**, 131, 966-986
- <sup>43</sup> Ishihama, Y., Proteomic LC–MS systems using nanoscale liquid chromatography with tandem mass spectrometry, *Journal of Chromatography A*, **2005**, 1067, 73-83.
- <sup>44</sup> Yamashita, M. and Fenn, J. B., Negative ion production with the electrospray ion source, *J. Phys. Chem.*, **1984**, 88, 4671–4675.
- <sup>45</sup> Fenn, J. B., Tanaka, K. Press Release: The Nobel Prize in Chemistry 2002, *The Nobel Foundation*, **2002**.
- <sup>46</sup> Pozniak, B. P. and Cole, R. B., Current measurements within the electrospray emitter, *J. Am. Soc. Mass Spectrom.*, **2007**, 18, 737-748
- <sup>47</sup> Taylor, G., Disintegration of water drops in an electric field, *Proc R Soc Lond A Math Phys Sci*, **1964**, 280, 383.
- <sup>48</sup> Fernandez de, I. M., The fluid dynamics of Taylor cones, *Annu. Rev. Fluid Mech.*, **2007**, 39, 217-243.
- <sup>49</sup> Li, K., Tu, H. and Ray, A. K., Charge limits on droplets during evaporation, *Langmuir*, **2005**, 21, 3786-3794
- <sup>50</sup> Kebarle, P. and Verkerk, U. H., Electrospray: from ions in solution to ions in the gas phase, *Mass Spectrom. Rev.*, **2009**, 28, 898-917
- <sup>51</sup> Marginean, I., Page, J.S., Tolmachev, A.V., Tang, K., Smith, R.D. Achieving 50% ionization efficiency in subambient pressure ionization with nanoelectrospray. *Anal Chem.* **2010**, 82(22), 9344–9349
- <sup>52</sup> America, A. H. P. and Cordewener, J. H. G., Comparative LC-MS: A landscape of peaks and valleys, *Proteomics*, **2008**, 8, 731–749.
- <sup>53</sup> Kamal, S., Sahani Kamal, C. and Sharad, W., Step-up in liquid chromatography from HPLC to UPLC: A comparative and comprehensive review, *Pharma Innov. J.*, **2018**, 7, 342–347.
- <sup>54</sup> El Jadid, S., Touahni, R. and Moussa, A., Strategies of Peptide Identification Using Tandem Mass Spectrometry, *Assoc. Comput. Mach.*, **2019**, 4, 1–5.
- <sup>55</sup> Dass, C., Tandem Mass Spectrometry in *Fundamentals of Contemporary Mass Spectrometry*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2006**, pp. 119–150.
- <sup>56</sup> Lambropoulou, D. A. and Ofrydopoulou, A., High Resolution Mass Spectrometry in Food Analysis, *Fingerprinting Tech. Food Authentication Traceability*, **2018**
- <sup>57</sup> Lakshmanan, R. and Loo, J. A., Top-down protein identification using a time-of-flight mass spectrometer and data independent acquisition, *Int. J. Mass Spectrom.*, **2019**, 435, 136–144.
- <sup>58</sup> Nys, G., Cobraiville, G., Servais, A.-C. and Fillet, M., (Nano)-LC coupled to Ion Mobility Q-TOF for an improved sensitivity and proteome coverage. **2018**
- <sup>59</sup> Thomson, B. A., Douglas, D. J., Corr, J. J., Hager, J. W. and Jolliffe, C. L., Improved Collisionally Activated Dissociation Efficiency and Mass Resolution on a Triple Quadrupole Mass Spectrometer System, *Anal. Chem.*, **1995**, 67, 1696–1704.

- 
- <sup>60</sup> Yost, R. A. and Enke, C. G., Selected Ion Fragmentation with a Tandem Quadrupole Mass Spectrometer, *J. Am. Chem. Soc.*, **1978**, *100*, 2274–2275.
- <sup>61</sup> Anderson, L. and Hunter, C. L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins, *Mol. Cell. Proteomics*, **2006**, *5*, 573–588.
- <sup>62</sup> March, R. E. and Todd, J. F., *Quadrupole ion trap mass spectrometry*, John Wiley & Sons, **2005**, vol. 165.
- <sup>63</sup> March, R. E., An introduction to quadrupole ion trap mass spectrometry, *J. Mass Spectrom.*, **1997**, *32*, 351–369.
- <sup>64</sup> Schwartz, J. C., Senko, M. W. and Syka, J. E. P., A two-dimensional quadrupole ion trap mass spectrometer, *J. Am. Soc. Mass Spectrom.*, **2002**, *13*, 659–669.
- <sup>65</sup> Weinstock, R., On a fallacious proof of Earnshaw's theorem, *Am. J. Phys.*, **1976**, *44*, 392–393.
- <sup>66</sup> Bate, D. J., Dholakia, K., Thompson, R. C. and Wilson, D. C., Ion oscillation frequencies in a combined trap, *J. Mod. Opt.*, **1992**, *39*, 305–316.
- <sup>67</sup> Lanucara, F; Holman, S.W; Gray, C.J; Evers, C.E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nature Chemistry*, **2014**, *6*, 281-294
- <sup>68</sup> Laphorn, C; Pullen, F; Chowdhry, B.Z., Ion Mobility Spectrometry-Mass Spectrometry (IMS-MS) of Small Molecules: Separating and Assigning Structures to Ions. *Mass Spectrometry Reviews*, **2013**, *32*, 43-71
- <sup>69</sup> Fenn, L.S; Kilman, M; Mahsut, A; Zhao, S.R; McLean, J.A., Characterizing ion mobility-mass spectrometry conformation space for the analysis of complex biological samples. *Anal. and Bioanal. Chem.* **2009**, *349*, 235-244
- <sup>70</sup> McLean, J.A., The mass-mobility correlation redux: The conformational landscape of anhydrous biomolecules. *J. Am. Soc. Mass Spectrom* **2009**, *10*, 1775-1781
- <sup>71</sup> Karpas, Z; Cohen, M.J; Stimac, R.M; Wernlund, R.F On the effects of structure and charge distribution on the mobility of ions. *Int. J. Mass Spec. and Ion Processes*, **1986**, *74*, 153-159
- <sup>72</sup> Shvartsburg, A. A. and Smith, R. D., Fundamentals of traveling wave ion mobility spectrometry, *Anal. Chem.*, **2008**, *80*, 9689-9699
- <sup>73</sup> Smith, D., Knapman, T., Campuzano, I., Malham, R., Berryman, J., Radford, S. and Ashcroft, A., Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies, *European Journal of Mass Spectrometry*, **2009**, *15*, 113-130
- <sup>74</sup> Paglia, G., Williams, J. P., Menikarachchi, L., Thompson, J. W., Tyldesley-Worster, R., Halldórsson, S., Rolfsson, O., Moseley, A., Grant, D., Langridge, J., Palsson, B. O. and Astarita, G., Ion mobility derived collision cross sections to support metabolomics applications, *Anal. Chem.*, **2014**, *86*, 3985–3993.
- <sup>75</sup> Ahmed, A., Cho, Y. J., No, M., Koh, J., Tomczyk, N., Giles, K., Yoo, J. S. and Kim, S., Application of the Mason–Schamp Equation and Ion Mobility Mass Spectrometry To Identify Structurally Related Compounds in Crude Oil, *Anal. Chem.*, **2011**, *83*, 77–83.
- <sup>76</sup> Guevremont, R., High-field asymmetric waveform ion mobility spectrometry (FAIMS), *Can. J. Anal. Sci. Spectrosc.*, **2004**, *49*, 105–113.
- <sup>77</sup> Ahmed, A., Cho, Y. J., No, M., Koh, J., Tomczyk, N., Giles, K., Yoo, J. S. and Kim, S., Application of the Mason–Schamp Equation and Ion Mobility Mass Spectrometry To Identify Structurally Related Compounds in Crude Oil, *Anal. Chem.*, **2011**, *83*, 77–83.

- 
- <sup>78</sup> Giles, K., Wildgoose, J., Pringle, S., Langridge, D., Nixon, P., Garside, J. and Carney, P., Characterising a t-wave enabled multi-pass cyclic ion mobility separator, *63th ASMS Conf. Mass Spectrom. Allied Top.*, **2015**, 72, 2015.
- <sup>79</sup> Hongli, L; Giles, K; Bendiak, B; Kaplan, K; Siems, W.F; Herbert, H.H Resolving structural isomers of monosaccharide methyl glycosides using drift tube and traveling wave ion mobility mass spectrometry. *Anal. Chem.* **2012**, 84, 3231-3239
- <sup>80</sup> Campuzano, I., Bush, M. F., Robinson, C. V., Beaumont, C., Richardson, K., Kim, H. and Kim, H. I., Structural characterization of drug-like compounds by ion mobility mass spectrometry: Comparison of theoretical and experimentally derived nitrogen collision cross sections, *Anal. Chem.*, **2012**, 84, 1026–1033.
- <sup>81</sup> Wyttenbach, T., Von Helden, G., Batka, J. J., Carlat, D. and Bowers, M. T., Effect of the long-range potential on ion mobility measurements, *J. Am. Soc. Mass Spectrom.*, **1997**, 8, 275–282.
- <sup>82</sup> Shvartsburg, A. A. and Jarrold, M. F., An exact hard-spheres scattering model for the mobilities of polyatomic ions, *Chem. Phys. Lett.*, , DOI:10.1016/0009-2614(96)00941-4.
- <sup>83</sup> Mesleh, M.F; Hunter, J.M; Shvartsburg, A.A; Schatz, G.C; Jarrold, M.F Structural Information from Ion Mobility Measurements: Effects of Long-Range Potential *J. Phys. Chem.* **1996**, 100, 16082-16086
- <sup>84</sup> Takaya, K., Kaneko, T., Tanuma, H., Nishide, T., Sugiyama, H., Nakano, N., Nagashima, H. and Seto, Y. Model calculation for ion mobility in air using the MOBCAL program. *International Journal for Ion Mobility Spectrometry.* **2016**, 19(4), 227-232
- <sup>85</sup> Bisswanger, H., Enzyme Kinetics: Principles and Methods, *John Wiley & Sons*, **2017**.
- <sup>86</sup> Rivera-Sagredo A, Cañada FJ, Nieto O, Jimenez-Barbero J, M.-L. M., Substrate specificity of small-intestinal lactase Assessment of the role of the substrate hydroxyl groups, *Eur J Biochem*, **1992**, 422, 415–422.
- <sup>87</sup> Khersonsky, O., Roodveldt, C. and Tawfik, D. S., Enzyme promiscuity : evolutionary and mechanistic aspects, *Curr. Opin. Chem. Biol.*, **2006**, 10, 498–508.
- <sup>88</sup> Bachan Upadhyay, L. S. and Verma, N., Enzyme Inhibition Based Biosensors: A Review, *Anal. Lett.*, **2012**, 46, 225–241.
- <sup>89</sup> Suprun, E. V., Shumyantseva, V. V. and Archakov, A. I., Protein Electrochemistry: Application in Medicine. A Review, *Electrochim. Acta*, **2014**, 140, 72–84.
- <sup>90</sup> DREYER, W. J. and NEURATH, H., The activation of chymotrypsinogen; isolation and identification of a peptide liberated during activation., *J. Biol. Chem.*, **1955**, 217, 527–39.
- <sup>91</sup> Demirjian, D. C., Morís-Varas, F. and Cassidy, C. S., Enzymes from extremophiles, *Curr. Opin. Chem. Biol.*, **2001**, 5, 144–151.
- <sup>92</sup> Ventosa, A., Nieto, J. J. and Oren, A., Biology of Moderately Halophilic Aerobic Bacteria, *Microbiol. Mol. Biol. Rev.*, **1998**, 62, 504 LP – 544.
- <sup>93</sup> Norris, P. R., Burton, N. P. and Foulis, N. A. M., Acidophiles in bioreactor mineral processing, *Extremophiles*, **2000**, 4, 71–76.
- <sup>94</sup> Horikoshi, K., Alkaliphiles: Some Applications of Their Products for Biotechnology, *Microbiol. Mol. Biol. Rev.*, **1999**, 63, 735 LP – 750.
- <sup>95</sup> Daly, M. J., Engineering radiation-resistant bacteria for environmental biotechnology, *Curr. Opin. Biotechnol.*, **2000**, 11, 280–285.



- <sup>96</sup> Graham, D., Pereira, R., Barfield, D. and Cowan, D., Nitrile biotransformations using free and immobilized cells of a thermophilic *Bacillus* spp., *Enzyme Microb. Technol.*, **2000**, 26, 368–373.
- <sup>97</sup> Ryu, K., Kim, J. and Dordick, J. S., Catalytic properties and potential of an extracellular protease from an extreme halophile, *Enzyme Microb. Technol.*, **1994**, 16, 266–275.
- <sup>98</sup> Schmidt-Dannert, C., Rúa, M. L., Atomi, H. and Schmid, R. D., Thermoalkalophilic lipase of *Bacillus thermocatenulatus*. I. Molecular cloning, nucleotide sequence, purification and some properties, *Biochim. Biophys. Acta - Lipids Lipid Metab.*, **1996**, 1301, 105–114.
- <sup>99</sup> Holmes, M. L., Scopes, R. K., Moritz, R. L., Simpson, R. J., Englert, C., Pfeifer, F. and Dyall-Smith, M. L., Purification and analysis of an extremely halophilic beta-galactosidase from *Haloferax alicantei*, *Biochim. Biophys. Acta*, **1997**, 1337, 276–286.
- <sup>100</sup> Giles, K., Travelling wave ion mobility, *Int. J. Ion Mobil. Spectrom.*, **2013**, 16, 69.
- <sup>101</sup> E. Atherton, R.C. Sheppard, In Solid-Phase Peptide Synthesis: A Practical Approach IRL Press at Oxford University Press, *Oxford*, **1989**
- <sup>102</sup> Gaussian 09, Revision D.01, 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.
- <sup>103</sup> Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K., M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I. G.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollmann, P. A. In: AMBER 99, University of California: San Francisco, **1999**.
- <sup>104</sup> Hines, K. M., May, J. C., McLean, J. A. and Xu, L., Evaluation of Collision Cross Section Calibrants for Structural Analysis of Lipids by Traveling Wave Ion Mobility-Mass Spectrometry, *Anal. Chem.*, **2016**, 88, 7329–7336.
- <sup>105</sup> Duez, Q., Chiot, F., Liénard, R., Josse, T., Choi, C. M., Coulembier, O., Dugourd, P., Cornil, J., Gerbaux, P. and De Winter, J., Polymers for Traveling Wave Ion Mobility Spectrometry Calibration, *J. Am. Soc. Mass Spectrom.*, **2017**, 28, 2483–2491.
- <sup>106</sup> Biemann, K., Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation, *Meth. Enzymol.*, **1990**, 193, 455–479.
- <sup>107</sup> Tsang, C. W. and Harrison, A. G., Chemical Ionization of Amino Acids, *J. Am. Chem. Soc.*, **1976**, 98, 1301–1308.
- <sup>108</sup> Perkins, B. R., Chamot-Rooke, J., Yoon, S. H., Gucinski, A. C., Somogyi, Á. and Wysocki, V. H., Evidence of diketopiperazine and oxazolone structures for HA b 2+ ion, *J. Am. Chem. Soc.*, **2009**, 131, 17528–17529.

- 
- <sup>109</sup> Bythell, B. J., Maître, P. and Paizs, B., Cyclization and Rearrangement Reactions of an Fragment Ions of Protonated Peptides, *J. Am. Chem. Soc.*, **2010**, *132*, 14766–14779.
- <sup>110</sup> Oomens, J., Young, S., Molesworth, S. and Stipdonk, M. van, Spectroscopic Evidence for an Oxazolone Structure of the b<sub>2</sub> Fragment Ion from Protonated Tri-alanine, *J Am Soc Mass Spectrom*, **2009**, *20*, 334–339.
- <sup>111</sup> Grewal, R. N., El Aribi, H., Harrison, A. G., Siu, K. W. M. and Hopkinson, A. C., Fragmentation of Protonated Tripeptides: The Proline Effect Revisited, *J. Phys. Chem. B*, **2004**, *108*, 4899–4908.
- <sup>112</sup> Smith, L. L., Herrmann, K. A. and Wysocki, V. H., Investigation of gas phase ion structure for proline-containing b<sub>2</sub> ion, *J. Am. Soc. Mass Spectrom.*, **2006**, *17*, 20–28.
- <sup>113</sup> Colby, S. M., Thomas, D. G., Nuñez, J. R., Baxter, D. J., Glaesemann, K. R., Brown, J. M., Pirrung, M. A., Govind, N., Teegarden, J. G., Metz, T. O. and Renslow, R. S., ISiCLE: A Quantum Chemistry Pipeline for Establishing in Silico Collision Cross Section Libraries, *Anal. Chem.*, **2019**, *91*, 4346–4356.
- <sup>114</sup> Gupta, N., Balomajumder, C. and Agarwal, V. K., Enzymatic mechanism and biochemistry for cyanide degradation: A review, *J. Hazard. Mater.*, **2010**, *176*, 1–13.
- <sup>115</sup> Ezzi, M.; Lynch, J.; Biodegradation of cyanide by *Trichoderma* spp. and *Fusarium* spp., *Enzyme Microb. Technol.*, **2005**, *36*, 849–854.
- <sup>116</sup> Fernandes, B. C. M., Mateo, C., Kiziak, C., Chmura, A., Wacker, J., Van Rantwijk, F., Stolz, A. and Sheldon, R. A., Nitrile hydratase activity of a recombinant nitrilase, *Adv. Synth. Catal.*, **2006**, *348*, 2597–2603.
- <sup>117</sup> Andexer, J. N., Langermann, J. V., Kragl, U. and Pohl, M., How to overcome limitations in biotechnological processes - examples from hydroxynitrile lyase applications, *Trends Biotechnol.*, **2009**, *27*, 599–607.
- <sup>118</sup> Liu, Z.-Q., Zhou, M., Zhang, X.-H., Xu, J.-M., Xue, Y.-P. and Zheng, Y.-G., Biosynthesis of Iminodiacetic Acid from Iminodiacetonitrile by Immobilized Recombinant *Escherichia coli* Harboring Nitrilase, *J. Mol. Microbiol. Biotechnol.*, **2012**, *22*, 35–47.
- <sup>119</sup> Gong, J. S., Lu, Z. M., Li, H., Shi, J. S., Zhou, Z. M. and Xu, Z. H., Nitrilases in nitrile biocatalysis: Recent progress and forthcoming research, *Microb. Cell Fact.*, **2012**, *11*, 1.
- <sup>120</sup> Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y., Takahashi, S., Sato, M. and Tsukihara, T., Crystal structure of N-carbamyl-d-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases, *Structure*, **2000**, *8*, 729–738.
- <sup>121</sup> Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B., Industrial biocatalysis today and tomorrow, *Nature*, **2001**, *409*, 258–268.
- <sup>122</sup> Stepankova, V., Bidmanova, S., Koudelakova, T., Prokop, Z., Chaloupkova, R. and Damborsky, J., Strategies for stabilization of enzymes in organic solvents, *ACS Catal.*, **2013**, *3*, 2823–2836.
- <sup>123</sup> Doukyu, N. and Ogino, H., Organic solvent-tolerant enzymes, *Biochem. Eng. J.*, **2010**, *48*, 270–282.
- <sup>124</sup> Ru, M. T., Dordick, J. S., Reimer, J. A. and Clark, D. S., Optimizing the salt-induced activation of enzymes in organic solvents: Effects of lyophilization time and water content, *Biotechnol. Bioeng.*, **1999**, *63*, 233–241.
- <sup>125</sup> Karan, R., Capes, M. D. and DasSarma, S., Function and biotechnology of extremophilic enzymes in low water activity, *Aquat. Biosyst.*, **2012**, *8*, 4.
- <sup>126</sup> Wu, Z., Gao, W., Phelps, M. A., Wu, D., Miller, D. D. and Dalton, J. T., Favorable effects of weak acids on negative-ion electrospray ionization mass spectrometry, *Anal. Chem.*, **2004**, *76*, 839–847.

## Appendix

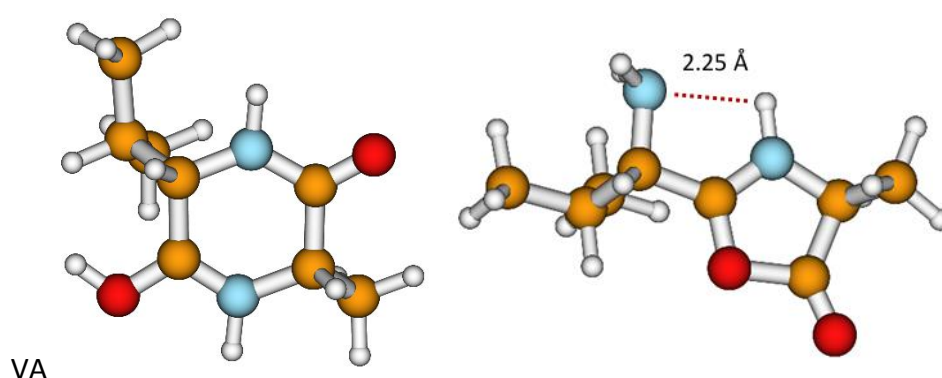
**Table 15. The computed minimum energies of the minimum structures of each  $b_2$  ion, with relative values between macrocyclic and oxazolone structures.** Energies were calculated at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Collisional cross sections were calculated using the helium drift gas parameters of Mobcal.

<i>ion</i>	<i>MacroCyclic CCS He (<math>\text{\AA}^2</math>)</i>	<i>Oxazolone CCS He (<math>\text{\AA}^2</math>)</i>	<i>MacroCyclic Energy (Hartree)</i>	<i>Oxazolone Energy (Hartree)</i>	<i>MacroCyclic Relative Energy (kcal/mol)</i>	<i>Oxazolone Relative Energy (kcal/mol)</i>
VA	70.5	73.8	-573.380122	-573.378103	0.00	1.27
VF	93.6	97.1	-804.360620	-804.351430	0.00	5.77
VG	65.7	68.3	-534.087539	-534.083843	0.00	2.32
VH	90	89.6	-798.401219	-798.369658	0.00	19.80
VL	85.9	90	-691.238413	-691.236809	0.00	1.01
VP	77	78.8	-650.769947	-650.751531	0.00	11.56
VQ	85.6	90.2	-781.371707	-781.352467	0.00	12.07
VS	72.5	75.8	-648.594979	-648.580351	0.00	9.18
VT	76	81.1	-687.886868	-687.873681	0.00	8.27
VY	95.9	97	-879.574966	-879.564560	0.00	6.53
DA	69	71.7	-683.374000	-683.345653	0.00	17.79
PA	67.6	70.4	-572.197547	-572.179738	0.00	11.18
QA	77.4	80.9	-702.799177	-702.795027	0.00	2.60
SA	63.5	66.3	-570.021838	-570.013770	0.00	5.06
CT	75.9	78.6	-1007.498921	-1007.480429	0.00	11.60
ET	80.2	84.3	-837.166607	-837.143787	0.00	14.32
TC	75.8	78.9	-1007.493545	-1007.48203	0.00	7.23
FC	92.6	95.2	-1123.971308	-1123.957712	0.00	8.53
GC	63.9	66.7	-853.702399	-853.693960	0.00	5.30
MR	107.8	105.2	-1254.314794	-1254.294747	0.00	12.58
YG	81.8	84.6	-761.655213	-761.698027	26.87	0.00
LA	77.7	80.6	-612.665033	-612.663979	0.00	0.66
FA	84.6	85.7	-725.787744	-725.778045	0.00	6.09
FL	100.7	101.9	-843.645555	-843.636569	0.00	5.64
AC	69.1	71.7	-892.995255	-892.985348	0.00	6.22
AD	69	71.6	-683.374000	-683.358401	0.00	9.79
AF	84.6	87.6	-725.787744	-725.779029	0.00	5.47
AL	77.7	80.7	-612.665033	-612.664283	0.00	0.47
AP	67.6	69.5	-572.197547	-572.179063	0.00	11.60
AQ	77.4	80.7	-702.799177	-702.780578	0.00	11.67
AS	63.5	66.4	-570.021838	-570.007989	0.00	8.69
AV	70.5	74.2	-573.380122	-573.379680	0.00	0.28
FV	94.2	95.4	-804.357210	-804.352104	0.00	3.20
GV	65.7	69	-534.087539	-534.088524	0.62	0.00
LV	85.9	90.2	-691.238413	-691.236089	0.00	1.46

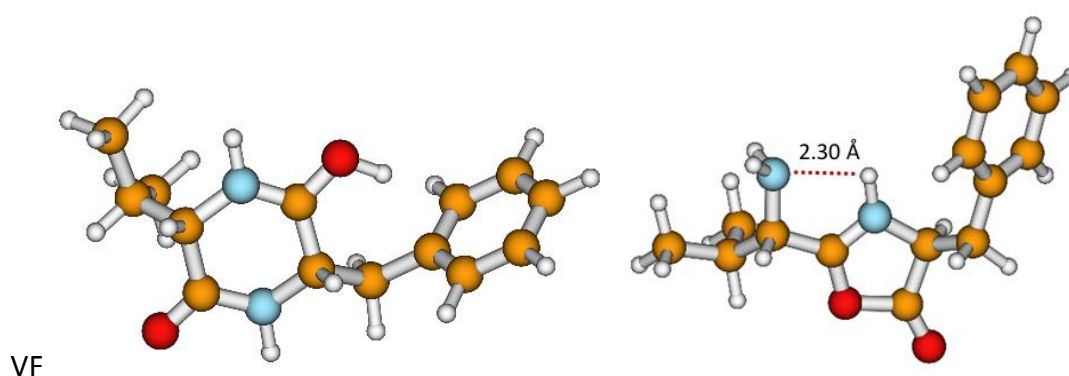
PV	77	80	-650.769948	-650.761025	0.00	5.6
SV	72.5	76.1	-648.594979	-648.583482	0.00	7.21
TV	76	81	-687.888537	-687.876498	0.00	7.55
VV	78.9	83.5	-651.953061	-651.952234	0.00	0.52
WV	104.2	103.2	-935.909400	-935.897938	0.00	7.19
YV	95.9	98	-879.574965	-879.565386	0.00	6.01
DL	84.9	88.2	-801.231723	-801.214385	0.00	10.88
FL	100.7	101.9	-843.645555	-843.636569	0.00	5.64
GL	72.5	75.4	-573.371141	-573.373096	1.23	0.00
HL	97.3	100	-837.684656	-837.654921	0.00	18.66
LL	92.4	97	-730.522329	-730.522843	0.32	0.00
PL	83.1	87.4	-690.055896	-690.044817	0.00	6.95
QL	93.1	97.5	-820.656315	-820.641824	0.00	9.09
SL	79.3	82.6	-687.879847	-687.868068	0.00	7.39
TL	83.7	87.7	-727.173437	-727.160992	0.00	7.81
AG	57.3	58.8	-455.513579	-455.511124	0.00	1.54
DG	63.9	65.6	-644.081516	-644.066483	0.00	9.43
FG	79.5	80.2	-686.49504	-686.483867	0.00	7.01
GG	52.4	53.7	-416.220234	-416.219564	0.00	0.42
HG	75.9	74.6	-680.538633	-680.495302	0.00	27.19
PG	63.2	64.9	-532.905577	-532.891366	0.00	8.92
QG	72.5	75.5	-663.508152	-663.493507	0.00	9.19
SG	58.4	60.8	-530.729077	-530.715129	0.00	8.75
TG	62.8	65.8	-570.022885	-570.008196	0.00	9.22
GA	57.3	59.1	-455.513579	-455.514153	0.36	0.00
GD	63.9	66.9	-644.081516	-644.067181	0.00	9.00
GF	79.5	82.5	-686.49504	-686.487929	0.00	4.46
GH	75.9	75.2	-680.538633	-680.506385	0.00	20.24
GP	63.2	64.4	-532.905577	-532.888702	0.00	10.59
GQ	72.5	75.6	-663.508152	-663.4843	0.00	14.97
GS	58.4	61.2	-530.729077	-530.716714	0.00	7.76
GT	62.8	66.5	-570.022885	-570.010101	0.00	8.02
PF	90.7	91	-803.177366	-803.159614	0.00	11.14
QF	94.4	101.6	-933.772994	-933.756077	0.00	10.62
DF	86.4	94.7	-914.348497	-914.333732	0.00	9.27
FD	86.4	93.2	-914.348497	-914.330652	0.00	11.20
HF	104.7	105.1	-950.801514	-950.770935	0.00	19.19
SF	86.8	89.7	-800.996848	-800.982662	0.00	8.90
TF	91.2	94.7	-840.290254	-840.275598	0.00	9.20
FF	108.2	108.3	-956.762387	-956.751186	0.00	7.03
FH	104.7	104	-950.801514	-950.768642	0.00	20.63
FP	90.7	90.4	-803.177366	-803.151294	0.00	16.36
FQ	94.4	99.7	-933.772994	-933.751754	0.00	13.33
FS	86.8	87.9	-800.996848	-800.980141	0.00	10.48
FT	91.2	93.5	-840.290254	-840.273343	0.00	10.61

LD	84.9	88.1	-801.231723	-801.216409	0.00	9.61
LF	100.7	98.6	-843.645555	-843.630546	0.00	9.42
LG	72.5	75.1	-573.371141	-573.369812	0.00	0.83
LH	97.3	98.8	-837.684656	-837.65524	0.00	18.46
LP	83.1	86	-690.055896	-690.037041	0.00	11.83
LQ	93.1	97.3	-820.656315	-820.638187	0.00	11.38
LS	79.3	82.9	-687.879847	-687.866359	0.00	8.46
LT	83.7	88.1	-727.173437	-727.159515	0.00	8.74

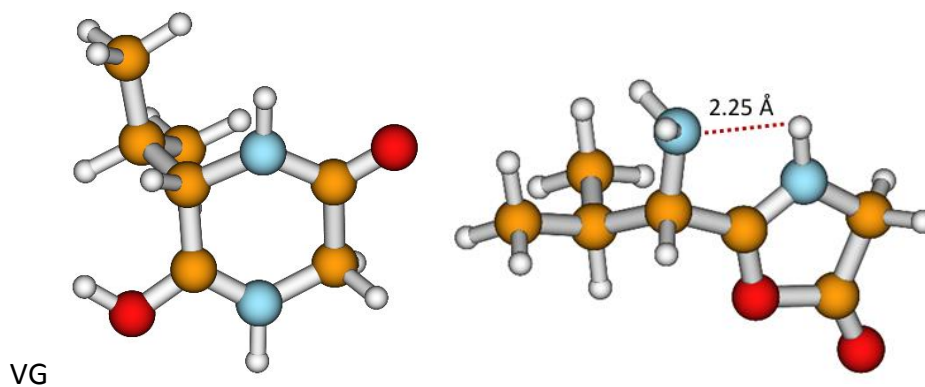
All minimum energy structures for computed macrocyclic and oxazolone structures.



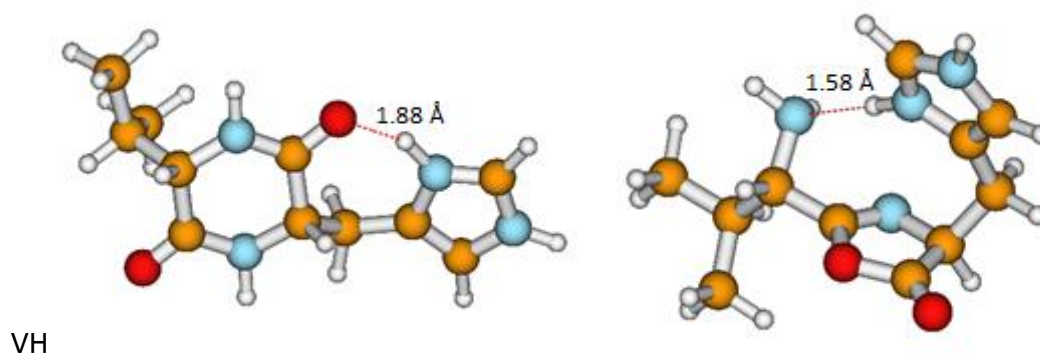
**Appendix Figure 1. The minimum energy structures of VA macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 70.5 Å<sup>2</sup> for the macrocyclic geometry and 73.8 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VA  $b_2$  ion was 71.93 Å<sup>2</sup>.



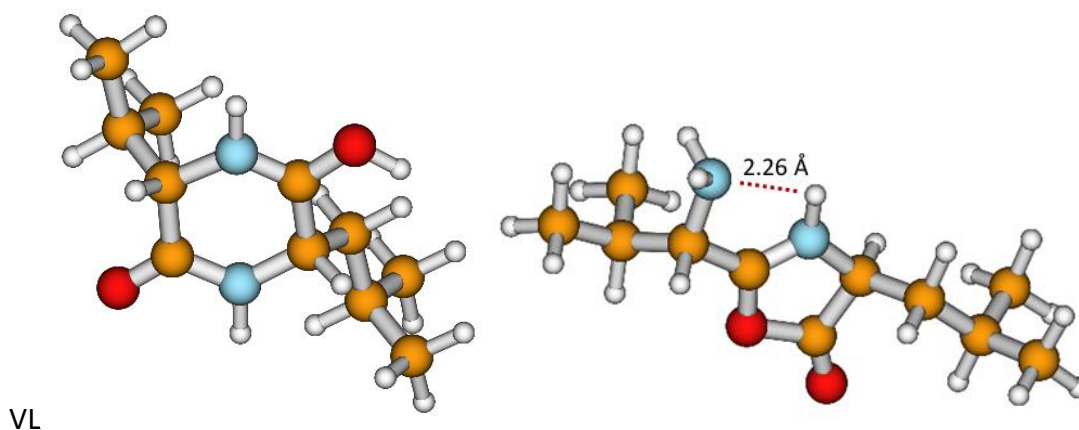
**Appendix Figure 2. The minimum energy structures of VF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 93.6 Å<sup>2</sup> for the macrocyclic geometry and 97.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VF  $b_2$  ion was 95.2 Å<sup>2</sup>.



**Appendix Figure 3. The minimum energy structures of VG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 65.7 Å<sup>2</sup> for the macrocyclic geometry and 68.3 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VG  $b_2$  ion was 68.01 Å<sup>2</sup>.

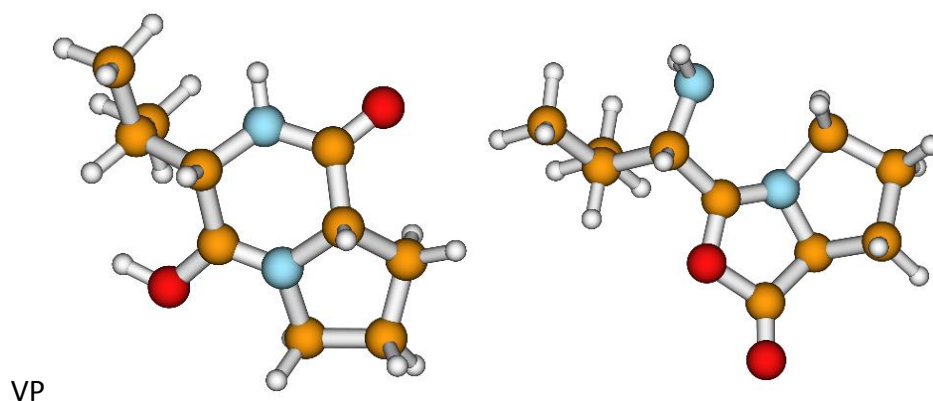


**Appendix Figure 4. The minimum energy structures of VH macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 90.0 Å<sup>2</sup> for the macrocyclic geometry and 89.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VH  $b_2$  ion was 90.9 Å<sup>2</sup>.

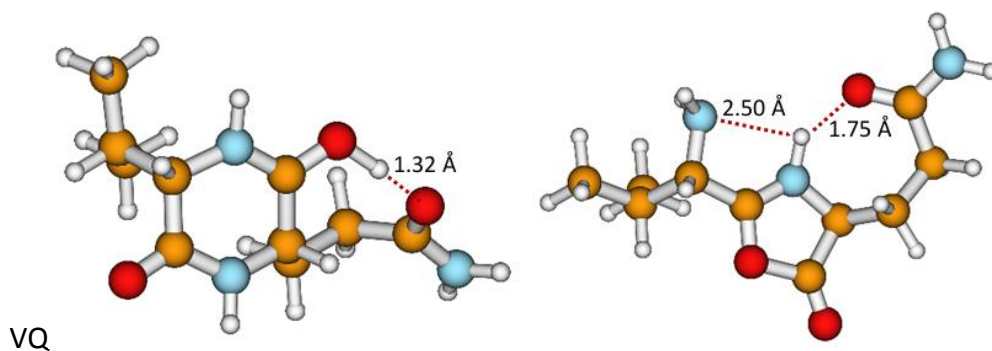


**Appendix Figure 5. The minimum energy structures of VL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was

used to generate a CCS value of  $85.9 \text{ \AA}^2$  for the macrocyclic geometry and  $90.0 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VL  $b_2$  ion was  $87.93 \text{ \AA}^2$ .

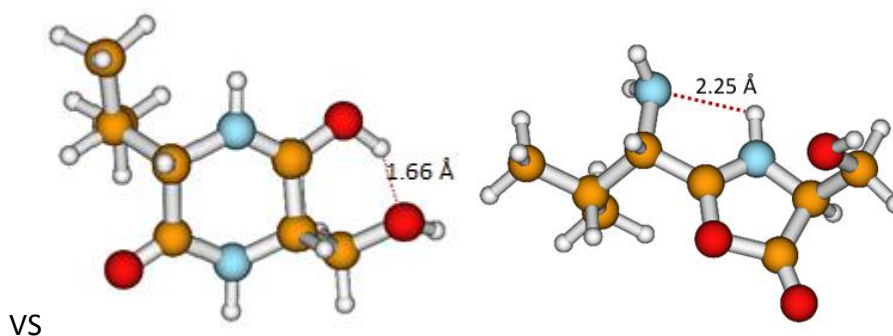


**Appendix Figure 6. The minimum energy structures of VP macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $77.0 \text{ \AA}^2$  for the macrocyclic geometry and  $78.8 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VP  $b_2$  ion was  $80.74 \text{ \AA}^2$ .

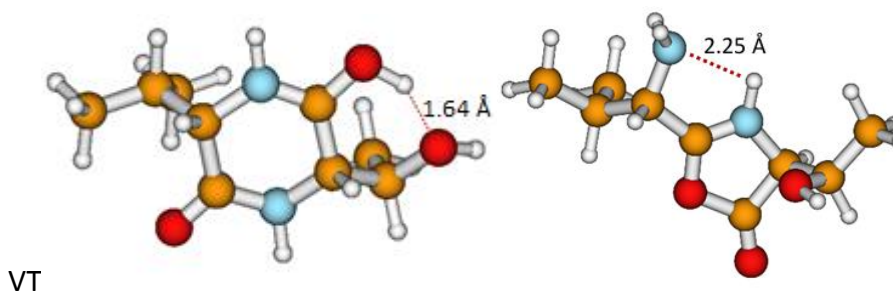


**Appendix Figure 7. The minimum energy structures of VQ macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $85.6 \text{ \AA}^2$  for the macrocyclic geometry and  $90.2 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VQ  $b_2$  ion was  $88.8 \text{ \AA}^2$ .

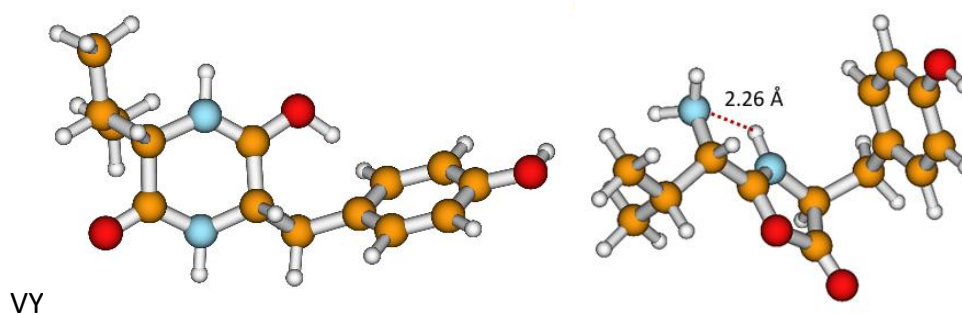




**Appendix Figure 8. The minimum energy structures of VS macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 72.5 Å<sup>2</sup> for the macrocyclic geometry and 75.8 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VS  $b_2$  ion was 80.05 Å<sup>2</sup>.

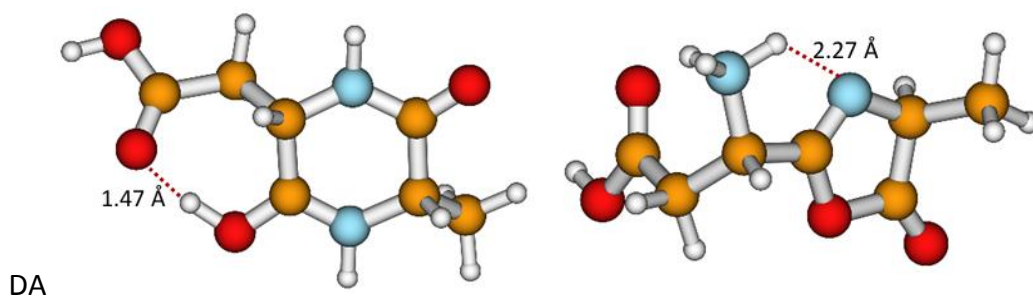


**Appendix Figure 9. The minimum energy structures of VT macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 76.0 Å<sup>2</sup> for the macrocyclic geometry and 81.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VT  $b_2$  ion was 81.96 Å<sup>2</sup>.

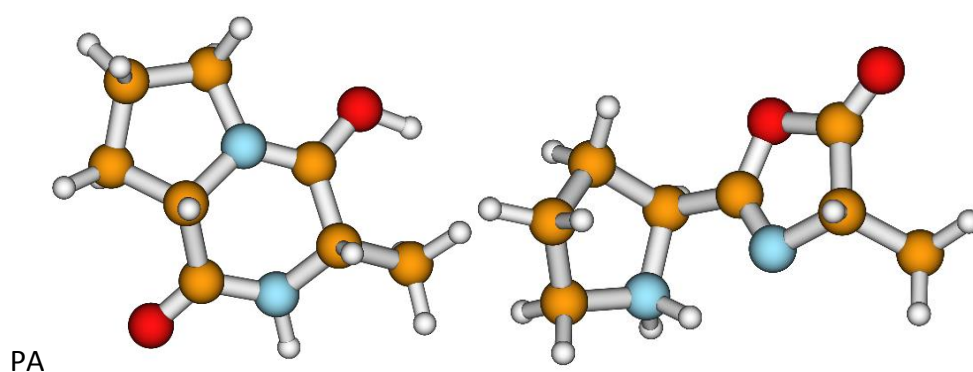


**Appendix Figure 10. The minimum energy structures of VY macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 95.9 Å<sup>2</sup> for the macrocyclic geometry and 97.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VY  $b_2$  ion was 97.96 Å<sup>2</sup>.

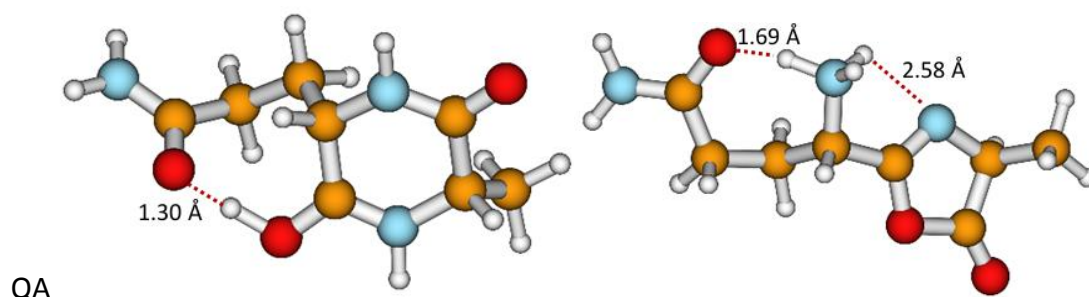




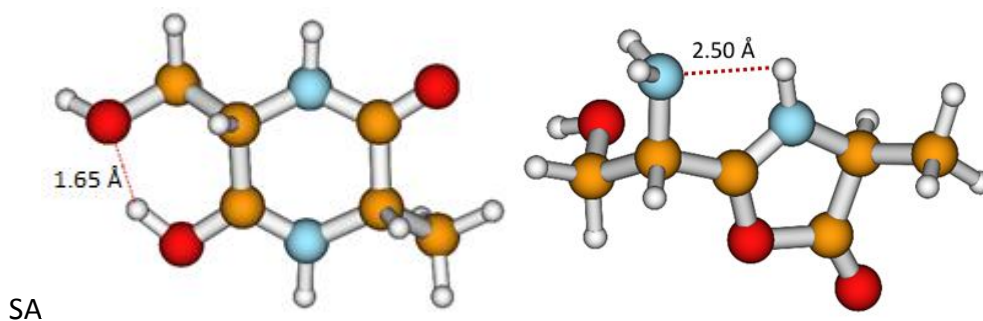
**Appendix Figure 11. The minimum energy structures of DA macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $69.0 \text{ \AA}^2$  for the macrocyclic geometry and  $71.7 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a DA  $b_2$  ion was not obtained.



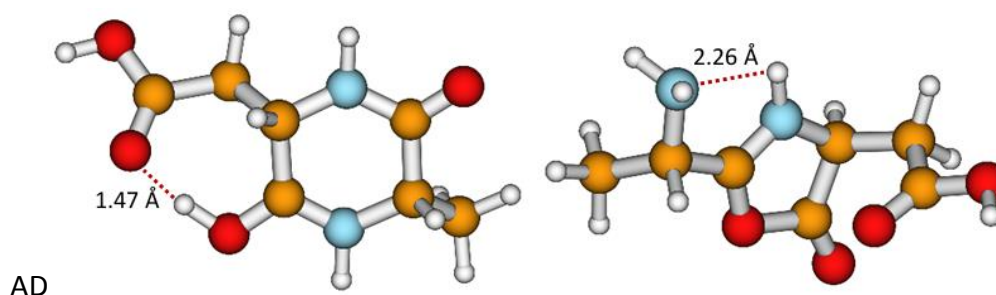
**Appendix Figure 12. The minimum energy structures of PA macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $67.6 \text{ \AA}^2$  for the macrocyclic geometry and  $70.4 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a PA  $b_2$  ion was not obtained.



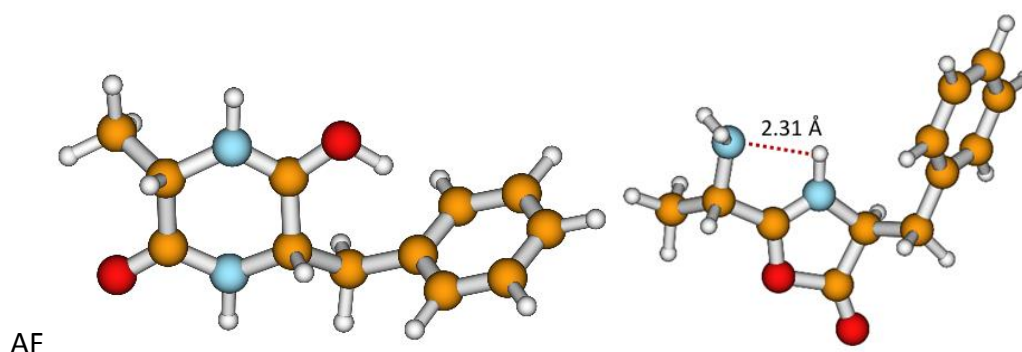
**Appendix Figure 13. The minimum energy structures of QA macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $77.4 \text{ \AA}^2$  for the macrocyclic geometry and  $80.9 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a QA  $b_2$  ion was not obtained.



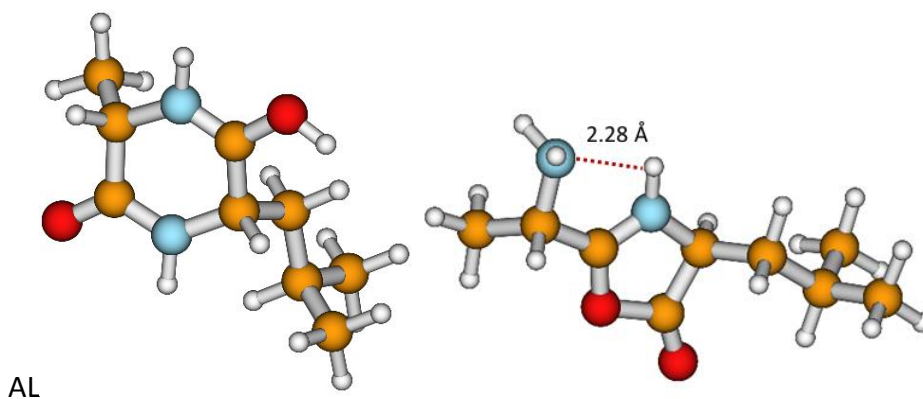
**Appendix Figure 14. The minimum energy structures of SA macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $63.5 \text{ \AA}^2$  for the macrocyclic geometry and  $66.3 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a SA  $b_2$  ion was  $68.9 \text{ \AA}^2$ .



**Appendix Figure 15. The minimum energy structures of AD macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $69.0 \text{ \AA}^2$  for the macrocyclic geometry and  $71.6 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for an AD  $b_2$  ion was not obtained.

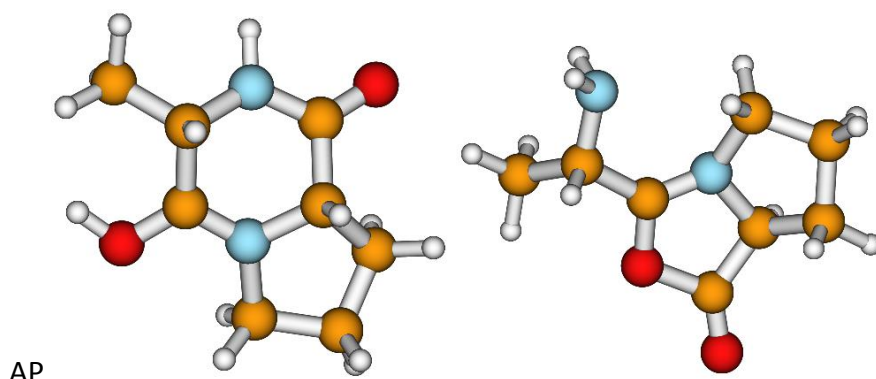


**Appendix Figure 16. The minimum energy structures of AF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $84.6 \text{ \AA}^2$  for the macrocyclic geometry and  $87.6 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for an AF  $b_2$  ion was not obtained.



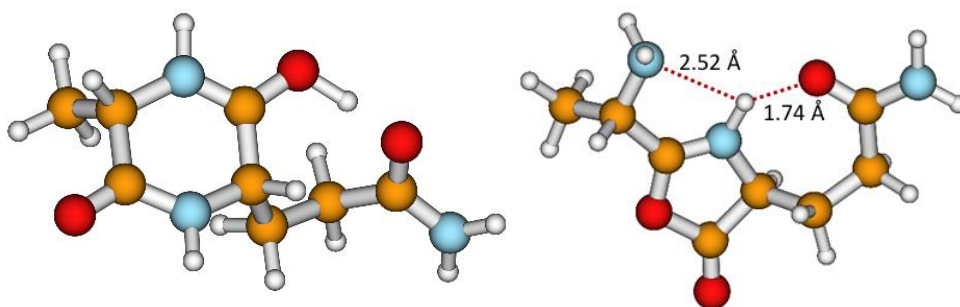
AL

**Appendix Figure 17. The minimum energy structures of AL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 77.7 Å<sup>2</sup> for the macrocyclic geometry and 80.7 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for an AL  $b_2$  ion was not obtained.



AP

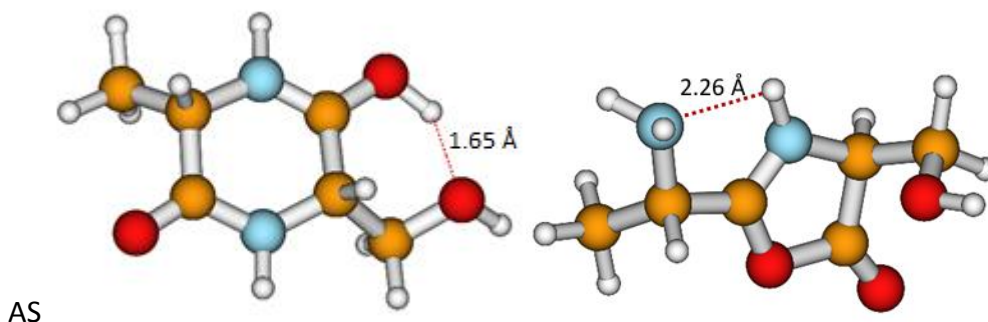
**Appendix Figure 18. The minimum energy structures of AP macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 67.6 Å<sup>2</sup> for the macrocyclic geometry and 69.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for an AP  $b_2$  ion was not obtained.



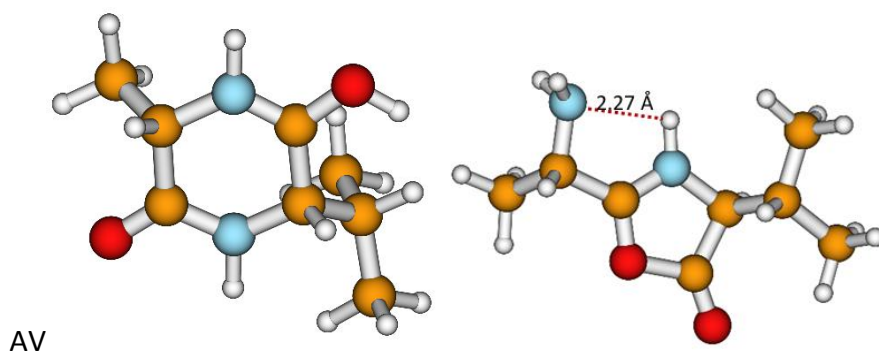
AQ

**Appendix Figure 19. The minimum energy structures of AQ macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal

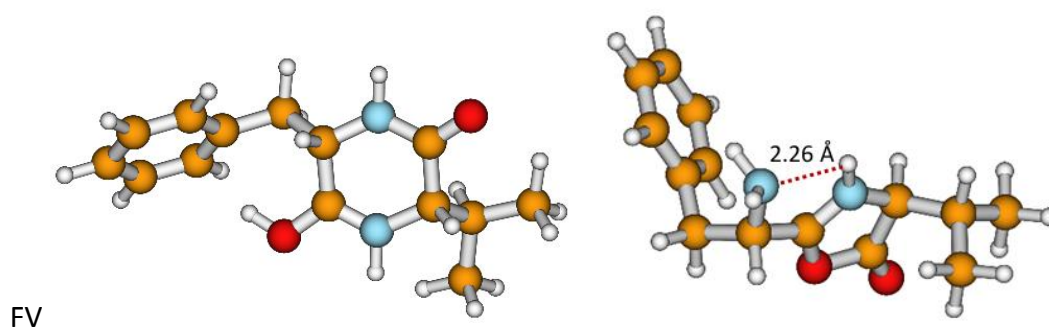
was used to generate a CCS value of  $77.4 \text{ \AA}^2$  for the macrocyclic geometry and  $80.7 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for an AQ  $b_2$  ion was not obtained.



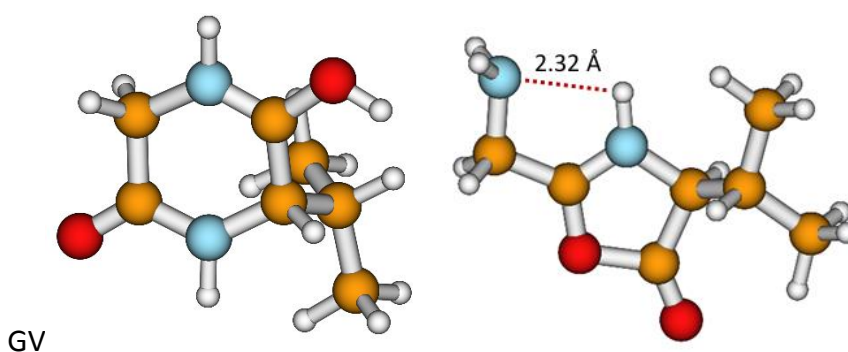
**Appendix Figure 20. The minimum energy structures of AS macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $63.5 \text{ \AA}^2$  for the macrocyclic geometry and  $66.4 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for an AS  $b_2$  ion was not obtained.



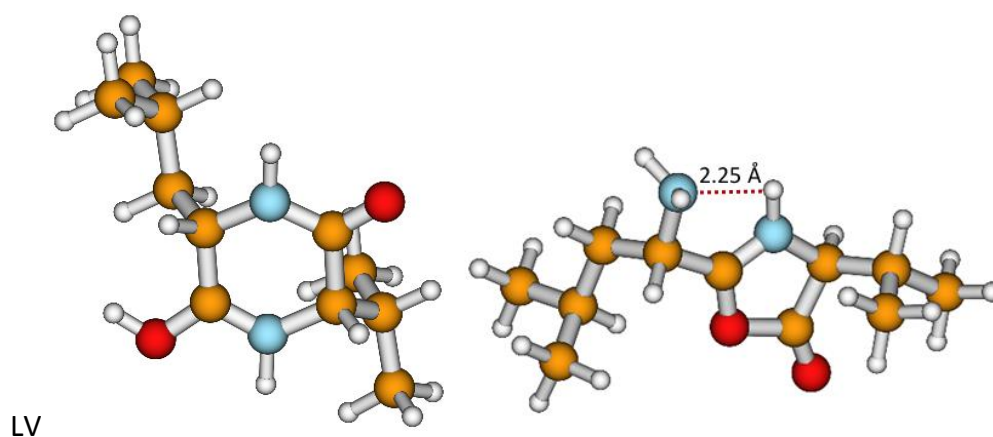
**Appendix Figure 21. The minimum energy structures of AV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of  $70.5 \text{ \AA}^2$  for the macrocyclic geometry and  $74.2 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for an AV  $b_2$  ion was  $72.15 \text{ \AA}^2$ .



**Appendix Figure 22. The minimum energy structures of FV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 94.2 Å<sup>2</sup> for the macrocyclic geometry and 95.4 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FV  $b_2$  ion was 94.92 Å<sup>2</sup>.



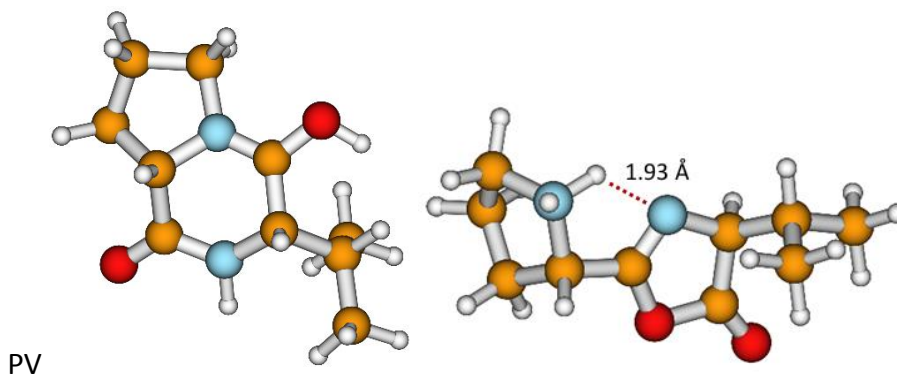
**Appendix Figure 23. The minimum energy structures of GV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 65.7 Å<sup>2</sup> for the macrocyclic geometry and 69.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GV  $b_2$  ion was 65.8 Å<sup>2</sup>.



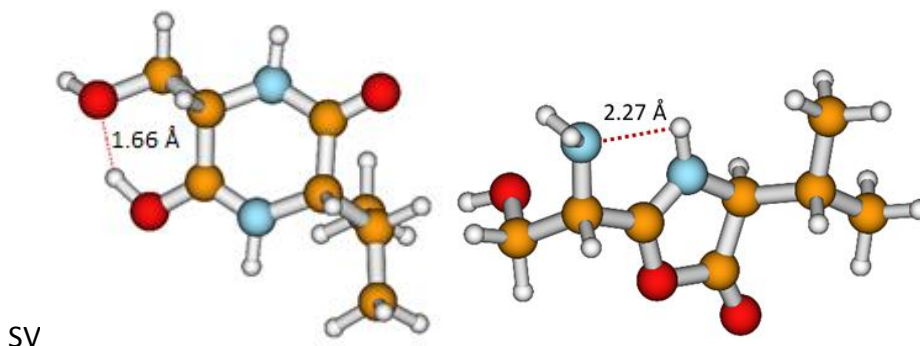
**Appendix Figure 24. The minimum energy structures of LV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 85.9 Å<sup>2</sup> for the macrocyclic geometry and 90.2 Å<sup>2</sup> for the



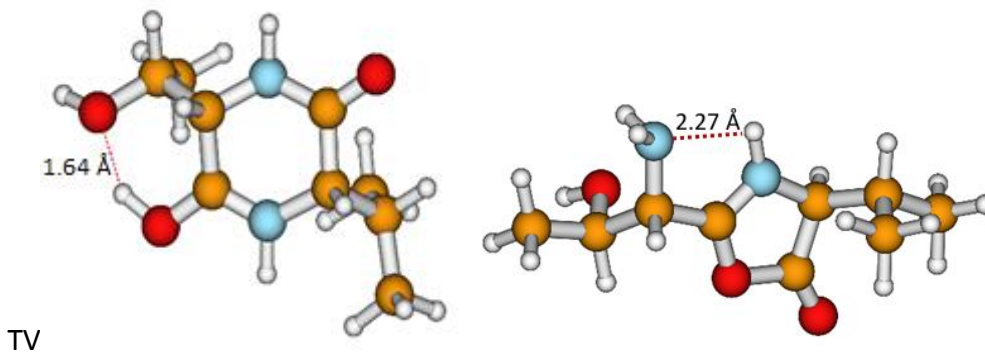
oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LV  $b_2$  ion was  $88.17 \text{ \AA}^2$ .



**Appendix Figure 25. The minimum energy structures of PV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $77.0 \text{ \AA}^2$  for the macrocyclic geometry and  $80.0 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a PV  $b_2$  ion was  $81.64 \text{ \AA}^2$ .

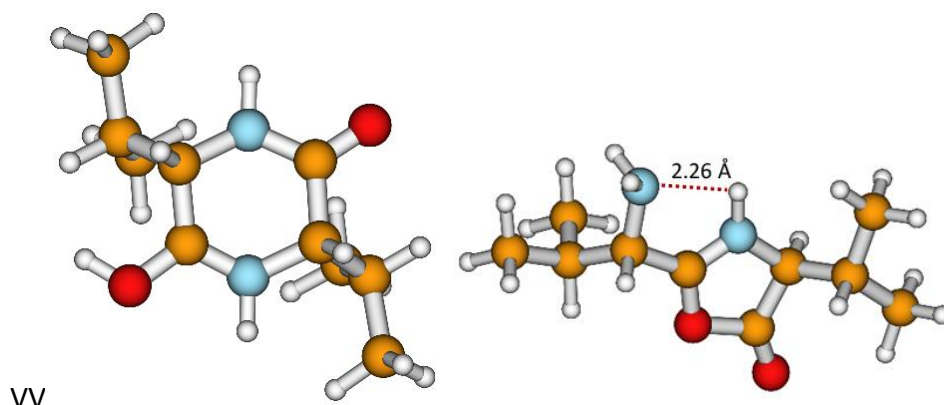


**Appendix Figure 26. The minimum energy structures of SV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $72.5 \text{ \AA}^2$  for the macrocyclic geometry and  $76.1 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a SV  $b_2$  ion was  $80.93 \text{ \AA}^2$ .

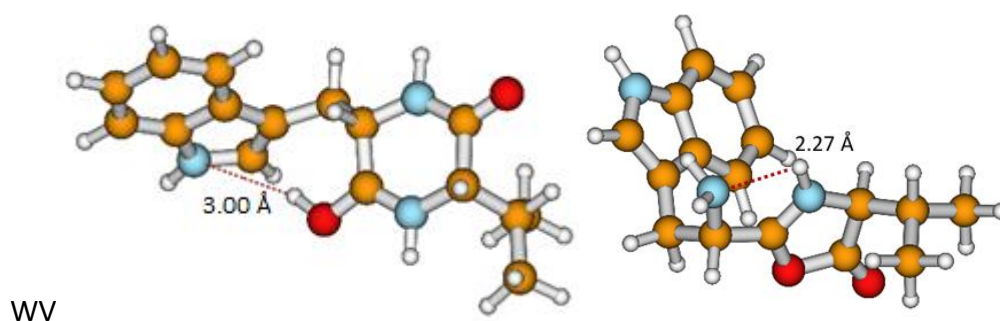


**Appendix Figure 27. The minimum energy structures of TV macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal

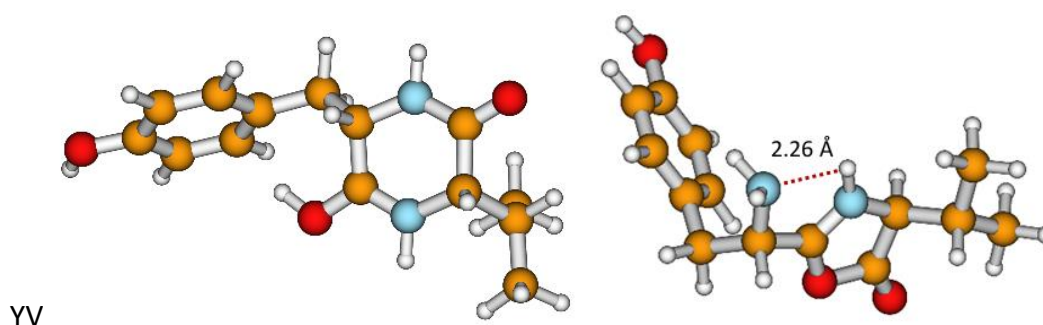
was used to generate a CCS value of 76.0 Å<sup>2</sup> for the macrocyclic geometry and 80.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a TV b<sub>2</sub> ion was 81.56 Å<sup>2</sup>.



**Appendix Figure 28. The minimum energy structures of VV macrocyclic (left) and oxazolone (right) b<sub>2</sub> ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 78.9 Å<sup>2</sup> for the macrocyclic geometry and 83.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a VV b<sub>2</sub> ion was 80.4 Å<sup>2</sup>.

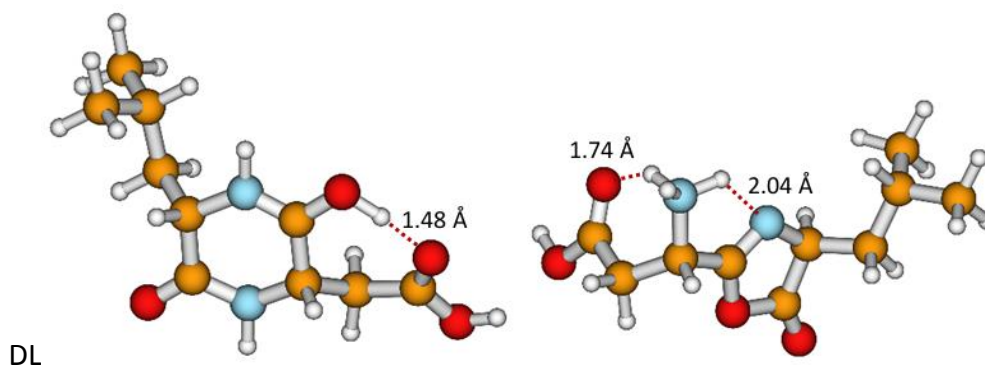


**Appendix Figure 29. The minimum energy structures of WV macrocyclic (left) and oxazolone (right) b<sub>2</sub> ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 104.2 Å<sup>2</sup> for the macrocyclic geometry and 103.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a WV b<sub>2</sub> ion was 104.33 Å<sup>2</sup>.

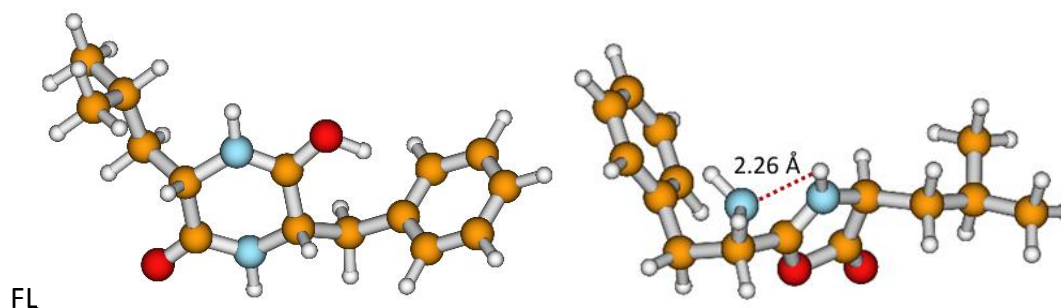


**Appendix Figure 30. The minimum energy structures of YV macrocyclic (left) and oxazolone (right) b<sub>2</sub> ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal

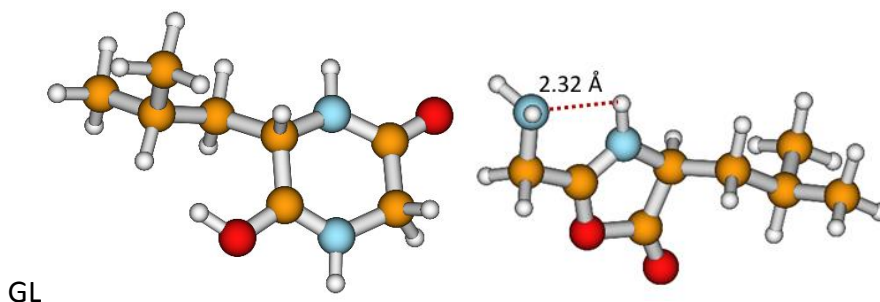
was used to generate a CCS value of 95.9 Å<sup>2</sup> for the macrocyclic geometry and 98.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a YV b<sub>2</sub> ion was 99.21 Å<sup>2</sup>.



**Appendix Figure 31. The minimum energy structures of DL macrocyclic (left) and oxazolone (right) b<sub>2</sub> ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 84.9 Å<sup>2</sup> for the macrocyclic geometry and 88.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a DL b<sub>2</sub> ion was not obtained.



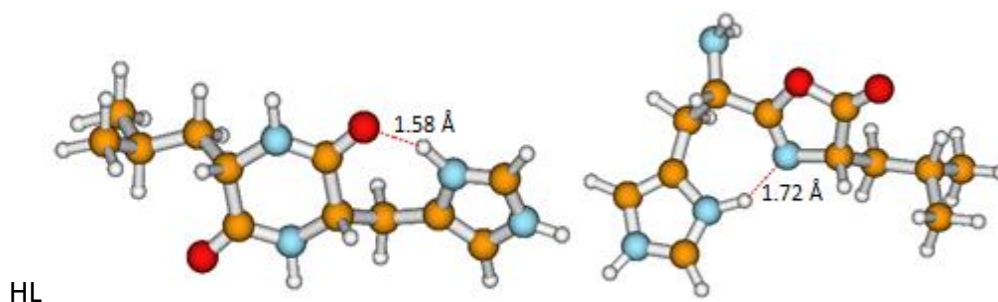
**Appendix Figure 32. The minimum energy structures of FL macrocyclic (left) and oxazolone (right) b<sub>2</sub> ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 100.7 Å<sup>2</sup> for the macrocyclic geometry and 101.9 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a FL b<sub>2</sub> ion was not obtained.



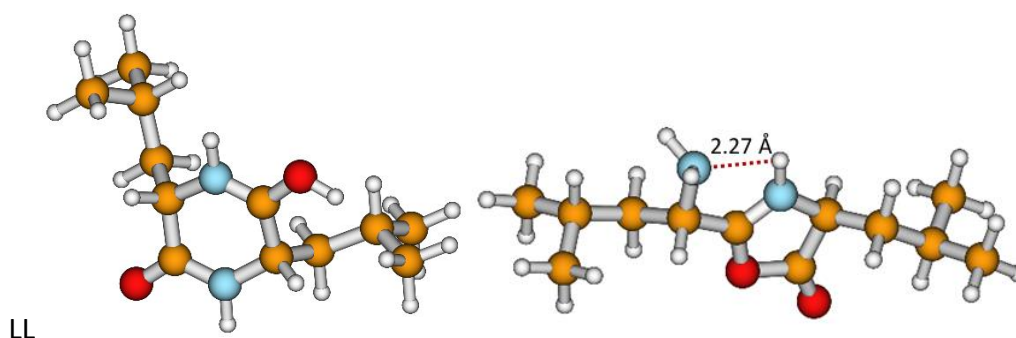
**Appendix Figure 33. The minimum energy structures of GL macrocyclic (left) and oxazolone (right) b<sub>2</sub> ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 72.5 Å<sup>2</sup> for the macrocyclic geometry and 75.4 Å<sup>2</sup> for the



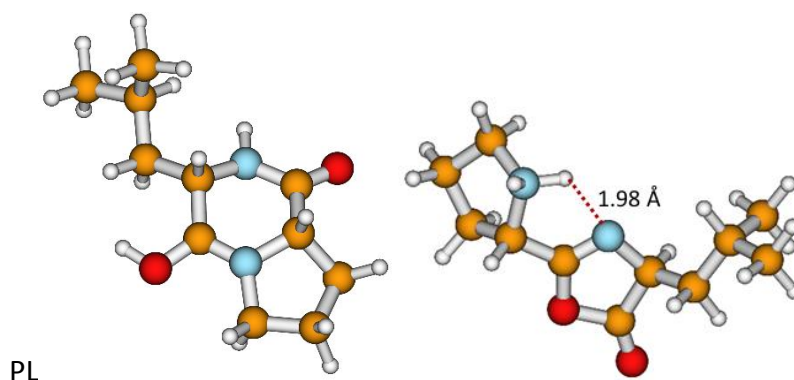
oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a GL  $b_2$  ion was not obtained.



**Appendix Figure 34. The minimum energy structures of HL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 97.3 Å<sup>2</sup> for the macrocyclic geometry and 100.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a HL  $b_2$  ion was not obtained.

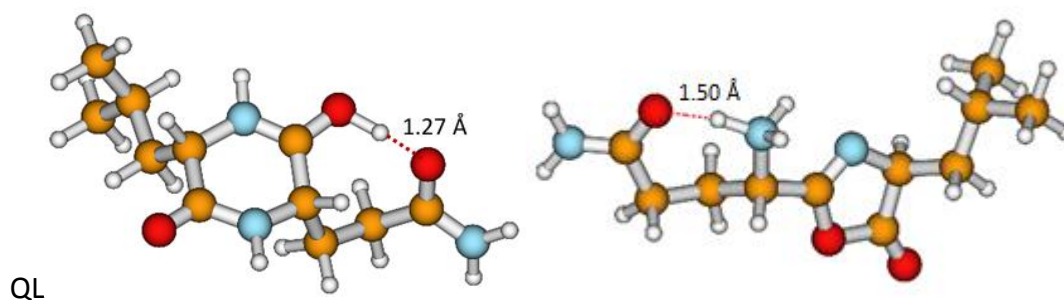


**Appendix Figure 35. The minimum energy structures of LL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 92.4 Å<sup>2</sup> for the macrocyclic geometry and 97.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a LL  $b_2$  ion was not obtained.

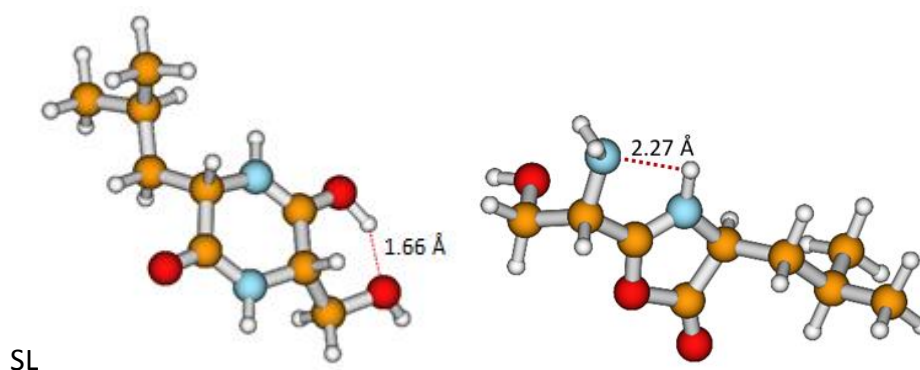


**Appendix Figure 36. The minimum energy structures of PL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 83.1 Å<sup>2</sup> for the macrocyclic geometry and 87.4 Å<sup>2</sup> for the

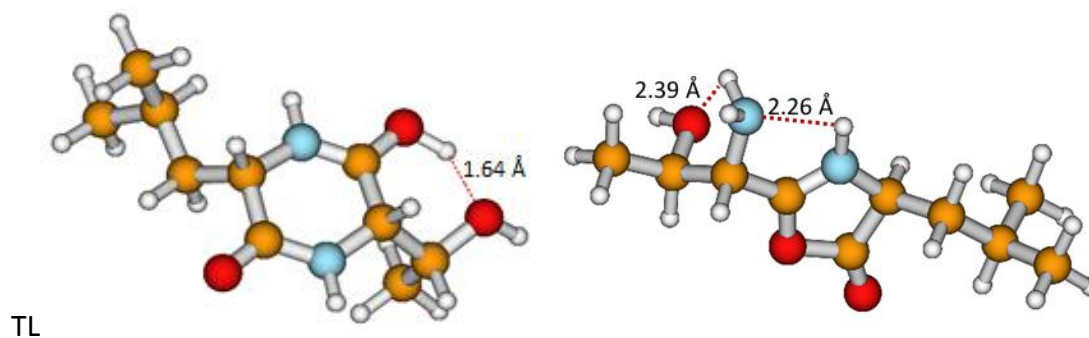
oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a PL  $b_2$  ion was not obtained.



**Appendix Figure 37. The minimum energy structures of QL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 93.1 Å<sup>2</sup> for the macrocyclic geometry and 97.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a QL  $b_2$  ion was not obtained.

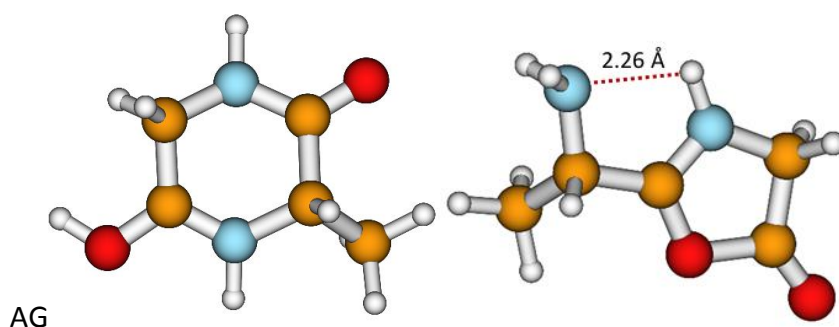


**Appendix Figure 38. The minimum energy structures of SL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 79.3 Å<sup>2</sup> for the macrocyclic geometry and 82.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a SL  $b_2$  ion was not obtained.

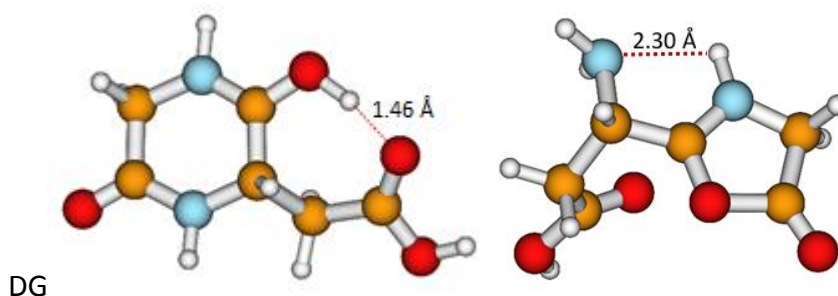


**Appendix Figure 39. The minimum energy structures of TL macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 83.7 Å<sup>2</sup> for the macrocyclic geometry and 87.7 Å<sup>2</sup> for the

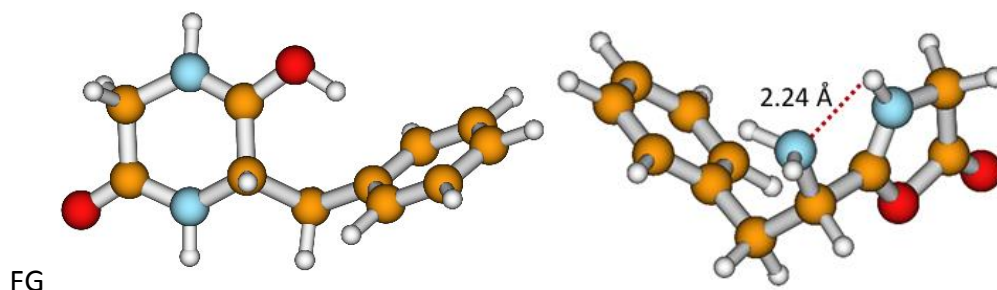
oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a TL  $b_2$  ion was not obtained.



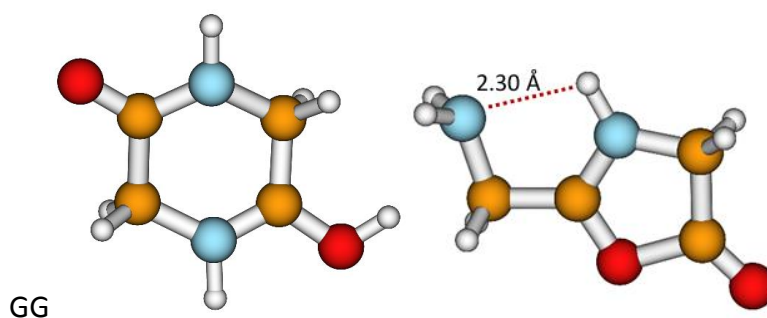
**Appendix Figure 40. The minimum energy structures of AG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 57.3 Å<sup>2</sup> for the macrocyclic geometry and 58.8 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for an AG  $b_2$  ion was 53.76 Å<sup>2</sup>.



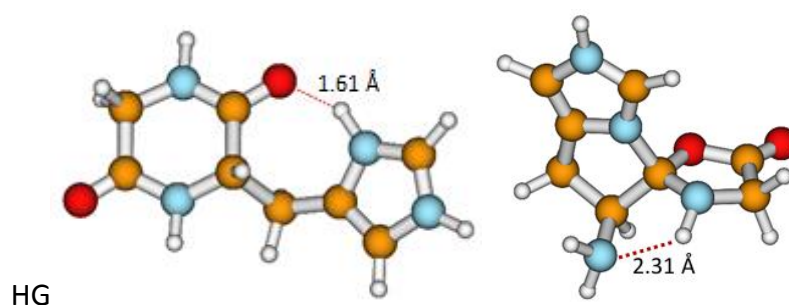
**Appendix Figure 41. The minimum energy structures of DG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 63.9 Å<sup>2</sup> for the macrocyclic geometry and 65.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a DG  $b_2$  ion was 71.59 Å<sup>2</sup>.



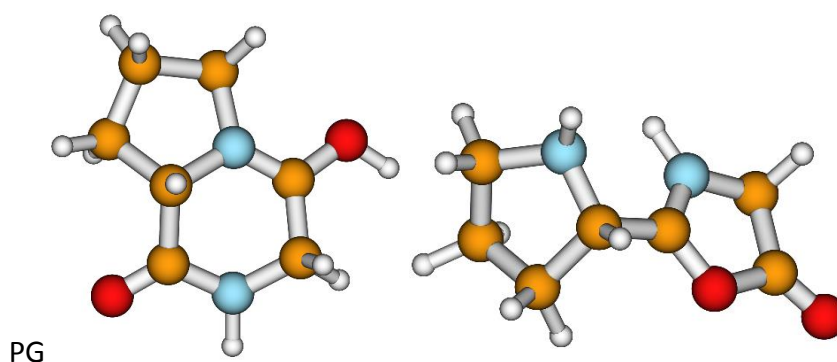
**Appendix Figure 42. The minimum energy structures of FG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 79.5 Å<sup>2</sup> for the macrocyclic geometry and 80.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FG  $b_2$  ion was 82.93 Å<sup>2</sup>.



**Appendix Figure 43. The minimum energy structures of GG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 52.4 Å<sup>2</sup> for the macrocyclic geometry and 53.7 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GG  $b_2$  ion was 54.43 Å<sup>2</sup>.

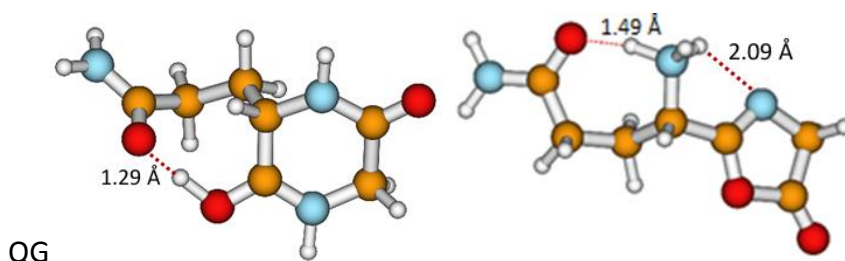


**Appendix Figure 44. The minimum energy structures of HG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 75.9 Å<sup>2</sup> for the macrocyclic geometry and 74.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a HG  $b_2$  ion was 80.86 Å<sup>2</sup>.

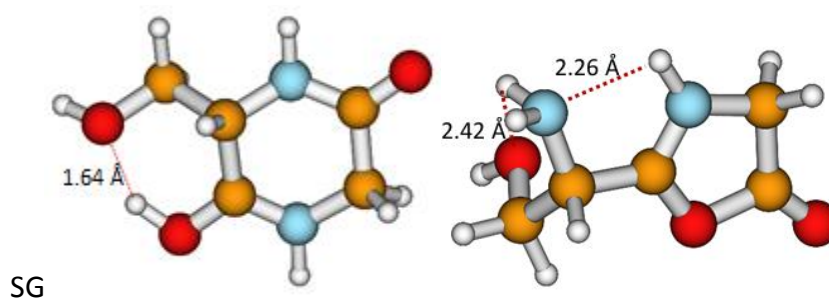


**Appendix Figure 45. The minimum energy structures of PG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 63.2 Å<sup>2</sup> for the macrocyclic geometry and 64.9 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a PG  $b_2$  ion was 64.61 Å<sup>2</sup>.

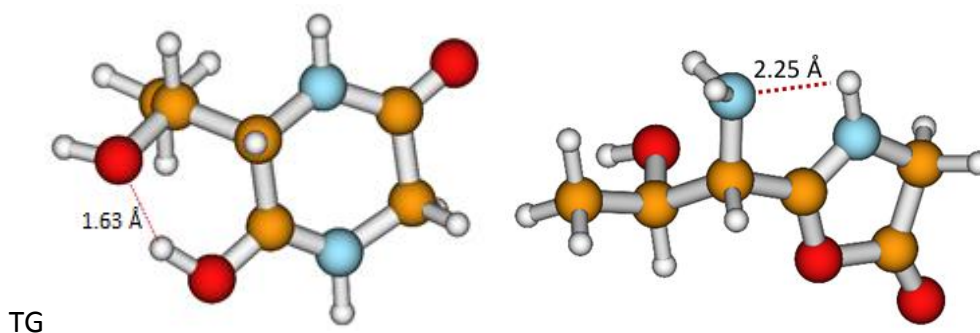




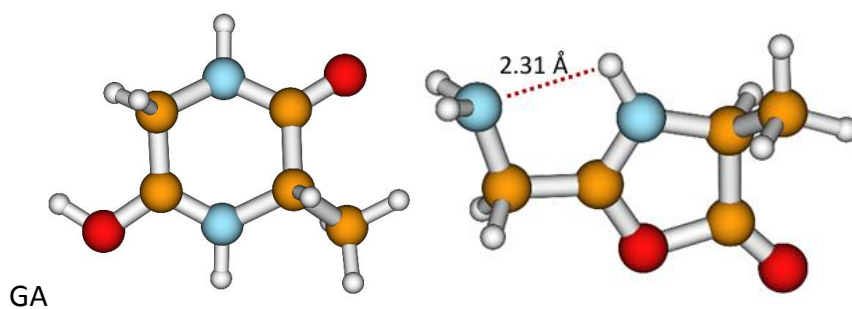
**Appendix Figure 46. The minimum energy structures of QG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 72.5 Å<sup>2</sup> for the macrocyclic geometry and 75.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a QG  $b_2$  ion was 82.93 Å<sup>2</sup>.



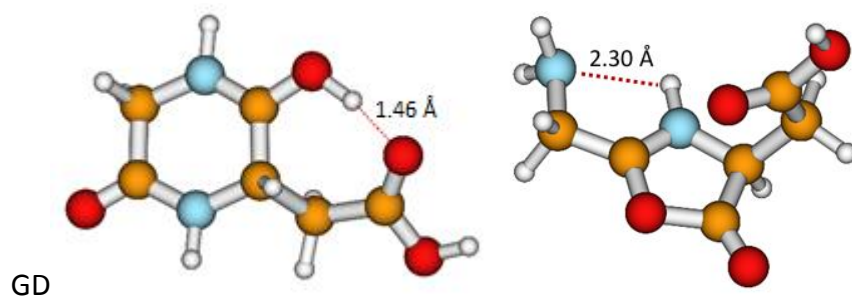
**Appendix Figure 47. The minimum energy structures of SG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 58.4 Å<sup>2</sup> for the macrocyclic geometry and 60.8 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a SG  $b_2$  ion was 56.19 Å<sup>2</sup>.



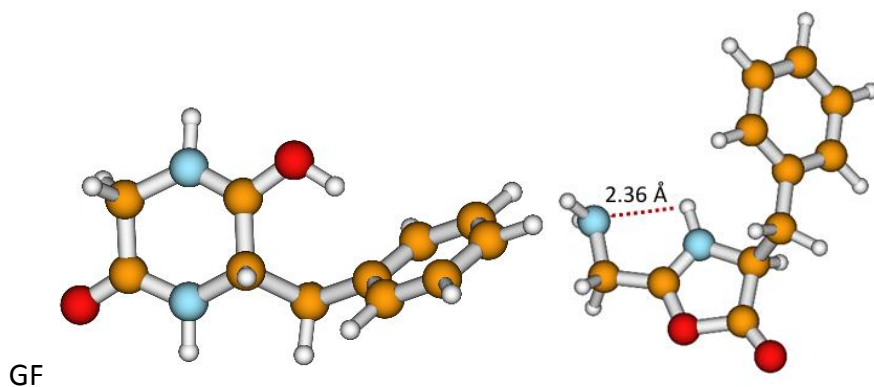
**Appendix Figure 48. The minimum energy structures of TG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 62.8 Å<sup>2</sup> for the macrocyclic geometry and 65.8 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a TG  $b_2$  ion was 62.11 Å<sup>2</sup>.



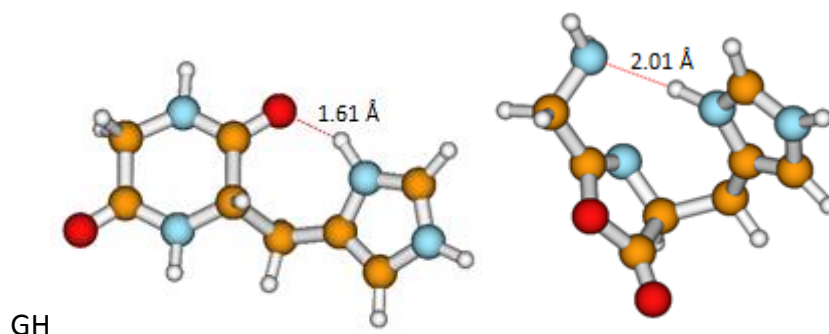
**Appendix Figure 49. The minimum energy structures of GA macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 57.3 Å<sup>2</sup> for the macrocyclic geometry and 59.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. An experimental CCS value for a GA  $b_2$  ion was not obtained.



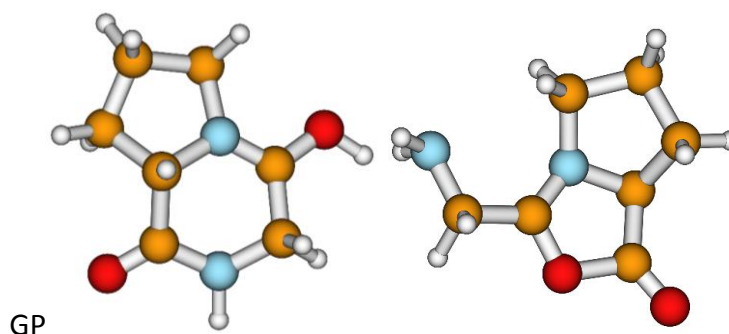
**Appendix Figure 50. The minimum energy structures of GD macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 63.9 Å<sup>2</sup> for the macrocyclic geometry and 66.9 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GD  $b_2$  ion was 71.1 Å<sup>2</sup>.



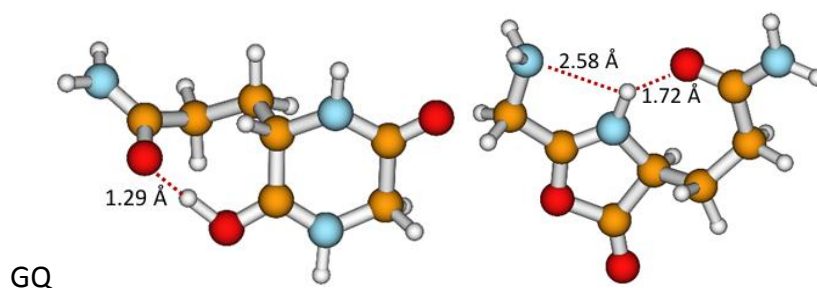
**Appendix Figure 51. The minimum energy structures of GF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 79.5 Å<sup>2</sup> for the macrocyclic geometry and 82.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GF  $b_2$  ion was 81.52 Å<sup>2</sup>.



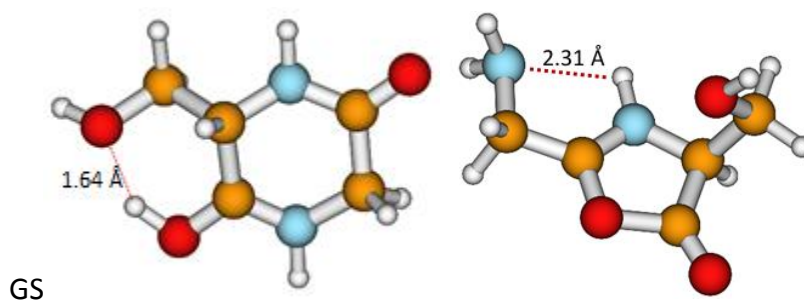
**Appendix Figure 52. The minimum energy structures of GH macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 75.9 Å<sup>2</sup> for the macrocyclic geometry and 75.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GH  $b_2$  ion was 77.21 Å<sup>2</sup>.



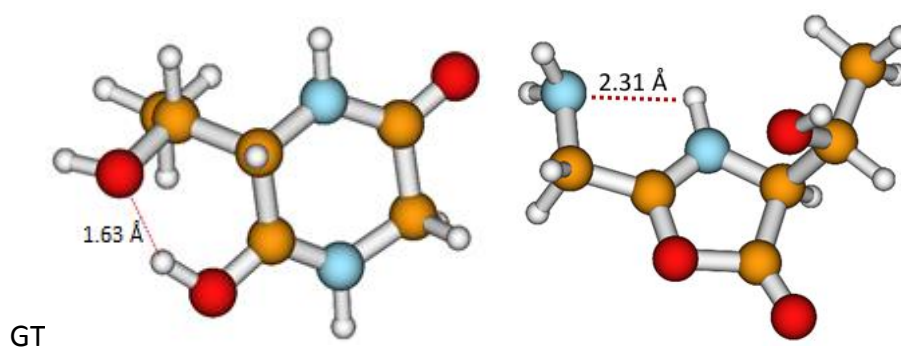
**Appendix Figure 53. The minimum energy structures of GP macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 63.2 Å<sup>2</sup> for the macrocyclic geometry and 64.4 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GP  $b_2$  ion was 62.47 Å<sup>2</sup>.



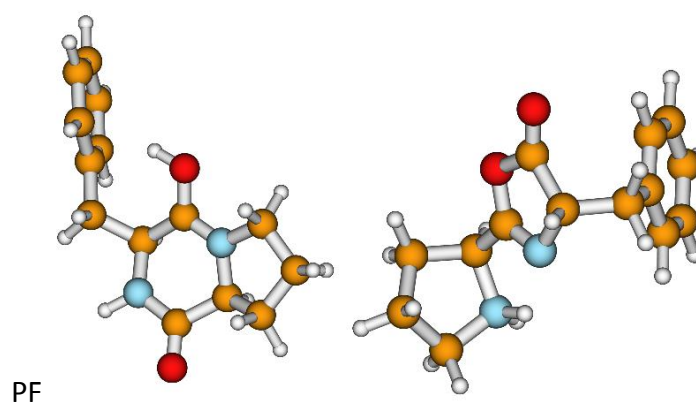
**Appendix Figure 54. The minimum energy structures of GQ macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 72.5 Å<sup>2</sup> for the macrocyclic geometry and 75.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GQ  $b_2$  ion was 82.66 Å<sup>2</sup>.



**Appendix Figure 55. The minimum energy structures of GS macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 58.4 Å<sup>2</sup> for the macrocyclic geometry and 61.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GS  $b_2$  ion was 58.79 Å<sup>2</sup>.

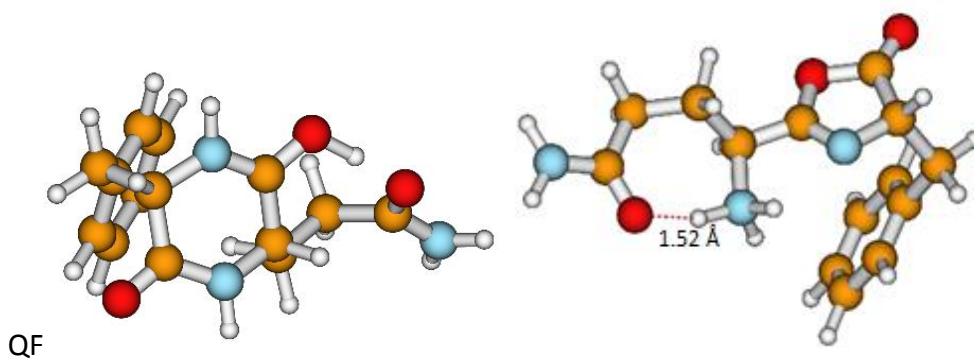


**Appendix Figure 56. The minimum energy structures of GT macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 62.8 Å<sup>2</sup> for the macrocyclic geometry and 66.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a GT  $b_2$  ion was 63.55 Å<sup>2</sup>.

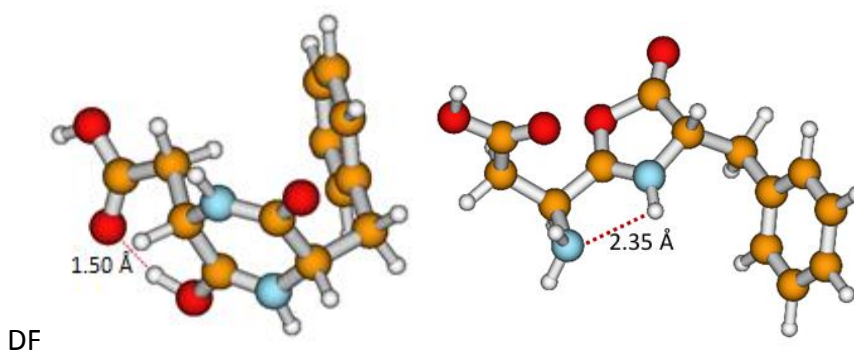


**Appendix Figure 57. The minimum energy structures of PF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 90.7 Å<sup>2</sup> for the macrocyclic geometry and 91.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a PF  $b_2$  ion was 92.45 Å<sup>2</sup>.

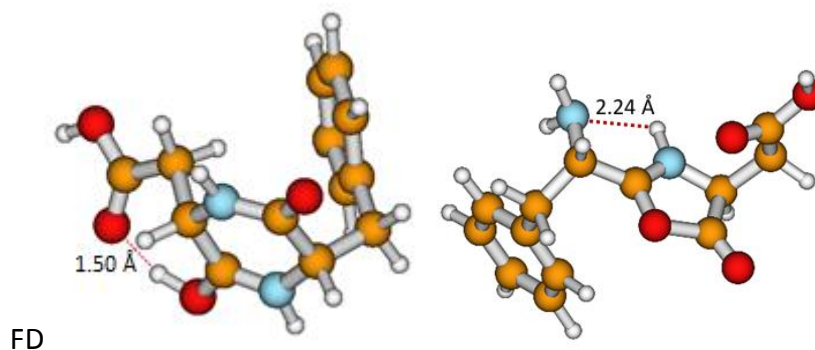




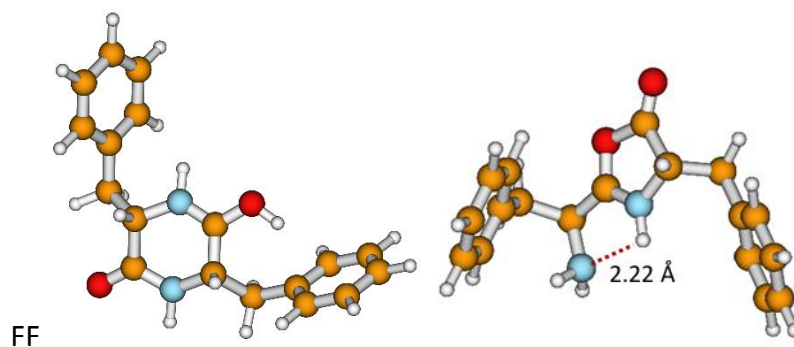
**Appendix Figure 58. The minimum energy structures of QF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 94.4 Å<sup>2</sup> for the macrocyclic geometry and 101.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a QF  $b_2$  ion was 101.22 Å<sup>2</sup>.



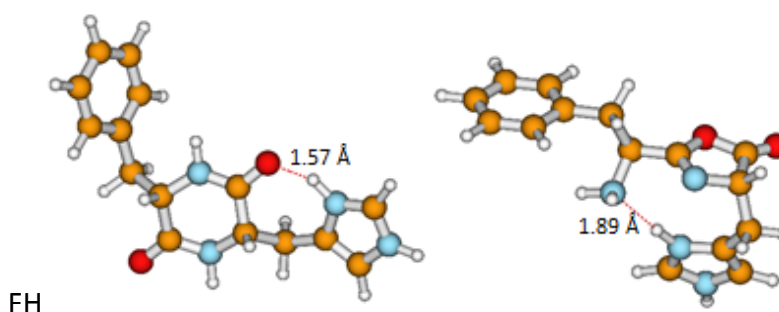
**Appendix Figure 59. The minimum energy structures of DF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 86.4 Å<sup>2</sup> for the macrocyclic geometry and 94.7 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a DF  $b_2$  ion was 97.26 Å<sup>2</sup>.



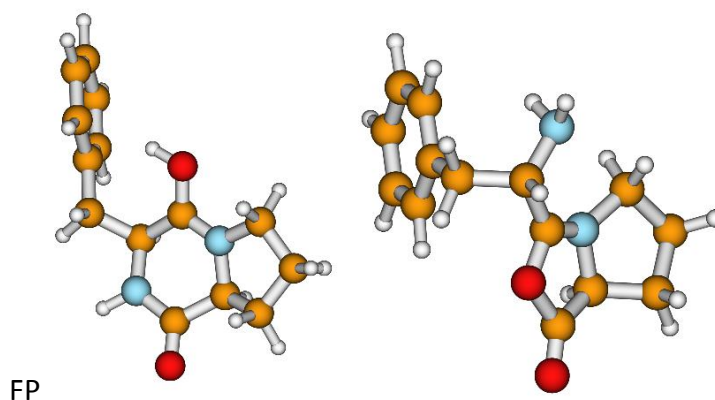
**Appendix Figure 60. The minimum energy structures of FD macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 86.4 Å<sup>2</sup> for the macrocyclic geometry and 93.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FD  $b_2$  ion was 100.57 Å<sup>2</sup>.



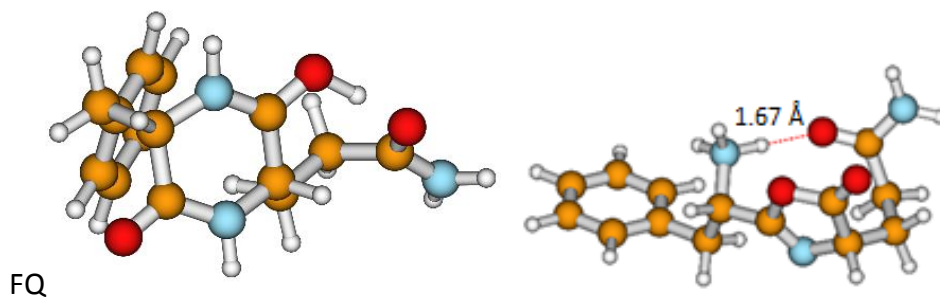
**Appendix Figure 61. The minimum energy structures of FF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 108.2 Å<sup>2</sup> for the macrocyclic geometry and 108.3 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FF  $b_2$  ion was 105.64 Å<sup>2</sup>.



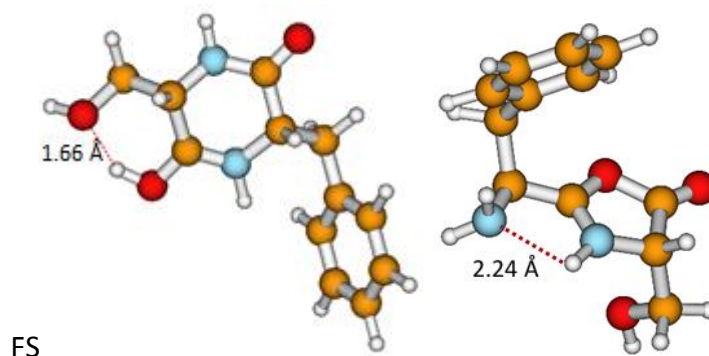
**Appendix Figure 62. The minimum energy structures of FH macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 104.7 Å<sup>2</sup> for the macrocyclic geometry and 104.0 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FH  $b_2$  ion was 101.88 Å<sup>2</sup>.



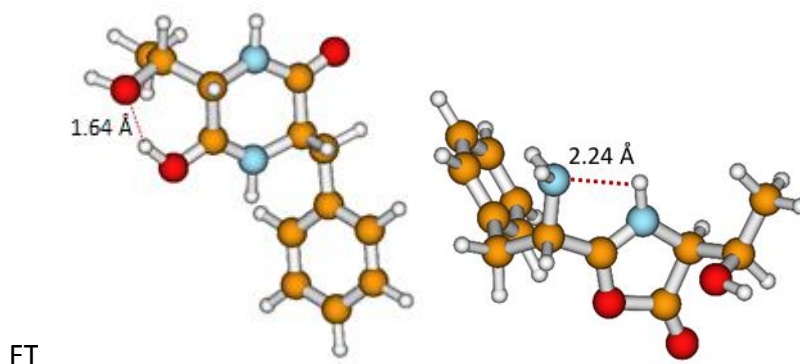
**Appendix Figure 63. The minimum energy structures of FP macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 90.7 Å<sup>2</sup> for the macrocyclic geometry and 91.2 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FP  $b_2$  ion was 92.52 Å<sup>2</sup>.



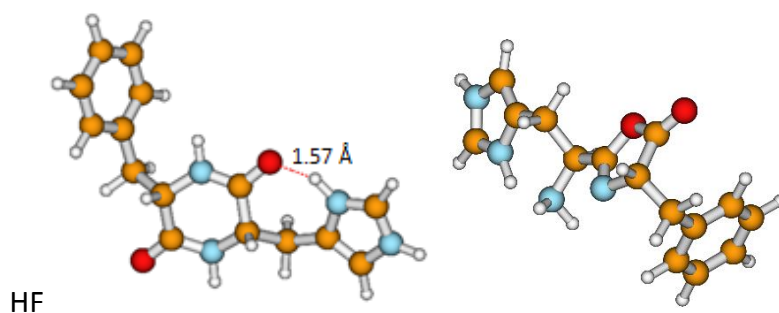
**Appendix Figure 64. The minimum energy structures of FQ macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 94.4 Å<sup>2</sup> for the macrocyclic geometry and 99.7 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FQ  $b_2$  ion was 101.3 Å<sup>2</sup>.



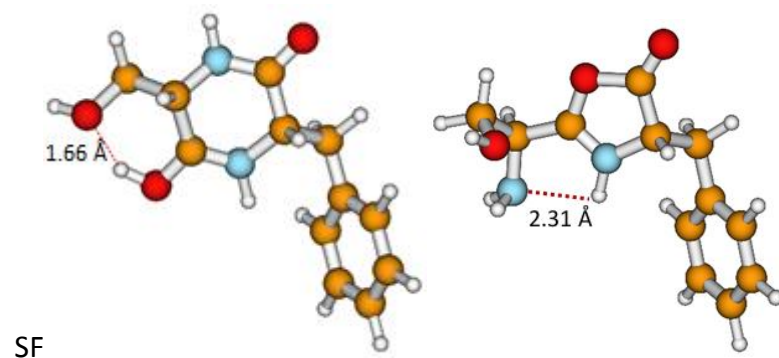
**Appendix Figure 65. The minimum energy structures of FS macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 86.8 Å<sup>2</sup> for the macrocyclic geometry and 87.9 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FS  $b_2$  ion was 92.1 Å<sup>2</sup>.



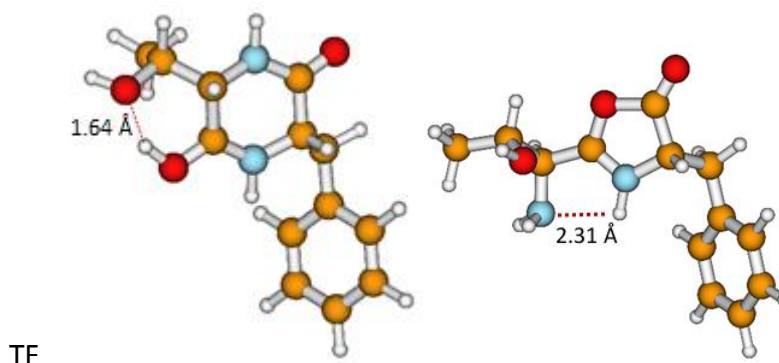
**Appendix Figure 66. The minimum energy structures of FT macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 91.2 Å<sup>2</sup> for the macrocyclic geometry and 93.5 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a FT  $b_2$  ion was 94.89 Å<sup>2</sup>.



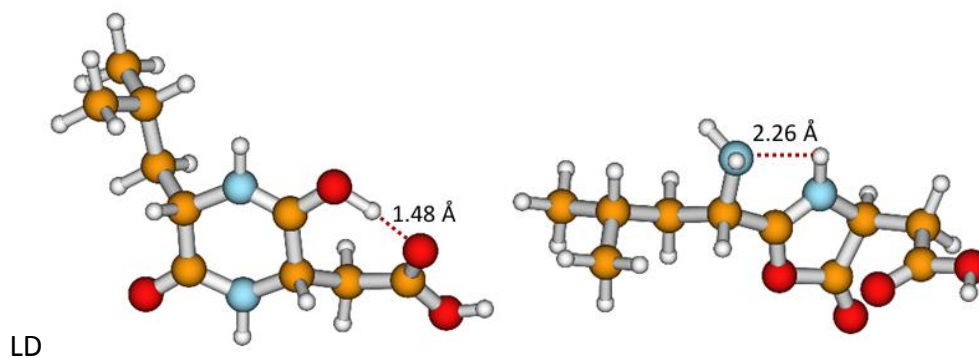
**Appendix Figure 67. The minimum energy structures of HF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 104.7 Å<sup>2</sup> for the macrocyclic geometry and 105.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a HF  $b_2$  ion was 101.72 Å<sup>2</sup>.



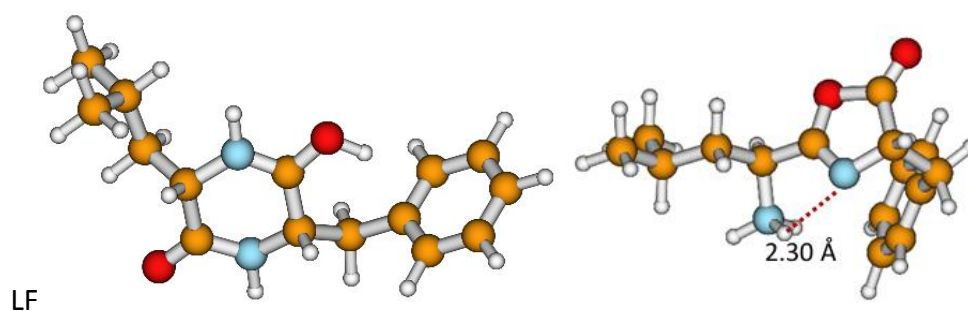
**Appendix Figure 68. The minimum energy structures of SF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 86.8 Å<sup>2</sup> for the macrocyclic geometry and 89.7 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a SF  $b_2$  ion was 90.61 Å<sup>2</sup>.



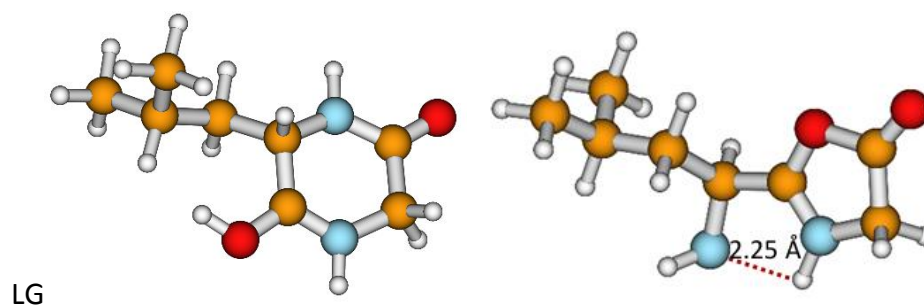
**Appendix Figure 69. The minimum energy structures of TF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Guassian09. Mobcal was used to generate a CCS value of 91.2 Å<sup>2</sup> for the macrocyclic geometry and 94.7 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a TF  $b_2$  ion was 94.6 Å<sup>2</sup>.



**Appendix Figure 70. The minimum energy structures of LD macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 84.9 Å<sup>2</sup> for the macrocyclic geometry and 88.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LD  $b_2$  ion was 88.85 Å<sup>2</sup>.

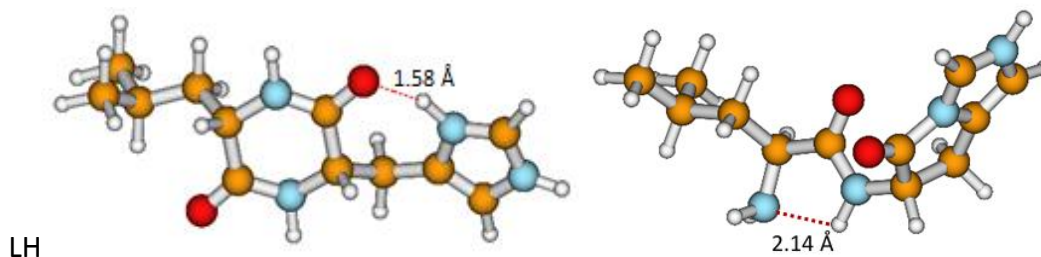


**Appendix Figure 71. The minimum energy structures of LF macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 100.7 Å<sup>2</sup> for the macrocyclic geometry and 98.6 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LF  $b_2$  ion was 98.87 Å<sup>2</sup>.

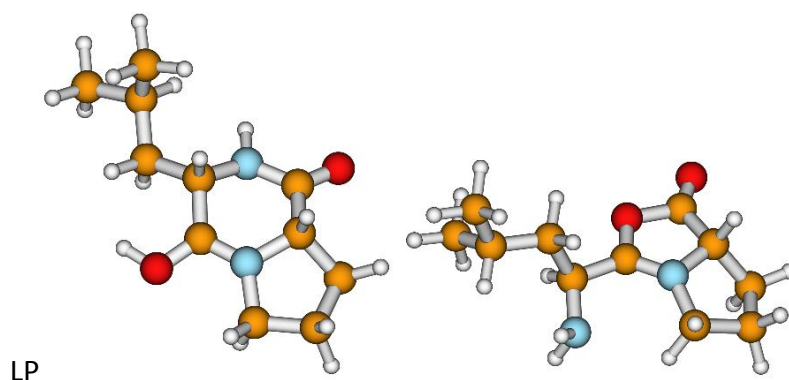


**Appendix Figure 72. The minimum energy structures of LG macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 72.5 Å<sup>2</sup> for the macrocyclic geometry and 75.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LG  $b_2$  ion was 76.98 Å<sup>2</sup>.

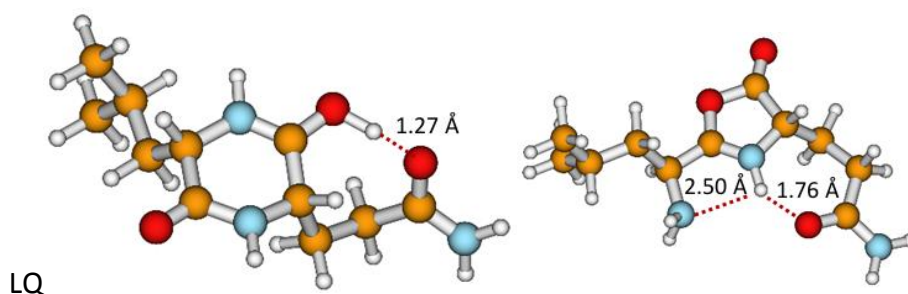




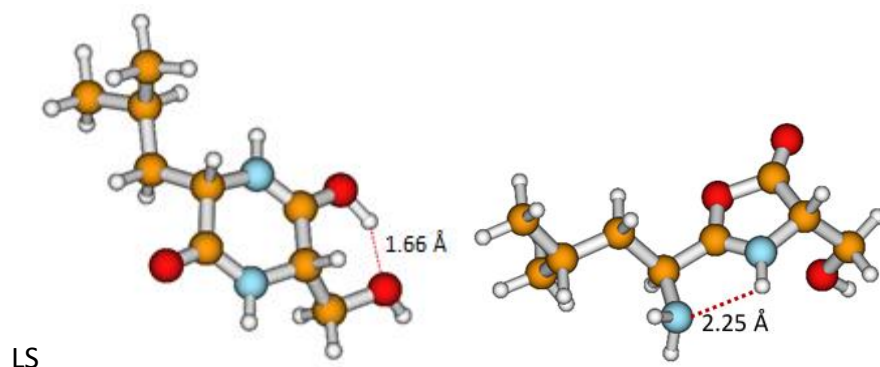
**Appendix Figure 73. The minimum energy structures of LH macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $97.3 \text{ \AA}^2$  for the macrocyclic geometry and  $98.8 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LH  $b_2$  ion was  $96.28 \text{ \AA}^2$ .



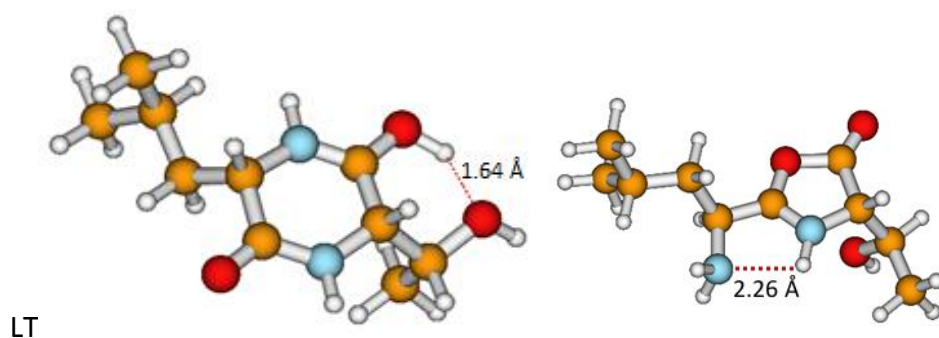
**Appendix Figure 74. The minimum energy structures of LP macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $83.1 \text{ \AA}^2$  for the macrocyclic geometry and  $86.0 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LP  $b_2$  ion was  $86.05 \text{ \AA}^2$ .



**Appendix Figure 75. The minimum energy structures of LQ macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of  $93.1 \text{ \AA}^2$  for the macrocyclic geometry and  $97.3 \text{ \AA}^2$  for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LQ  $b_2$  ion was  $93.58 \text{ \AA}^2$ .



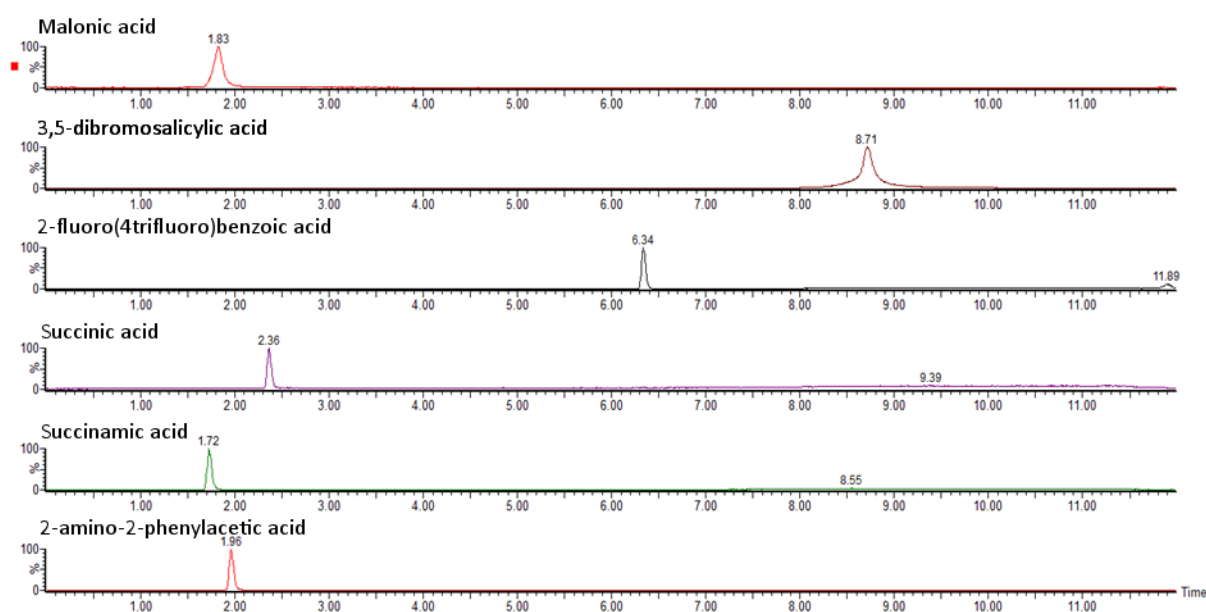
**Appendix Figure 76. The minimum energy structures of LS macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 79.3 Å<sup>2</sup> for the macrocyclic geometry and 82.9 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LS  $b_2$  ion was 82.56 Å<sup>2</sup>.



**Appendix Figure 77. The minimum energy structures of LT macrocyclic (left) and oxazolone (right)  $b_2$  ions.** Optimised at the B3LYP level of theory and 6-31+G(d,p) basis set using Gaussian09. Mobcal was used to generate a CCS value of 83.7 Å<sup>2</sup> for the macrocyclic geometry and 88.1 Å<sup>2</sup> for the oxazolone geometry using the trajectory model with a Helium drift gas. The respective experimentally determined CCS value for a LT  $b_2$  ion was 85.78 Å<sup>2</sup>.

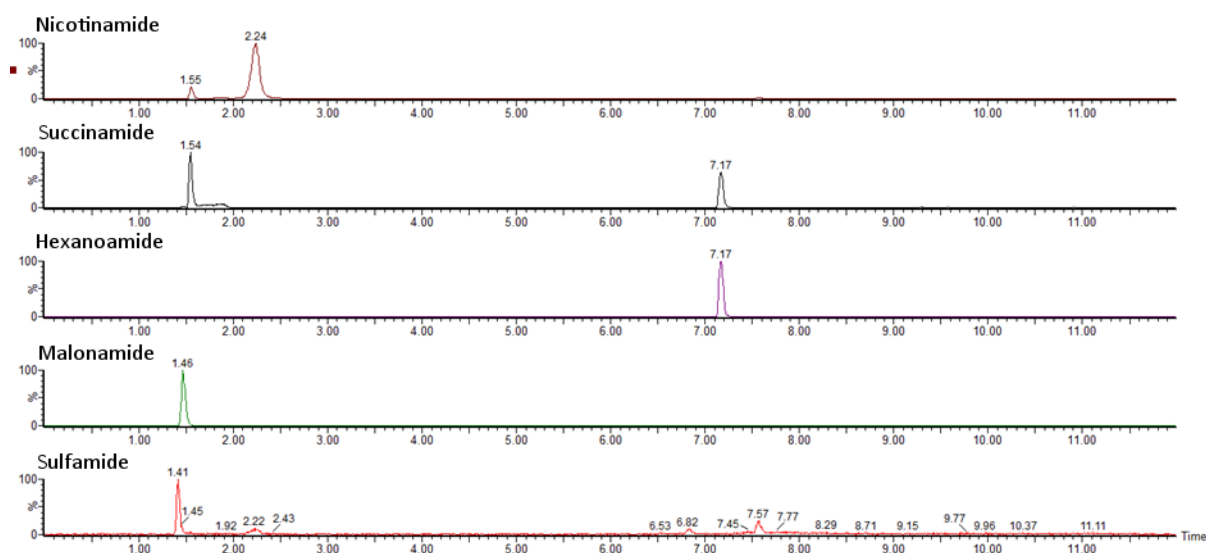
## Appendix 1.2: Detection of commercially available standards by UPLC-MS/MS

In the first step, each product of interest was separated via the aforementioned UPLC method and each listed MRM method was applied individually to each respective product. The products were split into five groups and measured in pools to ensure that there is sufficient retention time separation between products in each group. The chromatograms for each group are displayed in figures 29 to 33.

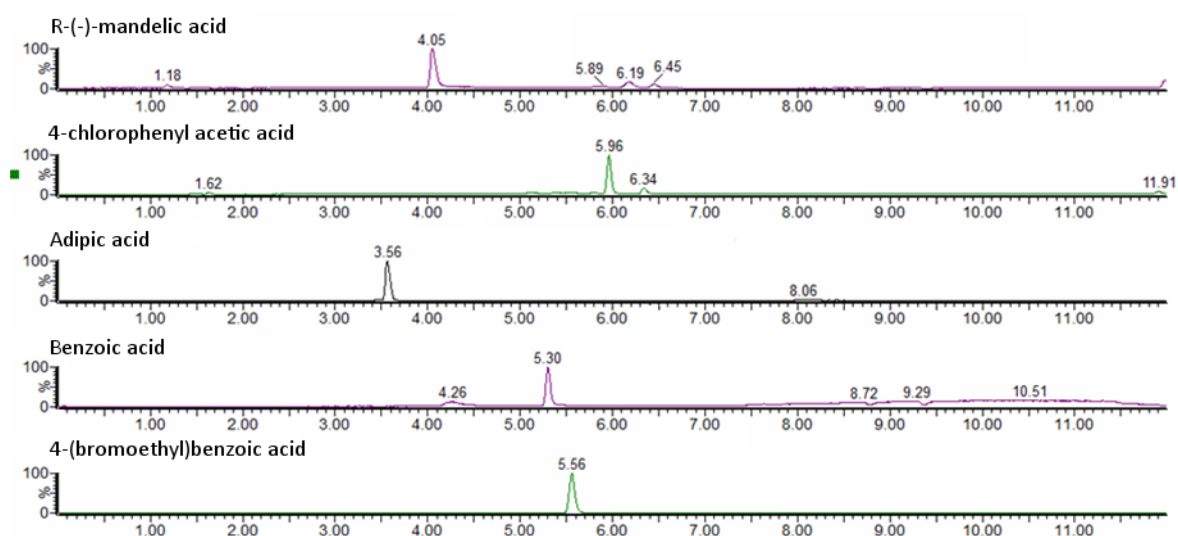


**Figure 38. Normalised chromatograms for standards.** Top to bottom: Malonic acid, 3,5-dibromosalicylic acid, 2-fluoro-4-(trifluoromethyl)benzoic acid, succinic acid, succinamic acid and 2-amino-2-phenylacetic acid. Produced from a mixture of standards. Separated using an Ascentis Express C18 column (150 mm × 2.1 mm, 2.7 µm) and an Acquity I-Class UPLC system. Acquired using a Waters Xevo TQ-XS triple quadrupole mass spectrometer.

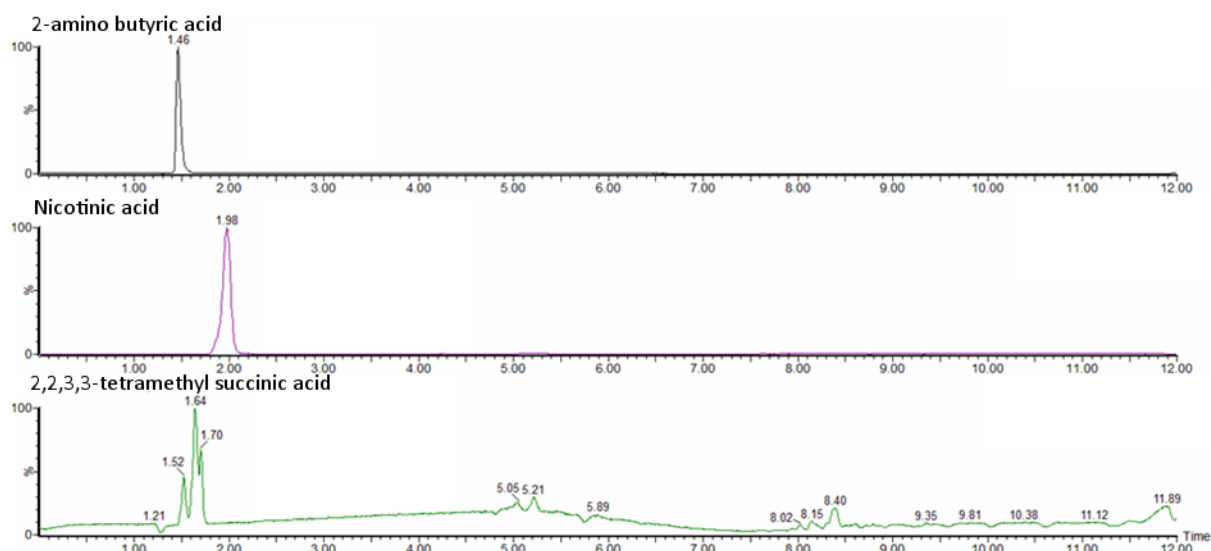




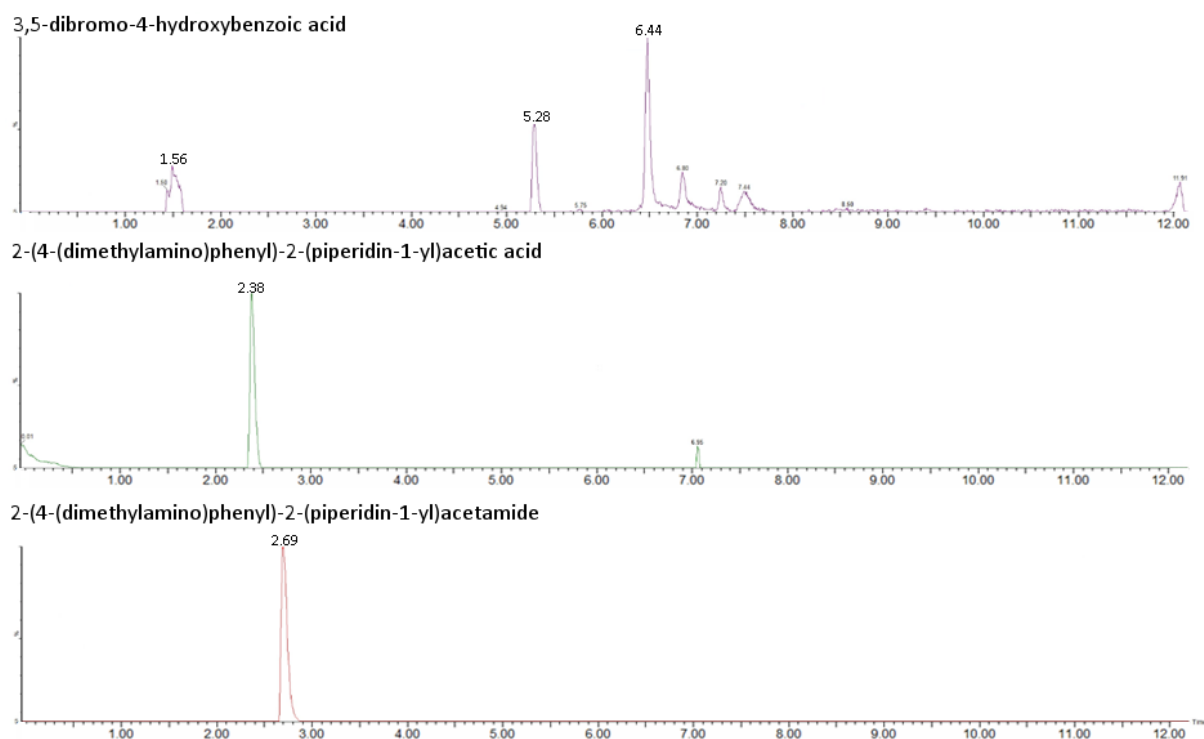
**Figure 39. Normalised chromatograms for standards.** Top to bottom: Nicotinamide, succinamide, hexanoamide, malonamide and sulfamide. Produced from a mixture of standards. Separated using an Ascentis Express C18 column (150 mm × 2.1 mm, 2.7 μm) and an Acquity I-Class UPLC system. Acquired using a Waters Xevo TQ-XS triple quadrupole mass spectrometer.



**Figure 40. Normalised chromatograms for standard.** Top to bottom: R-(-)-mandelic acid, 4-chlorophenyl acetic acid, adipic acid, benzoic acid and 4-(bromoethyl)benzoic acid. Produced from a mixture of standards. Separated using an Ascentis Express C18 column (150 mm × 2.1 mm, 2.7 μm) and an Acquity I-Class UPLC system. Acquired using a Waters Xevo TQ-XS triple quadrupole mass spectrometer.



**Figure 41. Normalised chromatograms for standards.** Top to bottom: 2-amino butyric acid, nicotinic acid and 2,2,3,3-tetramethylsuccinic acid. Produced from a mixture of standards. Produced from a mixture of standards. Separated using an Ascentis Express C18 column (150 mm  $\times$  2.1 mm, 2.7  $\mu$ m) and an Acquity I-Class UPLC system. Acquired using a Waters Xevo TQ-XS triple quadrupole mass spectrometer.



**Figure 42. Normalised chromatograms for standards.** Top to bottom: 3,5-dibromo-4-hydroxybenzoic acid, 2-(4-(dimethylamino)phenyl)-2-(piperidin-1-yl)acetic acid and 2-(4-(dimethylamino)phenyl)-2-(piperidin-1-yl)acetamide. Produced from a mixture of standards. Produced from a mixture of standards. Separated using an Ascentis Express C18 column (150 mm  $\times$  2.1 mm, 2.7  $\mu$ m) and an Acquity I-Class UPLC system. Acquired using a Waters Xevo TQ-XS triple quadrupole mass spectrometer.