Anion-induced Zwitterion Formation in Gas-phase Arginine

Edward M. Milner

A thesis submitted to the University of York for the degree of Master of Science by Research

University of York

Department of Chemistry

April 2010

Abstract

Gas-phase studies of biological molecules provide important information not available in solution. In particular, all naturally occurring amino acids are known to be zwitterionic in solution, whereas most adopt canonical structures in the gas-phase. However, complexation with solvent molecules and cations can induce zwitterion formation.

This thesis presents experimental and theoretical results that provide evidence for the stabilisation of the zwitterionic structure of gas-phase arginine by simple anions in the gas-phase. Mass spectrometry (MS) experiments conducted in a quadrupole ion trap (QIT) are presented. In these experiments, the loss of ammonia as opposed to water from the Br^-•Arg cluster suggests that the arginine within the cluster is present in its zwitterionic form. Computational calculations also suggest that the lowest-energy gas-phase structure of Br^-•Arg is one in which the arginine is present in its zwitterionic form. Thus, the experimental and theoretical data both support the idea that the zwitterionic structure of gas-phase arginine may be induced by complexation with anions. This work provides the first experimental evidence for anion-induced zwitterion formation in a gas-phase amino acid within a simple anion-molecule complex.

In addition, experimental results are presented for a series of deprotonated dicarboxylic acids, $RCOO^-$, complexed to neutral arginine in the gas-phase. The results from collisionally induced dissociation MS experiments in a QIT suggest an

enhancement in stability of the $RCOO^-$•Arg cluster for acids that are a similar length to arginine.

CONTENTS

LIST OF FIGURES

Chapter 1: Introduction

Chapter 2: Experimental Methods

Chapter 3: Anion-Induced Zwitterion Formation in Gas-phase Arginine

Chapter 4: Computational Study of Br-•Arg

Appendix 1: XY_n^{z-} solvent clusters in a QIT

Appendix 2: Computational Methods

LIST OF TABLES

Acknowledgements

I would like to thank my supervisor Dr Caroline Dessent for her supervision and financial support, and help in writing and correcting this thesis. I would also like to thank my family and housemates for supporting me throughout this course.

Thanks go to the Centre of Excellence in Mass Spectrometry (CoEMS) at York for use of their instruments. The CoEMS is supported by Science City York and Yorkshire Forward using funds from the Northern Way Initiative.

Special thanks go to Mike Nix, and Chris Taylor for help in mass spectrometry experiments. Finally, I would also like to thank the University of York mechanical workshops and Ed Bergstrom for assistance with maintenance of the LCQ.

Author's Declaration

I certify that the work presented in this thesis was conducted by myself at the university of York, United Kingdom of Great Britain and Ireland.

Signed _____ Date _____

Chapter 1: Introduction

1.1 Studying Biological Ions in the Gas-phase

The ultimate goal of gas-phase studies of biological ions is to compare and extrapolate their properties with those in the solution phase, which can give a wealth of physical information not available in solution. One of the most detailed means of obtaining information that can be related to gas-phase structure is through the use of Infrared (IR) and Ultraviolet (UV) spectroscopy, which can be applied to neutral and ionic molecules. Of course, before gas-phase biological ions or neutrals can be probed, they need to be generated. In this regard, ionisation methods such as Electrospray Ionisation (ESI) and Matrix-assisted Laser Desorption Ionisation (MALDI) are commonly used for the generation of gas-phase ions,[1] whereas supersonic jet expansions, for example, can be used to take neutrals into the gas-phase.[2] More recent advances include the use of He nanodroplets doped with solute molecules.[3,4,5] Due to the nature of the He 'solvent', spectra of the solute molecules are largely unperturbed, meaning that measurements done in this medium are of a similar quality to those obtained in the gas-phase.

UV spectra of neutral biological clusters in the gas-phase can be obtained by the use of one- or two-colour Resonantly Enhanced Multi-Photon Ionisation (REMPI).[6] The first photon excites the cluster to its intermediate excited state, after which there is an enhanced probability that the second photon ionises the molecule. The first laser can then be tuned through the region of interest in order to obtain the UV excitation

spectrum of the molecule, as only resonant transitions into the excited state will allow the second laser to ionise. The technique is mass resolved, as the resulting fragment can then be detected in a mass spectrometer. Alternatively, the more sensitive Laser Induced Fluorescence (LIF) technique can be used to detect the decay of the excited state, though this method lacks mass selectivity. For example, T. S. Zwier took UV spectra of tryptamine and 3-indole propionic acid using one-colour REMPI spectroscopy.[7] The UV spectrum of the latter exhibits peaks due to out-of-plane configurations of the carbonyl side-chain, assigned by UV-UV hole-burning.

REMPI can also be combined with IR for use in the IR-UV double resonance technique, which is a means of obtaining conformer-specific IR spectra.[7,8,9,10,11] An initial UV spectrum is obtained by REMPI which contains transitions due to all conformers. An IR 'pump' laser can then be scanned followed by a UV 'probe' laser fixed on a conformer-specific UV transition in the original REMPI spectrum. The signal due to the UV laser then becomes depleted when the IR laser is resonant with a ground-state vibrational transition. This method thus indirectly gives the conformer-specific ground-state vibrational spectrum of a neutral biomolecule. For example, Y. Inokuchi et al. took the LIF spectrum of jet-cooled L-tyrosine, followed by UV-UV and IR-UV hole-burning to determine conformer specific IR spectra for each of 8 different isomers.[12] These IR spectra suggested the presence of OH•••N interactions between the COOH and NH_2 groups of L-tyrosine for certain isomers. Thus, highly specific conformational information was obtained.

IR spectra of gas-phase ions can be taken with the use high-intensity tunable lasers, such as those discussed above, though the use of IRMPD is necessary to effect the

dissociation of molecules and strongly bound complexes, such as cationic clusters, by bombardment with many resonant photons.[2,13] Anions, separately to cations, can be probed with techniques such as photo-detachment spectroscopy and photo-electron spectroscopy.[14] Both anions and cations are charged, which facilitates their detection in a wide-range of mass spectrometers, such as magnetic sector, time-of-flight (TOF), quadrupolar and Ion Cyclotron Resonance (ICR) mass spectrometers.

Ion mobility experiments can be applied to conformational studies of large gas-phase ions, such as peptides and proteins, and offer an alternative to laser spectroscopy.[15,16] In Ion mobility, a gas-phase ion moves though a buffer gas contained within in a drift tube under the influence of a uniform electric field. The time that it takes the ion to migrate from one end of the drift tube to the other will then depend on the ions mass, charge, size and shape, and thus mass resolved information on molecular conformation can be obtained. For example, Bowers and co-workers obtained the ion mobility spectrum of the protonated gas-phase peptide bradykinin,[17] and from the mobility measurements, were able to establish that this ion adopts a tightly folded conformation in which the peptide folds up around the positive charge which sits on the guanidinium side chain group of one of the arginine residues.

Apart from ion mobility and laser spectroscopy of gas-phase ions, some of the tools available in commercial mass spectrometers, such as Sustained Off-Resonance Irradiation Collision Activated Dissociation (SORI CAD) in a Fourier Transform Ion-Cyclotron Resonance Mass Spectrometer (FTICR-MS),[13] or Collision Induced dissociation (CID) in a Quadrupole Ion Trap Mass Spectrometer (QIT-MS),[18,19] can be used to probe the lowest-energy pathways of unimolecular decay of gas-phase ions,

and thus give indirect information on gas-phase structure. In both SORI CAD in an FTICR and CID in a QIT, the ion is increasingly accelerated in a buffer gas causing an increase in the number of collisions that it undergoes, resulting in a slow deposition of internal energy that eventually leads to dissociation.

With regard to the relevance of gas-phase studies of biological molecules to biological processes, Perdita Barran and co-workers have examined the gas-phase conformations of the peptide gonadotropin-releasing hormone (GnRH) and structural analogues in order to determine their relevance in a real biological system.[20] GnRH binds to a receptor in the anterior pituitary, and historically, thousands of structural analogues of this peptide have been compared, and so the significance of each amino acid in its sequence to the ability of the peptide to bind to its receptor has been extensively examined. It is found that a chiral amino acid at residue 6 significantly reduces binding, which, in turn, is thought to be due to an inability to form a tight $II^1\beta$ turn that configures the peptide for binding. Perdita Barran and co-workers used gas-phase ion mobility measurements in conjunction with molecular dynamics calculations on GnRH and a range of structural analogues. They found that those structural analogues with chiral amino acids at the 6^{th} residue adopted more open structures, whereas GnRH and those without exhibited the tight $II^1\beta$ turn that is a thought to be a prerequisite for binding to the GnRH receptor.

This illustrates the potential importance of studies of gas-phase structure to biological processes. The topic of the research presented in this thesis concentrates on the gas-phase structure of the amino acid arginine, and whether it adopts a canonical or

zwitterionic structure in the gas-phase. This will be discussed in detail in the next section.

1.2 Amino Acids in Solution and in the Gas-phase

The set of naturally occurring amino acids are able to exist as either canonical (neutral) or zwitterionic structures, and in solution most adopt the latter, zwitterionic structure at neutral pH.[21] In the gas-phase however, all of the amino acids adopt the canonical form as the lowest-energy structure.[22,23,24] Thus, the presence of solvent appears to stabilise the zwitterionic state. If amino acids can be introduced into the gas-phase with various water molecules, cations, or anions attached, then their structures can be determined and related to the exact numbers or type of these species that induce zwitterion formation, and ultimately provide insight into the factors that control the balance between whether a particular amino acid adopts a canonical or zwitterionic form under a specific set of conditions.

Theory and experiment predict that the lowest-energy structure of gas-phase glycine is a canonical structure.[22,24,25,26,27,28] M. J. Locke et al. investigated the acid-base properties of glycine using electron impact ionisation, which yielded results that were consistent with this molecule being in its canonical structure in the gas-phase.[23] Further ab initio and millimetre wave spectroscopy experiments confirmed this and predicted an “extended” as opposed to a “cyclic” canonical structure.[25,26,27,28] Further to this, Y. Ding and K. Krogb-Jespersen ruled out the possibility of the zwitterionic structure being a stable minimum on the Potential Energy Surface (PES) with detailed electronic structure calculations.[24]

Initial experiments by Price et al. suggested a zwitterionic structure for arginine in the gas-phase, based upon dissociation kinetics of proton bound arginine dimers.[29] However, IR spectra subsequently obtained by cavity ring-down spectroscopy implied a canonical structure for this amino acid,[30] and numerous subsequent ab initio studies predict the canonical structure to be ~15 kJ mol^{-1} lower in energy than the zwitterionic structure in the gas phase.[31,32,33,34]

Apart from the results for glycine and arginine, millimetre wave spectroscopy on alanine also indicates a canonical structure as the lowest-energy gas-phase structure.[35] Thus, there is strong evidence to support the hypothesis that isolated aliphatic amino acids tend to adopt canonical structures. Indeed, the aromatic amino acids, tryptophan and phenylalanine, have been identified as canonical structures in the gas-phase, according to results obtained using UV and IR ion-dip spectroscopy in investigations by Snoek et al.[36,37]

1.3 Water-Amino Acid Clusters in the Gas-phase

A theoretical study on glycine complexed to three and eight waters revealed a change from the lowest-energy structure being canonical to it being zwitterionic [38] (An earlier computational study has indicated that the zwitterionic structure of glycine could be stabilised by just two water molecules in the gas-phase).[39] Using quite a different approach, K. H. Bowen and co-workers explored the hydration of glycine and tryptophan molecules in clusters of amino acids with water molecules and an excess electron.[40] By binding an excess electron, and increasing the level of hydration, any

increase in the Vertical Detachment Energy (VDE) detected in the Photon-Electron Spectrum (PES) observed in the experiment is due to the interaction of this excess electron with the dipole of a zwitterion. The results indicated that five water molecules are required to transform glycine into its zwitterion, while four are required for both phenylalanine and tryptophan. However, it is possible that the excess electron may be making a contribution to the zwitterion stability, so these numbers are probably underestimates for effecting the canonical to zwitterionic conversion. Later work by E. G. Diken et al. explored the photoelectron spectroscopy of glycine complexed to one and two water molecules (K. H. Bowen and co-workers were not able to synthesise these clusters),[40] and confirmed a non-zwitterionic core for these clusters.[41,42]

Complexation of a single water molecule to arginine is predicted to lower the energy of the energy zwitterionic structure by ~8 kJ mol-1 below that of the canonical structure.[43] Similarly, S. Park et al. carried out calculations for complexation of alanine to 1 to 3 water molecules.[44,45] They found that 2 water molecules were necessary to bring the zwitterionic structure to lower energy than the canonical structure. Thus, much of the evidence points towards the ability of solvent to lower the energy of the zwitterionic structure relative to the canonical structure for various amino acids.

1.4 Cationic Complexation of Amino Acids in the Gas-phase

Apart from solvation, there is an extensive body of work concerning the complexation of amino acids to metal ions in the gas-phase. For glycine, the effect of complexation

to the series of group 1 metal cations was explored using ab initio computational methods by Jensen et al.[46] They examined the lowest-energy structures glycine complexed to H^+, Li^+ and Na^+ by ab initio, and found the clusters preferentially adopted canonical structures. C. Kapota et al. later confirmed this result with IR spectra of glycine complexed to Na^+.[47] Further investigation of metal cation binding to glycine was carried out by Williams and co-workers by theoretical methods.[48] They complexed the series of group 2 metal cations with glycine, and found that for every metal ion in the series, the zwitterionic structure was more stable by 5-12 kcal mol^{-1}, apart from for the smallest cation, Be^+, where the canonical structure of arginine in the cluster was between 5-8 kcal mol^{-1} below the zwitterionic structure. Thus, for glycine, the experimental and computational results seem to indicate that larger metal cations with greater positive charge preferentially stabilize the zwitterionic structure in the gas-phase. This is in agreement with data for complexation of Co^+ with glycine which gives a bidentate canonical structure, whereas complexation with Co^{2+} results in glycine adopting a zwitterionic structure.[49] The same trend was observed for the transition from Cu^+ to Cu^{2+},[50] and Zn^{2+} was also found to lower the zwitterionic structure below the canonical structure according to calculation.[51]

Similar calculations and IR spectra of lysine complexed with the group 1 Li^+, Na^+ and K^+ cations indicate that it is non-zwitterionic in these complexes.[52,53] IRMPD spectra of serine complexed with group 1 metal cations reveal a transition to a zwitterionic structure on complexation with the largest metal cation, Cs^+,[54] whereas the analogous study carried out on threonine reveals a small population of the zwitterionic structure on complexation with Cs^+.[55] The zwitterionic structure of proline is stabilised by Na^+, which is the smallest metal cation, so the preferential adoption of a canonical or

zwitterionic state clearly depends heavily on the amino acid in question.[47] For the aromatic amino acids, IRMPD spectra obtained for tyrosine and phenylalanine complexed to K^+ indicate that these amino acids adopt canonical structures,[56] whereas IRMPD spectra of the much larger and more highly charged Ba^{2+} complexed to tryptophan indicated a zwitterionic structure.[57] Aromatic histidine is similarly shown to be zwitterionic on complexation with Ba^{2+}, though complexation with Ca^{2+} results in a canonical structure.[58] Ba^{2+} because of its large size and high positive charge is an excellent candidate for potentially stabilising zwitterionic structures. Indeed, E. R. Williams and co-workers took IRMPD spectra of Ba^{2+} complexed to arginine, glutamine, proline, serine and valine, and found them all to adopt zwitterionic structures.[59] In a similar fashion, all of glycine, alanine, valine, leucine, iso-leucine, serine, and proline have their zwitterionic structures stabilised when complexed with various methyl substituted ammonium ions, which are larger and more effective stabilisers than even Ba^{2+}.[60]

Arginine is an excellent candidate for potential zwitterion stabilisation by metal cations on account of it having one of the lowest gas-phase acidities,[61] thus the proton affinity (PA) of the group that protonates to form the zwitterion is high. This has been shown to be an important contributing factor to the stability of the zwitterion in cationized valine and glycine clusters.[62,63,64] The zwitterionic and canonical structures of arginine are given in Figure 1.1. Williams and co-workers have complexed arginine to group 1 metal cations, and examined the series Li^+ to Cs^+, using SORI CAD.[65] They observed a change of ammonia loss to water loss on going down the series between the Na^+ and K^+ clusters, and suggested that arginine changes from an initially canonical structure to a zwitterionic structure on going from Na^+ to K^+. They attribute

this to a size effect, due to the favourable interactions that result from the canonical structure being able to effectively envelope small metal cations, such as Li^+ (the canonical structure is 0.7 kcal mol^{-1} lower in energy than the zwitterionic one), which are interactions that are not present in clusters with the larger metal cations, such as Cs^+. Larger metal cations have the effect of raising the energy of the canonical structure relative to the zwitterionic structure (e.g. the zwitterionic structure of arginine is 3.3 kcal mol^{-1} lower in energy than the canonical structure of arginine in the Cs^+•Arg cluster).

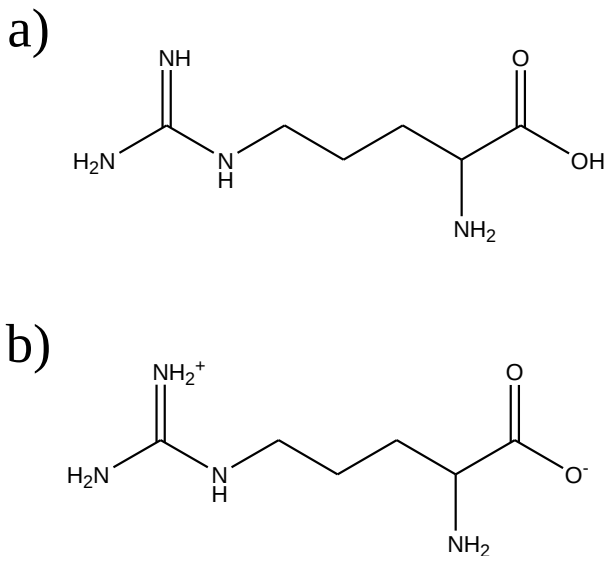

Figure 1. 1 Chemical diagrams of the different structures of arginine, where a) is the canonical structure and b) is the zwitterionic structure.

Further investigation using IRMPD in an FTICR confirm this result (for complexes of arginine with Li^+, Na^+ and K^+).[66] Williams et al. also investigated the group 2 metal cations of Ca^{2+}, Sr^{2+}, and Ba^{2+} complexed to arginine using IRMPD.[59] As the metal ion size is increased, the spectra and calculations indicate that the zwitterionic structure is increasingly stabilised relative to the canonical structure. The spectra

indicate that zwitterionic structures are adopted for arginine complexed with Sr^{2+} and Ba^{2+}, and all three metal cation clusters are predicted to adopt zwitterionic structures according to the calculations.

1.5 Anion Complexation of Amino Acids in the Gas-phase

In contrast to the wealth of information available for complexation with cations, there is very little work that has been carried out examining the effect of anion complexation to amino acids. Calculations by S. R. Kass indicate that zwitterionic structures for glycine are stable when complexed to carboxylate dianions of oxalic and malonic acids.[67] In addition, G. Yang et al. have explored the complexation of glycine with various pairs of anions and dianions using computational methods, and have proposed that stabilisation of zwitterionic glycine occurs when the PA of the anionic species that are complexed with it are in the range of 1567.0-1983.6 kJ mol^{-1} (or the differences of the two binding sites (ΔPA) of the dianion are less than 104.2 kJ mol^{-1}).[68] They point out, however, that if the proton affinity of the anionic stabiliser is too large, it will cause complex instability of the cluster.

With regard to arginine, stabilization of the zwitterion by dipole binding an excess electron has been reported in a computational study by Skuski et al.[69] They show that addition of an excess electron stabilizes the zwitterionic structure of arginine by 31 kJ mol^{-1} and the non-zwitterionic structure by only 15 kJ mol^{-1}, leading to quasi-degeneracy of the two structures, which then differ by less than 1.7 kJ mol^{-1} in terms of energy. Bowen et al. later reported a photo-electron spectrum of arginine bound to an excess electron, consistent with a zwitterionic structure.[70]

Separately from the electron binding studies, it is already known that arginine aggregates can be stabilized by anions.[71] According to Quadrupole Ion Trap Mass Spectrometry (QIT-MS) experiments and Density Functional Theory (DFT) calculations carried out by Beauchamp and co-workers, higher-order clusters of arginine molecules clustered with various anions have zwitterionic arginine forming head-to-tail cyclic arrays focused around a central anion.

1.6 Summary and Thesis Aims

As outlined above, the complexation of amino acids to cations has been explored extensively, but there remains limited information on the effect of anion complexation on the structure of amino acids. To date, there have been no investigations of the basic series of single monoanions complexed to arginine, which is an important gap in the literature. The primary aim of this study is therefore to conduct an analogous study on monomeric arginine complexed to group 7 anions instead of group 1 cations, in order to elucidate the structure of arginine present in simple anion-anion clusters.

L-arginine is characterized as a semi-essential amino acid in the body, though it is the primary source for the generation of nitric oxide by NO synthase, which is important in neuron transmission, the immune system and blood vessel dilation.[72,73] The structural state of arginine in any solvent-free regions might impact on the function of the enzyme NO synthase. More generally, many drugs are present as zwitterionic structures and as a consequence can form strong electrostatic interactions with protein receptors, impacting their function dramatically. Thus, fundamental studies of the

interaction of amino acids with anions might be important and should contribute to our understanding of ion-induced zwitterion stabilization.[68]

This thesis is divided into 7 sections with the next chapter, Chapter 2, describing the experimental methodology employed in Chapters 3 and 5. Chapter 3 presents an experimental study of gas-phase arginine complexed to various anionic species, and Chapter 4 presents computational work to support the experimental results presented in Chapter 3. Chapter 5 describes an experimental study of gas-phase arginine complexed to singly-charged carboxylic acid anions. Finally, Appendices A1 and A2 detail some preliminary work relating to the solvent clustering of small metal anions within a QIT, and details of the computational techniques employed in this thesis, respectively.

References

[1] Desfrancois, C.; Carles, S.; Schermann, J. P. *Chem. Rev.* **2000**, *100*, 3943

[2] Polfer, N. C.; Oomens, J. *Mass Spectrom. Rev.* **2009**, *28*, 468

[3] Choi, M. Y.; Douberly, G. E.; Falconer, T. M.; Lewis, W. K.; Lindsay, C. M.; Merritt J. M. *Int. Rev. Phys. Chem.* **2006**, *25*, 15

[4] Ferreira da Silva, F.; Bartl, P.; Denifl, S.; Mark, T. D.; Ellis, A. M.; Scheier, P. *Chem. Phys. Chem.* **2010**, *11*, 90

[5] Ferreira da Silva, F.; Denifl, D.; Mark, T. D.; Doltsinis, N. L.; Ellis, A. M.; Scheier, P. *J. Phys. Chem A* **2010**, *114*, 1633

[6] Dessent, C. E. H.; Muller-Dethlefs, K.; *Chem. Rev.* **2000**, *100*, 3999

[7] Zwier, T. S. *J. Phys. Chem. A* **2001**, *105*, 8827

[8] Stearns, J. A.; Seaiby, C.; Boyarkin, O. V.; Rizzo, T. R. *PCCP* **2009**, *11*, 125

[9] Baquero, E. E.; James III, W. H.; Choi, T. H.; Jordan, K. D.; Zwier, T. S. *J. Phys. Chem. A* **2008**, *112*, 11115

[10] de Vries, M. S.; Hobza, P. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585

[11] Florio, G. M.; Longarte, A.; Zwier, T. S. *J. Phys. Chem. A* **2003**, *107*, 4032

[12] Inokuchi, Y.; Kobayashi, Y.; Ito, T.; Ebata, T. *J. Phys. Chem. A* **2007**, *111*, 3209

[13] Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. *Analyst* **2005**, *130*, 18

[14] Simons, J. *J. Phys. Chem. A* **2008**, *112*, 6401

[15] Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J.; Noreen, S.; Hill, C. A.; Paul Thomas, C. L. *Analyst*, **2004**, *129*, 984

[16] Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577

[17] Wyttenbach, T.; Helden, G. V.; Bowers, M. T. *J. Am. Chem. Soc.* **1996**, *118*, 8355

[18] Colorado, A.; Brodbelt, J. *J. Am. Chem. Soc. Mass Spectrom.* **1996**, *7*, 1116

[19] Hart, K. J.; McLuckey, S. A. *J. Am. Chem. Soc. Mass. Spectrom*. **1994**, *5*, 250

[20] Barran, P. E.; Polfer, N. C.; Campopiano, D. J.; Clarke D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan J. R. W.; Maxwell A.; Dorin, J. R.; Millar, R. P.; Bowers, M. T. *Int. J. Mass Spectrom.* **2005**, *240*, 273

[21] Stryter, L.; Tymoczko, J. L.; Berg, J. M. *Biochemistry*; 5th ed, W. H. Freeman & Co: New York, NY, 2002

[22] Locke, M. J.; Hunter, R. L.; McIver, R. T. *J. Am. Chem. Soc.* **1979**, *101*, 272

[23] Locke, M. J.; McIver, R. T. *J. Am. Chem. Soc.* **1983**, *105*, 4226

[24] Ding, Y.; Kroug-Jesperson, K.; *Chem. Phys. Lett.* **1992**, *199*, 261

[25] Vishveshwara, S.; Pople, J. A. *J. Am. Chem. Soc.* **1977**, *99*, 2422

[26] Suenram, R. D.; Lovas, F. J. *J. Am. Chem. Soc.* **1978**, *72*, 372

[27] Sellers, H. L.; Schafer, L. *J. Am. Chem. Soc.* **1978**, *100*, 7728

[28] Schafer, L.; Sellers, H. L.; Lovas, F. J.; Suenram, R. D. *J. Am. Chem. Soc.* **1980**, *102*, 6566

[29] Price, W. D.; Jockusch, R. A.; Williams, E. R. *J. Am. Chem. Soc.* **1997**, *119*, 11988

[30] Chapo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K..; Saykally, R. J. *J. Am. Chem. Soc.* **1998**, *120*, 12956

[31] Maksic, Z. B.; Kavacevic, B. *J. Chem. Soc. Perkin Trans. 2* **1999**, 2623

[32] Rak, J.; Skurski, P.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11695

[33] Skurski, P.; Gutowski, M.; Barrios, R.; Simons, J. *Chem. Phys. Lett.* **2001**, *337*, 143

[34] Gdanitz, R. J.; Cardoen, W.; Windus, T. L.; Simons, J. *J. Phys. Chem. A* **2004**, *108*, 515

[35] Godfrey, P. D.; Firth, S.; Hatherley, L. D.; Brown, R. D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1993**, *115*, 9687

[36] Snoek, L. C.; Kroemer, R. T.; Hockridge, M. R.; Simons, J. P. *PCCP* **2001**, *3*, 1819

[37] Snoek, L. C.; Robertson, E. G.; Kroemer, R. T.; Simons, J. P. *Chem. Phys. Lett.* **2000**, *321*, 49

[38] Brandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2002**, *116*, 5023

[39] Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1995**, *117*, 8159

[40] Xu, S.; Nilles, J. M.; Bowen, K. H. *J. Chem. Phys.* **2003**, *119*, 10696

[41] Diken, E. G.; Hammer, N. I.; Johnson, M. A. *J. Chem. Phys.* **2004**, *120*, 9899

[42] Diken, E. G.; Headrick, J. M.; Johnson, M. A. *J. Chem. Phys.* **2005**, *122*, 224317

[43] Im, S.; Jang, S.; Lee, S.; Lee, Y.; Kim, B. *J. Phys. Chem. A* **2008**, *112*, 9767

[44] Park, S.; Ahn, D.; Lee, S. *Chem. Phys. Lett.* **2003**, *371*, 74

[45] Ahn, D.; Park, S.; Jeon, I.; Lee, M.; Kim, N.; Han, Y.; Lee, S. *J. Phys. Chem. B* **2003**, *107*, 14109

[46] Jensen, F. *J. Am. Chem. Soc.* **1992**, *114*, 9533

[47] Kapota, C.; Lemaire, J.; Maitre, P.; Ohanessian, G. *J. Am. Chem. Soc.* **2004**, *126*, 1836

[48] Strittmatter, E. F.; Lemoff, A. S.; Williams, E. R. *J. Phys. Chem. A* **2000**, *104*, 9793

[49] Constantino, E.; Rodriguez-Santiago, L.; Sodupe, M.; Tortajada, J. *J. Phys. Chem. A* **2005**, *109*, 224

[50] Bertran, J.; Rodriguez-Santiago, L.; Sodupe, M. *J. Phys. Chem. B* **1999**, *103*, 2310

[51] Rogalewicz, F.; Ohanessian, G.; Gresh, N. *J. Comp. Chem.* **2000**, *21*, 963

[52] Bush, M. F.; Forbes, M. W.; Jockusch, R. A.; Oomens, J.; Polfer, N. C.; Saykally, R. J.; Williams, E. R. *J. Phys. Chem. A*, **2007**, *111*, 7753

[53] Lemoff, A. S.; Bush, M. F.; O'Brie, J. T.; Williams, E. R. *J. Phys. Chem. A* **2006**, *110*, 8433

[54] Armentrout, P. B.; Rodgers, M. T.; Oomens, J.; Steill, J. D. *J. Phys. Chem. A*, **2008**, *112*, 2248

[55] Rodgers, M. T.; Armentrout, P. B.; Oomens, J.; Steill, J. D. *J. Phys. Chem. A*, **2008**, *112*, 2258

[56] Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; Helden, G.; Oomens, J. *J. Am. Chem. Soc.* **2005**, *127*, 8571

[57] Dunbar, R. C.; Polfer, N. C.; Oomens, J. *J. Am. Chem. Soc.* **2007**, *129*, 14562

[58] Dunbar, R. C.; Hopkinson, A. C.; Oomens, J.; Siu, C.; Siu, K. W. M.; Steill, J. D.; Verkerk, U. H.; Zhao, J. *J. Phys. Chem. B* **2009**, *113*, 10403

[59] Bush, M. F.; Oomens, J.; Saykally, R. J.; Williams, E. R.; *J. Am. Chem. Soc.* **2008**, *130*, 6463

[60] Wu, R.; McMahon, T. B. *J. Am. Chem. Soc.* **2008**, *130*, 3065

[61] O'Hair, R. A. J.; Bowie, J. H.; Gronert, S.; *Int. J. Mass Spectrom. Proc.***1992**, *117*, 23

[62] Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Am. Chem.. Soc.* **2003**, *125*, 13576

[63] Wyttenbach, T.; Witt, M.; Bowers, M. T. *J. Am. Chem. Soc.* **2000**, *122*, 3458

[64] Julian, R. R.; Jarrold, M. F. *J. Phys. Chem.. A* **2004**, *108*, 10861

[65] Jockusch, R. A.; Price, W. D.; Williams, E. R. *J. Phys. Chem. A* **1999**, *103*, 9266

[66] Bush, M. F.; O'Brien, J. T.; Prell, J. S.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc* **2007**, *129*, 1612

[67] Kass, S. R. *J. Am. Chem. Soc.* **2005**, *127*, 13098

[68] Yang, G.; Zu, Y.; Liu, C.; Fu, Y.; Zhou, L. *J. Phys. Chem. B* **2008**, *112*, 7104

[69] Skurski, P.; Rak, J.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11073

[70] Xu, S.; Zheng, W.; Radisic, S.; Bowen, K. H. *J. Chem. Phys.* **2005**, *122*, 091103

[71] Julian, R. R.; Hodyss, R.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2001**, *123*, 3577

[72] Unknown author *Wikipedia entry for nitric oxide synthase* [updated 19 Feb 2010; cited 20 Jan 2010]. Available from http://en.wikipedia.org/wiki/Nitric_oxide_synthase

[73] Remko, M.; Fitz, D.; Rode, B. M. *J. Phys. Chem. A*, **2008**, *112*, 7652

Chapter 2: Experimental Methods

2.1 Electrospray Ionisation Mass Spectrometry

2.1.1 Introduction

In order to study gas-phase ions generated from solution in a mass spectrometer, there must be a means by which these gas-phase ions are transferred from solution into the gas-phase. Electrospray Ionisation Mass Spectrometry (ESI-MS) developed by Fenn. and co-workers,[1] is a highly sensitive atmospheric pressure ionisation technique which confers three important advantages for the study of molecules. The first is that it is able to produce multiply charged anions, which allows the analysis of high molecular weights. The second is that it is a liquid-phase ion source, which allows the study of ions present in solution, and thirdly, and perhaps most importantly, it is a "soft" ionisation technique that can completely preserve the non-covalent interactions of ions that existed in solution.[2] In this work, it is the second and third of these advantages that offers a wealth of possibilities for the study of non-covalent interactions present between the selected gas-phase ions generated from solution.[2]

In electrospray ionisation, a solution is "electrosprayed" through a metal needle into a region held at atmospheric pressure. This electrospray is produced by applying a voltage between a needle tip and a heated metal capillary counter electrode, which is heated to high-temperature (typically 250 °C) in order to prevent the adiabatic formation of clusters. A dry, nitrogen counter-current gas is also applied to prevent

cluster formation. ESI is carried out at atmospheric pressure in air, which prevents gas discharges due to the positive electron affinity of oxygen which is able to mop-up free electrons present in the source region. However, as mass analysers are usually held under vacuum, this necessitates the use of several stages of gating lenses and pumps in order to affect a change from atmospheric pressure to vacuum.[3]

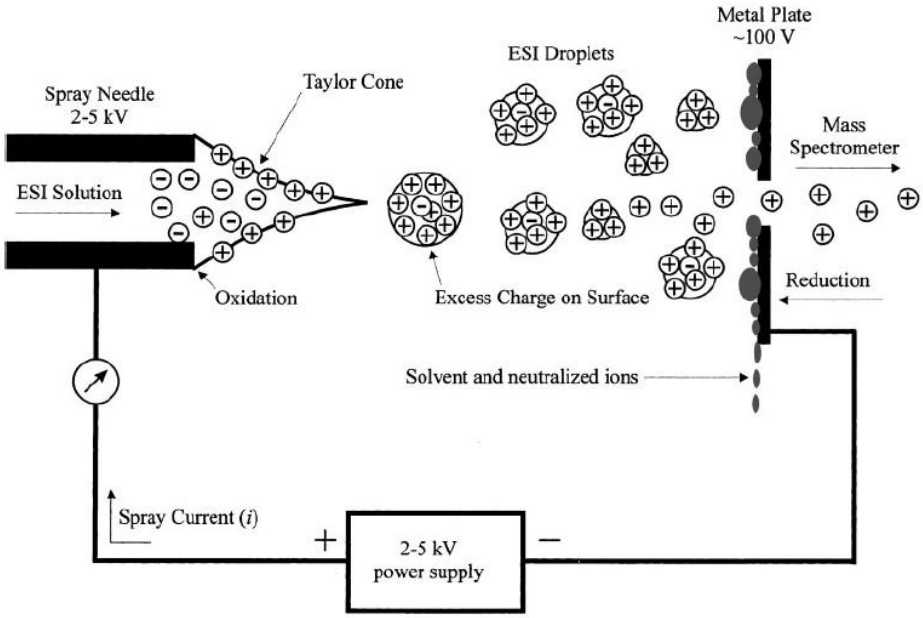

Figure 2. 1 Schematic of the ESI source. A high voltage is applied to the tip of the needle which causes the formation of a Taylor cone containing charged ions at its surface. As the voltage increases, droplets move towards the counter-electrode, losing mass due to evaporation as they go. Coulombic fission events contribute to the eventual formation of charged gas-phase ions that pass into the mass spectrometer. Diagram taken from ref. [4]

The voltage that is applied across the needle tip causes the formation of a Taylor cone on the tip of the needle, which at sufficiently high voltage, deforms up to a point where it emits charged droplets. Evaporation then occurs from these charged droplets as they travel through the ESI region, increasing the charge-to-solvent ratio of each droplet, up to the Rayleigh limit. When a droplet reaches this limit, it deforms and

emits charged jets, or Rayleigh jets, in a process referred to as Rayleigh fission. As a consequence of Rayleigh fission, each charged droplet loses relatively more charge than mass, until finally it turns into a bare gas-phase ion. At present, the exact mechanism for the final production of gas-phase ions is not fully understood, but most agree that there are two possible mechanisms: The Ion Evaporation Model (IEM) and the Charged Residue Model (CRM).

2.1.2 Taylor Cone Charged Droplet Formation

The liquid present at the tip of a needle or metal capillary has a natural surface tension that gives it a characteristic spherical shape, and the needle is necessarily very thin in order to affect a high electric field in the air at its tip.[2] The voltage that is applied sets up a potential difference between this needle tip and a heated capillary, counter-electrode. If the counter-electrode is negative, ions within the solution at the tip will move to cancel this applied field, that is, positive ions will move towards the meniscus of the liquid, whereas negative ions will drift downfield. This process affects the build-up of an electric double layer at the liquid tip surface. As charge gathers at the surface, the repulsive interaction between like charges overcomes the surface tension of the liquid, and the surface expands, forming a Taylor Cone. If the voltage is high enough, a jet is emitted from the surface of this cone which breaks up into small droplets. Thus, if a negative counter-electrode is used, a Taylor cone with positive ions forms which will divide into droplets with excess positive charges at their surfaces, and vice-versa for a positive counter-electrode. The electric potential (V_{on}) that causes instability of the Taylor cone is given by: [2]

$$V_{on} = 2 \times 10^5 \gamma {r_c}^{1/2} \ln(4d/r_c) \qquad [2.1]$$

Where γ is the surface tension, r_c is the radius of the capillary, and d is the distance between the capillary tip and the counter-electrode. Solvents that have high surface tensions, like H_2O, require a higher onset voltage. The onset voltage required for water is so high, in fact, that there is significant risk of electric discharge, making it unsuitable for ESI. When the Taylor cone becomes unstable, it emits charged droplets that have radiuses that are proportional to the flow rate, V_f, and conductivity, κ, of the solution.[2]

2.1.3 Rayleigh Fission of Charged Droplets

The charged droplets produced from the spray shrink owing to extensive solvent evaporation whilst the charge remains constant. Solvent evaporation gets its energy from the thermal energy of the air within the source, and can be aided by the use of highly volatile solvents such as MeCN and MeOH. The point at which the electrostatic repulsion between like charges of a droplet equals the force due to the surface tension of the liquid is given by the Rayleigh stability limit: [2]

$$q_{R_y} = 8\pi(\epsilon_0 \gamma R^3)^{1/2} \qquad [2.2]$$

Where γ is the surface tension of the liquid, R is the droplet radius, and q is the total charge of the droplet. The droplet becomes unstable when its R value and q value satisfy the Rayleigh equation. When they get close to the stability limit, these droplets undergo Coulombic or Rayleigh fission, resulting in the droplet emitting a fine jet of droplets. This fission, however, is uneven, and occurs many times over as the solvent continues to evaporate from the products causing further fission. An initial droplet of

a certain radius and number of elementary charges gives a predictable number of fission cycles. [2]

2.1.4 Ion Evaporation and Charged Residue Models

The observation of droplets smaller than 1 micrometer has not been observed at present,[2] so it is difficult to verify whether fission occurs below this radius. Because of this uncertainty, there are currently two different theories that suggest routes to the final production of gas-phase ions. These are the Charged Residue model (CRM) and the Ion Evaporation Model (IEM) or Iribarne-Thompson Model.

CRM simply predicts that evaporation and fission cycles continue to occur until we have obtained ion-solvent clusters of only one ion on average. The solvent then evaporates from this cluster leaving the naked ion. In contrast, IEM proposes that as the size of the droplet decreases, the field strength at the droplets surface becomes large enough to assist the field desorption of solvated ions. The rate of emission, k_I, for this process is given below: [2]

$$k_I = \frac{k_B T}{h} e^{-\Delta G^{\#}/k_B T} \qquad [2.3]$$

Where T is the temperature, k_B and h are constants, and the free energy of activation, $\Delta G^{\#}$, is a sum of two potential terms. The first is the electric field due to repulsion of the charges remaining on the surface with the escaping ion, and the second term expresses the force of attraction of the escaping charge with the polarisable droplet surface. Thus, the rate of emission is controlled by $\Delta G^{\#}$, and so is dependent on the

balance between the forces of repulsion and attraction of the escaping ion with the droplet surface. [2]

2.2 Quadrupole Ion Trap

2.2.1 Introduction

A Quadrupole Ion Trap (QIT) is a mass spectrometer that makes use of a quadrupolar trapping potential to trap ions in stable orbits before mass selection and analysis. The quadrupole ion trap is constructed from a central hyperbolic ring electrode sandwiched between two hyperbolic end-cap electrodes.[5] A fundamental RF potential comprising a direct and alternating part is normally applied to this ring electrode, which causes the trapping potential. Ions trapped in this potential can then be mass analyzed by either scanning the direct or alternating potentials, or by applying a separate resonant frequency to the end-cap electrodes. Ions within the trap repel each other, which causes their trajectories to expand, necessitating the use of a He buffer gas which quenches the excess energy of these ions and collisionally dampens their motions back to the centre of the trap. Ion losses are significantly reduced as a result.[3] A cross-section of such an ion trap is given in the Figure 2.2 below, labelled according to the key dimensions of the ring electrode radius, r_0, and the distance between the two end-cap electrodes, z_0.

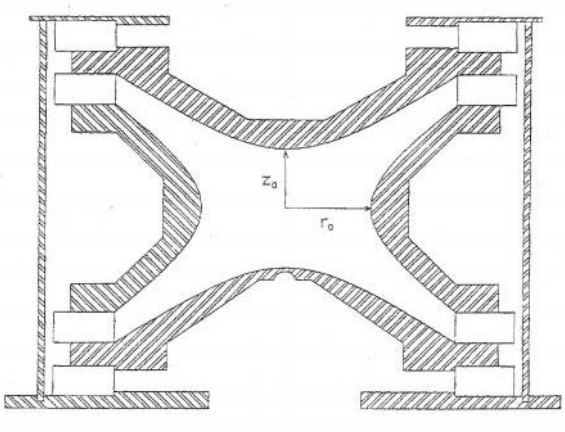

Figure 2. 2 A cross-section of a QIT. The parameters z_0 and r_0 define the dimensions of the hyperbolic ring and end-cap electrodes. Diagram taken from ref. [6]

2.2.2 Basic Theory

The motion of the ions in the trapping field is described by the solutions to the Mathieu equations, expressed in terms of the dimensions of the trap, $u = r_0$ or z_0:[3]

$$q_u = q_z = -2q_r = \frac{8zeV}{m(r_0^2+2z_0^2)\omega^2} \qquad [2.4]$$

$$a_u = a_z = -2a_r = \frac{-16zeU}{m(r_0^2+2z_0^2)\omega^2} \qquad [2.5]$$

Where z represents the charge of the ion, m the mass of the ion, U is the direct potential, and V is the "zero-to-peak" amplitude of the alternating potential. The solutions to the Mathieu equation can be plotted in a_u and q_u space, giving a diagram in which stable trajectories are defined for ions that lie within the triangular region. If $a_z = 0$, for example, ions become unstable when their q_z values are equal to 0.908.

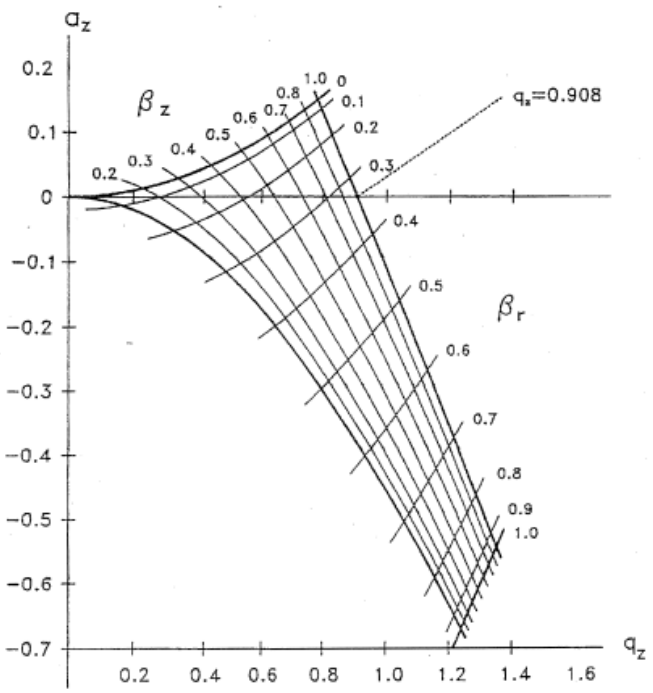

Figure 2. 3 The stability diagram in q(z) and a(z) space. At the edge of the diagram, q(z) equals 0.908, which is the stability limit. Diagram taken from ref. [6]

When the RF potential is applied to the end-cap electrodes of the trap, the ions do not oscillate at the frequency of this potential, but at a reduced "secular" frequency, given by: [3]

$$f_z = \beta_z \nu/2 \qquad [2.6]$$

Where f_z is the secular frequency, ν is the fundamental frequency of the applied potential, and β_z is a dimensionless parameter related to a_z and q_z. The secular frequency of an ion is important because it is utilised in the method of mass analysis by resonant ejection, described below.

2.2.3 Mass Analysis by Resonant Ejection

The simplest method of mass analysis is by reducing the direct voltage, U, to zero, and by varying the alternating potential, V. Thus, the trap is operated along the $a_z = 0$ line and all the masses are spread out over this axis according to their specific q_z values at a certain voltage. Because q_z is proportional to 1/(m/z), lighter ions have higher q_z values and are closer to the instability point where $q_z = 0.908$, and so when the voltage is ramped, all these ions become unstable first. When an m/z becomes unstable, 50% of the ions are expelled along the z-axis and hit the detector, and so calibration of the detector with the voltage ramp then allows the calculation of the mass spectrum. V is variable up to a certain maximum value, which gives us the upper mass limit for detection in any QIT.[3]

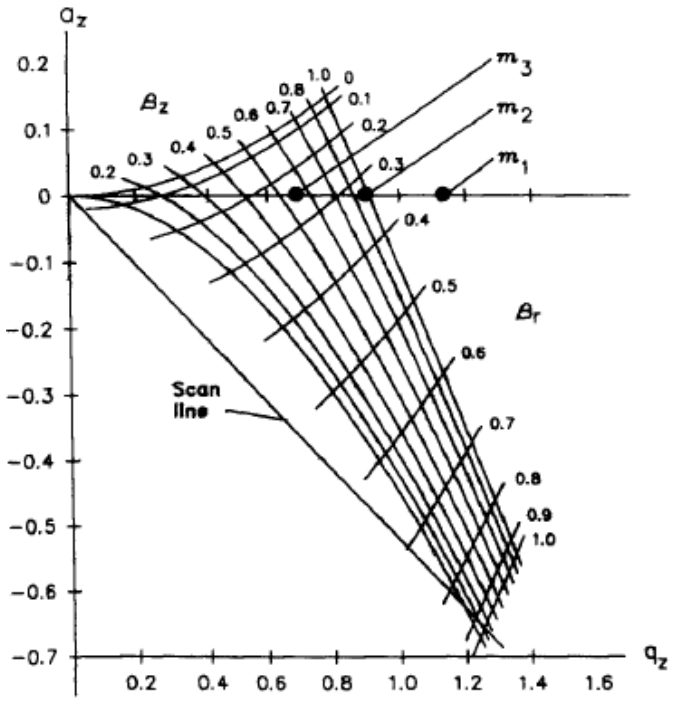

Figure 2. 4 Diagram showing ejection of ions at the stability limit. Ions of different m/z are spaced out along the a_z axis according to their q_z values. In the diagram, $m_3 > m_2 > m_1$, and so m_1 has already been ejected. Increasing the voltage, V, has the effect of moving all ions to higher q(z), and provides a means of ejecting ions from the trap. Diagram taken from ref. [7]

In most modern traps, however, mass analysis is carried out by resonant ejection, which involves applying a voltage to the end-cap electrodes at the secular frequency of the ion. As described previously, each ion oscillates at its own secular frequency in the trap. If an additional voltage is applied with a frequency equal to the secular frequency of a particular ion, then this ion's motion will come into resonance with the applied frequency and the amplitude of its oscillations will increase. If the amplitude of this frequency is large enough, the ion's oscillations will be so large that it becomes unstable and is ejected from the trap. Importantly, it offers a means of mass selection, because we can apply a waveform that ejects all ions apart from the desired one.[3]

2.2.4 Collision Induced Dissociation

A range of smaller amplitude resonant frequencies can be applied that excite the ion of interest as opposed to ejecting it, in a process known as resonant excitation. To do this, all other ions are expelled from the trap by resonant ejection, and then the voltage, V, is adjusted so as to bring the secular frequency of the selected m/z to the same frequency as the RF generator used to resonantly excite. Over a set excitation period, the RF generator voltage amplitude is ramped to increasingly accelerate/decelerate the targeted ion, which increases the number of collisions that it undergoes with the He buffer gas already present in the trap. On collision with a slow-moving He atom, a portion of the kinetic energy of the fast moving ion is transferred into internal energy. Thus, if the amplitude of the resonant frequency is increased, the number of collision increases, which raises the internal energy of the ion.[8,9] As each collision only imparts a very small amount of energy, this results in a slow deposition of internal energy (analogous to a slow heating technique such as Blackbody Infra-

Red Dissociation (BIRD)),[8] that can be used to sample the lowest-energy unimolecular dissociation pathways of the ion. This technique is known as collision Induced Dissociation (CID), and in QITs it can impart between 1-100 eV of internal energy.[10] The minimum energy necessary for the dissociation is known as the critical energy (CE) of the ion, and it is this value that can be estimated from the CID curve.[9] An example of a typical CID or MS/MS sequence is given in Figure 2.5.

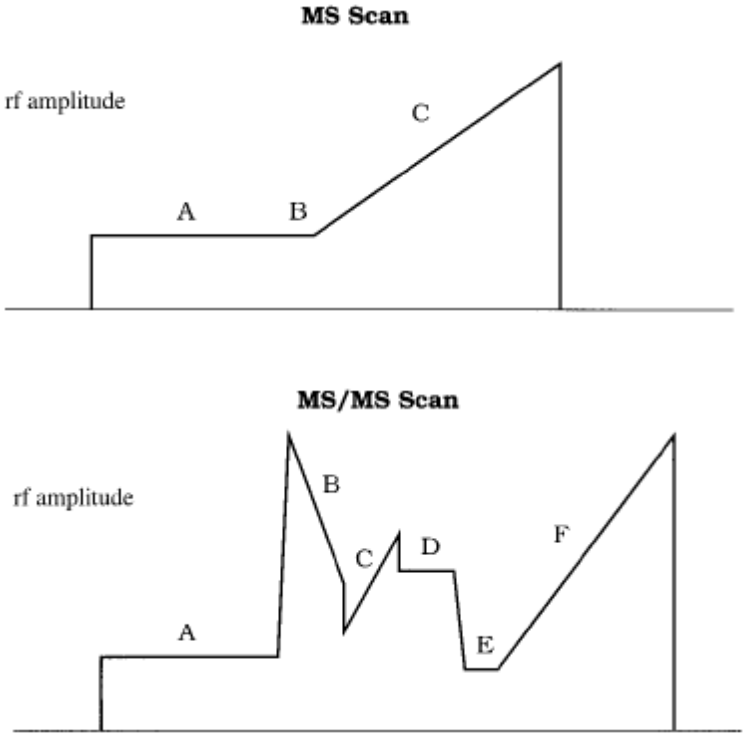

Figure 2. 5 A typical sequence for an MS and an MS/MS scan. For the MS scan, A = ionisations period, B = cooling time and C = Resonance ejection ramp. The MS/MS scan involves, A = ionisation period, B = reverse scan to eject high mass ions, C = forward scan to eject low mass ions, D = resonance excitation period, E = cooling time and F = resonance ejection ramp. Diagram taken from ref. [11]

There are, however, a number of issues that must be considered when taking CID: The meaning of a CID energy, the presence of other variables, and the presence of

products from the excitation. Addressing the first, the CID energy can only be used as a relative measure of the CE of the ion. This is because the kinetic energy imparted to the ions in a QIT is not well defined for all masses and charges due to differences in the degree of acceleration by the voltage ramp, which makes an accurate assessment of the amount of energy imparted during the CID process problematic. It also means that only energies between ions of similar mass (<100 amu) and the same charge should be compared.[12] With regards to the second issue, resonance excitation is also both a function of the excitation resonance duration (typically 30ms in an LCQ) and the pressure of the collision gas. Both need to be kept constant during the course of a run, but the former also needs to be kept low in order to avoid unwanted ion losses which reduce signal strength. Finally, because excitation is carried out at the secular frequency of the precursor ion, this means that dissociation of any products typically does not occur, although this is dependent on the product ions being sufficiently cool to not fragment any further.[3]

In an LCQ QIT, the CID voltage can be varied between 0 and 2.5 V, and a plot of normalised peak intensity of any pairs of peaks related by a mutual fragmentation channel against the end-cap voltage can be produced. CID curves have been used by various groups to assess gas-phase fragmentation routes of ions,[8,9,13,14,15] and an example of such a plot is given in the Figure 2.6.

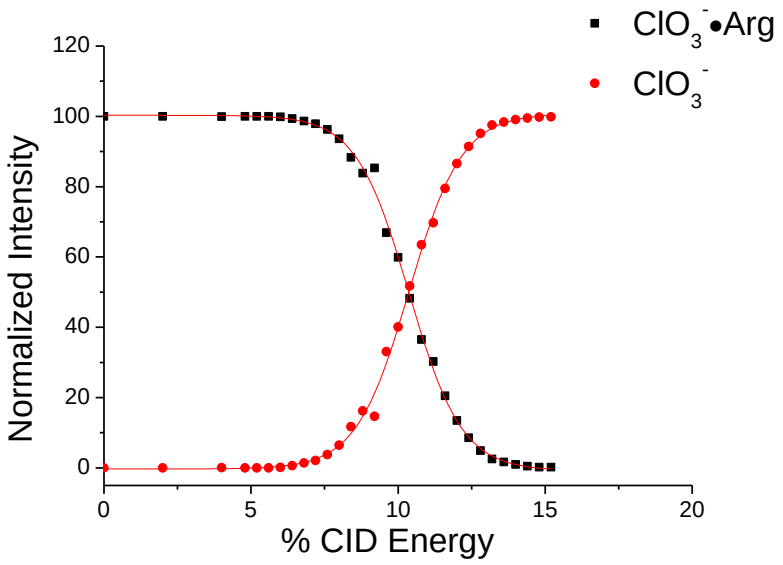

Figure 2. 6 Decay of the precursor ion ClO_3^-·Arg and production of ClO_3^- ion plotted as a normalized percentage of total ion count, against percentage CID energy. Each point is an average of approximately 100 points, recorded by the instrument over a period of 1 minute.

In the above example, the precursor ion, ClO_3^-·Arg, fragments into the product ions, Arg and ClO_3^-, according to the following equation:

$$ClO_3^-\cdot Arg \rightarrow Arg + ClO_3^- \qquad [2.7]$$

The CID voltage is quoted as a percentage of the maximum voltage, and is termed the CID energy. Thus, fragmentation of 50% of the precursor ion ClO_3^-·Arg to ClO_3^- plus Arg requires a 10% CID energy, which corresponds to an excitation voltage of 0.25 V in the instrument.

With regard to the typical errors of estimated CE values for this technique, Brodbelt and co-workers established linear calibration curves for threshold activation voltages

against literature CE values for ions with well-defined CE values, at various excitation times and q_z values.**Error! Bookmark not defined.** From these calibration curves, estimated CE values of new ions could then be made with an accuracy to within ± 6 kcal mol^{-1} for most ions (for CE values typically ranging from 30-50 kcal mol^{-1}). However, they found it necessary to use separate calibration curves for their series of proton-bound dimers (hydrogen-bonded complexes) and radical ions (covalently bonded). This latter point has significance, as it implies that CID curves for ions that are bonded differently should not normally be compared.

References

[1] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

[2] Cole, R. B. *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*; John Wiley Sons, INC: New York, NY, 1997

[3] Hoffmann, E. D.; Stroobant, V. *Mass Spectrometry: Principles and Applications*; 2nd Ed, John Wiley & Sons, LTD: Chichester, 2002

[4] Cech, N. B. *Mass Spectrom. Rev.* **2001**, *20*, 362

[5] March, R. E. *J. Mass Spectrom.* **1997**, *32*, 351

[6] March R. E. *Int. J. Mass Spectrom.* **2000**, *200*, 285

[7] Todd, J. F. J. *Mass Spectrom. Rev.* **1991**, *10*, 3

[8] Hart, K. J.; McLuckey, S. A. *J. Am. Chem. Soc. Mass. Spectrom.* **1994**, *5*, 250

[9] Colorado, A.; Brodbelt, J. *J. Am. Chem. Soc. Mass Spectrom.* **1996**, *7*, 1116

[10] Sleno, L.; Volmer, D. A. *J. Mass Spectrom.* **2004**, *39*, 1091

[11] Jonscher, K. R.; Yates, J. R. *Anal. Biochem.* **1997**, *244*, 1

[12] Satterfield, M.; Brodbelt, J. S. *Inorg. Chem.* **2001**, *40*, 5393 and references therein

[13] Satterfield, M.; Broadbelt, J. S. *Inorg. Chem.* **2001**, *40*, 5393

[14] Rogalewicz, F.; Hoppilliard, Y.; Ohanessian, G. *J. Mass Spectrom.* **2000**, *195/196*, 565

[15] McLuckey, S. A.; Glish, G. L.; Berkel, G. J. *Int. J. Mass. Spectrom. Ion Proc.* **1991**, *106*, 213

Chapter 3: Anion-Induced Zwitterion Formation in Gas-phase Arginine

3.1 Introduction

As previously described in the introduction to this thesis, studies of gas-phase biological ions by techniques such as IR spectroscopy, Ion mobility or CID coupled with mass spectrometry can provide important information relating to gas-phase structure of these ions, which can ultimately be related to the structures they adopt in the body or in biological processes.[1] Of particular interest are the factors that affect zwitterion formation in amino acids, given that most solution phase amino acids are zwitterionic,[2] whereas most bare, neutral amino acids are predicted or found to adopt their canonical structures in the gas-phase.[3,4,5,6,7,8] Complexation of neutral amino acids with metal cations [9,10,11,12,13,14,15,16,17] or water molecules [18,19,20,21] can then bring the zwitterionic structure of the amino acid to lower energy than the canonical structure within the cluster. Complexation with anions or electrons may also have the same effect, though there is comparatively a more limited range of information available in the literature in contrast to the situation with metal cations or complexation with water molecules.[22,23,24,25]

In view of this, it was decided to investigate the complexation of an amino acid with a series of monoanions, in an analogous fashion to the cation complexation studies of Williams and co-workers.[26] A suitable choice for an amino acid to study is arginine,

given that it is one of the most basic amino acids,[27] meaning that the proton affinity of the group that accepts a proton to become a zwitterion is high, which has been shown to be an important contributing factor to the zwitterionic stabilisation of various amino acids by cations.[28,29,30] In the gas-phase, arginine it is found to adopt the canonical structure by experiment, and its canonical structure is calculated to be 15 kJ mol^{-1} lower in energy than its zwitterionic structure (both structures are presented in Figure 3.1 below).[6,31]

This study will be conducted using Electro Spray Ionization (ESI), a technique developed by Fenn and co-workers,[32] coupled to a Quadrupole Ion Trap (QIT), as detailed in Chapter 2. The fragmentation patterns observed upon Collision Induced Dissociation (CID) for the X^-•Arg system will be investigated as they could provide a signature as to whether arginine is canonical or zwitterionic within the cluster. Additionally, by increasing the CID energy, important information about the lowest-energy fragmentation channels available to these molecular clusters can be obtained.[33,34]

Thus, CID of the singly charged X^-·Arg series, with a number of anions ranging in gas-phase basicity and size where X = F, Cl, Br, I, ClO_3 and NO_3 are presented. This series exhibits a deliberate change in proton affinity (PA), as the PA or ΔPA between two monoanions or two binding sites on the same dianion is predicted to be the determining factor in the stabilization of glycine zwitterions by two monoanions or a dianion.[24] A computational study will be presented in the next chapter to complement the experimental work presented here, following the method for dealing with conformationally complex molecules, previously developed by Burke et al.[35]

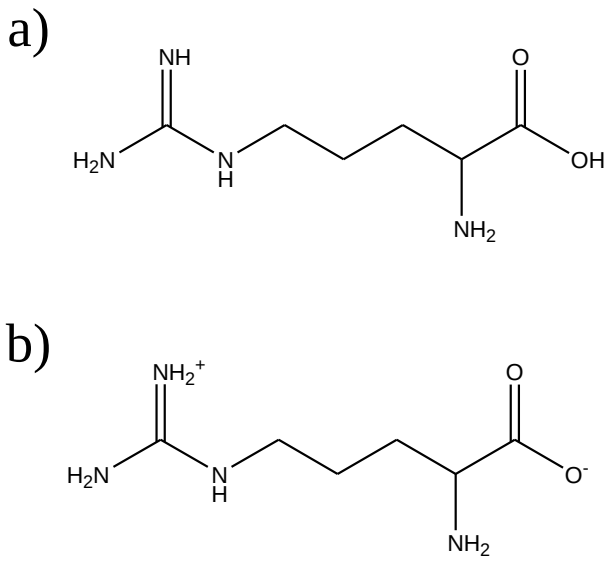

Figure 3. 1 Chemical diagrams of the different structures of arginine, where a) is the canonical structure and b) is the zwitterionic structure.

3.2 Experimental Section

The experiments were carried out on the apparatus described in the experimental methods section of this thesis. The anions were prepared by electrospraying 100ml solutions of the anion salt and L-arginine (both at 10^{-4} M concentration) in 100% MeOH. Simple salts such as Na^+X^- and K^+X^- were used for all anions, apart from F^- which was introduced using tetrabutylammoniumfluoride (TBA-F). For F^-·Arg, a few ml of water was added to the methanol solution to help dissolve the TBA-F. All chemicals were purchased from Sigma Aldrich and required no further purification.

Mass spectrometry experiments were carried out on a Bruker Esquire 6000 QIT, with the nebulising gas set at a pressure of 10.0 psi, drying gas set at a flow rate of 8 $lmin^{-1}$,

the capillary temperature set to 100 °C, and an excitation time of 40 ms. It is worth noting that the spectral resolution of the Bruker is approximately 0.3 mass units. CID graphs were recorded typically over 20-30 minutes, with each step in voltage (CID Energy) given a 1 minute time-bin. These time-bins of data points were then averaged to produce an individual intensity point. The intensity for all peaks present was then normalized and plotted against the CID voltage converted to a % CID Energy.

3.3 Results and Discussion

CID mass spectra for all the species studied are presented in Figure 3.2 (each at a particular collision voltage), with Table 3.1 summarising the data of Figure 3.2. Figure 3.3 gives the CID curves for each of the species under study. Table 3.2 lists the corresponding gas-phase basicities of all species under study, and Table 3.3 lists the radii and binding energies to water of all of the anions in the study.

In comparison to the analogous cationic clusters, the CID of the anionic clusters are more complicated, due to the possibility of proton abstraction, the products of which are readily observed in the mass spectra. For example, the complexes have the potential to fragment into $[Arg - H]^-$ (i.e. equivalent to $[Arg - H^+]^-$, which is deprotonated arginine, and denoted as Arg^- throughout the text and figures) and HX, and so if we assume that charge was initially on X^-, then a proton has been abstracted from arginine and HX has fragmented from the cluster:

$$X^-\bullet Arg \rightarrow HX + [Arg - H]^- \qquad [3.1]$$

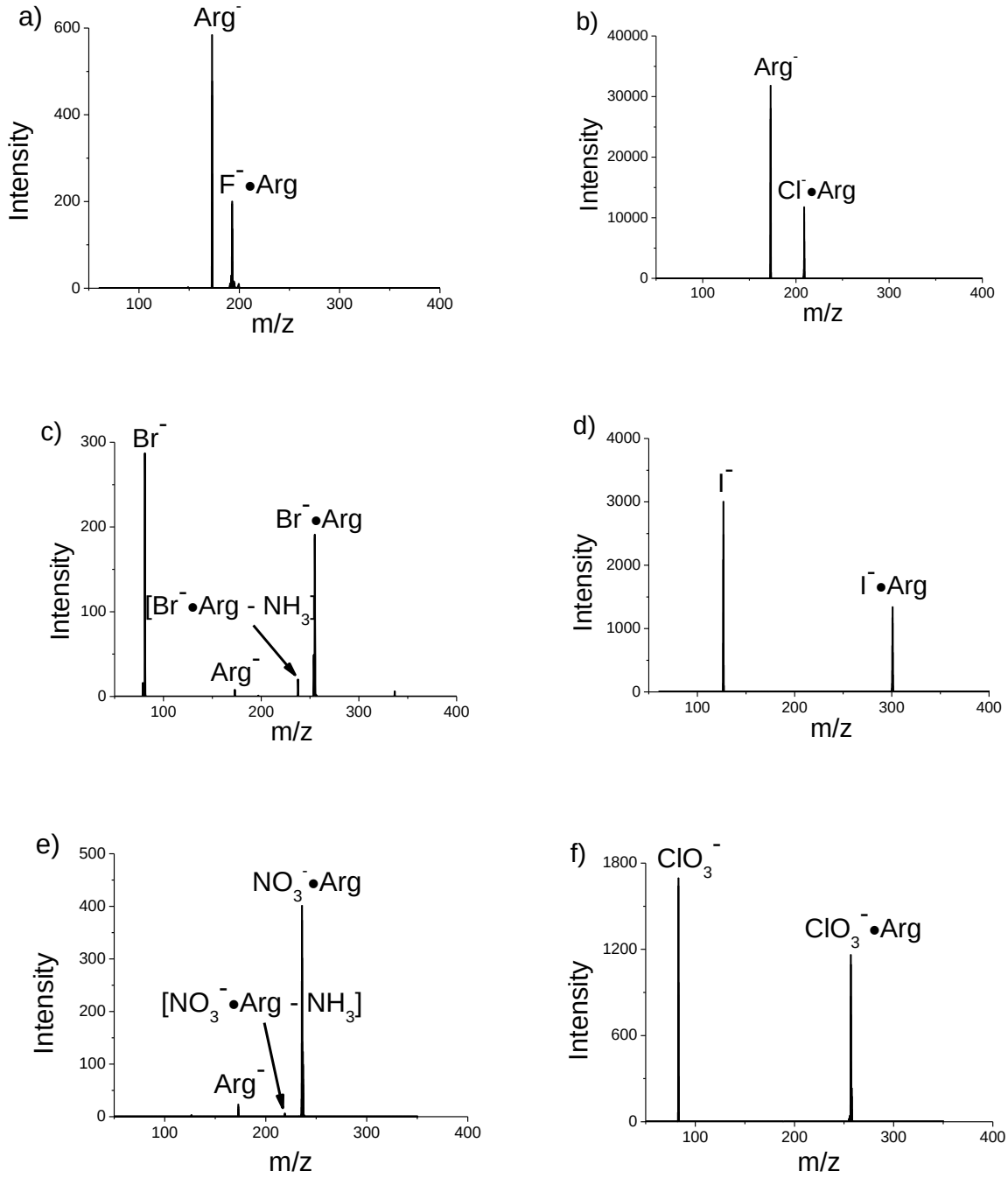

Figure 3. 2 Negative ion ESI-CID mass spectra obtained from collisionally activated dissociation of X^-•Arg for a) X = F at 0.10 V collision energy, b) X = Cl at 0.30 V, c) X = Br at 0.33 V, d) X = I at 0.28 V, e) X = NO_3 at 0.30 V and f) X = ClO_3 at 0.26 V.

Additionally, the anionic systems appear to be much more unstable to dissociation within the QIT, as the signal for most anions decayed rapidly as the CID voltage was raised. The same experiment performed on the related cationic clusters leads to little signal degradation over the same time period. It is possible that the signal loss observed for the anions studied can occur as a result of electron loss. Given this intrinsic instability, anion m/z peaks of interest were optimized using the optics optimization software available on the Bruker Esquire to increase the signal intensity prior to the CID experiment.

As can be seen from Figure 3.2 and Table 3.1, we observe a number of product ions from which we can infer a number of different decay routes, given the charges on the products and initial charge of the complex. These are:

$$X^-\bullet Arg \rightarrow HX + [Arg - H]^- \qquad [3.2]$$

$$X^-\bullet Arg \rightarrow X^- + Arg \qquad [3.3]$$

$$X^-\bullet Arg \rightarrow [X^-\bullet Arg - H_2O] + H_2O \qquad [3.4]$$

$$X^-\bullet Arg \rightarrow [X^-\bullet Arg - NH_3] + NH_3 \qquad [3.5]$$

Equation 3.2 represents loss of HX from the cluster, with production of Arg^-. Assuming that the excess charge is initially on the anion, this corresponds to proton abstraction from arginine. Alternatively, equation 3.3 represents loss of the anion X^-, corresponding to simple fragmentation of the cluster. Equations 3.4 and 3.5 then

correspond to loss of a molecule of H_2O and NH_3 respectively, with X^- staying bound to the remaining fragment of arginine.

Table 3. 1 Product fragment ions present in the CID-mass spectra of the X^-•Arg clusters displayed in Figure 5.2 a-e.

Cluster	Fragmentation channel			
	1. HX loss	2. X^- loss	3. H_2O loss	4. NH_3 loss
a) F^-•Arg	✓	×	×	×
b) Cl^-•Arg	✓	×	×	×
c) Br^-•Arg	✓	✓	×	✓
d) I^-•Arg	×	✓	×	×
e) NO_3^-•Arg	✓	×	×	✓
f) ClO_3^-•Arg	×	✓	×	×

The appearance of Arg^- and X^- are important in this set of fragmentation processes, since these represent a balance between the strength of binding of X^- to arginine, and the ability of X^- to proton abstract from arginine. Fragmentation channels [3.4] and [3.5] correspond to NH_3 and H_2O loss respectively, and will be discussed later in this chapter in the context of the canonical versus zwitterionic structure of arginine within the X^-•Arg clusters.

The data presented in Figure 3.2 illustrate that both F^-•Arg and Cl^-•Arg decay primarily via fragmentation channel [3.2], Br^-•Arg gives a very small amount of this proton abstraction product, whereas it is not observed in I^-•Arg. This simple trend can be straightforwardly related to the basicity of the anion (Table 3.2).

Table 3. 2 The Gas-Phase Acidity (GPA) of the acids of the anions studied in this chapter. The Gas-Phase Basicity (GPB) and Proton Affinity (PA) of arginine are also given.[1]

Species	GPA		PA	GPB
	ΔH (kJ mol^{-1})	ΔG (kJ mol^{-1})	$-\Delta H$ (kJ mol^{-1})	$-\Delta G$ (kJ mol^{-1})
HF [36]	1554.0 ± 8.8	1529.0 ± 8.4	-	-
HCl [37]	1394.9 ± 0.42	1372.8 ± 0.84	-	-
arginine [27,38]	1388 ± 13	1359 ± 13	1051.0	1006.6
HNO_3 [39]	1357.7 ± 0.84	1329.7 ± 0.84	-	-
HBr [40]	1353.5 ± 0.42	1331.8 ± 0.84	-	-
HI [41]	1315.0	1293.7 ± 0.84	-	-
$HClO_3$ [42]	1200 ± 25	-	-	-

The data presented in Table 3.2 allows us to predict whether X^- should proton abstract from neutral arginine (Note that the measurements are made for canonical arginine, because they are gas-phase measurements and arginine is expected to be canonical in the gas-phase). The trend is that F^- and Cl^- are predicted to proton abstract as both HF and HCl have gas-phase acidities above that for arginine, whereas Br^- and I^- are predicted not to proton abstract as both HBr and HI have gas-phase acidities that lie below that for arginine. Indeed, this is in agreement with the results observed in the mass spectra. A small amount of the protonation product is observed for Br^- as stated above, but CID is a collisionally activated process meaning that pathways that are slightly higher in energy may still be allowed to some small degree. Thus, if both loss channels are observed, they must be similar in energy to one another, but it is the product in greatest abundance at equilibrium that is lowest in energy (which is production of Br^- for Br^-•Arg).

[1] Note: Gas-Phase Acidity (GPA) is for $AH \rightarrow A^- + H^+$, Proton Affinity (PA) is for $A^- + H^+ \rightarrow AH$ or $B + H^+ \rightarrow BH^+$, and Gas-Phase Basicity (GPB) is for $A^- + H^+ \rightarrow AH$ or $B + H^+ \rightarrow BH^+$.

This leads to a discussion of fragmentation route [3.3], which is simple fragmentation of the cluster. Experimentally, this corresponds to the observation of X^- as a CID product. This is the dominant pathway that occurs for Br^- and I^-, which do not have a high enough proton affinity to proton abstract from arginine, so that proton abstraction must occur at a higher energy compared to simple fragmentation for these complexes, and therefore, is not observed.

CID within a QIT gives the lowest-energy pathway available to a cluster, and is also collisionally activated, so the products observed from CID will correspond to those loss channels that are similar in energy, with the lowest-energy loss channel corresponding to the fragment seen in highest abundance.[43] Thus, any loss channels that are too high in energy will not be observed. It also means that the CID energy can be used comparatively between species as a measure of relative energy.[33,44,45] For example, if two molecules, 1 and 2, are both predicted to decay by routes A and B, and yet only A is observed for 1 and only B observed for 2, then in case 1, B cannot be observed because A<B, and similarly, in case 2, A cannot be observed because B<A. Thus, the relative energies of routes A and B change on going from molecule 1 to molecule 2, so much so that they crossover. In a similar vein, if both molecules were to decay exclusively by route A, then the energy of route A could be compared relatively between the two. The CID curves for each species are presented in Figure 3.3 below, and given the above argument, various comparisons can be made.

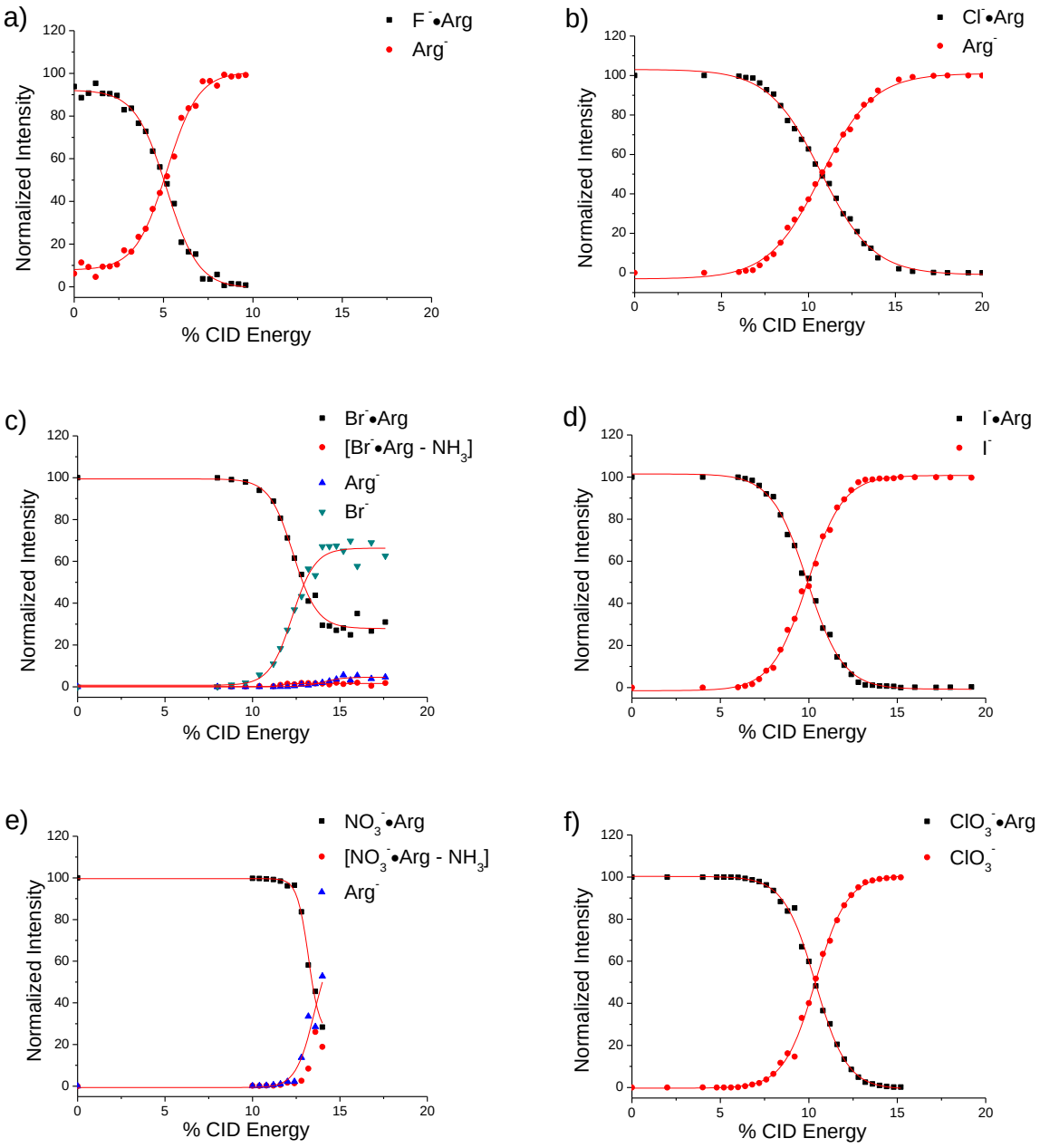

Figure 3. 3 % Fragmentation curves (CID) for decay of X^-.Arg, where a) X = F, b) X = Cl, c) X = Br, d) X = I, e) X = NO_3 and f) X = ClO_3 into X^-, Arg^- and [X^-•Arg-NH_3] product ions, obtained by tuning on the corresponding m/z value in the main ESI mass spectrum for each species. Note that for c), the spectrum has been tuned on the Br^{81} isotope, as isolating on the averaged mass leads to inefficient excitation of both isotopes. Similarly, the corresponding Cl^{35} isotope is isolated in the case of chloride and chlorate complexes, for the same reason.

However, it must be noted that there is a mass limitation to comparison for CID curves, given that larger masses will experience slightly different fields. This mass limitation is that the anions must be of the same charge and have a mass difference of less than 100 amu, for the effect to be negligible (larger masses experience singly higher collision energies).[46] The largest mass difference in the study is between F^-•Arg and I^-•Arg, which is ~108 amu. This is slightly above the limit, but not significantly.

Comparison of the CID energies for the halides reveals an interesting trend in the energy between routes [3.2] and [3.3] (Figure 3.4). Both F^-•Arg and Cl^-•Arg can be compared, as the route of proton abstraction is lower for F^-•Arg than Cl^-•Arg, in keeping with F^-•Arg having the higher basicity. Br^-•Arg and I^-•Arg both go by the cluster fragmentation channel [3.3], so can also be compared with each other. It is now the weaker base I^-, that fragments at lower energy than Br^-, in keeping with the formers larger size and hence weaker binding in the cluster. Thus, the overall trend in the CID energies is a supposition of two different loss channels and is F^-•Arg < I^-•Arg < Cl^-•Arg < Br^-•Arg.

NH_3 loss, route [3.5], is only observed for Br^-•Arg, and the relationship of proton abstraction to fragmentation plays an important role in this. Since Br^- has a lower basicity than Cl^-, but also binds more strongly than I^- to the arginine molecule, NH_3 loss, route [3.5], is observed. This is because NH_3 loss is of a similar energy to both pathways of proton abstraction and fragmentation in Br^-•Arg, whereas either one of these processes is comparatively more efficient with the other halogen complexes.

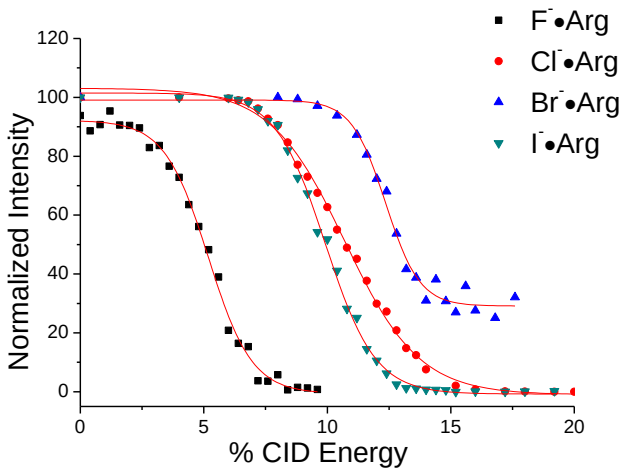

Figure 3. 4 Combined % Fragmentation curves (CID) for decay of the parent ion X^-•Arg ion for the series of halogens from Figure 3.3.

With this in mind, it was of interest to repeat the measurements using anions that have proton affinities that lie below that of Arg^- like that of Br^-, to try and avoid the deprotonation pathway, and that were of similar or smaller ionic radius, to try and avoid the X^- loss pathway, in an effort to observe NH_3/H_2O loss. Both NO_3^- and ClO_3^- fit the criteria of having similar basicities to Br^-. Experimentally, NO_3^- (Figure 3.3 e) gave a small amount of the deprotonation product and a small amount of the NH_3 loss product, as with Br^-, but the fragmentation product was absent. ClO_3^- (Figure 3.3f) dissociated exclusively by the simple fragmentation route to give ClO_3^- as the single product ion.

Clearly, fragmentation is the dominant pathway for anions larger than Br^-, and to examine this trend, radii of the anions and binding energies of the anions with H_2O are compiled in Table 3.3. For the series of halogens, the order of radii going from smallest to largest is $F^- < Cl^- < Br^- < I^-$, which is mirrored in the order of their binding energies with H_2O, $F^- > Cl^- > Br^- > I^-$. Thus, the observation that I^- exclusively

fragments with no NH_3 loss product is perhaps not surprising, given that it has the lowest binding energy with H_2O and largest ionic radius of the series. Both ClO_3^- and NO_3^- are predicted not to proton abstract from arginine (Table 3.2), but the NH_3 loss channel is only observed for NO_3^-. NO_3^- has a similar radius to Br^-, but binds more strongly to H_2O, on account of it being bidentate. Thus the combination of strong binding energy and low basicity make the NH_3 loss channel competitive in comparison to the other anion clusters, and that is why it is observed.

Table 3. 3 Radii and binding energies with H_2O for all anions in the study. Ionic radii have been used where appropriate,[47] and the radii of chlorate and nitrate anion were estimated.[1]

Anion	Radii in Å	ΔH for $X^- + H_2O \rightarrow X^-\bullet H_2O$ (kJ mol^{-1})
F^-	1.33 (47)	114.6 ± 2.1 (48)
Cl^-	1.81 (47)	61.5 ± 4.2 (49)
Br^-	1.96 (47)	49.0 ± 1.7 (50)
I^-	2.2 (47)	43.1 ± 1.3 (50)
NO_3^-	1.92 (51)	59.0 ± 6.3 (52)
ClO_3^-	2.34(51)	55.2 ± 6.3 (52)

However, ClO_3^- has a lower binding energy to water than NO_3^-, but it is still higher than for both Br^- and I^-. Given this, it is interesting then that this anion exclusively gives the fragmentation product at the same E_{half} (the CID energy at which 50% of the precursor ion has decayed) as I^-, making the NH_3 loss channel uncompetitive. This obviously does not fit the trend. However, the chlorate-arginine interactions present in

[1] An approximation to the radii for nitrate and chlorate was made by adding the atomic radius of oxygen (0.66 Å, ref. [49]) to the N/Cl-O bond length, calculated by a geometry optimization on the anion at the B3LYP 6-311++G** level of theory calculated in SPARTAN.

the ClO_3^-•Arg cluster may be abnormally weak in comparison to its interactions with H_2O (i.e. ClO_3^- may exhibit a significant change in how it binds on going from H_2O to arginine, possibly due to it having the largest radii in the series of anions), thus making its binding energy with H_2O a bad indicator for its binding energy to arginine.

Finally, for all of the clusters in this study, the formation of Arg^-•••XH, is unlikely, because such a species would probably be weakly bound and would therefore fragment at low energy, leading to spectra dominated by Arg^-. Of course, both F^-•Arg and Cl^-•Arg give Arg^- at low CID energy, and certainly for these two anions, this possibility cannot be ruled out.

3.3.1 Summary

In summary, the CID loss routes available to each X^-•Arg cluster can be discussed using a supposition of the ideas discussed in the above. For F^-•Arg and Cl^-•Arg**,** low-energy proton abstraction masks fragmentation and any other possible NH_3/H_2O loss routes upon CID. Comparing F^-•Arg to Cl^-•Arg, the proton abstraction loss channel is lower in energy for the former in keeping with the higher gas-phase basicity of the F^- anion than the Cl^- anion. However, there exists the possibility that either one of them is present as Arg^-•HX, so what we are in fact observing is low energy fragmentation and not proton abstraction.

Br^-•Arg has the highest energy CID curve, in keeping with the fact that it does not proton abstract or fragment particularly efficiently, with decay of the parent anion occurring at ~12.5 % CID energy. Thus, NH_3 loss is observable with this anion, and not with the other halides. NH_3 loss is suggestive of the zwitterion, and will be

discussed in the next section. I^-•Arg fragments at lower energy to Br^-•Arg, which is consistent with its larger size and weaker binding energy to H_2O. Thus, any possible NH_3 loss channel is too high in energy to be competitive, despite the fact that proton abstraction is predicted to be particularly unfavourable.

NO_3^-•Arg is analogous to Br^-•Arg, in that it gives the NH_3 loss product in its CID spectrum. A small amount of proton abstraction by NO_3^- is observed as with Br^-•Arg, although fragmentation is completely absent. This proton abstraction occurs at sufficiently high CID energy (12.5 %), thus opening up the NH_3 loss channel. Examining the tables of radii and gas-phase basicity, it is clear why NH_3 loss is observed: NO_3^- is approximately the same size as Br^- and has a larger binding energy with H_2O, but has a gas-phase basicity that is lower than Arg^-. The larger binding energy for NO_3^- than Br^- may explain why fragmentation is completely absent in the former when compared to the latter.

ClO_3^-•Arg decays exclusively via simple cluster fragmentation, and at very similar CID energy to I^-•Arg. This result is difficult to rationalize, given its stronger binding energy to H_2O than both I^- and Br^-, and lower basicity than Arg^-: This anion cluster should fragment at higher energy in comparison to Br^-•Arg, which would make the NH_3/H_2O loss product observable. However, ClO_3^- is the same size as I^- and bidentate, so it may be that the mode of binding of ClO_3^- to water might be different than to arginine, leading to abnormally weak binding in the latter cluster.

3.4 Further Discussion

Experimentally, 4 different collision activated reaction channels are observed, given by equations [3.2] → [3.5]. However, it is NH_3/H_2O loss that is of the most significance, being indicative of a rearrangement involving arginine (equations [3.4] and [3.5]). To further investigate the loss of NH_3 from the Br^-•Arg and NO_3^-•Arg complexes, a CID spectrum was recorded for $[Arg•H]^+$ (i.e. protonated neutral arginine. Throughout the text it will be abbreviated to Arg^+) and one for Arg^-. It was found that Arg^- loses cyanamide and ammonia, whereas Arg^+ loses both water and ammonia. Obviously, water loss cannot occur from Arg^-, but ammonia loss can occur from both Arg^+ and Arg^-. Although NH_3 loss should come from the guanidinium group in both cases, the mechanism of loss is predicted to be different in each case, given that the guanidinium group is protonated in Arg^+ and not in Arg^-. The CID spectra for both species are presented in Figure 3.5.

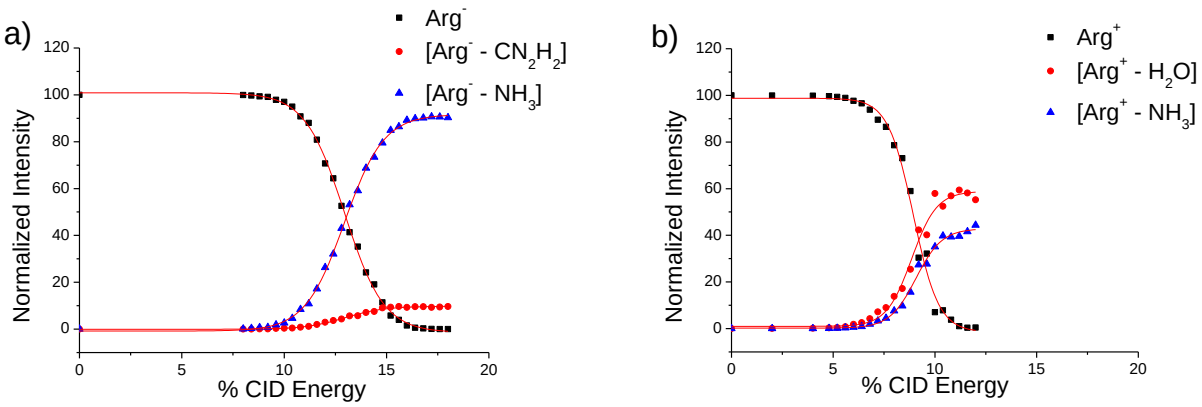

Figure 3. 5 % Fragmentation curves (CID) for decay of a) Arg^- and b) Arg^+, obtained by tuning on the corresponding m/z value in the main ESI mass spectrum for each species.

With regard to NH_3 and CN_2H_2 loss from Arg^-, both can be lost by the same intramolecular rearrangement, and the two pathways are related by symmetry (Figure

3.6). This is likely to be the case, given that they are both present as products in the same CID and occur at similar energy, though the relative proportions indicate that ammonia loss is favoured.

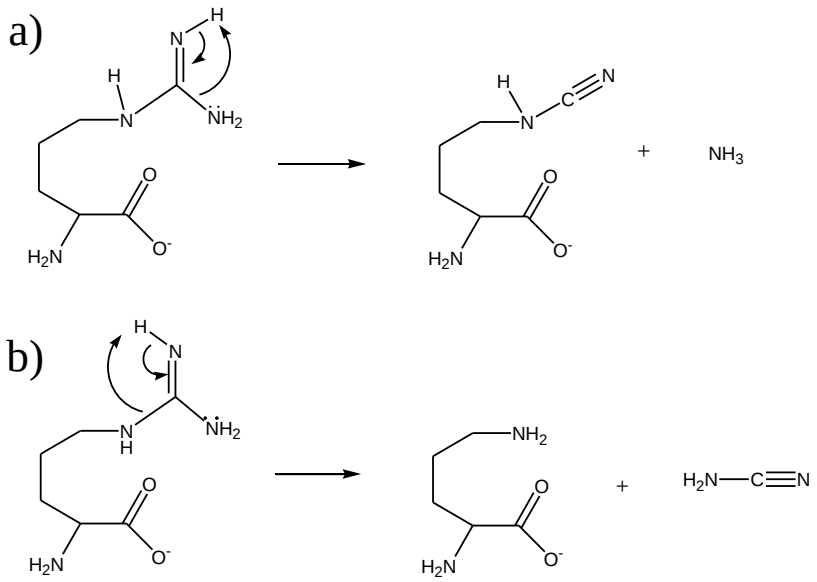

Figure 3. 6 Proposed mechanism for a) ammonia loss and b) cyanamide loss from Arg^-.

For Arg^+, Suhai and co-workers predict that ammonia loss occurs with production of a 7-membered ring,[53] whereas water loss most probably occurs by an intramolecular ring closing reaction involving OH of the carboxylate and the NH of the guanidinium group. These mechanisms are presented below in Figure 3.7.

Coming back to the CID curves presented in Figure 3.5, ammonia loss for Arg^+ occurs at ~9 % CID energy, whereas it occurs at ~12.5 % CID energy for Arg^-, and water loss for Arg^+ occurs at slightly lower energy, at ~8 % CID energy. In addition, the mechanisms for ammonia and water loss from the canonical and zwitterionic forms of arginine are predicted to be similar to those observed for Arg^- and Arg^+

(Figure 3.8). Thus, NH_3 loss from the canonical form occurs by a similar mechanism to Arg^-, whereas that of H_2O loss is similar to Arg^+, and NH_3 loss from the zwitterion occurs by a similar mechanism to Arg^+.

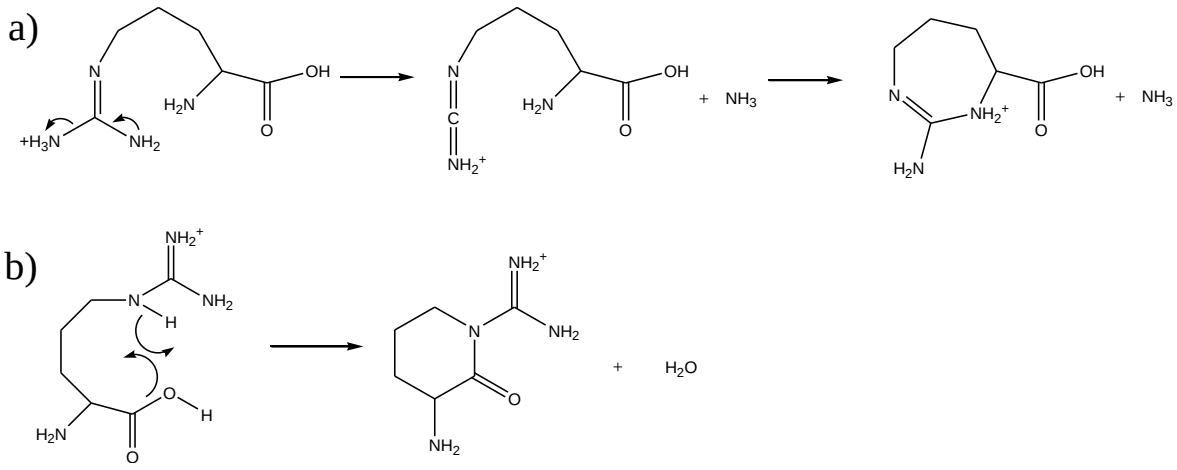

Figure 3. 7 Proposed mechanisms for a) ammonia loss and b) water loss from Arg^+

Taking these probable mechanisms together with the CID energies, suggests that if arginine were canonical, we would expect to see H_2O loss occurring at low energy (~8%) like that for Arg^+, with any NH_3 loss occurring at higher energy (12.5% CID energy) (Thus, we would expect to see water loss on its own). In the cases of both Br^-•Arg and NO_3^-•Arg, NH_3 loss starts to occur at around 12.5 %, along with all the other products, which is much higher in energy than that required for H_2O loss in Arg^+. If water loss occurs by a similar mechanism to that in Arg^+, then we would expect to see it in our spectrum at lower energy than NH_3 loss. The exclusive high-energy NH_3 loss observed for Br^- and NO_3^- complexed with arginine strongly suggests that arginine is not canonical but zwitterionic in these complexes, as the zwitterion cannot lose H_2O, by virtue of the negative charge on its carboxylate terminus.

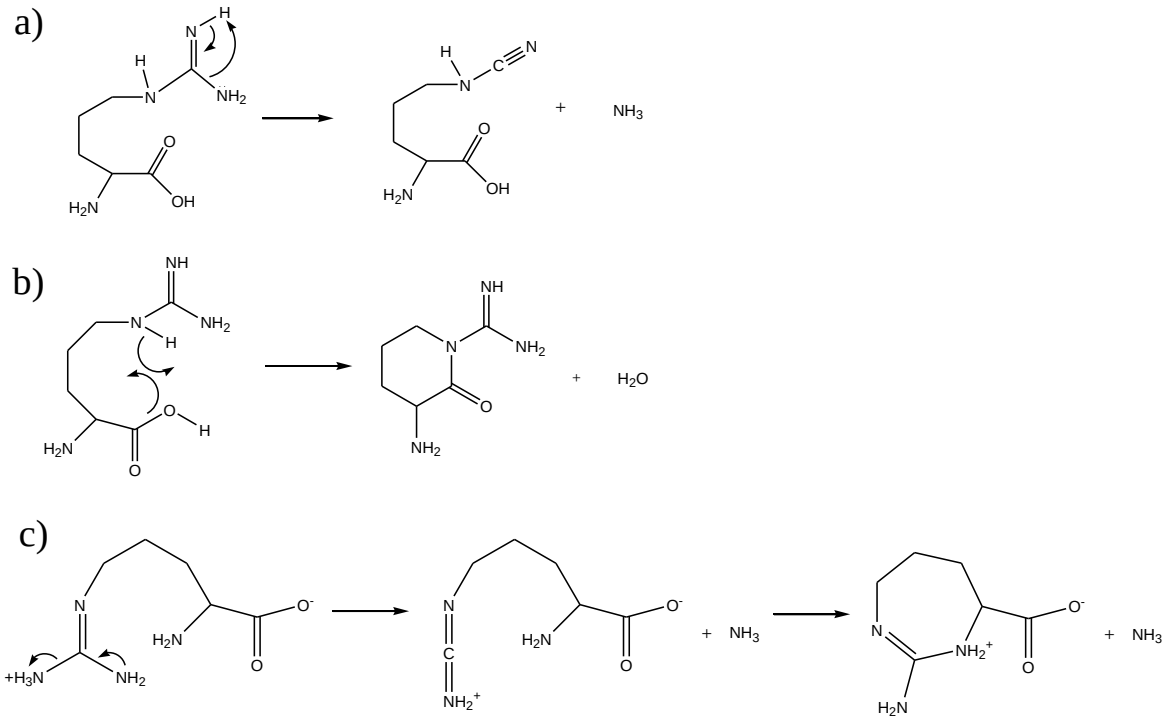

Figure 3. 8 Proposed mechanisms for a) ammonia loss and b) water loss from the canonical form of arginine, and for c) ammonia loss from the zwitterionic form of arginine.

3.5 Conclusion

In the present study, the experimental evidence points towards Br^- being able to stabilize the zwitterionic structure of Arginine. Given this conclusion, it may be that both F^- and Cl^- also stabilise the zwitterionic form, given previous studies focussing on a single electron.[22,25] Equally, there may be a size effect at play, as observed with cationic clusters of arginine (See Introduction, *Cationic Complexation of Amino Acids in the Gas-phase*), so that only larger anions are able to stabilise the zwitterionic structure of arginine effectively. To further reinforce this conclusion, it was deemed necessary to embark on a computational study of the Br^-•Arg system. This is reported on in the next chapter.

Of course, inherent in this work is the assumption that there is no barrier, or at least, a small barrier to conversion between the zwitterionic and canonical structures. A significant barrier between the two structures would create ambiguity around the structure that is being observed in the gas-phase. This is because arginine is zwitterionic in solution. On transfer to the gas-phase by ESI, it may simply retain this structure if there is a sufficiently large barrier to conversion between the two structures.

References

[1] Barran, P. E.; Polfer, N. C.; Campopiano, D. J.; Clarke D. J.; Langridge-Smith, P. R. R.; Langley, R. J.; Govan J. R. W.; Maxwell A.; Dorin, J. R.; Millar, R. P.; Bowers, M. T. *Int. J. Mass Spectrom.* **2005**, *240*, 273

[2] Stryter, L.; Tymoczko, J. L.; Berg, J. M. *Biochemistry*; 5th ed, W. H. Freeman & Co: New York, NY, 2002

[3] Locke, M. J.; Hunter, R. L.; McIver, R. T. *J. Am. Chem. Soc.* **1979**, *101*, 272

[4] Locke, M. J.; McIver, R. T. *J. Am. Chem. Soc.* **1983**, *105*, 4226

[5] Ding, Y.; Krogb-Jespersen, K.. *Chem. Phy. Lett.* **1992**, *199*, 261

[6] Chapo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J. *J. Am. Chem. Soc.* **1998**, *120*, 12956

[7] Godfrey, P. D.; Firth, S.; Hatherley, L. D.; Brown, R. D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1993**, *115*, 9687

[8] Snoek, L. C.; Robertson, E. G.; Kroemer, R. T.; Simons, J. P. *Chem. Phys. Lett.* **2000**, *321*, 49

[9] Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A. *J. Phys. Chem. A*, **2007**, *111*, 11759

[10] Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; Helden, G.; Oomens, J. *J. Am. Chem. Soc.* **2005**, *127*, 8571

[11] Bush, M. F.; Oomens, J.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc.* **2008**, *130*, 6463

[12] Heaton, A. L.; Bowman, V. N.; Oomens, J.; Steill, J. D.; Armentrout, P. B. *J. Phys. Chem. A* **2009**, *113*, 5519

[13] Strittmatter, E. F.; Lemoff, A. S.; Williams, E. R. *J. Phys. Chem. A* **2000**, *104*, 9793

[14] Jensen, F. *J. Am. Chem. Soc.* **1992**, *114*, 9533

[15] Bush, M. F.; Forbes, M. W.; Jockusch, R. A.; Oomens, J.; Polfer, N. C.; Saykally, R. J.; Williams, E. R. *J. Phys. Chem. A*, **2007**, *111*, 7753

[16] Kapota, C.; Lemaire, J.; Maitre, P.; Ohanessian, G. *J. Am. Chem. Soc.* **2004**, *126*, 1836

[17] Armentrout, P. B.; Rodgers, M. T.; Oomens, J.; Steill, J. D. *J. Phys. Chem. A*, **2008**, *112*, 2248

[18] Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1995**, *117*, 8159

[19] Brandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2002**, *116*, 5023

[20] Im, S.; Jang, S.; Lee, S.; Lee, Y.; Kim, B. *J. Phys. Chem. A* **2008**, *112*, 9767

[21] Park, S.; Ahn, D.; Lee, S. *Chem. Phys. Lett.* **2003**, *371*, 74

[22] Skurski, P.; Rak, J.; Simons, J.; Gutowski, M. *J. Am. Chem. Soc.* **2001**, *123*, 11073

[23] Kass, S. R. *J. Am. Chem. Soc.* **2005**, *127*, 13098

[24] Yang, G.; Zu, Y.; Liu, C.; Fu, Y.; Zhou, L. *J. Phys. Chem. B* **2008**, *112*, 7104

[25] Xu, S.; Zheng, W.; Radisic, D.; Bowen, K. H. *J. Chem. Phys.* **2005**, *122*, 091103

[26] Jockusch, R. A.; Price, W. D.; Williams, E. R. *J. Phys. Chem. A* **1999**, *103*, 9266

[27] O'Hair, R. A. J.; Bowie, J. H.; Gronert, S.; *Int. J. Mass Spectrom. Proc.***1992**, *117*, 23

[28] Lemoff, A. S.; Bush, M. F.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 13576

[29] Wyttenbach, T.; Witt, M.; Bowers, M. T. *J. Am. Chem. Soc.* **2000**, *122*, 3458

[30] Julian, R. R.; Jarrold, M. F. *J. Phys. Chem. A* **2004**, *108*, 10861

[31] Gdanitz, R. J.; Cardoen, W.; Windus, T. L.; Simons, J. *J. Phys. Chem. A* **2004**, *108*, 515

[32] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64

[33] Colorado, A.; Brodbelt, J. *J. Am. Chem. Soc. Mass Spectrom.* **1996**, *7*, 1116

[34] Hart, K. J.; McLuckey, S. A. *J. Am. Chem. Soc. Mass. Spectrom.* **1994**, *5*, 250

[35] Burke, R. M.; Dessent, C. E. H. *J. Phys. Chem. A*, **2009**, *113*, 2683

[36] Bierbaum, V.M.; Schmidt, R.J.; DePuy, C.H.; Mead, R.H.; Schulz, P.A.; Lineberger, W.C. *J. Am. Chem. Soc.* **1981**, *103*, 6262

[37] Trainham, R.; Fletcher, G.D.; Larson, D.J. *J. Phys. B* **1987**, *20*, L777

[38] Hunter, E.P.; Lias, S.G. *J. Phys. Chem.***1998**, *27*, 413

[39] Davidson, J.A.; Fehsenfeld, F.C.; Howard, C.J. *Int. J. Chem. Kinet.* **1977**, *9*, 17

[40] Blondel, C.; Cacciani, P.; Delsart, C.; Trainham, R. *Phys. Rev. A* **1989**, *40*, 3698

[41] Hanstorp, D.; Gustafsson, M. *J. Phys. B* **1992**, *25*, 1773

[42] Wang, X.B.; Wang, L.S. *J. Chem. Phys.* **2000**, *113*, 10928

[43] Sleno, L.; Volmer, D. A. *J. Mass Spectrom.* **2004**, *39*, 1091

[44] Hart, K. J.; McLuckey, S. A. *J. Am. Chem. Soc. Mass. Spectrom.* **1994**, *5*, 250

[45] Satterfield, M.; Brodbelt, J. S. *Inorg. Chem.* **2001**, *40*, 5393

[46] Satterfield, M.; Brodbelt, J. S. *Inorg. Chem.* **2001**, *40*, 5393 and references therein

[47] Shannon, R. D. *Acta Cryst.* **1976**, *A32*, 751-767

[48] Weis, P.; Kemper P. R.; Bowers, M. T.; Xantheas, S. S. *J. Chem. Phys.* **1999**, *121*, 3531

[49] Hiraoka, K.; Mizuse, S. *J. Chem. Phys.* **1987**, *118*, 457

[50] Hiraoka, K.; Mizuse, S.; Yamabe, S. *J. Chem. Phys.* **1988**, *92*, 3943

[51] Allen, H.; Kennard, O.; Watson, D. G.; Brammer, l.; Orpen, A. G.; Taylor, R. *J. Chem. Soc. Perkin Trans.* **1987**, *2*, S1–S19

[52] Blades, A. T.; Klassen, J. S.; Kerbarle, P. *J. Am. Chem. Soc.* **1995**, *117*, 10563

[53] Csonka, I. P.; Paizs, B.; Suhai, S. *J. Mass Spectrom.* **2004**, *39*, 1025

Chapter 4: Computational Study of Br^-•Arg

4.1 Introduction

The experimental findings reached in Chapter 3 regarding the zwitterionic structure of Arg being more stable than the canonical structure when complexed to Br^- in the gas-phase were further investigated by a computational study. This study aimed to find as many of the conformations for the zwitterionic and canonical structures of Br^-•Arg using molecular mechanics, and then to assess the energies of these conformations with higher order single-point quantum mechanics (QM) calculations. Conformers deemed chemically relevant can then be optimized with higher order QM calculations. This is one possible strategy for locating the global minimum for a conformationally complex molecule. Comparison of the lowest-energy conformers from optimization for both the zwitterionic and canonical structures of Br^-•Arg (Figure 4.1) should then give an indication as to which structure is more stable, and therefore which structure the complex is likely to adopt when in the gas-phase.

The challenge is thus to try and effectively sample as much relevant conformational space as possible, and then to pick the most probable lowest-energy conformers from the sample generated for further optimization. In this study, conformational sampling was carried out using the MMFF94 force-field as implemented in SPARTAN 08.[1] Thus, the conformer distribution function gives a selection of chemically relevant conformers that can be classified into families.[2-7] Calculation of the single-point

energies (SPEs) at a higher level of theory then allows the selection of the lowest-energy conformer from each family for optimization at the same level of theory, in order to further refine the energy ordering of the families. The sampling is done quickly (on the order of seconds) using molecular mechanics, which is therefore an efficient way to examine the potential energy surface (PES).

The strategy described above saves time in calculating the energies of geometrically similar structures that are likely to have similar energies, and is a strategy that has been adopted in many forms throughout the literature.[2-7] Burke et al. used an almost identical approach on the conformationally flexible Na^+•$[ATP-2H]^{2-}$ dianion, which had 4 tautomeric structures.[2] They found 50 low-energy conformers for each of these at the MMFF94 level that could then be grouped into 5 families for further optimisations at the B3LYP/6-31+G* level of theory. The same approach was also used on the multiply- and singly- charged anions of adenosine 5'-triphosphate and adenosine 5'-diphosphate.[3] Similarly, Williams and co-workers have studied the cationic clusters of M^+•Arg by a method involving conformational searching with the AMBER and MMFF94 force-fields, followed by division into families and subsequent higher level optimisations at the B3LYP/LACVP** and B3LYP/LACVP* or B3LYP/LACVP++** levels of theory respectively.[4,5] Wyttenbach et al. obtained 100 model conformers of each of several nucleotide structures by a two-stage simulated annealing procedure using the AMBER force-field.[6] They then grouped and optimised the lowest-energy conformer of each group at the B3LYP/6-31G level of theory. Finally, Toroz et al. describe possible “stepwise rotation” and “hierarchical selection” schemes for analysing the conformationally complex tyrosine-glycine dipeptide.[7] The former scheme involves taking sequential steps in a specified dihedral

angle, followed by optimisations at the B3LYP/6-31+G* level of theory of each of the resultant conformers. The next dihedral is then varied as before for each of these conformers, resulting in further conformations for optimisation. The authors described this approach as thorough but one that does not exhaust the conformational space, and in addition, the large angle step size necessary to keep the number of conformations to a manageable level can potentially miss low-energy conformations.

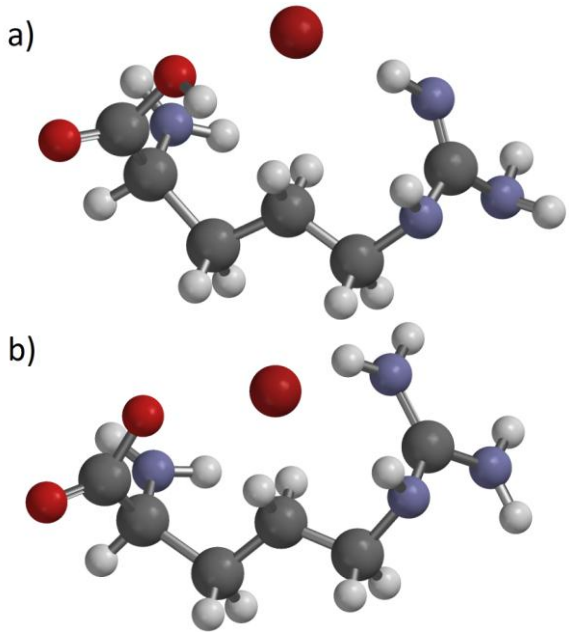

Figure 4. 1 Illustrative chemical structures of the a) canonical and b) zwitterionic structures of Br^-•Arg.

4.2 Method

Conformational searches were performed using the MMFF94 molecular force-field as implemented in SPARTAN 08 using the default parameters.[1] The default parameters include a starting temperature of 5000K and an end temperature of 300K, with the 100 lowest-energy conformers being saved. Conformational searches were conducted on the zwitterionic structure and one of the two possible isomers of the canonical

structure of Br⁻•Arg. The two possible isomers of the canonical structure of Br⁻•Arg differ by the site of deprotonation of the guanidinium side-chain group on Arg, and are depicted in Figure 4.2. It was decided that isomer a) of Figure 4.2 was the canonical structure isomer most likely adopted by Br⁻•Arg. Thus, calculations on that isomer are presented in this chapter, and not on canonical structure isomer b).

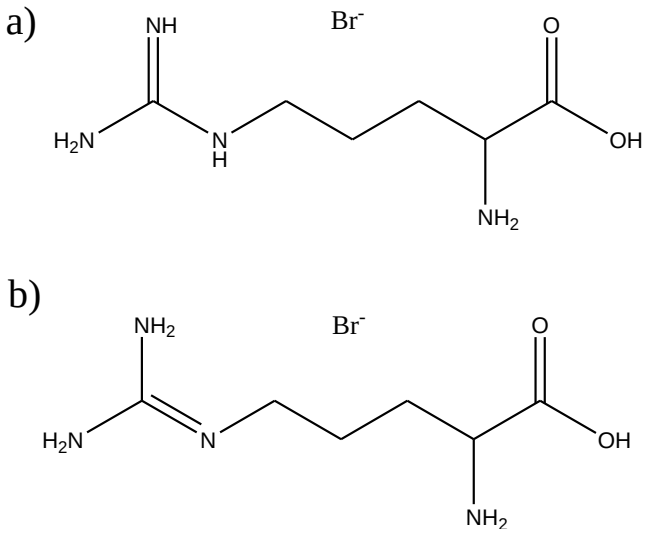

Figure 4. 2 Chemical diagrams of the 2 possible isomers of the canonical structure of Br⁻•Arg, isomers a) and b).

SPEs for the 50 lowest-energy conformers for the canonical structure of Br⁻•Arg were calculated at the B3LYP/6-31+G** and MP2/6-31+G** levels of theory using SPARTAN 08. The same was done for the 30 conformers generated for the Br⁻•Arg zwitterionic structure. The 6-31+G** basis set was chosen for a number of reasons. Firstly, the 6-31G split valence basis set gives a good computational performance to cost ratio, over a minimal basis set, such as STO-3G (See Appendix 2, *Linear Combination of Atomic Orbitals Basis Sets*).[8] Secondly, Polarisation functions, are generally required to allow small shifts of electron density away from atomic centres. Conventionally, a single set of 6d orbitals on 2nd row elements is usually sufficient

(*), but additional polarisation functions were required for the polarisable H-bonds on arginine that interact with Br^-, corresponding to an extra set of 3p orbitals on Hydrogens (**). Thirdly, diffuse functions (+), are a requirement for anions, to allow for the displacement of the weakly bound excess electron from the atomic centre (i.e. the displacement of significant electron density away from the nucleus).

For both structural forms, important interactions between Br^- and select atoms on Arg were decided upon, and then each conformer was assessed to see whether it possessed any of the named interactions, allowing it to be a classified into a family of similar conformers. Further subdivision of these families into subfamilies could then be made by identifying additional, less chemically important interactions. An interaction was considered to be present if the distance between the select atoms of Arg and Br^- was less than 3.3 Å for any conformer of both the canonical and zwitterionic structures. This was an arbitrary interatomic distance decided on so as to be greater than a typical hydrogen bond (< 2 Å), thus capturing all important possible interactions, and was used for both classification into families and subfamilies. In this way, geometrically similar conformers were grouped together for each structure.

Following conformer classification, the lowest-energy conformer from each subfamily was then further optimized at the B3LYP/6-31+G** level of theory. Where the MP2/6-31+G** (sp) and B3LYP/6-31+G** (sp) levels of theory did not give the same conformer as the lowest-energy conformer for that subfamily, both conformers were fully optimized. In addition to these optimizations, frequency calculations were conducted to ensure that the structures corresponded to minima on the PES. It was not possible to run the frequency calculations at the B3LYP/6-31+G** level of theory,

due to memory constraints in SPARTAN 08, therefore these frequency calculations had to be run at the B3LYP 6-31+G* level of theory.

4.3 Results and Discussion

Tables 4.1 and 4.2 describe the classification system used for the conformers for the canonical and zwitterionic structures of Br^-•Arg respectively. Figures 4.3a/b and 4.4 represent the contents of Tables 4.1 and 4.2 graphically for the canonical and zwitterionic structures, respectively. Figure 4.5 is a general comparison of the energy for the canonical and zwitterionic structures, Figure 4.6 gives the energies for the canonical families, and Figure 4.7 gives the energies for the zwitterionic families. Finally, Tables 4.3 and 4.4 give general comparisons of energies for the zwitterionic and canonical structures, and Figure 4.8 presents a comparison of the optimized energies for the zwitterionic and canonical structures.

4.3.1 Description of the Conformer Families for the Canonical Structure of Br^-•Arg

Key interactions (< 3.3 Å) of 4 nominated atoms of Arg with Br^- for the 50 lowest conformers of the canonical structure of Br^-•Arg are listed in Table 4.1. This classification results in 8 major families labelled A-H, with the presence of other various minor interactions allowing further subdivision into subfamilies. These minor interactions largely involve nitrogen atoms of arginine with hydrogens of arginine.

For those structures belonging to family A, there were three interactions of Br^- to the most significant atoms of Arg (i.e. those belonging to the amino and carboxylate

terminal groups, and the guanidinium side chain group). Family B, then lacked the important H1 interaction with Br^- (H1 of the O group) when compared to family A, and family C was missing the perhaps less significant HE interaction with Br^- (HE of the N terminus), which compresses the alkane backbone of Arg. Family D exhibited a solitary interaction of Br^- with the H1 of the OH group (There is no interaction of Br^- with the amino terminus in these conformers). Family E was missing the 1HH1 interaction with the nitrogen terminus as compared to family A, leading the nitrogen terminus to protrude out more. Family F conformers were similar to the family A conformers, but with an alternative HH2 interaction instead of a 1HH1/2HH1 interaction (i.e. the guanidinium side-chain group has twisted around). Family G conformers exhibited a solitary interaction of Br^- with 1HH1/2HH1. Finally, the family H conformers were analogous to the family B conformers, but with alternative HH2 interactions to Br^- instead of 1HH1/2HH1 interactions to Br^- (i.e. the guanidinium side-chain group has again twisted around).

Most of these conformer families were further subdivided into sub-families based on the presence of a further numbers of interactions that were considered to be less important, but significant enough to merit further subdivision (Note that the importance of such interactions can only properly be assessed by comparing the conformer energies at a reliable level of theory). For example, the difference between A1 and A2 is that the NH_2 group of the carboxylate end has rotated around to interact with Br^- in the latter. Subfamily A1 contained the largest number of conformers, whereas family D, for example, contained many different sub-families corresponding at the most to a couple of conformers. The classification is shown clearly in Table 4.1 and Figures 4.3a/b below.

Table 4. 1 A table of the geometric interactions present for the canonical structure of Br^-•Arg. The distance between the atoms needed to be less than 3.3 Å for the interaction to be considered relevant for the classification.

	Major interactions				Minor interactions				
Family	H1-Br	HE-Br	1HH1/2HH1-Br	HH2-Br	H/H2-Br	N-HE	N-H1	N-1HH1/2HH1	H1-NE
A1	✓	✓	✓	×	×	×	×	×	×
A2	✓	✓	✓	×	✓	✓	✓	×	×
A3	✓	✓	✓	×	✓	×	✓	×	×
B1	×	✓	✓	×	✓	×	✓	×	×
B2	×	✓	✓	×	×	×	✓	×	×
B3	×	✓	✓	×	×	×	×	×	✓
B4	×	✓	✓	×	✓	✓	✓	×	×
C1	✓	×	✓	×	✓	✓	✓	×	×
C2	✓	×	✓	×	×	×	×	×	×
C3	✓	×	✓	×	×	✓	×	×	×
C4	✓	×	✓	×	✓	×	✓	×	×
C5	✓	×	✓	×	✓	✓	×	✓	×
D1	✓	×	×	×	×	×	×	×	×
D2	✓	×	×	×	✓	×	×	×	×
D3	✓	×	×	×	×	×	×	✓	×
D4	✓	×	×	×	×	✓	×	✓	×
D5	✓	×	×	×	×	✓	×	×	×
D6	✓	×	×	×	✓	✓	✓	✓	×
E	✓	✓	×	×	×	×	×	×	×
F	✓	✓	×	✓	×	×	×	×	×
G	×	×	✓	×	✓	×	✓	×	×
H	×	✓	×	✓	✓	×	✓	×	×

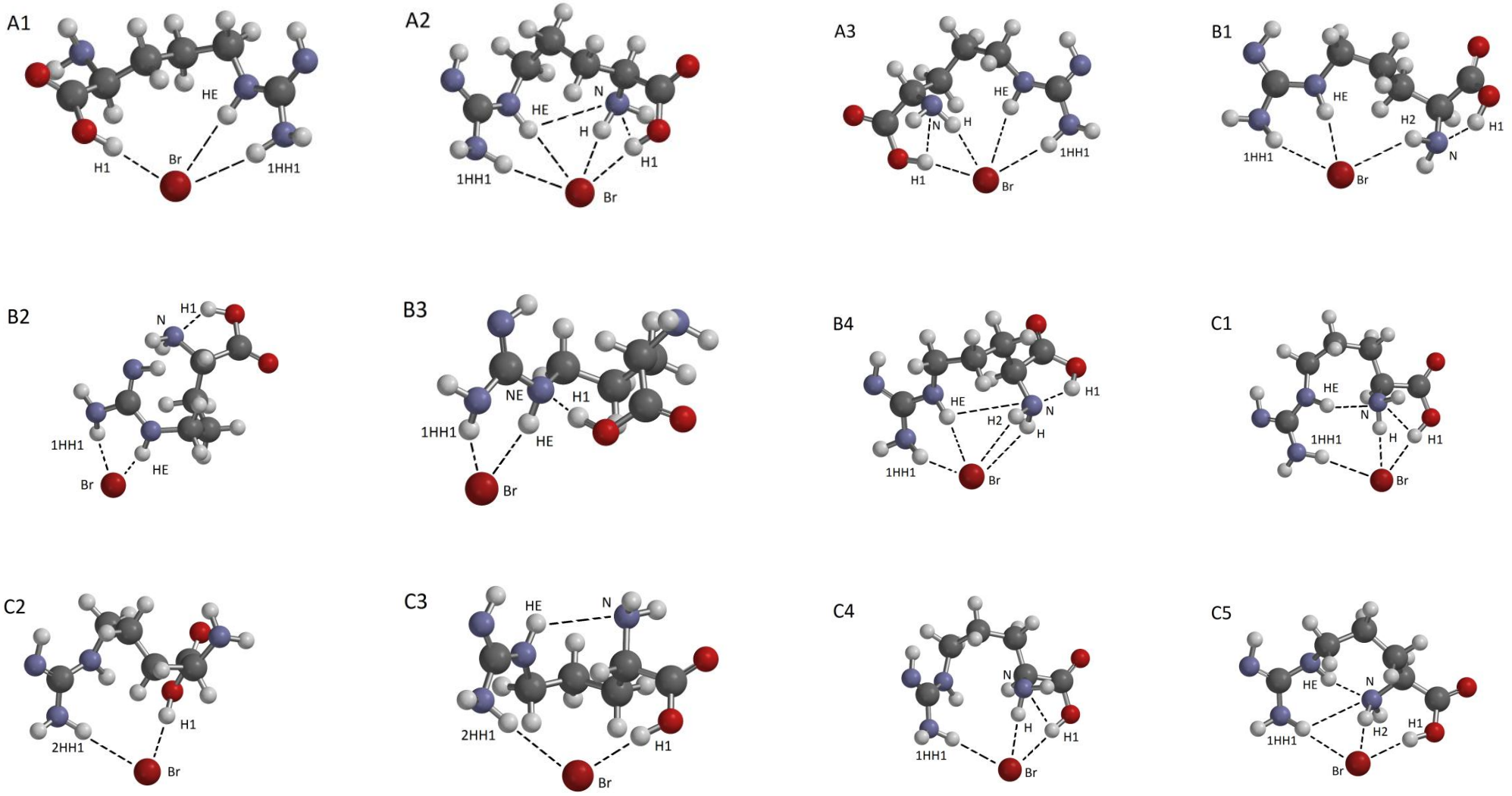

Figure 4. 3a Illustration of the different interactions present for families A to C for the canonical structure of Br^-•Arg.

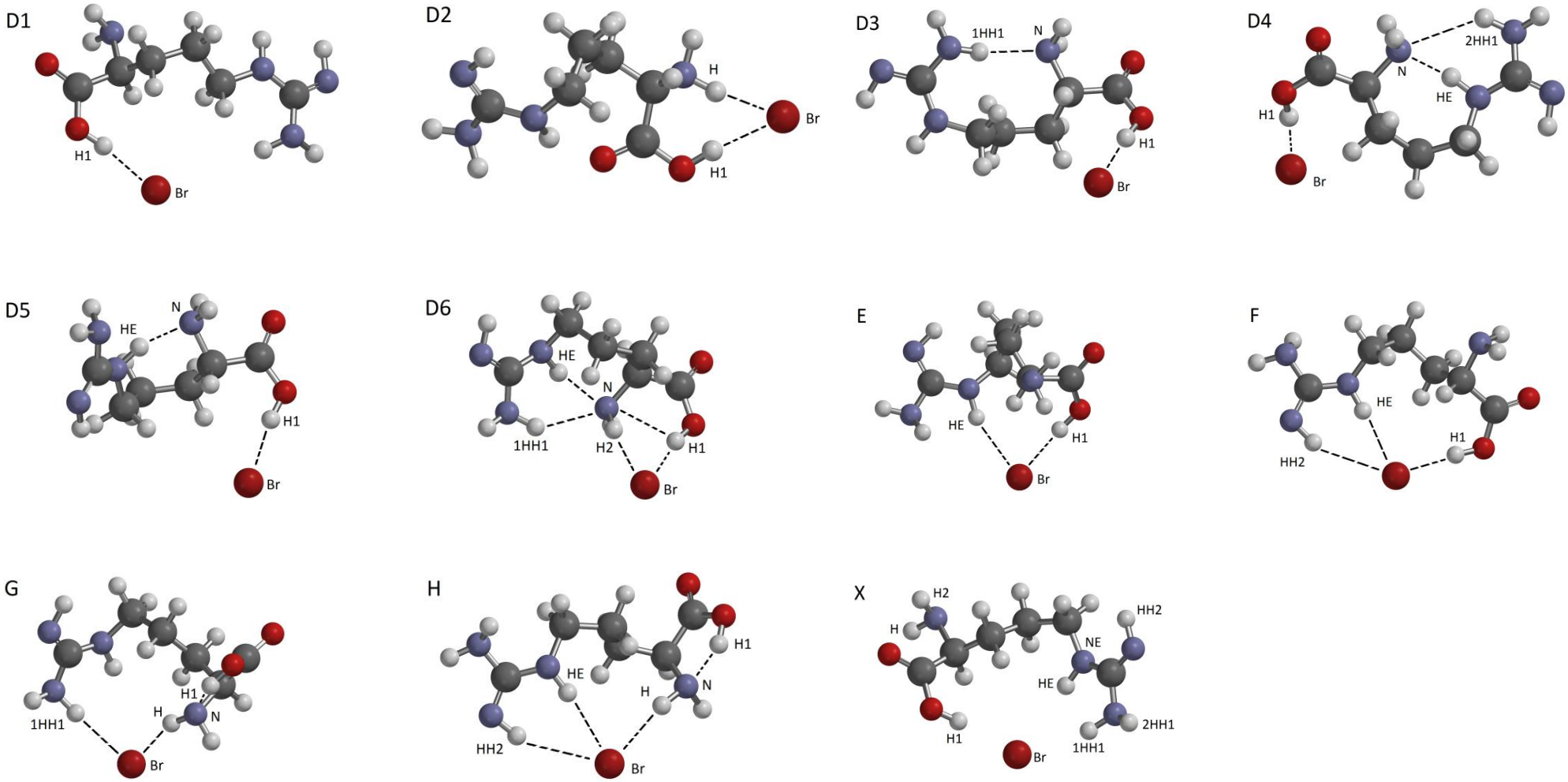

Figure 4. 3b Illustration of the different interactions present for families D to H for the canonical structure of Br^-•Arg (diagram X illustrates the atoms used in the classification for both Figures 4.2a and 4.2b).

4.3.2 Description of the Conformer Families for the Zwitterionic Structure of Br^-•Arg

Classification for the zwitterionic structure of Br^-•Arg is based upon the key and minor interactions listed in Table 4.2 (< 3.3 Å) and presented in Figure 4.4. For the zwitterionic structure, the SPARTAN 08 program found far fewer structures (only 30 from a search that aimed to locate the lowest 100), which could be due to there being either fewer, deeper minima, or to inadequacies in the force-field. Classification was considerably simpler for the zwitterionic structure of Br^-•Arg. Families A and B only differed by the two N's of the guanidinium side-chain group that Br^- interacted with, thus the guanidinium side-chain group is rotated around for family B compared to family A. The sub-families of both were then classified into differences between the prominent interactions of the two O's of the negatively charged carboxylate group with one, two or both N's of the opposite side of the guanidinium side-chain group. The structures of family C then just displayed a single interaction of Br^- with the H3 of the positive nitrogen of the guanidinium side-chain group, and this family could be subdivided in exactly the same way as for families A and B.

Table 4. 2 A table of the geometric interactions present for the zwitterionic structures of Br^-•Arg. The distance between the atoms needed to be less than 3.3 Å for the interaction to be included in the classification.

	Major Interactions					Minor Interactions							
	Br-H5	Br-H4	Br-HE	Br-H1	Br-H3	HE-O1	HE-O	H3-O1	H3-O	H1-O	H1-O1	H5-O	H5-O1
A1	✓	✓	×	×	×	✓	×	×	×	×	✓	×	×
A2	✓	✓	×	×	×	×	✓	×	×	✓	×	×	×
A3	✓	✓	×	×	×	✓	✓	×	×	✓	✓	×	×
A4	✓	✓	×	×	×	✓	✓	×	×	×	✓	×	×
B1	×	×	✓	✓	×	×	×	✓	×	×	×	×	✓
B2	×	×	✓	✓	×	×	×	×	✓	×	×	✓	×
B3	×	×	✓	✓	×	×	×	✓	✓	×	×	✓	✓
C1	×	×	×	×	✓	✓	✓	×	×	✓	✓	×	×
C2	×	×	×	×	✓	✓	×	×	×	×	✓	×	×
C3	×	×	×	×	✓	×	✓	×	×	✓	×	×	×

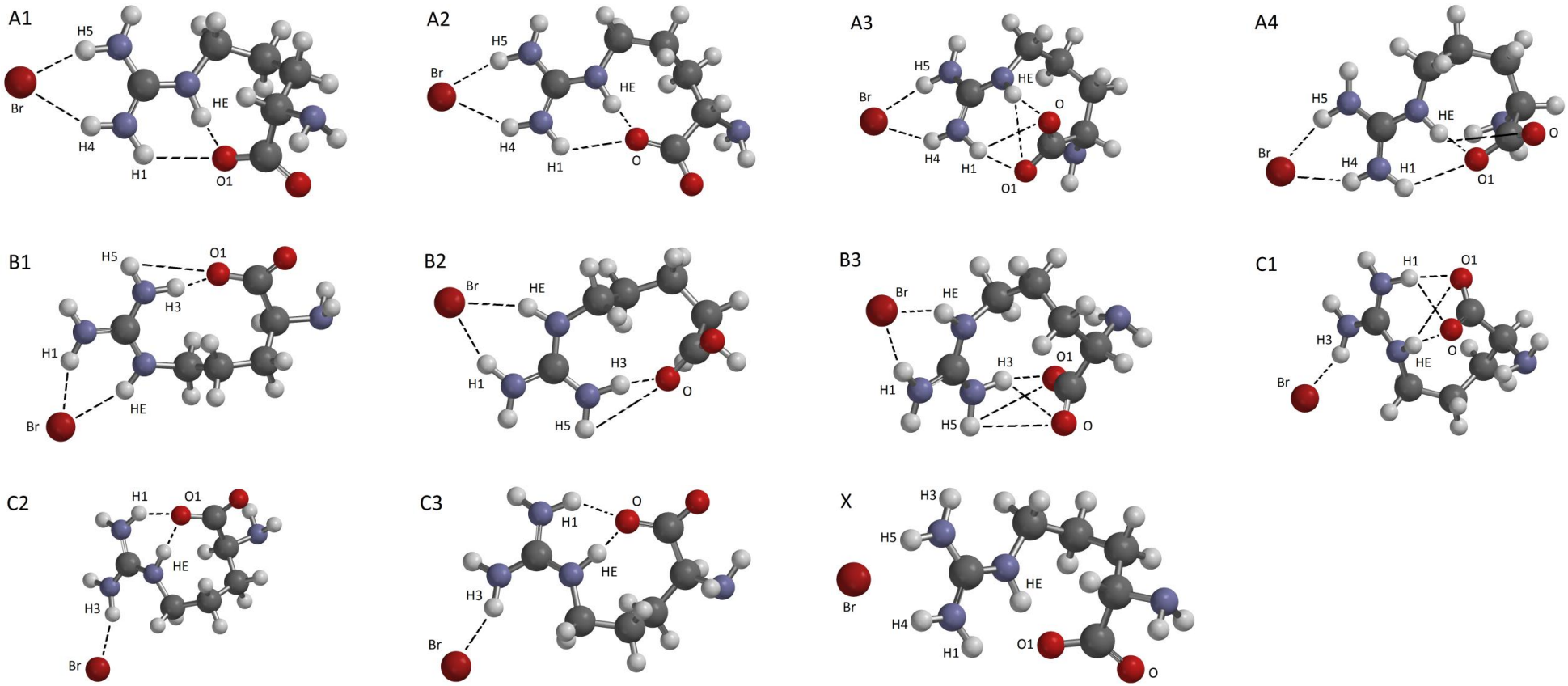

Figure 4. 4 Illustration of the zwitterionic families of Br^-•Arg. Diagram X illustrates the atoms used to classify the conformers.

4.3.3 General Comparisons of the Single-Point Energies for the Zwitterionic and Canonical Structures of Br^-•Arg

Figure 4.5 gives all the single-point energies calculated in this study, where graph a) is the MMFF94 (sp) energies for the 30 zwitterionic conformers and 100 canonical conformers. The energies are all considerably lower for the zwitterionic structure, the difference being ~3.75 eV, though caution must be exercised when comparing different tautomers using the same force-field.[3]

On going to the 50 lowest-energy conformers for the canonical structure and all of the 30 zwitterionic conformers at the B3LYP/6-31+G** (sp) level of theory (Figure 4.5 b), the lowest-energy zwitterionic structure is still lower than the lowest-energy canonical structure, but they are now much closer in energy, the difference being ~0.1 eV. The range for the zwitterionic structure is ~0.7 eV, whereas that for the canonical structure is ~0.4 eV, despite there being more canonical conformers. For the same conformers at the MP2/6-31+G** (sp) level of theory (Figure 4.5 c), the lowest-energy conformer for the canonical structure is < 0.05 eV lower in energy than that for the zwitterionic structure. The range of the canonical conformers is ~0.475 eV, whereas that for the zwitterion is ~0.65 eV. The MP2/6-31+G** (sp) calculation has had the effect of compressing the energy spread across the two tautomers.

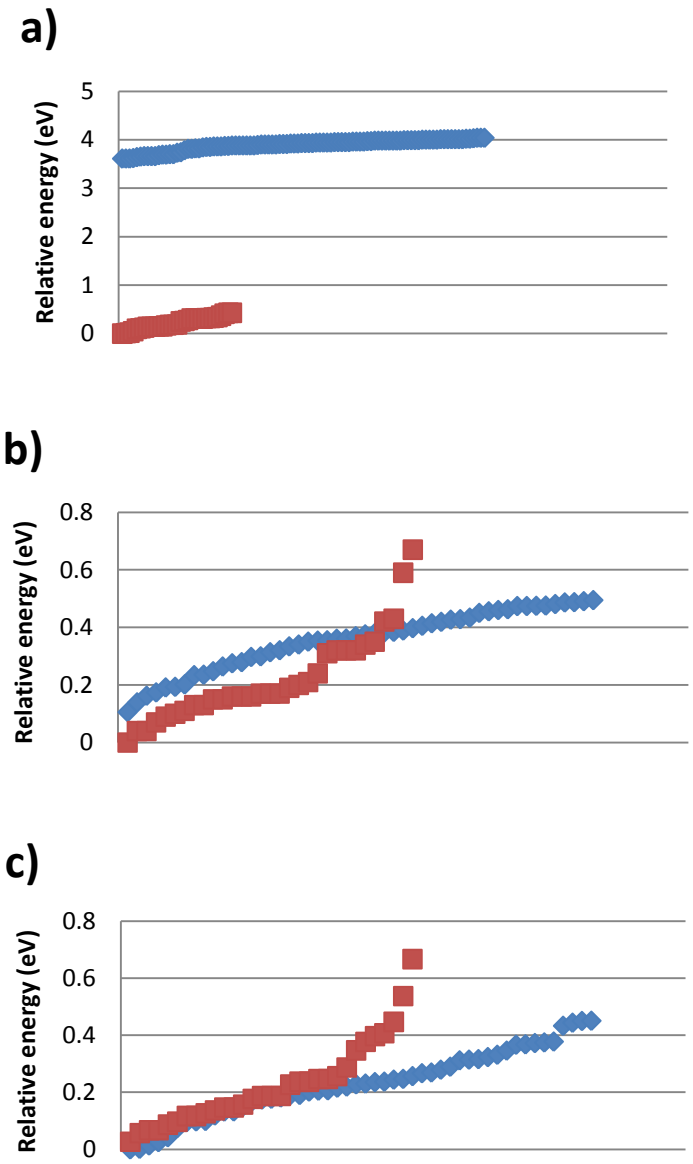

Figure 4. 5 Figure giving the SPEs at all levels of theory of all of the conformers generated using the MMFF94 force-field. The 100 lowest-energy conformers for the canonical structure (◆) and the 30 conformers for the zwitterionic structure (■) at the MMFF94 (sp) level of theory are presented in a). The 50 lowest-energy conformers for the canonical structure and the 30 conformers for the zwitterionic structure at the MMFF94 (sp) level of theory, were then recalculated at the b) B3LYP/6-31+G** (sp) and c) MP2/6-31+G** (sp) levels of theory.

4.3.4 Energies of the Conformer Families for the Canonical Structure of Br^-•Arg

The initial ordering of the families for the canonical structure of Br^-•Arg using the MMFF94 force-field (Figure 4.5 a) is B/A < C < G < F < H < D < E. The conformers belonging to family A are the most numerous and exhibit the largest energy variation of ~0.4 eV, giving this family the lowest and highest conformers of all of the conformers examined. Conformers belonging to family B vary over the range ~0.325 eV, those belonging to family C vary over 0.25 eV, and those belonging to family D vary over ~0.15 eV. All the other families exhibit ranges that are smaller than 0.15 eV. The lowest-energy conformers for families B, A, C and G are similar in energy, varying over a ~0.06 eV range. However, there is then a gap of ~0.125 eV between this low-energy group of families and the remaining higher-energy families, which vary over a range of ~0.08 eV.

The ordering of the families is A < B < C < E < G < D < F < H at the B3LYP/6-31+G** (sp) level of theory (Figure 4.5 b). Again the family A structures have a tendency to exhibit the largest energy range of ~0.325 eV, the range of families B and C are both ~0.3 eV, whereas the family D structures exhibit a smaller range of ~0.125 eV. The remaining families exhibit ranges that are all smaller than for family D. The lowest-energy conformers from each of families A, B and C vary over a small ~0.1 eV range, but approximately 30% of the structures of family A have energies that are lower than any other conformer from any other family. There is a gap of ~0.2 eV, then the higher-energy families have their lowest-energy conformers at higher relative energies, over a range of ~0.1 eV.

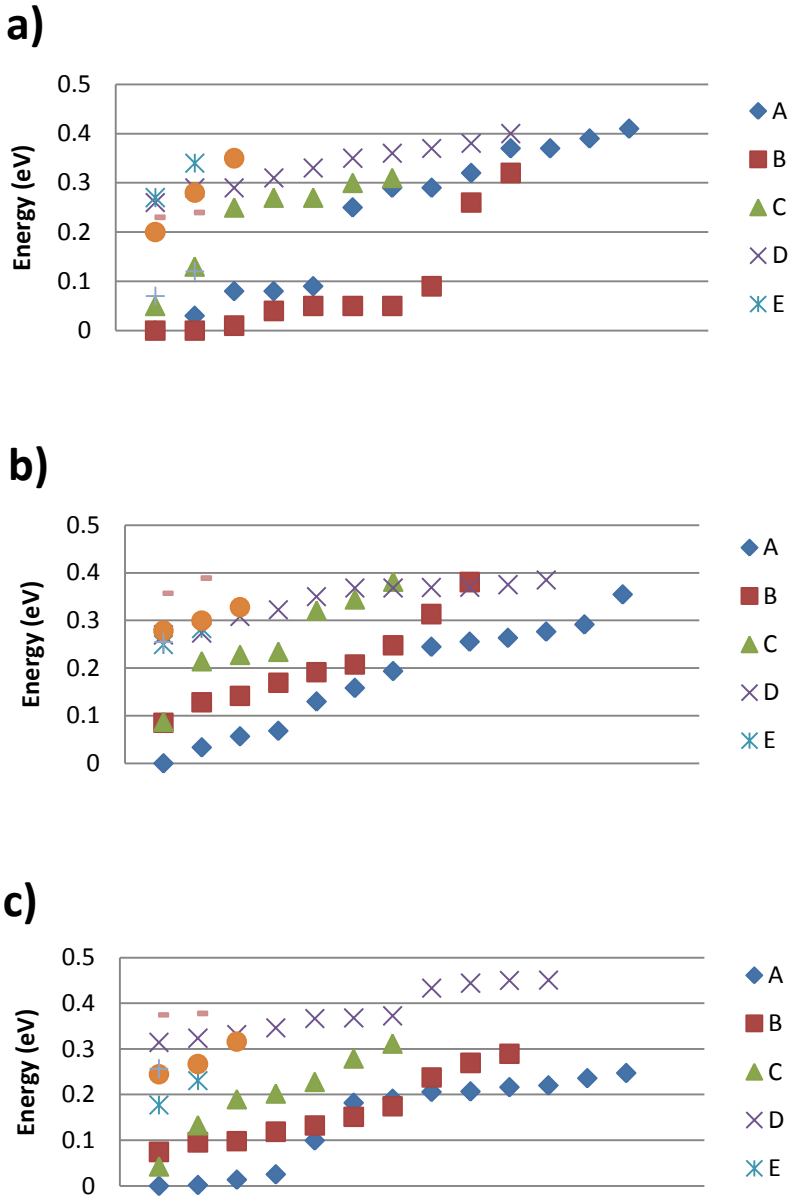

Figure 4. 6 SPE distribution plots of the 50 lowest-energy conformers calculated by the MMFF94 force-field for the canonical structure of Br^-•Arg, at the a) MMFF94 (sp), b) B3LYP/6-31+G** (sp) and c) MP2/6-31+G** (sp) levels of theory. At each level of theory, the conformers have been grouped into their respective families.

On going to the MP2/6-31+G** (sp) energies (Figure 4.5 c), the initial order is A < C < B, though 40% of the conformers belonging to family B appear at less than the second lowest-energy conformer belonging to family C. The remainder are in the order E < F < G < D < H. The A family has a range of ~0.25 eV, the B family has a

range of ~0.2 eV, the C family has a range of ~0.275 eV, the D family has a range of ~0.15 eV, and the remaining families have ranges less than that for family D. The lowest-energy conformers for families A, B and C have a range of ~0.1 eV, with the other families spaced out at higher-energy.

The families A, B and C always contain the lowest-energy conformers at all levels of theory, though these families also always exhibit the largest energy ranges. In the case of the other families, E, G, D, F and H, they tend to exhibit tighter energy spreads at higher relative energies at all levels of theory.

4.3.5 Energies of the Conformer Families for the Zwitterionic Structure of Br^-•Arg

The initial ordering of the families for the zwitterionic structure of Br^-•Arg using the MMFF94 force-field (Figure 4.6 a) is A < B < C, with 30% of the structures for family A existing at energies that are lower than the lowest-energy conformer belonging to family B. In this series, conformers belonging to family A vary over ~0.2 eV, those belonging to family B vary over ~0.4 eV, and those belonging to family C vary over a ~0.1 eV energy range. The lowest-energy conformers for families A and B are in a ~0.05 eV range, and then there is a gap of ~0.2 eV until the lowest-energy conformer belonging to family C.

The ordering of the families is A < B < C at the B3LYP/6-31+G** (sp) level of theory (Figure 5.6 b). However, approximately 80% of the conformers belonging to family A and approximately 75% belonging to B exist in the same ~0.2 eV range lower in energy than the lowest-energy conformer belonging to family C. In this series,

conformers belonging to family A vary over ~0.2 eV, with those belonging to family B varying over ~0.7 eV, and those belonging to family C varying over a ~0.2 eV range. The lowest-energy conformers for both families A and B are in a small ~0.05 eV range and then there is a gap of ~0.2 eV until the lowest-energy conformer belonging to family C.

The ordering of the families at the MP2/6-31+G** (sp) level of theory is A < B < C (Figure 5.6 c). However, approximately 45% of the conformers for family A and approximately 35% of the conformers for family B exist in the same ~0.1 eV range, lower in energy than the lowest-energy conformer belonging to family C. In this series, conformers belonging to family A vary over ~0.2 eV, those belonging to family B vary over ~0.6 eV, and those belonging to family C vary over a ~0.3 eV range. The lowest-energy conformers for both families A and B are in a small ~0.05 eV range and then there is a gap of ~ 0.1 eV until the lowest-energy conformers belonging to family C.

At both the MP2/6-31+G** (sp) and B3LYP/6-31+G** (sp) levels of theory, families A and B always contain the lowest-energy conformers, with family A containing conformers that are slightly lower in energy than B: family C appears to be a higher-energy family. The effect of MP2 is generally to compress the energies of the structures together: the MMFF94 (sp) and B3LYP/6-31+G** (sp) energies generally tend to exhibit larger ranges between the families and conformers within the same family.

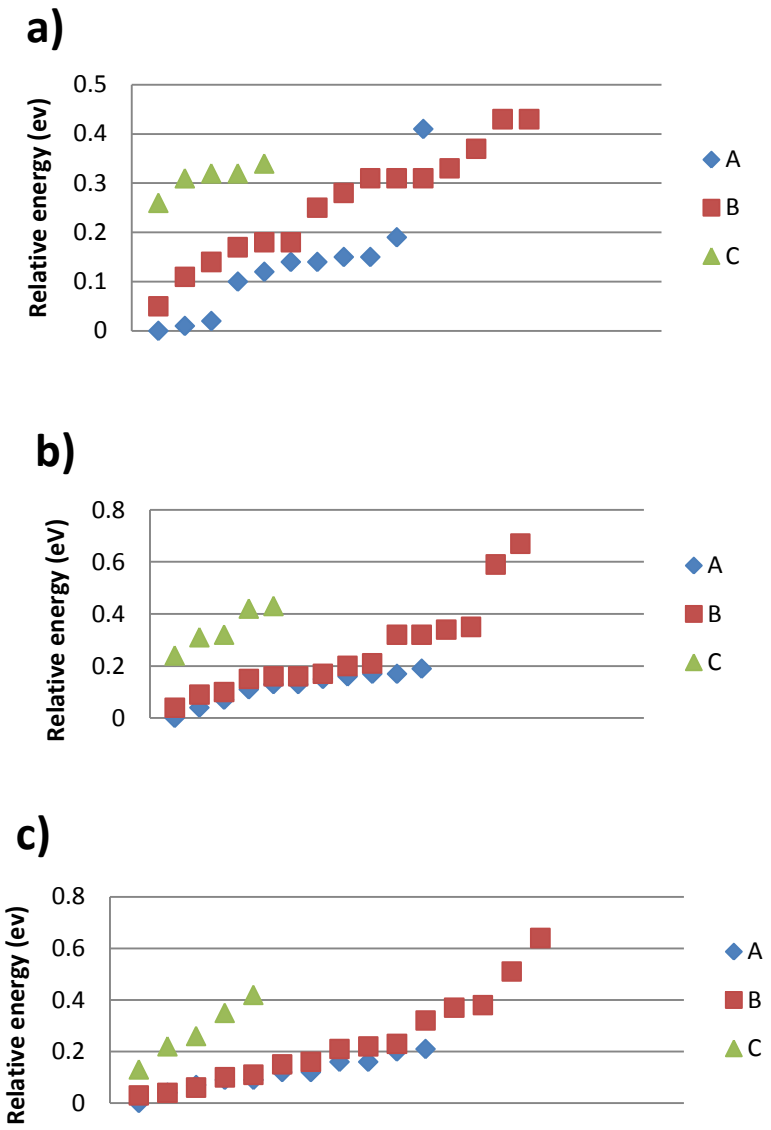

Figure 4. 7 SPE distribution plots of the 50 lowest-energy conformers calculated by the MMFF94 force-field for the zwitterionic structure of Br^-•Arg, at the a) MMFF94 (sp), b) B3LYP/6-31+G** (sp) and c) MP2/6-31+G** (sp) levels of theory. At each level of theory, the conformers have been grouped into their respective families.

4.3.6 Optimized Energies of the Selected Conformers of Br^-•Arg

As discussed in the methods section, the lowest SPE from each family was optimized at the B3LYP/6-31+G** level of theory in an effort to obtain structures closer to the true global minima. Where there were differences between the B3LYP/6-31+G** (sp)

and MP2/6-31+G** (sp) levels of theory over which geometry was the lowest-energy conformer for that family, both geometries were optimized. On visual inspection, the geometries of the selected conformers changed little on geometry optimization for both the zwitterionic and canonical structures. This is in line with the results of similar studies on related small, biological molecules.[3]

Table 4.3 gives the B3LYP/6-31+G** (sp), MP2/6-31+G** (sp) and B3LYP/6-31+G** energies of the selected conformers for both the canonical and zwitterionic structures. Table 4.4 then gives the order of the families according to their ΔE^a (the energy relative to the lowest-energy conformer for that structure) values taken from Table 4.3. For the canonical structure, the ordering of the families changes on going from the B3LYP/6-31+G** (sp) to the B3LYP/6-31+G** level of theory, but not too significantly. Comparing the same conformer at the two levels never results in a difference in energy that is anymore than 0.2 eV. These same changes in ordering are observed on going from B3LYP/6-31+G** (sp) to MP2/6-31+G** (sp), though the difference in energy of ΔE^a never exceeds 0.1 eV this time. Thus, in terms of the ordering of the families for the canonical structure, the MP2/6-31+G** (sp) and B3LYP/6-31+G** (sp) energies are in reasonable agreement. For the zwitterionic structure, on going from the B3LYP/6-31+G** (sp) to the B3LYP/6-31+G** level of theory, there is no change in the ordering, apart for the lowering in energy of the C3 conformer, but generally the differences in energy of ΔE^a are < 0.1 eV between the same conformer. In comparing the B3LYP/6-31+G** (sp) to the MP2/6-31+G** (sp) for the zwitterionic structure, there is some reordering, but again, the differences in energy of ΔE^a between the same conformer are < 0.1 eV.

Table 4. 3 Relative energies of the lowest-energy conformers of each family for both the zwitterionic and canonical structures of Br^-•Arg at all levels of theory (B3LYP/6-31+G** (sp), B3LYP/6-31+G** and MP2/6-31+G** (sp))

structure	family	B3LYP/6-31+G** (sp)		B3LYP/6-31+G**		MP2/6-31+G** (sp)	
		ΔE^a/eV	ΔE^b/eV	ΔE^a/eV	ΔE^b/eV	ΔE^a/eV	ΔE^b/eV
canonical	A1	0.000	0.219	0.000	0.000	0.000	93.367
	A2	0.068	0.288	0.047	0.047	0.002	93.368
	A3	0.130	0.349	0.092	0.092	0.099	93.466
	B1	0.085	0.305	0.014	0.014	0.074	93.441
	B2	0.313	0.533	0.232	0.232	0.269	93.636
	B3	0.381	0.600	0.275	0.275	0.290	93.656
	B4	0.142	0.361	0.004	0.004	0.119	93.485
	C1	0.088	0.307	0.056	0.056	0.042	93.409
	C2	0.215	0.434	0.168	0.168	0.133	93.499
	C3	0.228	0.447	0.181	0.181	0.202	93.569
	C4	0.234	0.454	0.055	0.055	0.190	93.557
	C5	0.344	0.564	0.121	0.121	0.279	93.646
	D1	0.274	0.493	0.229	0.229	0.346	93.713
	D2	0.368	0.587	0.297	0.297	0.373	93.739
	D3	0.308	0.528	0.252	0.252	0.331	93.698
	D4	0.270	0.489	0.175	0.175	0.314	93.681
	D5	0.370	0.589	0.277	0.277	0.367	93.733
	D6	0.375	0.594	0.162	0.162	0.323	93.690
	E	0.249	0.469	0.186	0.186	0.177	93.544
	F	0.279	0.498	0.147	0.147	0.244	93.611
	G	0.173	0.393	0.087	0.087	0.170	93.537
	H	0.357	0.576	0.223	0.223	0.375	93.741
zwitterionic	A1	0.000	0.351	0.000	0.000	0.000	93.631
	A2	0.070	0.421	0.090	0.090	0.090	93.721
	A3	0.130	0.481	0.156	0.156	0.040	93.671
	A4	0.170	0.521	0.225	0.225	0.160	93.791
	B1	0.040	0.391	0.072	0.072	0.030	93.661
	B2	0.090	0.441	0.123	0.123	0.060	93.691
	B3	0.150	0.501	0.182	0.182	0.040	93.671
	C1	0.240	0.591	0.255	0.255	0.130	93.761
	C2	0.420	0.771	0.307	0.307	0.350	93.981
	C3	0.430	0.781	0.273	0.273	0.420	94.051

*ΔE^a/eV is the energy relative to the lowest-energy conformer at that level of theory for the structure. $\Delta E^{\boldsymbol{b}}$/eV is the energy relative to the lowest-energy conformer for the structure amongst all levels of theory.

Table 4. 4 Families ordered according to increasing ΔE^a taken from Table 4.3

structure	B3LYP/6-31+G** (sp)		B3LYP/6-31+G**		MP2/6-31+G** (sp)	
	family	ΔE^a	family	ΔE^a	family	ΔE^a
canonical	A1	0.000	A1	0.000	A1	0.000
	A2	0.068	B4	0.004	A2	0.002
	B1	0.085	B1	0.014	C1	0.042
	C1	0.088	A2	0.047	B1	0.074
	A3	0.130	C4	0.055	A3	0.099
	B4	0.142	C1	0.056	B4	0.119
	G	0.173	G	0.087	C2	0.133
	C2	0.215	A3	0.092	G	0.170
	C3	0.228	C5	0.121	E	0.177
	C4	0.234	F	0.147	C4	0.190
	E	0.249	D6	0.162	C3	0.202
	D4	0.270	C2	0.168	F	0.244
	D1	0.274	D4	0.175	B2	0.269
	F	0.279	C3	0.181	C5	0.279
	D3	0.308	E	0.186	B3	0.290
	B2	0.313	H	0.223	D4	0.314
	C5	0.344	D1	0.229	D6	0.323
	H	0.357	B2	0.232	D3	0.331
	D2	0.368	D3	0.252	D1	0.346
	D5	0.370	B3	0.275	D5	0.367
	D6	0.375	D5	0.277	D2	0.373
	B3	0.381	D2	0.297	H	0.375
zwitterionic	A1	0.000	A1	0.000	A1	0.000
	B1	0.040	B1	0.072	B1	0.030
	A2	0.070	A2	0.090	A3	0.040
	B2	0.090	B2	0.123	B3	0.040
	A3	0.130	A3	0.156	B2	0.060
	B3	0.150	B3	0.182	A2	0.090
	A4	0.170	A4	0.225	C1	0.130
	C1	0.240	C1	0.255	A4	0.160
	C2	0.420	C3	0.273	C2	0.350
	C3	0.430	C2	0.307	C3	0.420

Figure 4.8 is a plot of the B3LYP/6-31+G** energies for both structures, and clearly shows that 2/3 of the conformers for the zwitterionic structure are lower in energy than the lowest-energy canonical structure. The difference between the lowest-energy zwitterionic conformer and the lowest-energy canonical conformer is ~0.2 eV. This

result suggests that at the B3LYP/6-31+G** level of theory, the global minimum for the zwitterionic structure is lower in energy than that for the canonical structure.

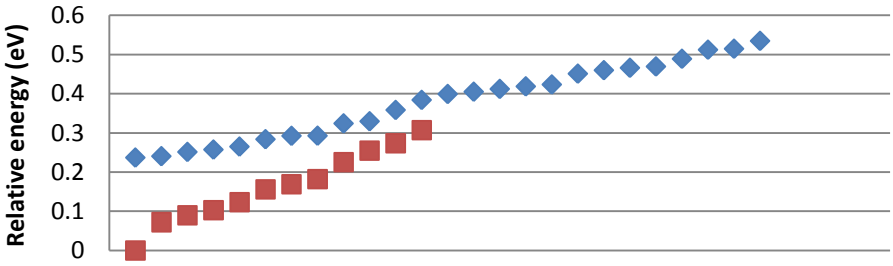

Figure 4. 8 Graph giving the relative optimized energies at the B3LYP/6-31+G** level of theory for the lowest-energy conformers from each family of the canonical (◆) and zwitterionic (■) structures of Br⁻•Arg at the B3LYP/6-31+G** (sp) and MP2/6-31+G** (sp) levels of theory.

4.4 Conclusions

As pointed out previously, the MP2/6-31+G** (sp) global minima for both tautomeric structures of Br⁻•Arg are < 0.05 eV apart and the B3LYP/6-31+G** (sp) global minima are ~0.1 eV apart. The B3LYP/6-31+G** level of theory gives a difference of ~0.25 eV between the global minima for the zwitterionic and canonical structures, which is made more significant given that 2/3 of the conformers for the zwitterionic structure are lower in energy than the lowest-energy conformer for the canonical structure of Br⁻•Arg.

Geometric representations of the global minima at the B3LYP/6-31+G** level of theory for a) the canonical, and b) the zwitterionic structures of Br⁻•Arg are presented in Figure 4.9. Both conformers belong to their respective A families. In the canonical

structure, Br^- is bound between the carboxylate group and the internal end of the guanidinium side chain, whereas in the zwitterionic structure, the Br^- is bound to the terminal end of the guanidinium side chain exclusively. Thus, it appears that an electrostatic interaction of Br^- with the positive end of the arginine zwitterion predominates over other possible H-bonding interactions in the zwitterionic Br^-•Arg cluster. These two conformations differ by 0.25 eV, so the question stands as to whether this is a large or small energy difference. R. M. Burke and C. E. H. Dessent reported the relative energies of the global minima of 3 different tautomeric structures to the lowest-energy tautomeric structure global minima of the Na^+•$[ATP-2H]^{2-}$ dianion (differing in the location of the two deprotonated oxygens of the triphosphate backbone group) at the B3LYP/6-31+G* level of theory as 0.010, 0.0167 and 0.822 eV.[3] Additionally, for 3 different tautomeric structure global minima relative to the lowest-energy tautomeric structure global minima of the $[ATP-2H]^{2-}$ dianion, they reported energies of 0.165, 0.227 and 0.280 eV.[3] As a further example, E. R. Williams and co-workers found small differences in energy between the global minima of the canonical and zwitterionic structures of M^+•Arg, which changed as a function of the size of M for M = Li, Na, K, Rb and Cs at the B3LYP/LACVP** level of theory.[5] Thus, for M = Li, the global minima for the canonical structure is 0.03 eV lower in energy than that for the zwitterionic structure, and for M = Na, K, Rb and Cs, the global minima for the zwitterionic structures are 0.03-0.16 eV lower in energy than those for the canonical structures. Thus, by comparison to the results presented in the papers discussed above, the difference in energy of 0.25 eV between the canonical and zwitterionic structure global minima for Br^-•Arg determined in this chapter can be interpreted as a fairly significant energy difference.

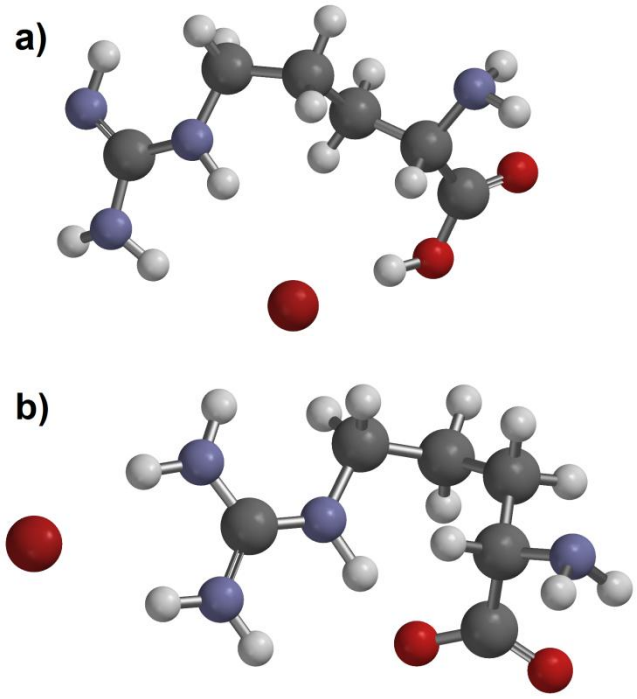

Figure 4. 9 The lowest-energy conformers of the a) canonical and b) zwitterionic structures of Br^-•Arg at the B3LYP/6-31+G** level of theory.

For Br^-•Arg, 2/3 of the zwitterionic structure conformer optimisations are lower in energy than the canonical structure conformer optimisations at the B3LYP/6-31+G** level of theory, with the difference in energy of the corresponding global minima being 0.25 eV. This result suggests that the global minima for the zwitterionic structure of Br^-•Arg is lower in energy than that for the canonical structure, and the interpretation of the experimental results given in the conclusion of Chapter 3. However, as stressed in the methods section to this chapter, calculations were not carried out on the other canonical structure isomer of Br^-•Arg (Figure 4.2 b). This is perhaps a significant isomer, and an analogous set of calculations carried out on this isomer would have been useful in completing the picture that has been built-up of the probable lowest-energy conformations for the canonical structure of Br^-•Arg.

The work presented here should be supplemented by performing further fully optimised MP2/6-31+G** calculations. As discussed in the analysis section, the MP2/6-31+G** (sp) energies were very similar for the conformers of both structures of Br^-•Arg, so the MP2/6-31+G** calculations would have been useful in determining whether the energy separation changes or remains the same between the two structures. For MP2, there is the possibility of using a triple zeta basis set such as cc-pVTZ, which is generally considered to give reliable results, or even CCSD(T) (coupled cluster) with the same basis set, but such calculations would take an unfeasibly long time for a system of the size of Br^-•Arg.

With regards to the basis set used throughout this chapter, there are improved, "triple zeta" basis sets, such as 6-311+G** or "general contracted" basis sets, such as cc-pVTZ (correlation-consistent polarised valence triple zeta, derived by Dunning and co-workers)[10,11] which could have been used, but the computational cost increases significantly for a relatively small increase in accuracy in terms of the small (< 0.1 eV) typical energy differences found between conformers for the zwitterionic and canonical structures of Br^-•Arg.[8,12] For example, for the mean unsigned errors in 11 relative glucose conformational energies for various basis sets at the HF level of theory, the 3-21G basis set gives an error of 0.087 eV, the 6-31G* basis set gives an error of 0.009 eV and the cc-pVDZ basis set gives an error of 0.004 eV.[13] Certainly, computational studies on similar systems have used basis sets of similar quality to 6-31+G** for their B3LYP calculations.[2,3,5]

It would have been possible to have used a different force-field to MMFF94 to generate the starting conformations. With a view to assessing the quality of various

organics force-fields used for organic systems and applied to amino acids such as arginine, Kaminsky and Jensen generated between 144-432 conformers for each of 4 model amino acid systems (glycine, alanine, serine, cysteine) by varying certain torsional angles by 30-180 degree steps, followed by full MP2 optimisations.[14] They found that the MMFF94 force-field found most of the MP2 conformations out of the traditional fixed-charged force-fields (MM2, MM3, CHARMM27, AMBER94, AMBER99, and OPLS) across all the model-peptides studied.

The computational methods used in this chapter compliment work carried out by R. M. Burke and C. E. H. Dessent on the 4 tautomeric structures of Na^+•$[ATP-2H]^{2-}$.[3] In that paper, the conformer distribution function in SPARTAN (5000K starting temperature) using the MMFF94 force-field was used to generate 100 conformers, for which the 50 lowest-energy conformers at the MMF94 level were grouped into 5 families and recalculated at the B3LYP/6-31+G* (sp) level of theory. The lowest-energy conformer from each family was then further optimized at the B3LYP/6-31+G* level of theory, and additional MP2/6-31+G* (sp) calculations on these conformers were then calculated in order to confirm the energy ordering between families. Primarily, that work differed to the work carried out in this chapter in that it used a slightly smaller basis set for the calculations (6-31+G* as opposed to 6-31+G**) and not as many families were found, leading to consequently fewer optimisations (5 families per structure as opposed to 10-22 families per structure). The computational methods used in this chapter and in the paper described above mirror earlier work on isolated adenosine 5'-triphosphate and adenosine 5'-diphosphate multiply- and singly- charged anions.[2]

Of perhaps more relevance to the Br^-•Arg system is the computational study carried out on the analogous cationic M^+•Arg systems by E. R. Williams and co-workers, which had the same hope of differentiating between zwitterionic and canonical structures.[5] For their two canonical structures and single zwitterionic structure, they used the AMBER force-field (optimised for biomolecules and organics)[13] and generated 120 conformers for each cluster in their series (M = Li, Na, K, Rb, Cs). This was done by using simulated annealing from 900K or less, to 200K, dependent on the identity of M in the M^+•Arg cluster, further optimising these conformers, and then using them as starting points for new molecular dynamics searches. The resultant 120 conformers for each structure could then be classified into families: 3 families for the zwitterionic structure and 2 families for each of the 2 canonical structures, giving a total of 7 different families for each M^+•Arg in the series. From each of these families, optimisations at the B3LYP/LACVP** level of theory were then carried out (The LACVP** basis set incorporates an effective core potential for K, Rb and Cs metal ions, but is 6-31G** for all other atoms).[15] In comparison to the work carried out in this chapter, the level of calculation is similar, but contains far fewer optimisations, due to the finding of considerably more families for the two structures of the Br^-•Arg system.

References

[1] *Spartan 08*; Wavefunction Inc.; Irvine, CA, 1991-2009

[2] Burke, R. M.; Pearce, J. K.; Boxford, W. E.; Bruckmann, A.; Dessent, C. E. H. *J. Phys Chem. A*, **2005**, *109*, 9775

[3] Burke, R. M.; Dessent, C. E. H *J. Phys. Chem. A* **2009**, *113*, 2683

[4] Bush, M. F.; O'Brien, J. S.; Prell, J. S.; Saykelly, R. J.; Williams, E. R. J. Am. Chem. Soc. **2007**, *129*, 1612

[5] Jockusch, R. A.; Price, W. D.; Williams, E. R. *J. Phys. Chem. A* **1999**, *103*, 9266

[6] Liu, D.; Wyttenbach, T.; Bowers, M. T. J. Am. Chem. Soc. **2006**, *128*, 15155

[7] Toroz, D.; Mourik, T. V. Mol. Phys. **2006**, *104*, 559

[8] Rassalov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *J. Comp. Chem.* **2001**, *22*, 976

[10] Dunning, Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007

[11] Woon, D. E.; Dunning, Jr., T. H. *J. Chem. Phys.* **1993**, *98*, 1358

[12] Hehre, W. J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*; Wavefunction, INC: Irvine, 2003

[13] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons, LTD: Chichester, UK, 2002

[14] Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774

[15] Hay, P. J.; Wadt, W. R. *J. Chem.. Phys.* **1985**, *82*, 270

Chapter 5: The Collision Induced Dissociation of $RCOO^-$•Arg in a Quadrupole Ion Trap

5.1 Introduction

Given the strong indication that Br^- stabilizes the zwitterionic form of arginine presented in Chapter 3, it was decided to extend this work and examine the effect of a range of mono-deprotonated dicarboxylic acids complexed to arginine, to see if the products observed under CID could be used to give a similar indication as to the tautomeric structure of arginine within $RCOO^-$•Arg gas-phase complexes. Previous work done on amino acid zwitterion stabilisation by carboxylic acids is limited, and has been theoretical in nature.[1,2] In the X^-•Arg work presented in Chapter 3, the proximity and size of the anion complexed to arginine had a large effect on the observed products and CID energies. Thus, in the current chapter, this trend is further explored in the series of mono-deprotonated dicarboxylic acids, primarily by varying the dicarboxylic acid chain length. The chain length controls the size of the molecule and is also likely to be proportional to the gas-phase basicity of the anion. This is because alkyl groups have an increased electron releasing affect in proportion to their size and can thus increase the electron density on an anionic charge or lone electron pair.[3] For the purposes of comparison, results for a selection of mono-carboxylic acid and aromatic carboxylic acid anions complexed to arginine were also obtained. Thus, CID spectra were obtained for arginine complexed to a series of dicarboxylic acids

(Figure 5.1 a), a selection of aromatic-carboxylic acids (Figure 5.1 b), and a selection of monocarboxylic acids (Figure 5.1 c).

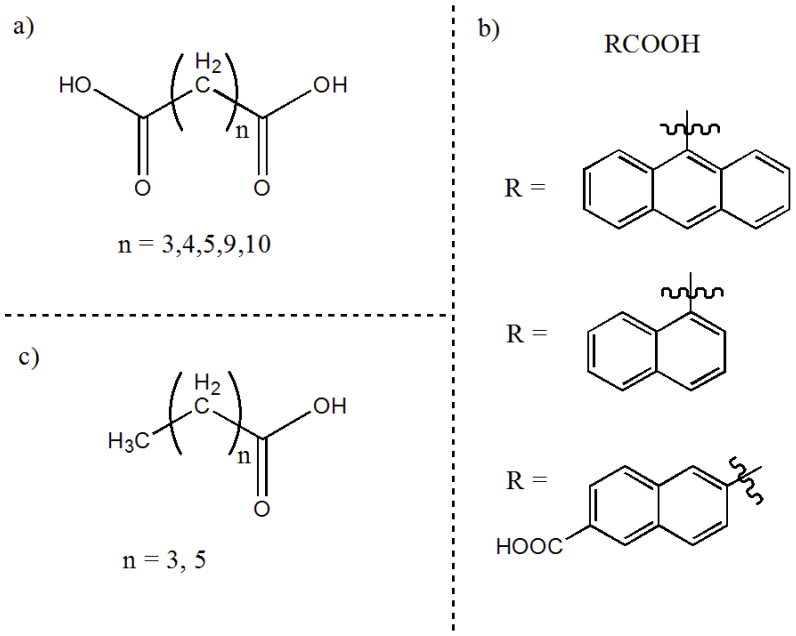

Figure 5. 1 Chemical structures of all of the carboxylic acids that were complexed to arginine in this study, where a) are dicarboxylic acids, b) are aromatic carboxylic acids and c) are monocarboxylic acids.

5.2 Experimental Section

The experiments were carried out as described in Chapter 2. The anions were prepared by electrospraying 100ml solutions of the carboxylic acid and L-arginine (both at 10^{-4} M concentration) in 100% MeOH. The solutions were made-up from powders of the carboxylic acids and arginine. All chemicals were purchased from Sigma Aldrich and required no further purification.

CID experiments were carried out on a Bruker Esquire 6000 QIT, as described in Chapter 2, with the nebulising gas set at a pressure of 10.0 psi, drying gas set at a flow rate of 8 $lmin^{-1}$, capillary temperature set to 100 $^{\circ}C$, and an excitation time of 40 ms.

The spectral resolution of the Bruker is approximately 0.3 mass units. CID graphs were recorded typically over 20-30 minutes, with each step in voltage (CID Energy) given a 1 minute time-bin. These time-bins of data points were then averaged to produce an individual intensity point. The intensity for all peaks present was then normalized and plotted against the CID voltage which was converted to a % CID Energy (as a % of the voltage limit of the trap).

5.3 Results and Discussion

Collision-induced dissociation mass spectra for all species studied are given in Figures 5.2 to 5.4 below for the various $RCOO^-$•Arg clusters of deprotonated acids complexed to arginine. Figure 5.2 gives the spectra for the series of dicarboxylic acids, R = $(CH_2)_n$COOH where n = 3,4,5,9,10. Figure 5.3 gives the spectra for the series of aromatic acids where R = $(C_{10}H_6)$COOH for 2,6,-dinaphthalene dicarboxylic acid, R = $C_{14}H_9$ for 9-anthroic acid, and R = $C_{10}H_7$ for 1-napththoic acid. Figure 5.4 gives the spectra for the series of monocarboxylic acids, R = $(CH_2)_nCH_3$ where n = 3, 5. Figure 5.5 displays an expanded region for R = $(CH_2)_n$COOH where n = 5, demonstrating the presence of various other species not observable in Figures 5.2 to 5.4. Table 5.1 gives a full list of all the peaks that are visible in the CID-MS for all the $RCOO^-$•Arg clusters. Figure 5.6 gives the normalised intensities of the products present at 0.40 V. Finally, Figures 5.7 to 5.11 give the corresponding CID curves for all of the spectra from Figures 5.2 to 5.4 respectively. Note that on throughout this chapter, Arg^- will be used to denote deprotonated arginine (i.e. Arg – H^+).

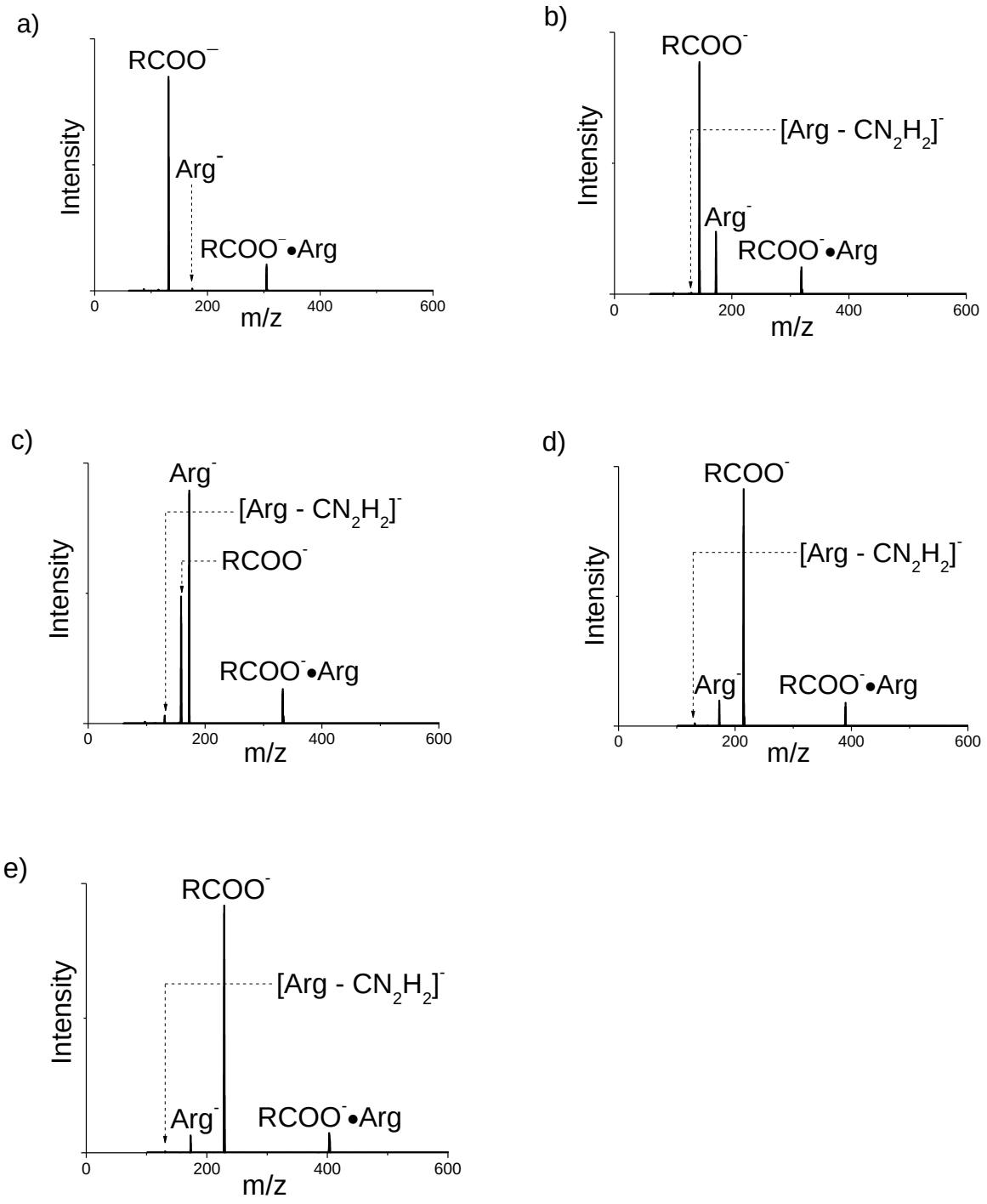

Figure 5. 2 Negative ion ESI-CID mass spectra obtained from collisionally activated dissociation of the dicarboxylic $RCOO^-$·Arg clusters, where R = $(CH_2)_nCOOH$ for a) n = 3 at 0.36 V,[1] b) n = 4 at 0.39 V, c) n = 5 at 0.41 V, d) n = 9 at 0.35 V and e) n = 10 at 0.33 V.

[1] Note loss of cyanamide, CN_2H_2, probably occurs for $COOH(CH_2)_3COO^-$•Arg, but the peak is the same mass as $COOH(CH_2)_3COO^-$ for this complex.

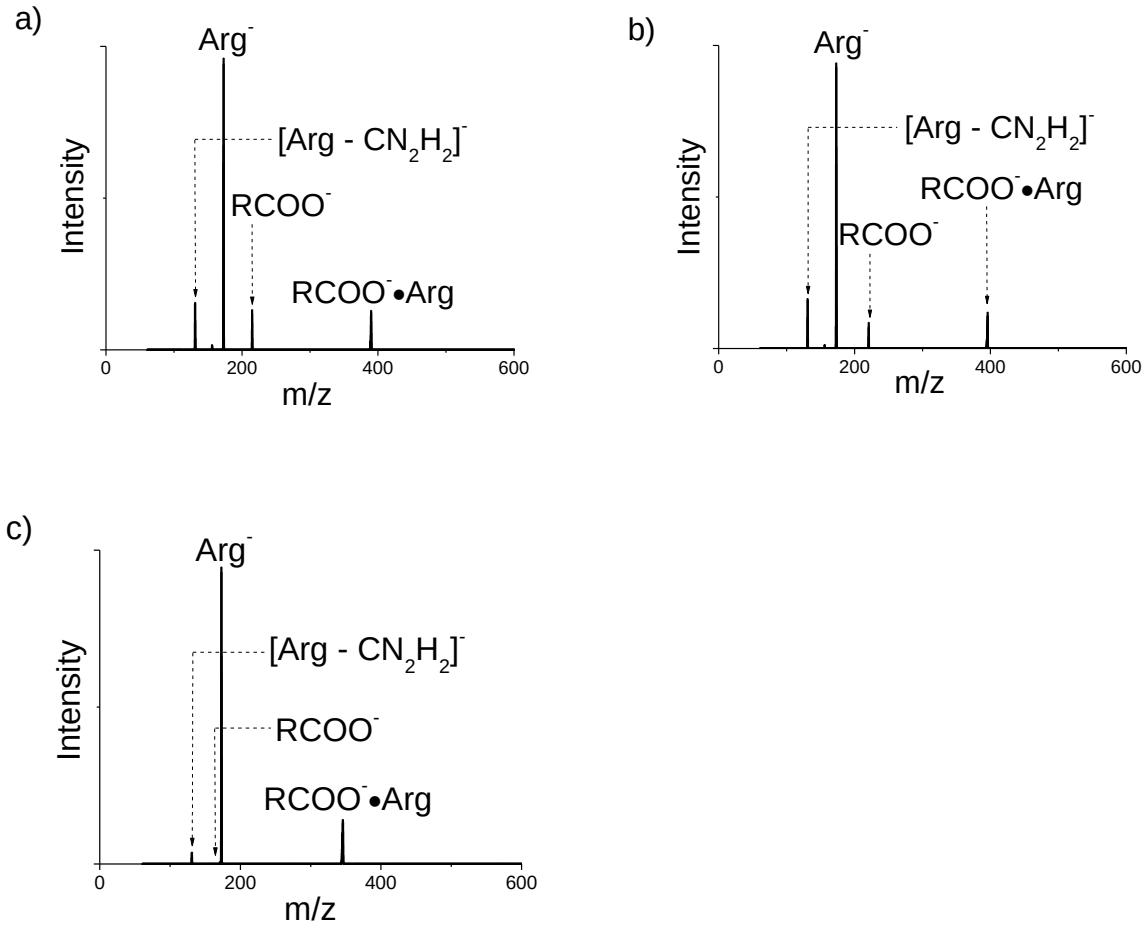

Figure 5. 3 Negative ion ESI-CID mass spectra obtained from collisionally activated dissociation of the aromatic $RCOO^-$·Arg clusters for a) R = $COOH(C_{10}H_6)$ at 0.34 V, b) R = $C_{14}H_9$ at 0.34 V and c) R = $C_{10}H_7$ at 0.39 V.

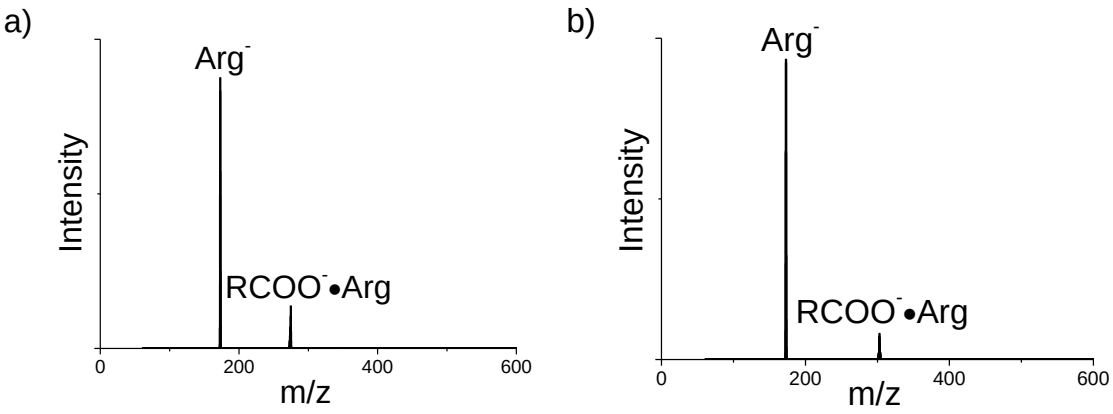

Figure 5. 4 Negative ion ESI-CID mass spectra obtained from collisionally activated dissociation of the monocarboxylic $RCOO^-$·Arg clusters where R = $CH_3(CH_2)_n$ for a) n = 3 at 0.23 V and b) n = 5 at 0.32 V.

As can be seen from the spectra, we observe the appearance of three different products, each corresponding to different collisionally activated dissociation reaction channels available to the cluster. These three channels are: deprotonation [5.1], fragmentation [5.2] and cyanamide loss [5.3]. All of the complexes studied exhibited all three of the following CID pathways, apart from the monocarboxylic acids, which only exhibited the deprotonation loss pathway, [5.1].

$$RCOO^-\bullet Arg \rightarrow RCOOH + Arg^- \quad [5.1]$$

$$RCOO^-\bullet Arg \rightarrow RCOO^- + Arg \quad [5.2]$$

$$RCOO^-\bullet Arg \rightarrow [Arg - CN_2H_2]^- + RCOOH \quad [5.3]$$

Loss channel [5.1] (deprotonation), corresponds to proton abstraction from arginine, if we assume that the charge is initially localized on the carboxylic acid. Deprotonation is expected since the pK_a of the α-carboxylic acid group of neutral arginine is 2.01,[4]

whereas the pK_a's of all acids in this study vary from 3.65 to 4.89.[6,10,13] The acid with the higher pK_a has the most basic conjugate anion, thus all of the carboxylic acid anions in the study are more basic than neutral arginine. Thus, the mass spectra of Arg^-•RCOOH would be dominated by the Arg^- product because Arg^- is less basic than $RCOO^-$, and so is unlikely to proton abstract from RCOOH i.e. a Arg^-•RCOOH complex is unlikely to exhibit fragmentation and deprotonation pathways together, which many of the clusters do. In addition, as discussed in Chapter 3, the formation of the cluster Arg^-•RCOOH is unlikely, because such a species would fragment at low energy (Arg^-•RCOOH would not be as stable as $RCOO^-$•Arg), leading to spectra dominated by Arg^-.

Loss channel [5.2] corresponds to simple fragmentation of the cluster, if it is assumed that the negative charge initially resides on the carboxylic acid anion, $RCOO^-$, in the $RCOO^-$•Arg cluster. Loss channel [5.3], however, corresponds to loss of cyanamide from the cluster, and occurs through a mechanism involving the guanidinium side-chain group of arginine, in an analogous fashion to ammonia loss (see Chapter 3, Figure 3.6 b). The presence of this product suggests a two-step mechanism in which a proton is abstracted by the dicarboxylic acid from either the protonated guanidinium N of the zwitterion or the carboxylate group of the canonical structure of arginine, giving Arg^-, which then loses cyanamide giving the $[Arg - CN_2H_2]^-$ product. Note, with the X^-•Arg clusters from Chapter 3, cyanamide loss is not observed, although both fragmentation and deprotonation loss products are observed. Also, note that products already present in the CID spectra are not themselves excited as the CID voltage is applied at the secular frequency of the parent cluster only.

The products corresponding to fragmentation pathways [5.1] → [5.3] are not, however, the only products, and peaks corresponding to $[Arg - NH_3]^-$, $[RCOO - H_2O]^-$, $[RCOO – CO_2]^-$ and $[RCOO – H_2O – CO_2]^-$, are observed for some of the clusters. An example of an expanded mass spectrum illustrating the presence of these minor peaks is given in Figure 5.5 below, along with a table (Table 5.1) denoting the presence or absence of these additional peaks for all the clusters studied. Generally, these peaks are indicative of additional loss routes that will not be discussed further in this work due to space and time constraints.

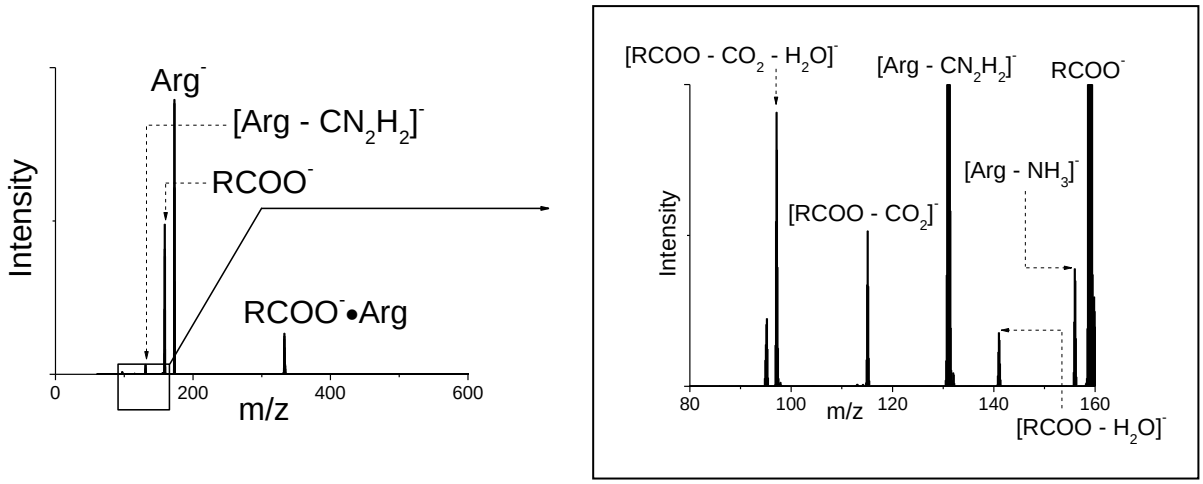

Figure 5. 5 Expanded negative ion ESI-CID mass spectrum obtained from collisionally activated dissociation of the dicarboxylic cluster $RCOO^-$·Arg, where R = $(CH_2)_5COOH$ at 0.50 V

Table 5. 1 Table of all of the major and minor product ions observed in the negative ion ESI-CID mass spectra obtained from collisionally activated dissociation of the parent cluster, $RCOO^-$·Arg for all clusters in the study.

	Major products			Minor products			
	$RCOO^-$	Arg^-	$[Arg - CN_2H_2]^-$	$[Arg - NH_3]^-$	$[RCOO - H_2O]^-$	$[RCOO - CO_2]^-$	$[RCOO - CO_2 - H_2O]^-$
$R = COOH(CH_2)_n$							
$n = 3$	✓	✓	×[1]	×	✓	✓	×
$n = 4$	✓	✓	✓	×	✓	✓	×
$n = 5$	✓	✓	✓	✓	✓	✓	✓
$n = 9$	✓	✓	✓	×	✓	✓	×
$n = 10$	✓	✓	✓	✓	✓	×	✓
$R = (C_{10}H_6)COOH$	✓	✓	✓	✓	×	×	×
$R = C_{14}H_9$	✓	✓	✓	✓	×	×	×
$R = C_{10}H_7$	✓	✓	✓	✓	×	×	×
$R = CH_3(CH_2)_n$							
$n = 3$	×	✓	×	×	×	×	×
$n = 5$	×	✓	×	×	×	×	×

[1] Same mass as $RCOO^-$ for this acid

5.4 Further Discussion

Figure 5.6 below gives the normalized intensities of the major products (loss channels [5.1] → [5.3]) for all species in this study at 0.40 V excitation energy in the QIT, extracted from the CID data.

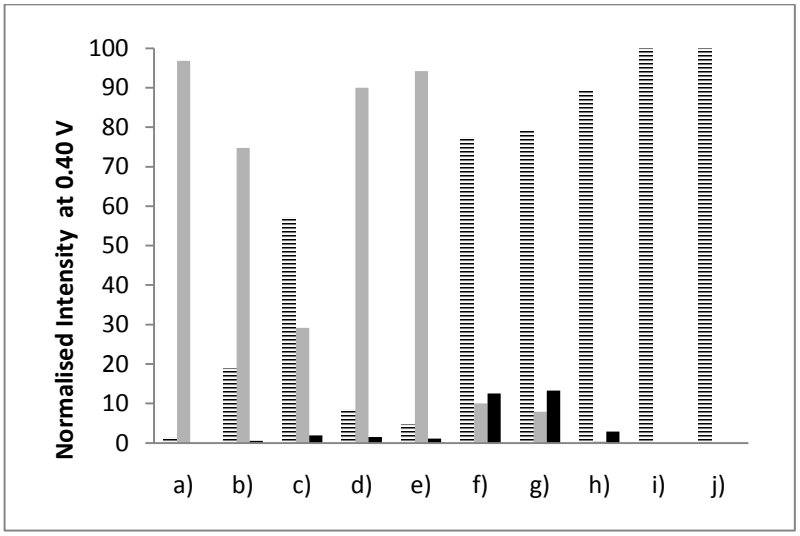

Figure 5. 6 The normalized intensities of Arg ⁻ (striped, corresponding to loss channel [5.1]), RCOO ⁻ (grey, corresponding to loss channel [5.2]) , and cyanamide loss (solid black, corresponding to loss channel [5.3]) obtained at 0.40 V CID energy from the parent ion, RCOO⁻•Arg. R = $(CH_2)_n$COOH for a)-e), where a) n = 3, b) n = 4, c) n = 5, d) n = 9 and e) n = 10. For f), R = $(C_{10}H_6)$COOH, for g) R = $C_{14}H_9$, and for h) R = $C_{10}H_7$. R = $(CH_2)_nCH_3$ for i)-j), where i) n = 3 and j) n = 5.

The most interesting feature of the plot displayed in Figure 5.6 is the clear minimum in the fragmentation products (grey), corresponding to loss channel [5.2], that occurs for the dicarboxylic acid RCOO⁻•Arg clusters, a)-e), which gives an indication of the stability of each of these clusters relative to one another. Each of the three sets of RCOO⁻•Arg clusters are discussed further below.

5.4.1 Dicarboxylic acids

Figure 5.7 gives the negative ion ESI-CID curves obtained from collisionally activated dissociation of the dicarboxylic $RCOO^-$·Arg clusters, where R = $(CH_2)_nCOOH$. These CID curves exhibit all of the major products of deprotonation, fragmentation and cyanamide loss. Additionally, Table 5.2 gives the real and predicted pK_a values for each of the RCOOH in this series. On increasing the alkyl chain length of R, the basicity of $RCOO^-$ increases due to an increased electron releasing effect.[3] This trend is observed in the experimental and calculated pKa values for each RCOOH in Table 5.2.

Table 5. 2 Table of experimental and calculated pKa values for RCOOH, where R = $(CH_2)_nCOOH$ for *n* = 3, 4, 5, 9 and 10

n	Exp. pKa in H_2O	Calculated pKa in H_2O[5]
3	4.34 [6]	4.51
4	4.42 [6]	4.65
5	4.48 [6]	4.71
9	-	4.74
10	-	4.75

Comparison can be made between the CID curves for the decay of the parent ion, $RCOO^-$·Arg, where R = $(CH_2)_nCOOH$ (Figure 5.7). E_{half} is the CID energy at which half of the precursor ions have decayed to product ions. E_{half} = 12% CID energy for *n* = 3, 12.5% for *n* = 4, 13% for *n* = 5, 10% for *n* = 9 and 11% for *n* = 10. Thus there is a maximum in E_{half} for decay of the parent ion on increasing the chain length, with that maximum occurring for *n* = 5. This trend is clearly illustrated by Figure 5.8, which combines these CID curves.

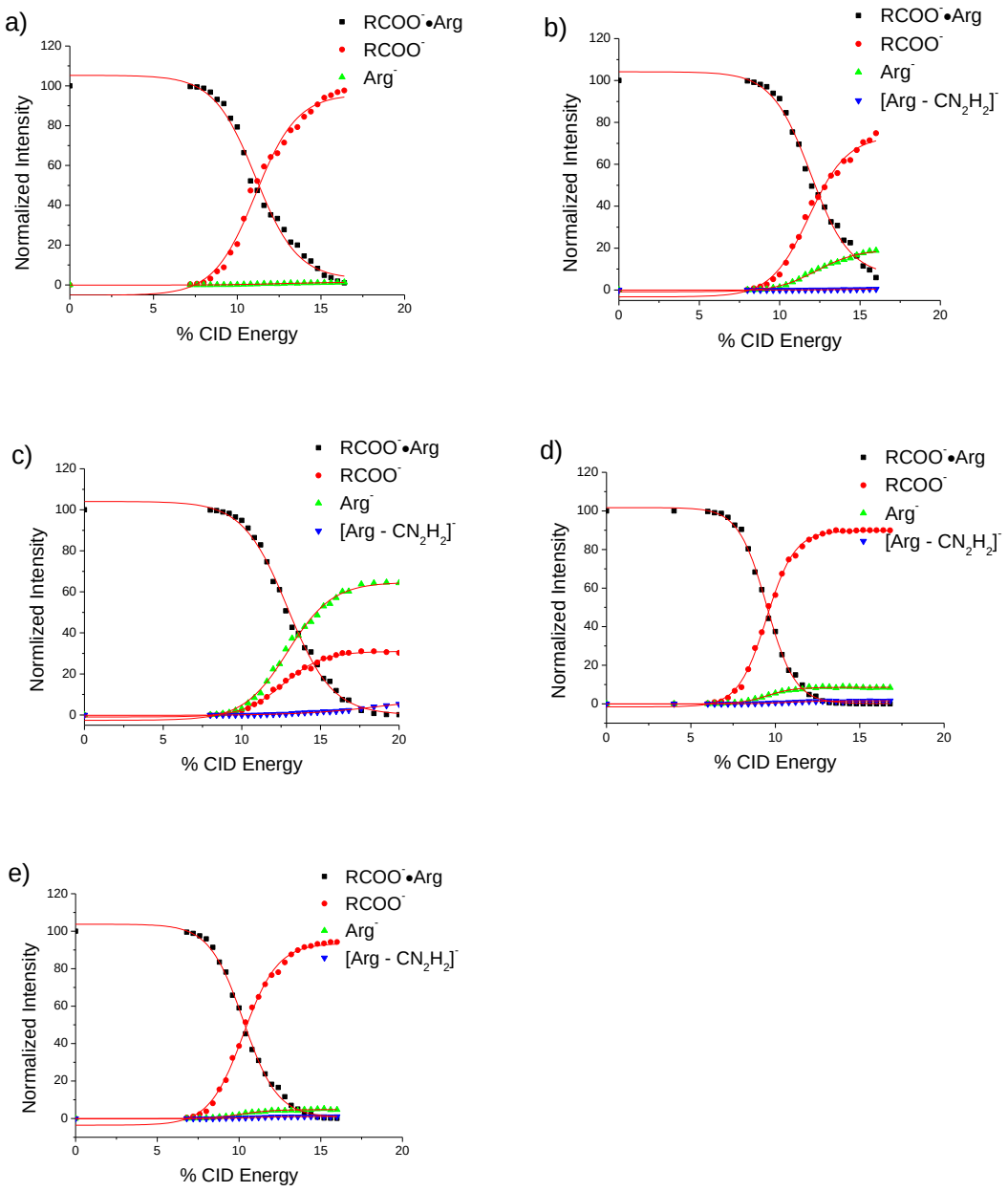

Figure 5. 7 % CID curves obtained from collisionally activated dissociation of the dicarboxylic RCOO⁻•Arg clusters, where R = $(CH_2)_n$COOH for a) n = 3, b) n = 4, c) n = 5, d) n = 9 and e) n = 10.

With reference to Figure 5.6 a)-e) and Figure 5.7 a)-e), the fragmentation product, $RCOO^-$, appears in greatest abundance for all RCOO⁻•Arg clusters, apart from for n = 5, which primarily gives the deprotonation product. However, because it has the

highest CID energy in the series, it's fragmentation channel must occur at higher energy than this. Thus, the trend in CID energy is a direct comparison of the energy of the fragmentation channel between all the $RCOO^-$•Arg clusters, where R = $(CH_2)_nCOOH$: The energy of the fragmentation loss channel increases from n = 3 and n = 4, then the deprotonation loss channel is the lowest energy for n = 5, but at the highest CID energy of the series thus indicating an even higher fragmentation loss channel, followed by lower energy fragmentation loss channels for n = 9 and 10.

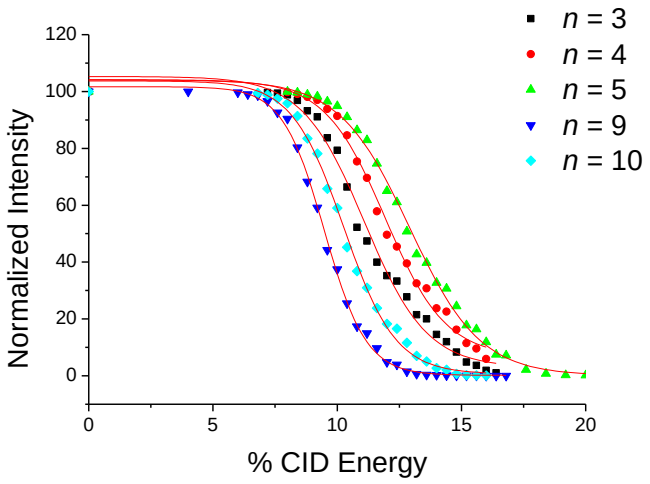

Figure 5. 8 Combined % CID curves from Figure 5.7, obtained from collisionally activated dissociation of the dicarboxylic $RCOO^-$·Arg clusters, where R = $(CH_2)_nCOOH$

This maximum in the relative energy of fragmentation is highly suggestive of there being a difference in the strength of binding of $RCOO^-$ in the $RCOO^-$•Arg cluster on traversing the series. Additionally, the assignment of this loss channel to fragmentation of an $RCOO^-$•Arg cluster (equation [5.2]) can be made with more confidence, as the pK_a values of RCOOH in water increase with increasing chain length (Table 5.2). Thus, we would expect a change in % CID energy in the same direction, not a maximum, if we were in fact observing deprotonation of RCOOH by

Arg^- in an Arg^-•RCOOH cluster. As discussed in relation to the X^-•Arg clusters (Chapter 3), an increase in the energy to fragment can be attributed to an increase in the binding energy of $RCOO^-$ to Arg in the $RCOO^-$•Arg cluster. It appears that the interactions present in the mid-length, n = 5, dicarboxylic acid are stronger than those present in the shorter and longer chain acids. The n = 5 dicarboxylic acid is a similar length to arginine, and might be a "better fit" than the other acids, which would explain the observed trend.

Table 5. 3 Table of estimated molecular lengths of the alkyl chain in Å of RCOOH for R = $CH_3(CH_2)_nCOOH$, where n = 3-10, calculated using geometry optimisations at the MMFF94 level of theory using SPARTAN 08. Also included is the estimated length of arginine.

RCOOH where R = $CH_3(CH_2)_nCOOH$	Length in Å from C to C of the two COOH groups of RCOOH[1]
n = 3	5.04
n = 4	6.35
n = 5	7.54
n = 9	12.59
n = 10	13.90
neutral arginine	6.07

To illustrate this point, estimated lengths of the acids n = 3 to 10 for RCOOH where R = $CH_3(CH_2)_nCOOH$ and arginine have been collected in Table 5.3. As can be seen from the table, neutral arginine has a similar length to the acid where n = 4. The estimate is not very accurate, but supports the idea that arginine should be a similar length to the n = 5 acid.

[1] In the case of arginine, the distance is between the C of the COOH terminus and the C between the 3 N's of the guanidinium side-chain group

There are similar studies that attribute changes in E_{half} to changes in binding energy of a resulting fragment to the remaining precursor ion. For example, C. J. Taylor et al. took QIT-MS CID spectra of $Re_2Cl_8^{2-}$ and $Re_2Br_8^{2-}$ which both decayed via ionic fragmentation to form $Re_2Cl_7^-$ + Cl^- and $Re_2Br_7^-$ + Br^- respectively.[7] They found that E_{half} for decay of the parent ion increased from ~3% to ~7% CID energy (2.5 V maximum voltage) on going from $Re_2Br_8^{2-}$ to $Re_2Cl_8^{2-}$, which they attributed to an increase in the binding energy of the Cl^- ion over the Br^- ion. Similarly, J. S. Brodbelt and M. Satterfield attribute the relative variation in CID energy in a QIT for a series of pyridyl ligand/metal complexes as being due to the relative strength of binding of the ligand to the metal centre.[8]

It must be noted that there exists a limitation in CID whereby the mass of the species under study can effect the CID energy imparted to the precursor ion. As a consequence, no two masses that are different by more than 100 m/z mass units, or have different charges, are usually compared.[9] The difference in mass between the lightest and heaviest ions in the above is 98 m/z, which is at the threshold at which the mass becomes significant. However, it is probable that this effect is not at play in the observed trend, as it would cause a lowering of the % CID energy on increasing the mass of the precursor, and not the clear maximum observed in the series.

5.4.2 Aromatic carboxylic acid clusters

The aromatic carboxylic acid clusters all primarily give the deprotonation product upon CID excitation. Their CID curves are given in Figure 5.9 below. In addition, chemical structures for these acids are shown in Figure 5.10.

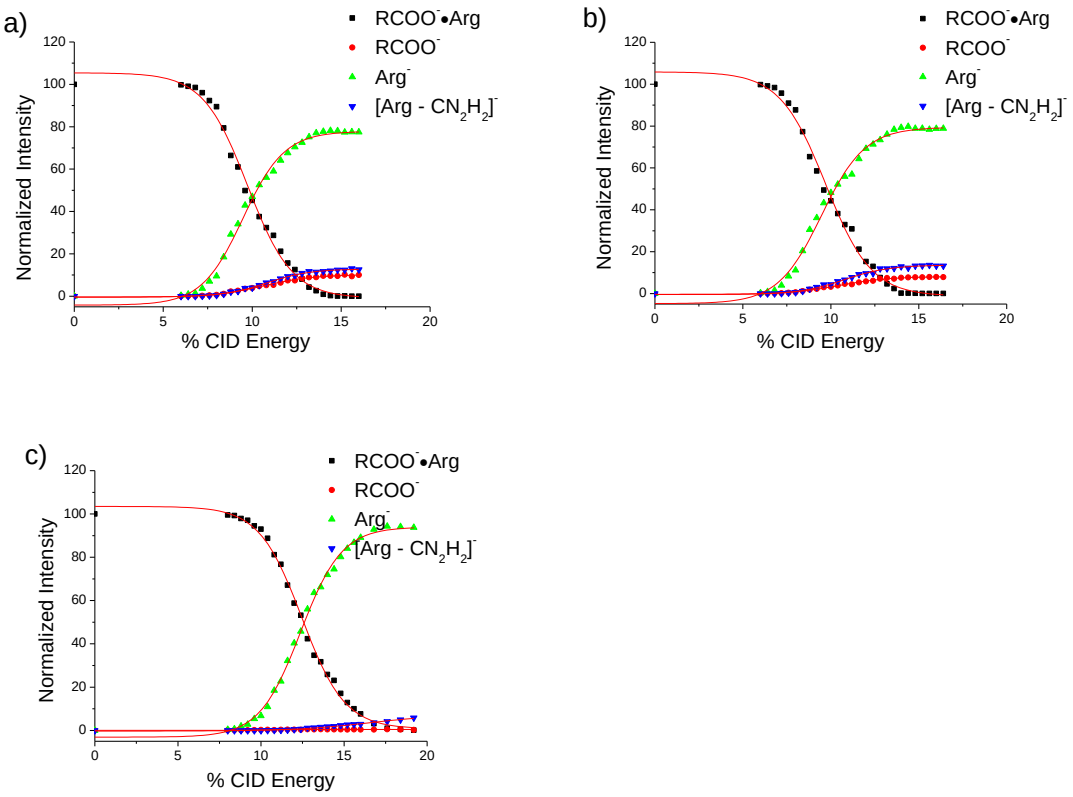

Figure 5. 9 % CID curves obtained from collisionally activated dissociation of the aromatic $RCOO^-$·Arg clusters for a) R = $COOH(C_{10}H_6)$, b) R = $C_{14}H_9$, and c) R = $C_{10}H_7$.

On examining these CID curves, E_{half} (the energy at which half of the precursor ions have decayed) is ~10% for R = $COOH(C_{10}H_6)$ and R = $C_{14}H_9$, whereas for R = $C_{10}H_7$ it is ~12.5 %. These observed E_{half} values give a direct comparison of the energy of the proton abstraction loss channel, [5.1], between these three clusters, as all three primarily decay by this route to give Arg^- as the most abundant product. Thus, the proton abstraction loss channel is lower energy for R = $COOH(C_{10}H_6)$ and R = $C_{14}H_9$, than for R = $C_{10}H_7$. As discussed in the Chapter 3, this can be related to the basicity of the anion, and a similar parallel will be made here.

RCOOH

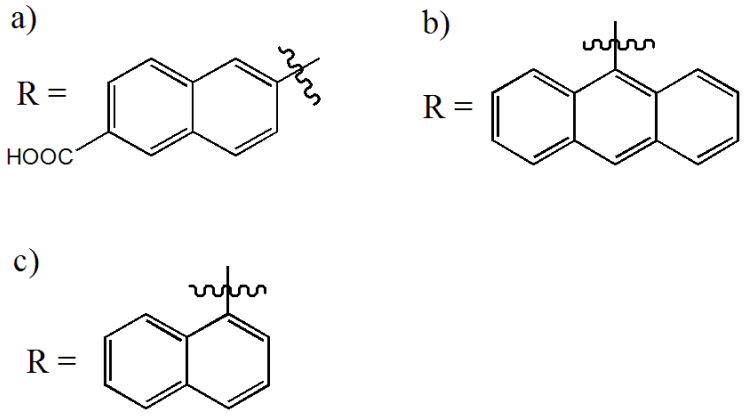

Figure 5. 10 Chemical structures of RCOOH for a) R = $COOH(C_{10}H_6)$, b) R = $C_{14}H_9$, and c) R = $C_{10}H_7$.

Comparing R = $COOH(C_{10}H_6)$ to R = $C_{10}H_7$, it is the addition of the second carboxylate group in the former that causes more effective charge delocalisation as compared to the latter, and thus a decrease in the basicity of $RCOO^-$. Similarly, when comparing R = $C_{14}H_9$ to R = $C_{10}H_7$, it is the additional ring in R = $C_{14}H_9$ that allows an increase in charge delocalisation, and thus reduction in basicity of $RCOO^-$. However, the increased anion basicity appears to have had the effect of increasing the energy required for the deprotonation loss channel, which is contrary to the result found in Chapter 3 for the X^-•Arg system, where the energy for the deprotonation loss channel decreased on going from X = Cl to X = F (i.e. to higher basicity).

It is interesting to note that the additional carboxylate group for R = $COOH(C_{10}H_6)$, and the additional ring for R = $C_{14}H_9$, as compared to R = $C_{10}H_7$, have a very similar effect to one another, as both molecules have almost identical decay curves. The pK_a in water for RCOOH where R = $C_{14}H_9$ is 3.65, whereas for R = $C_{10}H_7$ it is 3.69, thus

the latter is predicted to have a conjugate anion ($RCOO^-$) of higher basicity.[10] The pK_a for RCOOH where R = $(C_{10}H_6)COOH$ could not be found in the literature.

5.4.3 Monocarboxylic Acids

$RCOO^-$•Arg where R = $CH_3(CH_2)_n$ for n = 3 and n = 5, both give the de-protonation product, Arg^-, exclusively upon CID excitation, consistent with a GPA for n = 3 of $\Delta H = 1449 \pm 8.8$ kJ mol^{-1},[11] which is greater than that for arginine, where $\Delta H = 1388 \pm 13$ kJ mol^{-1}.[12] In Figure 5.12 below, the % CID energy of the deprotonation loss channel increases on going from n = 3 to n = 5 with the increased basicity of the $RCOO^-$ anion.

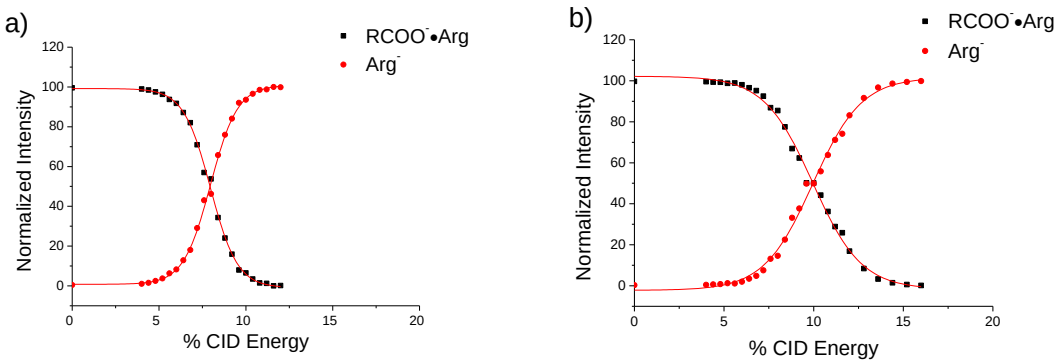

Figure 5. 11 % CID curves obtained from collisionally activated dissociation of the monocarboxylic $RCOO^-$·Arg clusters where R = $CH_3(CH_2)_n$ for a) n = 3 and b) n = 5.

The increased basicity of the anion is caused by the increased electron releasing effect of R as it increases in size. The increased basicity of the two anions, where n = 3 and n = 5, is observed in the increased pK_a values in H_2O of the corresponding RCOOH, which are 4.84 and 4.89 respectively.[13] This is in agreement with the result observed for the aromatic acids above, but contradicts the trend of increased basicity leading to

a lower proton abstraction loss channel seen for the X^-•Arg system of Chapter 3 (where X = F and Cl).

5.5 Conclusion

Clearly the type of substituent that "R" is in the $RCOO^-$•Arg clusters, has a large effect on the energies and relative proportions of the CID products observed upon resonance excitation. The energy of the deprotonation pathway for $RCOO^-$•Arg where R = $(CH_2)_nCH_3$ increases as the chain length is increased from *n* = 3 to *n* = 5. The energy of the deprotonation pathway for aromatic $RCOO^-$•Arg systems where R = $(C_{10}H_6)COOH$, $C_{14}H_9$ and $C_{10}H_7$ changes according to the stabilization of the charge on $RCOO^-$. Thus, the addition of a ring or carboxylic acid group stabilizes the excess anionic charge, reduces the basicity, and lowers the relative energy of the deprotonation loss-channel for R = $(C_{10}H_6)COOH$ and R = $C_{14}H_9$ as compared to R = $C_{10}H_7$. Perhaps, the most surprising result, however, is for R = $(CH_2)_nCOOH$, which gives a maximum in the energy to produce the $RCOO^-$ fragment, interpreted as a maximum in the binding energy of $RCOO^-$ to Arg in $RCOO^-$•Arg, which occurs at *n* = 5, as the chain length is increased through *n* = 3, 4, 5, 9 and 10. Unfortunately, this trend is not readily explained without the aid of a computational study, and such a study would take an unfeasibly long time for this range of molecular systems, some with considerable conformational flexibility.

It is, however, possible to envisage that the calculation would be of the lowest-energy conformers for both the zwitterionic and canonical structures of $RCOO^-$•Arg where R = $(CH_2)_nCOOH$ for each value of *n* in the series. The resulting global minima for each structure, could then provide an explanation for the overall trend in the series. As a

preliminary to a further computational study, an initial SPARTAN 08 MMFF94 conformer distribution calculation was carried out. The 100 lowest-energy conformers for each of the zwitterionic and canonical structures for n = 3, 5 and 10 were calculated, and the lowest-energy conformer taken for each of these (Figure 5.12). In Figure 5.12, the lowest-energy canonical structures are denoted IC, IIC, IIIC, and the lowest-energy zwitterionic structures are denoted IZ, IIZ, IIIZ, for n = 3, 5 and 10 respectively.

With reference to the calculated geometries in the Figure 5.12, it is instructive to first consider the case where three clusters could be zwitterionic: Thus, IZ zwitterion only interacts with one end of arginine, so is not particularly stable, as compared to IIZ zwitterion where the carboxylic acid has enough flexibility to interact with both ends of arginine. IIIZ also interacts with both ends but there is increased strain there resulting in a destabilization relative to IIZ. A similar argument might apply to IC, IIC and IIIC, though the simple visual inspection method presented above does not easily explain the trend in stabilities for these structures. Equally, there might be a transition from zwitterionic to canonical, or vice versa which would involve both type of structures being considered together.

It is currently unclear which structure arginine adopts in the cluster, however, the physical result appears to have its origin in the changing interactions of $RCOO^-$ where R = $(CH_2)_nCOOH$ through n = 3, 4, 5, 9 and 10 with arginine in either of its zwitterionic or canonical structures. What is clear is that the energy for binding increases until n = 5, then decreases on further increasing the chain length, which is

probably due to this acid being a particularly "good fit" to arginine, allowing arginine to interact with both COO^- and COOH ends of the dicarboxylic acid effectively.

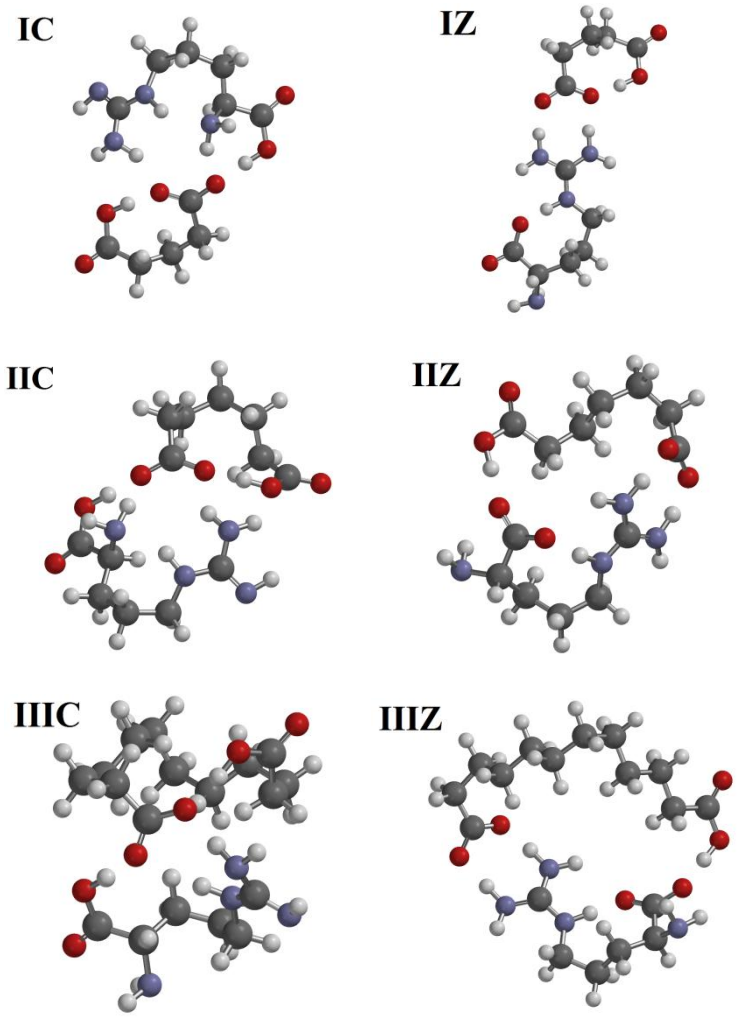

Figure 5. 12 Lowest-energy geometries from a SPARTAN conformer distribution calculation of the 100 lowest-energy conformers at the MMF94 level of theory for the canonical and zwitterionic structures of Arg in $RCOO^-$•Arg, where R = $(CH_2)_n$COOH. IC, IIC and IIC represent the canonical structure of Arg for n = 3, 5 and 10 respectively. IZ, IIZ and IIIZ represent the zwitterionic structure of Arg for n = 3, 5 and 10 respectively.

The correlation of the dicarboxylic acid chain length affecting the binding properties in gas-phase clusters has been observed before in the photoelectron spectra of L. S. Wang and corworkers.[14] They observed an increase in the second electron binding

energy for a series of $COO^-(CH_2)_nCOO^-$ dianions with increasing chain length, which they attributed to the change in Coulombic repulsion energy as a function of the distance of the two anionic centres from one another with increasing *n*. They found that the dianion with $n = 2$ is unstable however, but on solvation with one water molecule forming the $COO^-(CH_2)_2COO^-•H_2O$ cluster, becomes stable. Complexation with a second water molecule then increases the binding energy of the second electron further, which indicates that the two water molecules stabilize the two anionic centres separately.

A similar affect might be occurring for the $RCOO^-$•Arg series where R = $CH_3(CH_2)_nCOOH$, in which effective binding occurs when both the COOH and COO^- ends of the $RCOO^-$ anion can interact effectively with arginine in the $RCOO^-$•Arg cluster. This would then manifest itself as an increase in stability for IIZ/IIC i.e. the dicarboxylic acid where $n = 5$.

References

[1] Yang, G.; Zu, Y.; Liu, C.; Fu, Y.; Zhou, L. *J. Phys. Chem. B* **2008**, *112*, 7104

[2] Kass, S. R. *J. Am. Chem. Soc. Comm.* **2005**, *127*, 13098

[3] Clayden, J.; Greeves, N.; Warren, S.; Wothers, P.; *Organic Chemistry*, Oxford University Press, Oxford, 2001

[4] Stryter, L.; Tymoczko, J. L.; Berg, J. M. *Biochemistry*; 5th ed, W. H. Freeman & Co: New York, NY, 2002

[5] Calculation carried out using SPARC (v4.5)

[6] Braude, E. A.; Nachod F. C. *Determination of Organic structures by Physical Methods*, Academic Press, New York, 1955

[7] Taylor, C. J.; Wu, B.; Nix, M. G. D.; Dessent, C. E. H. *Chem. Phys. Lett.* **2009**, *479*, 184

[8] Satterfield, M.; Brodbelt, J. S. *Inorg. Chem.* **2001**, *40*, 5393

[9] Satterfield, M.; Brodbelt, J. S. *Inorg. Chem.* **2001**, *40*, 5393 and references therein

[10] Braude, E. A.; Nachod F. C. *Determination of Organic structures by Physical Methods,* Academic Press, New York, 1955

[11] Caldwell, G.; Renneboog, R.; Kebarle, P. *Can. J. Chem. Ecol.* **1989**, *67*, 661

[12] O'Hair, R.J.; Bowie, J.H.; Gronert, S. *Int. J. Mass Spectrom. Ion Proc.* **1992**, *117*, 23

[13] Perrin, D. D., *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworths, London, 1965; Supplement 1972

[14] Ding, C. F.; Wang, X. B.; Wang, L. S. *J. Phys. Chem. A* **1998**, *102*, 8634

Appendix 1: XY_n^{z-} solvent clusters in a QIT

This appendix details some uncompleted work on small metal ion solvent clusters generated in a QIT

A1.1 MeCN Collision Spectra

ESI has generally been optimized to produce bare (unsolvated) gas-phase ions.[1] However, for fundamental studies of solvation effects it would be useful to be able to produce small, solvated clusters of electrosprayed ions for further study. Thus, a method that introduces solvation after the ESI source, such as the trapping region of a QIT, would be advantageous. Indeed, there have been various reports of electrosprayed ions clustering to several water molecules,[2,3,4] and more generally, clustering in a number of different types of mass spectrometer with differing sources.[5,6,7] In light of this, the purpose of this work was to explore ways of forming solvated complexes of molecular ions (anionic and ion-pairs) within a QIT, and to subsequently investigate any loss channels of these clusters through the use of CID.

Some preliminary results for various XY_n^{z-} complexes, where X = Pt or Ir, Y = CN^-, Br^- or Cl^- and Z = charge state, complexed to MeCN were obtained by seeding the He buffer gas in an LCQ QIT gas with 1% MeCN (the experimental set-up is given in Figure A1.4). The spectra of these various species are reported in the figures below,

allowing observations to be made about the conditions under which ion peaks will cluster within a QIT.

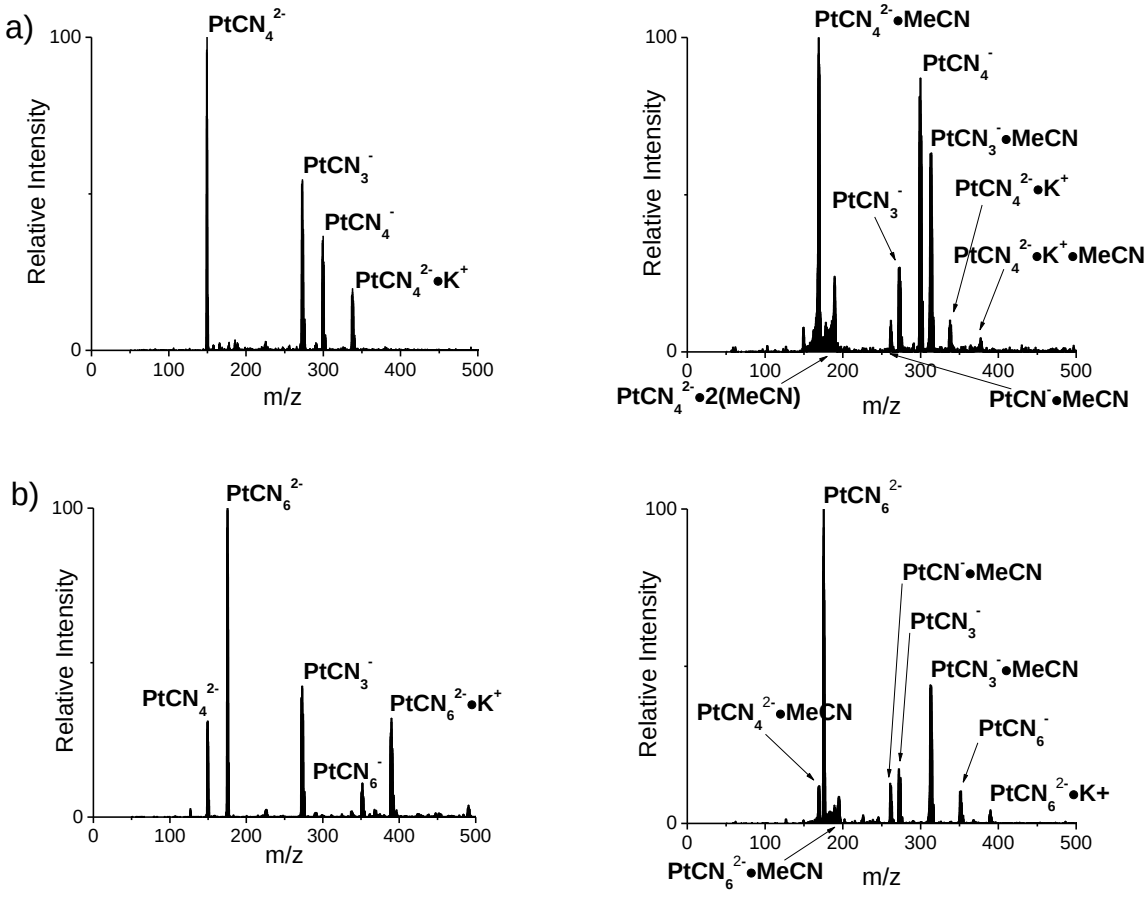

Figure A1. 1 ESI spectra of a) $PtCN_4^{2-}\cdot K^+$ and b) $PtCN_6^{2-}\cdot K^+$. For each species, the spectra has been recorded with MeCN seeded He collision gas (right) and with the normal He collision gas (left).

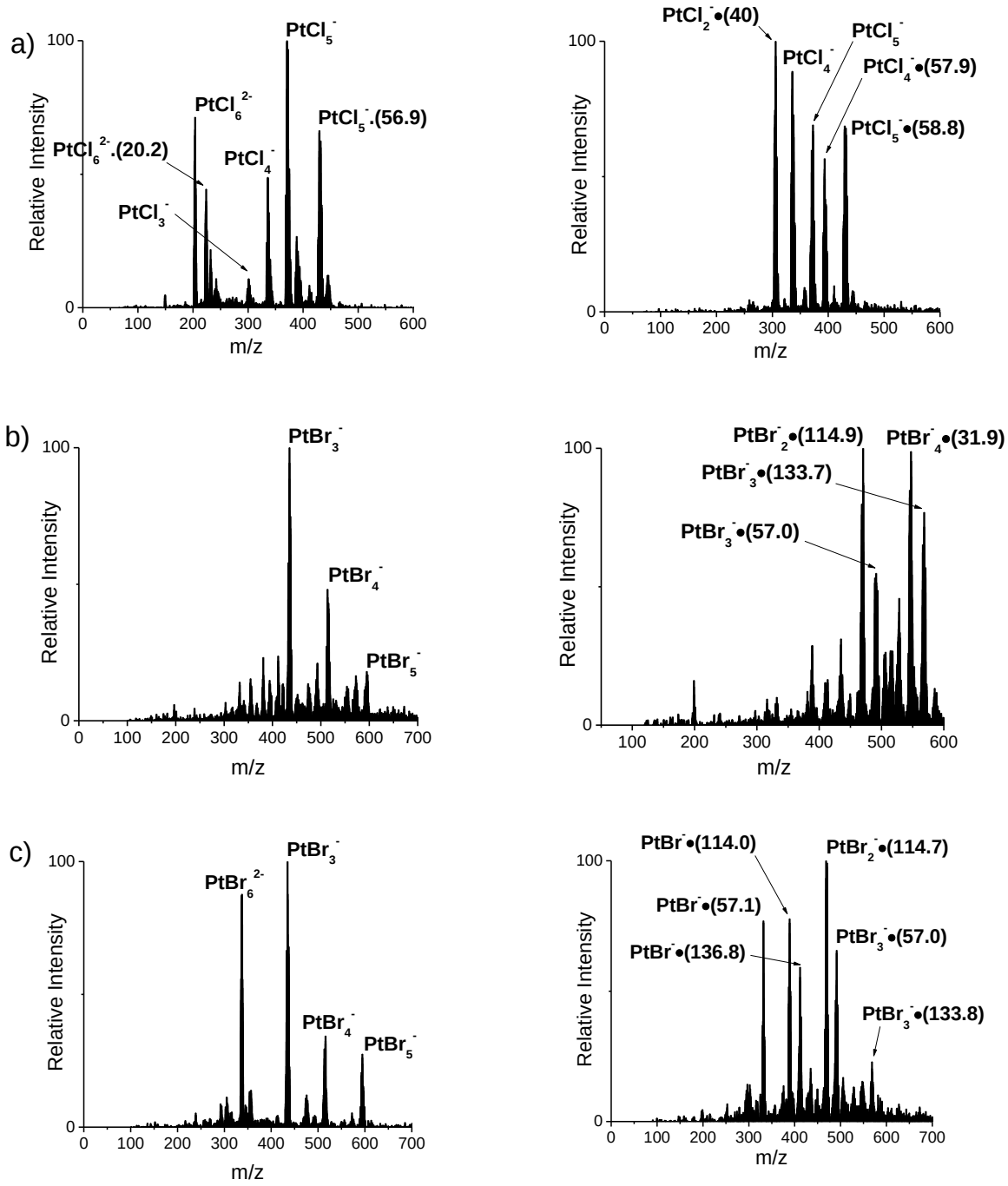

Figure A1. 2 ESI spectra of a) $PtCl_6^{2-}\cdot K^+$, b) $PtBr_4^{2-}\cdot K^+$ and $PtBr_6^{2-}\cdot K^+$. For each species, the spectra has been recorded with MeCN seeded He collision gas (right) and with the normal He collision gas (left).

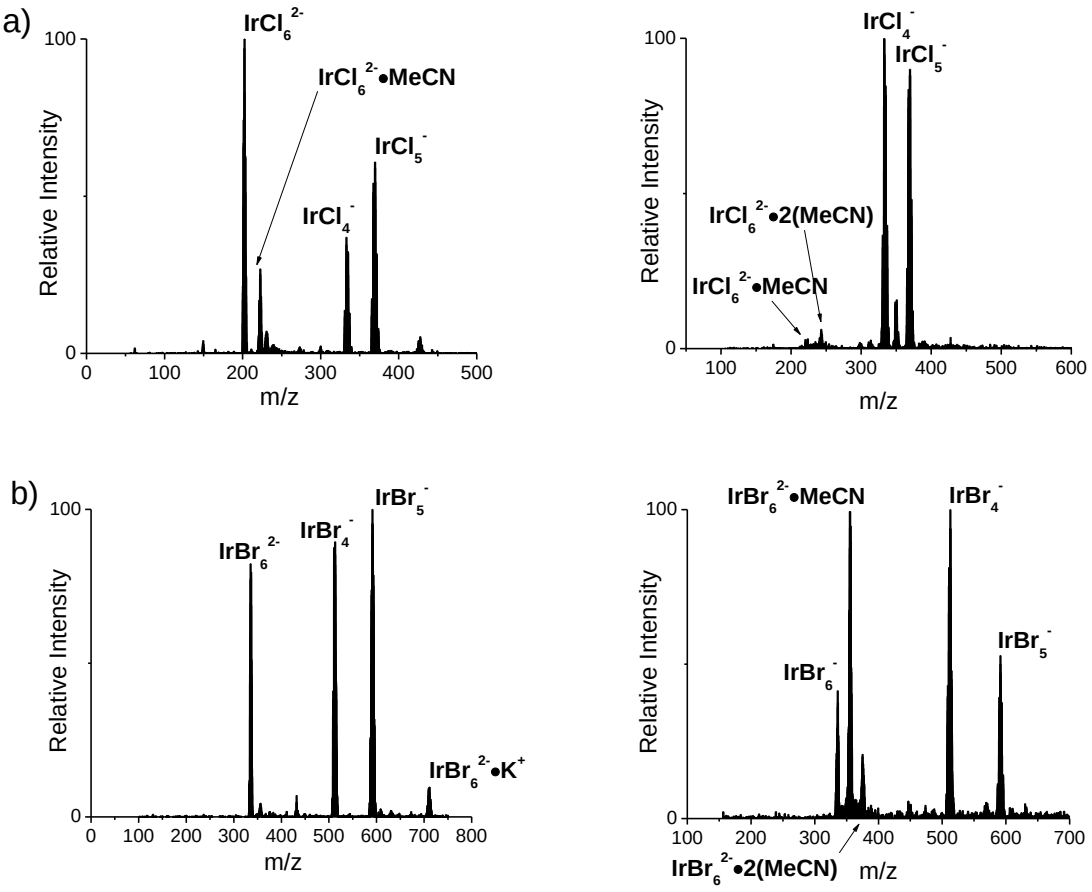

Figure A1. 3 ESI spectra of a) $IrCl_6^{2-}{\cdot}Na^+$, b) $IrBr_6^{2-}{\cdot}K^+$ and $PtBr_6^{2-}{\cdot}K^+$. For each species, the spectra has been recorded with MeCN seeded He collision gas (right) and with normal He collision gas (left).

Figure A1.1 displays the spectra obtained by spraying a solution $Pt(CN)_{4,6}^{2-}$ in MeCN. The baseline spectra shown on the left in the absence of 1% MeCN seeded He collision gas show peaks for $Pt(CN)^-$, $Pt(CN)_3^-$ $PtCN_4^-$ $PtCN_6^-$and $Pt(CN)_{4,6}^{2-}$ and $PtCN_4^{2-}{\cdot}K^+$. However, in the spectra with the collision gas, given to the right, $Pt(CN)^-$, $Pt(CN)_3^-$ and $Pt(CN)_{4,6}^{2-}$ all give clustered peaks, i.e. $PtCN_4^-$, $PtCN_6^-$ and $PtCN_4^{2-}{\cdot}K^+$ do not. Thus, solvation of a peak tends to occur when the platinum species is co-ordinately unsaturated and has a higher charge. $Pt(CN)_4^{2-}$ is less saturated and more densely charged than $Pt(CN)_6^{2-}$, and therefore binds to MeCN twice.

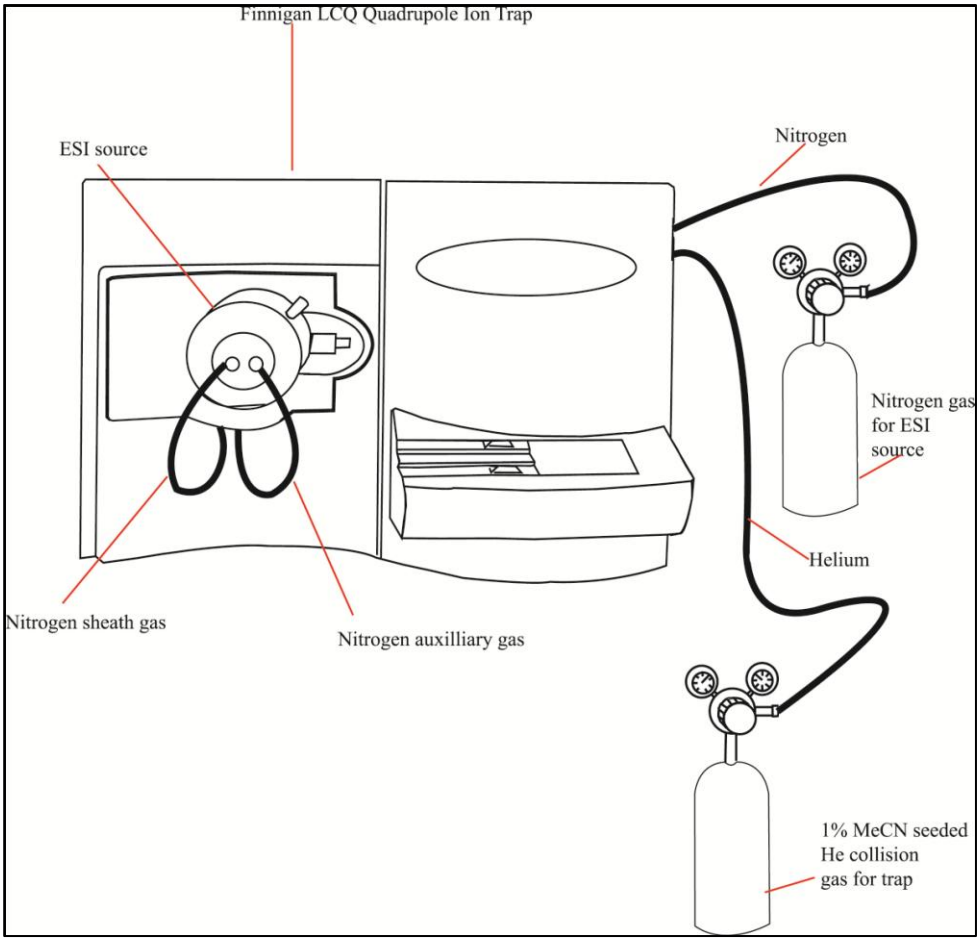

Figure A1. 4 Diagram of the experimental set-up used to achieve the MeCN collision spectra.

Figure A1.2 displays the baseline spectra and complexed spectra of some halogenated Pt analogues. Clearly, the primary/major ions have reacted instead of becoming solvated in this system, as the new peaks observed correspond to masses that are not simple additions of MeCN. They are assigned on the assumption that the isotopic distribution does not change on reaction with the solvent. Notice that some of the peaks show additions of approximately the same mass (57, 114 and 134 within a few mass units, as assignment is challenging). At present, it is difficult to account for these increases in masses, as they are not the result of simple additions of MeCN. It was possible, however, to isolate on some of these peaks to perform CID experiments,

which indicates possibly, along with the unusual masses, that a gas phase reaction has taken place.

Finally, Figure A1.3 displays the spectra obtained from the IrX_6^{2-} (X = Br or Cl) complex ions. These results mirror the findings from the two $Pt(CN)_{4,6}^{2-}$ complexes, with lower coordination and higher charge leading to the formation of solvated peaks.

The mass spectra presented above yield useful information about the propensity of various small metal anionic species to bind to solvent molecules. However, information about the energy of fragmentation of such species can only be obtained through isolation in the QIT followed by CID. As described previously, CID involves the scanning and isolation of ions by resonant ejection at the stability limit of the Mathieu stability diagram. In order to isolate, an additional voltage is applied to the end-cap electrode which causes acceleration of the ions leading to increased transfer of kinetic energy to internal energy, and eventual fragmentation. Thus, monitoring peak height as a function of CID voltage can give a measure of relative energy, and hence stability, of fragmentation between different species.

Unfortunately, the solvated clusters discussed above were not stable to isolation in the trap, apart from for the PtY_n^{z-} complexes (where Y = Cl or Br), which presumably had already reacted with the solvent. This may be due to the fact that the vibrationally excited (metastable) $PtCN_n^{z-}$•*m*MeCN (where *m* = the no. of solvent molecules) clusters readily dissociate as the isolation voltage is applied to the end-cap electrodes, or as a result of spontaneous dissociation due to the ions displaying lifetimes that are shorter than the timescale for resonant expulsion. Notably, a broad-shoulder effect is

observed on going to lower m/z on these solvated peaks. An example of this effect for the solvated $PtCN_4^{2-}$ peak is given in Figure A1.5. Interestingly, attempting to isolate on the original parent peak (i.e. optimizing the ion optics for the parent ion m/z), whether it is present or not, returns the solvated peak. In this way, solvated peaks can easily be identified in a mass spectrum.

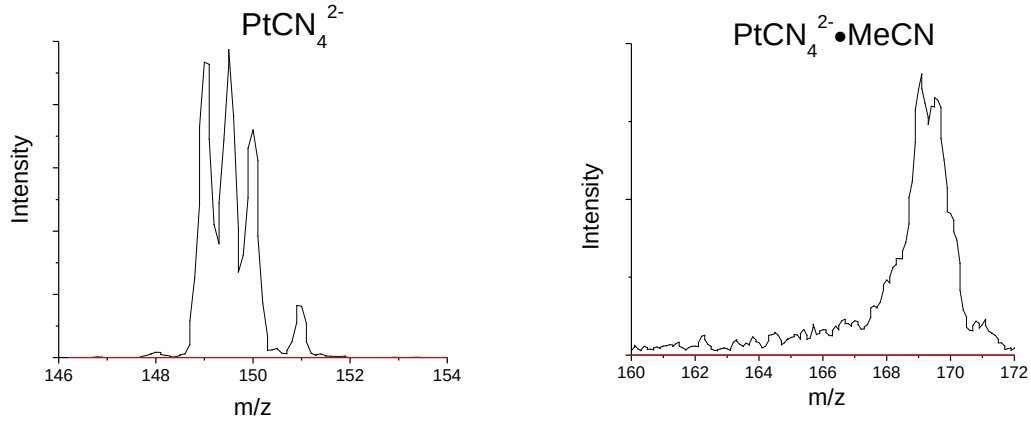

Figure A1. 5 Blown-up ESI spectra for $PtCN_4^{2-}$ (left) and $PtCN_4^{2-}$•MeCN (right) i.e. with normal He collision gas (left) and with 1% MeCN seeded He collision gas (right)

A1.2 Multiple Clustering with an ESI source leak valve

Given the lack of stability to isolation of solvated ESI clusters obtained by changing the collision gas, two other methods were tried in order to produce solvated metal anion complexes. The first was to lower the temperature on the post ESI source capillary to below 100 C, while the second was to pass the nitrogen auxiliary gas through a conical flask containing solvent before allowing it to flow into the ESI region. In addition, both were used in combination to generate the spectrum observed in Figure A1.7, which shows significant "multiple clustering" to both water (the

solvent used in the "bubbler") and acetonitrile (the solution solvent held in the syringe). The experimental set-up used to achieve multiple-clustering is given in Figure A1.6.

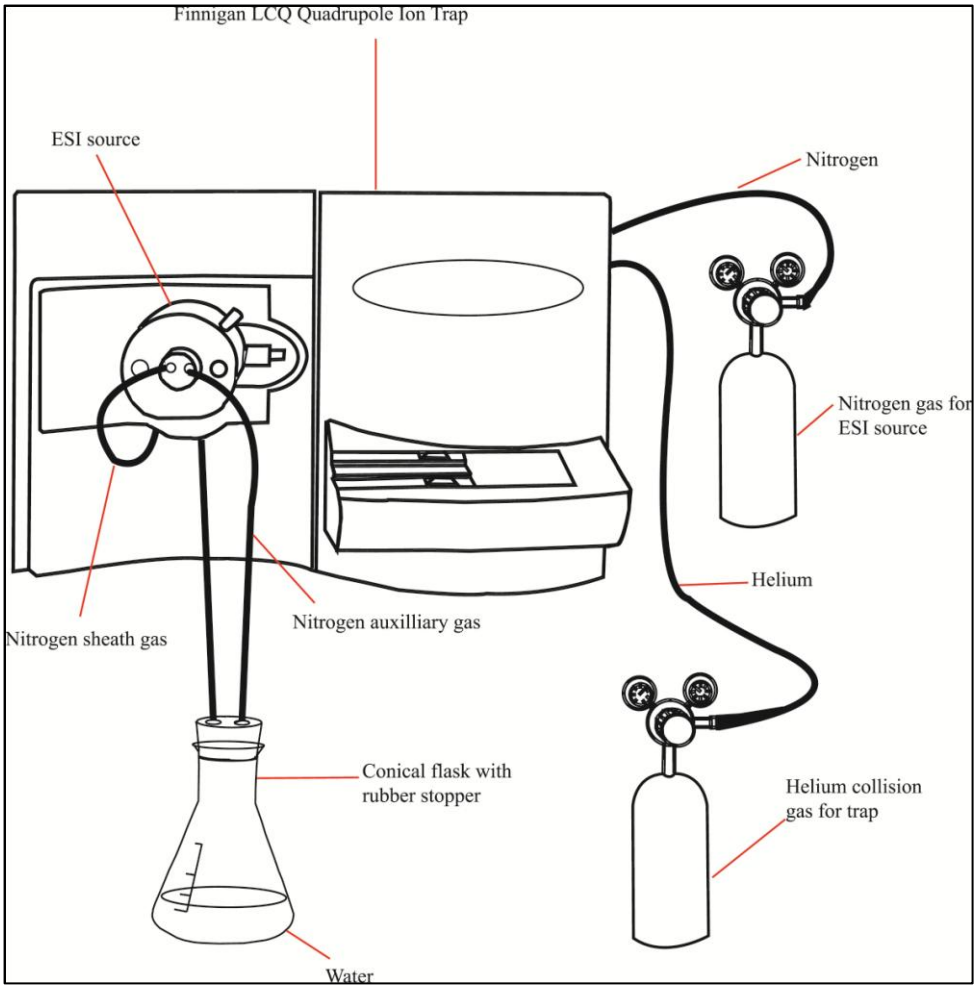

Figure A1. 6 Diagram used of the experimental set-up used to achieve multiple-clustering.

The majority of these clusters could be isolated within the trap, apart from the singly and doubly complexed water complexes. Interestingly, the singly complexed MeCN peak could be isolated in the trap, contrary to the difficulties discussed previously when using the MeCN collision gas introduced within the QIT. This allowed the CID experiment to be carried out, and so CID mass spectra at 15% collision energy were

obtained for solvated ion peaks of $Pt(CN)_4^{2-}$, which are displayed in Figure A1.8. Intriguingly, the fragment products are observed to change as the $Pt(CN)_4^{2-}$ dianion becomes increasingly solvated. Specifically, bare $Pt(CN)_4^{2-}$ and the less co-ordinately saturated solvent clusters, $Pt(CN)_4^{2-}$•MeCN, $Pt(CN)_4^{2-}$•2MeCN and $Pt(CN)_4^{2-}$•$3H_2O$, all give $Pt(CN)_3^-$ exclusively, whereas the more co-ordinately saturated solvent clusters, $Pt(CN)_4^{2-}$•$3H_2O$•MeCN and $Pt(CN)_4^{2-}$•$4H_2O$, give $Pt(CN)_3^-$ and $Pt(CN)_4^-$ as CID products. The $Pt(CN)_3^-$ CID product is evidence for ionic fragmentation, whereas the $Pt(CN)_4^-$ CID product is evidence for electron detachment, and both can be discussed in the context of the Repulsive Coulombic Barrier (RCB) of the $Pt(CN)_4^{2-}$ dianion, illustrated in Figure A1.9.

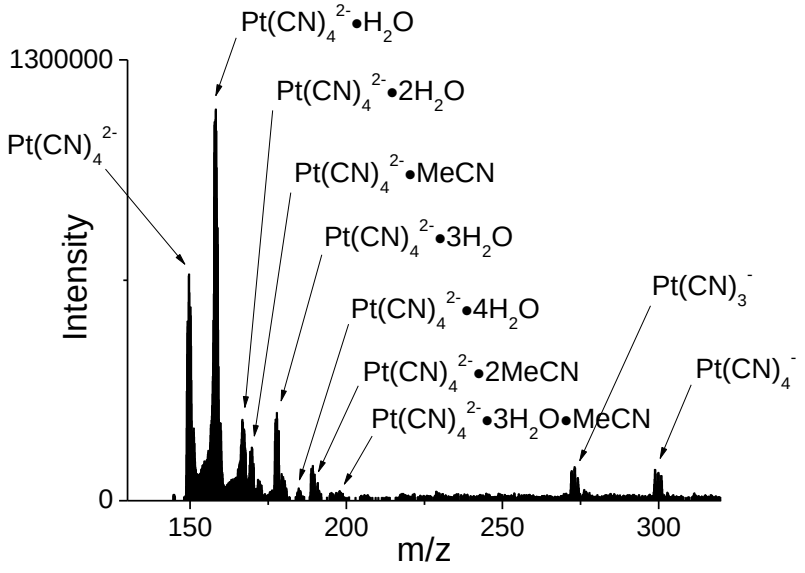

Figure A1. 7 Mass spectrum recorded for $PtCN_4^{2-}$ under conditions of low capillary temperature (70 °C), with a modified nitrogen auxiliary gas line flowing through a conical flask containing water.

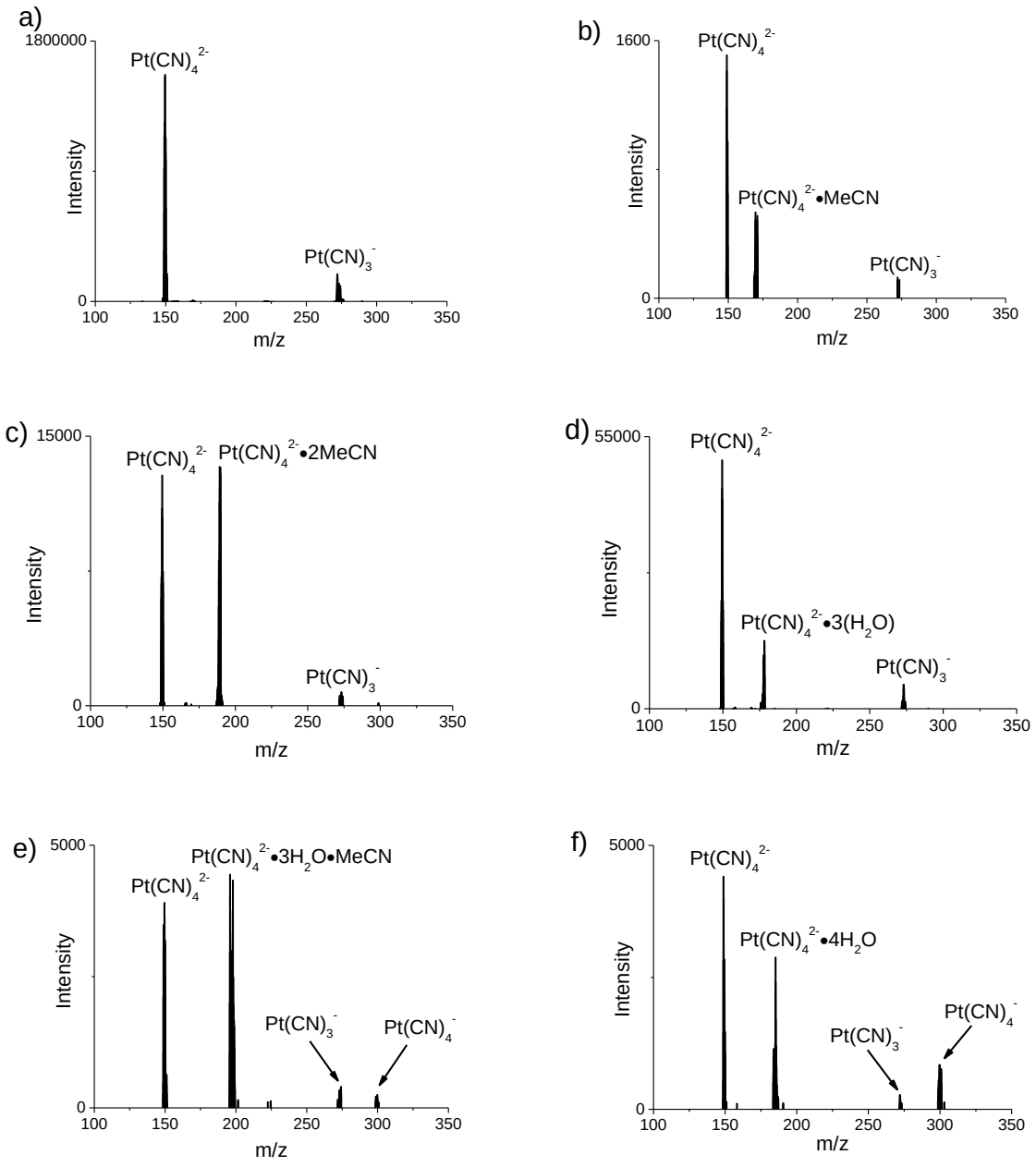

Figure A1. 8 Mass spectra recorded at 15% CID energy, with the ion optics isolated on peaks appearing in the main spectrum displayed in Figure A1.5. The parent ions isolated were: a) $PtCN_4^{2-}$, b) $PtCN_4^{2-}$·MeCN, c) $PtCN_4^{2-}$·2MeCN, d) $PtCN_4^{2-}$·$3H_2O$, e) $PtCN_4^{2-}$·$3H_2O$·MeCN and f) $PtCN_4^{2-}$·$4H_2O$. As more solvent is added, the electron detachment product, $Pt(CN)_4^-$, starts to appear.

Multiple Charged Anions (MCAs), such as the ones seen in the spectra above, can either decay by losing an electron, by losing an anionic ligand, or by loss of a neutral ligand (not observed in the above). The 1st loss channel is denoted as electron detachment, the 2nd as ionic fragmentation. The potential energy surfaces for both of these fragmentation channels contain Repulsive Coulomb Barriers (RCBs), which arise from the combination of the short-range, attractive binding force with a purely repulsive long-range Coulombic interaction. Thus, the RCB can be considered to arise from an intersection of the two curves.[8] A schematic representation of typical potential energy surfaces for electron detachment and ionic fragmentation with RCBs is given in Figure A1.9 below.

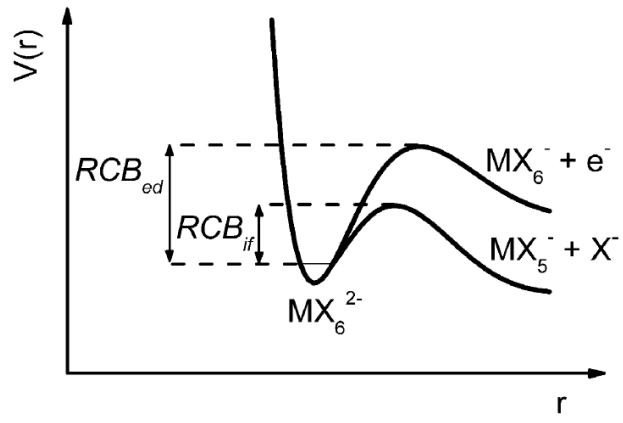

Figure A1. 9 Diagram of the potential energy surfaces for decay of an MX_6^{2-} dianion via ionic fragmentation and electron detachment, showing the RCB(inner) for electron detachment (RCB_{ed}) and ionic fragmentation (RCB_{if}). Diagram taken from ref. [9]

L. S. Wang and co-workers, who obtained photoelectron spectra (PES) of citrate dianions (CA^{2-}), estimated the magnitude of RCB_{ed}(outer) by combining the energy of the observed spectral bands for the two electronic states of CA^- with their known electron affinities.[10] This was found to be the equal to the Coulomb repulsion energy

for the two charges. In a later paper on $^{-}O_2C\text{-}(CH_2)_n\text{-}CO_2^{-}$ dianions (n = 3-10), RCB_{ed}(inner) was found to be equal to the binding energy of an electron to the carboxyl group.[11] Thus, the RCB can be thought of as a sum of the binding energy of the electron to the anion at short-range, and the Coulombic repulsion of those two charges at long-range.

Ionic fragmentation, can be treated in an analogous fashion to electron detachment, with RCB_{if}(inner) being associated with the purely attractive binding energy of Y^- to XY_n^- at short-range, whereas RCB_{if}(outer) is approximately equal to the Coulombic repulsion between the two charged fragments.[9] RCB(outer) for both electron detachment and ionic fragmentation can be equated, as an electron and a small anion can be thought to have a similar long-range Coulombic repulsion.[12] The propensity of various transition metal dianions to decay by either electron detachment or ionic fragmentation has been investigated extensively,[9,13,14,15,16] with the result that ionic fragmentation seems to represent the dominant decay pathway for dianions that can fragment to produce stable (closed shell) singly charged anions,[12] such as the systems studied here. $PtCN_4^{2-}$ dianion, in particular, is found to decay solely by ionic fragmentation.[13]

Solvation plays an important role in altering the barrier heights for both ionic fragmentation and electron detachment. For example, the calculated curves for $IrCl_6^{3-}\cdot H_2O_n$ reveal that RCB_{ed}(inner) and RCB_{if}(inner), both increase with increasing solvation, n, whereas RCB_{ed}(outer) and RCB_{if}(outer), both decrease with increased solvation due to the increased dielectric screening by the solvent, reducing the Coulombic repulsion.[17] The increase in RCB_{if}(inner) can be attributed to the increase

of favourable $Cl^-\cdots H_2O$ interactions that contribute to the purely attractive binding energy between Cl^- and $IrCl_6^{3-}\cdot nH_2O$ as n increases, and in exactly the same way, the rise of RCB_{ed}(inner) is due to the increase in the purely attractive binding energy of an electron with the cluster as n increases.[12]

These ideas can be readily applied to the observations seen in the above spectra. For $PtCN_4^{2-}$, equations for electron detachment and ionic fragmentation are given below:

$$PtCN_4^{2-} \rightarrow PtCN_4^- + e^- \qquad \text{A1.1}$$

$$PtCN_4^{2-} \rightarrow PtCN_3^- + CN^- \qquad \text{A1.2}$$

Thus, in the case of electron detachment, we expect to observe the $PtCN_4^-$ product, whereas in the case of ionic fragmentation, we would expect to see the $PtCN_3^-$ product. $PtCN_4^{2-}$ is known to decay exclusively via ionic fragmentation, even when bound to K^+ (albeit, at higher CID energy).[13] It is curious then, that upon sufficient solvation, the $PtCN_4^-$ product is also observed.

As discussed in previous chapters, CID directly probes the lowest-energy unimolecular pathway, so the presence of both $PtCN_3^-$ and $PtCN_4^-$ products in the extensively solvated clusters implies that both fragmentation pathways [A1.1] and [A1.2] are similar in energy. Thus, the barrier heights to fragmentation on the potential energy surfaces are at the same energy. In contrast, the presence of the single ionic fragmentation product, [A1.2], observed in the bare complex and less solvated complexes implies that that RCB_{if}(inner) is lower than RCB_{ed}(inner) (approximating

RCB_{ed}(outer) = RCB_{if}(outer)). Clearly, this demonstrates the increased importance of the binding energy of the electron, which controls RCB_{ed}(inner), to the cluster over any attractive $CN^-\ldots H_2O$ interactions, which control RCB_{if}(inner). It is interesting to note that even counterion complexation of $PtCN_4^{2-}$ with K^+ does not alter the relative energy of ionic fragmentation to electron detachment sufficiently to observe both products.

References

[1] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64

[2] Klassen, J. S.; Blades, A. T.; Kerbarle, P. *J. Phys. Chem.* **1995**, *99*, 15509

[3] Lee, S. W.; Freivogel, P.; Schindler, T.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 11758

[4] Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. *J. Am. Soc. Mass. Spectrom.* **1997**, *8*, 565

[5] Wickens, J. R.; Sleeman, R.; Keely, B. J. *Rap. Comm. Mass Spectrom.* **2007**, *21*, 2491

[6] Attygalle, A. B.; Kharbatia, N.; Bialecki, J.; Ruzicka, J.; Svatos, A.; Stauber, E. J. *Rap. Comm. Mass Spectrom.* **2006**, *20*, 2265

[7] Rajabe, K.; Easterling, M. L.; Fridgen, T. D., *J. Am. Soc. Mass Spectrom*. **2009**, *20*, 411

[8] Gas-phase MCAs - M. K. Scheller, R. N. Compton, L. S. Caderbaum, Science 1995, 270, 1160

[9] Boxford, W. E.; Pierce, J. K.; Dessent, C. E. H. *Chem. Phys. Lett.* **2004**, *399*, 465

[10] Wang, X. B.; Ding, C. F.; Wang, L. S. *Phys. Rev. Lett.* **1998**, *81*, 3351

[11] Wang, L.S.; Ding, C. F.; Wang, X. B.; Nicholas, J. B.; Nicholas, B. *Phys. Rev. Lett.* **1998**, *81*, 2667

[12] Boxford W. E.; Dessent, C. E. H. *Phys. Chem. Chem. Phys*. **2006**, *8*, 5151

[13] Burke, R. M.; Boxford, W. E.; Dessent, C. E. H. *J. Chem. Phys*. **2006**, *125*, 021105

[14] Boxford, W. E.; El Ghazaly, M. O. E.; Dessent, C. E. H.; Nielsen, S. B. *Int. J. Mass Spectrom*. **2005**, *244*, 60

[15] Taylor, C. J.; Wu, B.; Dessent, C. E. H. Int. *J. Mass Spectrom*. **2008**, *276*, 31

[16] Burke, R. M.; Boxford, W. E.; Dessent, C. E. H. *J. Chem. Phys.* **2007**, *126*, 064308

[17] Boxford, W. E.; Dessent, C. E. H. *J. Phys. Chem. A* **2005**, *109*, 5836

Appendix 2: Computational Methods

This appendix reviews the computational techniques employed in this dissertation

A2.1 Molecular Mechanics

A2.1.1 *Introduction*

Molecular mechanics (MM) provides a computationally inexpensive means of calculating molecular energies and geometries. Molecular mechanics programs such as SPARTAN use a force field that contains many different functions used to describe different atomic and molecular interactions. These functions are fitted to experimentally derived parameters in order to make the functions more reliable, in a process known as parameterisation. The energy often reported in force-field calculations is the force-field energy, and is calculated as the sum of the energy from all of the individual bond-length, angle, dihedral and non-bonded interactions: [1]

$$E_{tot} = E_l + E_\theta + E_\omega + E_{nb} \qquad [A2.1]$$

Molecular mechanics approximates each atom as a point charge connected by a spring to the next one, thus making an ensemble of point charges connected by springs. Bond stretches, bond angles, and dihedral angles are approximated using individual functions, which are each derived from classical Physics: bond stretches and bond angles are approximated by harmonic functions, dihedral angles by a Fourier series, and there are various other potentials, such as the Lennard-Jones potential, that are

used to model non-bonded interactions. However, molecular mechanics neglects an accurate representation of electrons based on Quantum theory, thus restricting it to a representation of molecular ground-states, though this short-fall is compensated by the fact that MM calculations take seconds on modern computers, which makes it very suitable for treating large ensembles of atoms. Additionally, much of the potential energy surface (PES) that does not apply to the ground-state is made redundant because of a parameterisation process that derives the various constants used in the functional forms from ground-state experimental data.[1]

A2.1.2 *Bond Stretch*

The Morse function is the most accurate approximation to the vibration or stretch of a bond: it is harmonic near the equilibrium bond length and exhibits dissociation as this bond length is increased. However, as this form is computationally expensive, most force-fields adopt a simpler, harmonic function such as: [2]

$$E_l = \sum k_l(l - l_0)^2 + k_l'(l - l_0)^3 + k_l''(l - l_0)^4 \qquad [A2.2]$$

Where k_l, k_l' and k_l'' are the 0^{th}, 1^{st} and 2^{nd} order force constants of the bond respectively, and the term $(l - l_0)$ is the deviation of the bond length, l, from the equilibrium bond length, l_0. The first term in the above expression models the harmonic behaviour of the bond near the equilibrium bond length, whilst the second and higher order terms approximate the anharmonicity of the potential at bond lengths that deviate significantly from equilibrium. As force fields are only applicable to ground states, bond lengths are not expected to differ significantly from their

equilibrium values, and the harmonic function above models bond stretching adequately.[2]

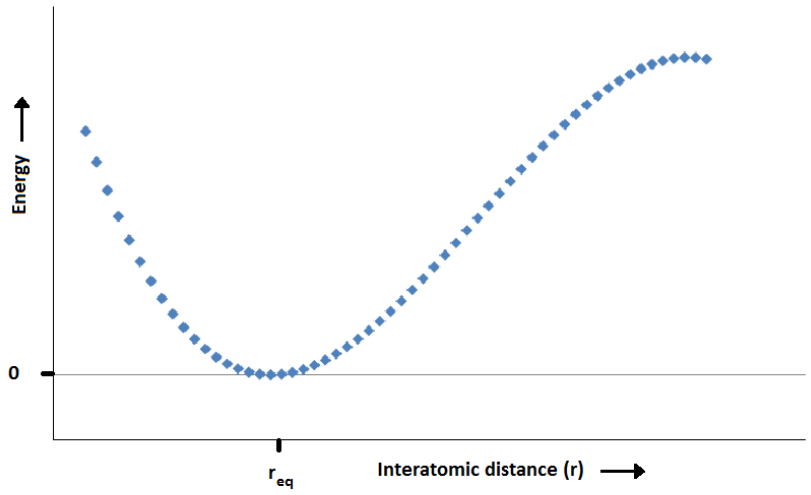

Figure A2. 1 Diagram of the bond-stretching potential energy. The equilibrium distance where the potential sums to zero is labelled. At long bond distances, the potential exhibits anharmonic behaviour.

A2.1.3 *Bond Angles*

Bond angles, dependent on the angle between 3 bonded atoms, can be modelled as simple harmonic functions in the same way as bond stretches, and suffer from the same difficulties away from equilibrium. No series of terms, however, is sufficient to model the behaviour at an angle close to 180^0, but none the less, the following series is sufficient at angles close to equilibrium: [2]

$$E_\theta = \sum k_\theta(\theta - \theta_0)^2 + k'_\theta(\theta - \theta_0)^3 + k''_\theta(\theta - \theta_0)^4 \qquad [A2.3]$$

Where k_θ, $k_\theta^{'}$ and $k_\theta^{''}$ are the 0^{th}, 1^{st} and 2^{nd} order bond bending constants respectively, θ is the bond angle, and θ_0 is that angle at equilibrium. There are 3 major problems associated with calculating bond angles: There is no series of terms that is sufficient

to describe the potential at angles close to linearity, in some systems multiple equilibrium values are possible for any one angle (e.g. for the trigonal bi-pyramidal complex of PCl_5, the Cl-P-Cl angle is stable axially, in the trigonal plane, and in the angle between the two), and out-of-plane bending is not considered (an issue in organic systems containing rings).[2]

A2.1.4 *Dihedral Angles*

Torsional rotation energy is technically a non-bonding interaction, but is modelled explicitly in all force-fields. It is the energy change that results for the molecule ABCD as rotation occurs about a bond BC between the bonds AB and CD. A cosine term is the most appropriate function that models such a periodic motion, and thus typically successive one-, two- and three-fold cosine terms can be summed together to give a Fourier series that models this torsion energy for molecules with a greater number of substituents. Thus, the Fourier series below models the behaviour for one torsion rotation: [2]

$$E(\omega_{ABCD}) = \frac{1}{2}\sum_{\{j\}_{ABCD}} V_{j,ABCD}\left[1 + (-1)^{j+1}cos\left(j\omega_{ABCD} + \psi_{j,ABCD}\right)\right] \qquad [A2.4]$$

ω is the angle that defines the rotation relative to a single point, V_j is the barrier height for rotation, j is the periodicity of rotation, and the phase angle, ψ, specifies the phase of each of the j terms relative to one another. The ½ cofactor limits V_j as the maximum value that any one term can take, and as most torsional energies are minimised for antiperiplanar arrangements, the $(-1)^{j+1}$ term allows the function to sum to zero when $\omega = \pi$, over all values for j. For example, in Figure 4.2 below, the series needs to include a one-fold term because the fluorine atom interacts only once

per revolution, a two-fold term because the two hydrogens interact twice per 360^0, and the three-fold term is necessary to maintain the underlying 3-fold periodicity. The one-fold term can be given a larger barrier height, V_j, so as to model the greater steric interaction of the fluorine atom.[2]

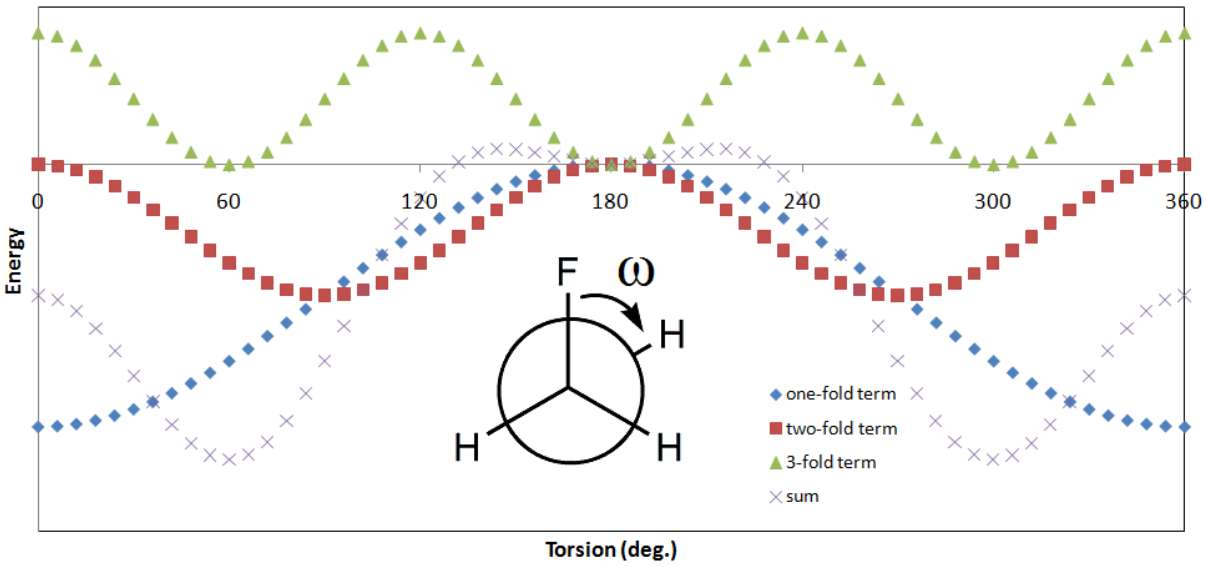

Figure A2. 2 Diagram of the variation in the dihedral angle of fluromethanol. The total energy is a Fourier sum (cross) of one (diamond) , two (square) and three-fold (triangle) terms that are all functions of the dihedral angle, ω. Diagram adapted from ref [2].

A2.1.5 *Non-bonded Van der Waals and Electrostatic Interactions*

The non-bonded, distance-dependent Van der Waals and electrostatic forces that exist between atoms within the same molecule are caused by induced-dipole/dipole interactions (electron correlation) and polarity respectively.

If two atoms, A and B, are brought into range, long-range attractive Van der Waals forces lower the total energy of the system. However, as they are brought closer together, beyond a certain point the electron densities of the atoms begin to mix

causing repulsion which increases the total energy. The "Lennard-Jones" potential models both of these attractive and repulsive behaviours: [2]

$$E_{vdw}(r_{AB}) = 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right] \qquad \text{[A2.5]}$$

Where r is the distance between A and B, σ is the interaction distance where the attractive and repulsive terms cancel each other out, and ε is the well-depth. Short-range repulsions are accounted for by the r^{-12} term, whilst long-range attractions are accounted for by the r^{-6} term. The Lennard-Jones potential is easy to calculate and thus good for force-fields targeted at larger systems, but some more computationally expensive functions can be used for smaller molecular systems. These include the Morse potential, described previously, and the Hill potential: [2]

$$E(r_{AB}) = D_{AB}\left[1 - e^{-\alpha_{AB}(r_{AB} - r_{AB,eq})}\right]^2 \qquad \text{[A2.6]}$$

$$E(r_{AB}) = \varepsilon_{AB}\left[\frac{6}{\beta_{AB}-6} exp\left(\beta_{AB}\frac{1-r_{AB}}{r^*_{AB}}\right) - \frac{\beta_{AB}}{\beta_{AB}-6}\left(\frac{r^*_{AB}}{r_{AB}}\right)^6\right] \qquad \text{[A2.7]}$$

In the Morse potential, D_{AB} is the dissociation energy of the bond and α is a fitting parameter. In the Hill potential, r^*_{AB} is the interatomic distance at the minimum of the potential well, and β is a fitting parameter. The Hill potential has a similar form to the Lennard-Jones potential, but the attractive term is now an exponential function, which models the attraction more realistically.[2]

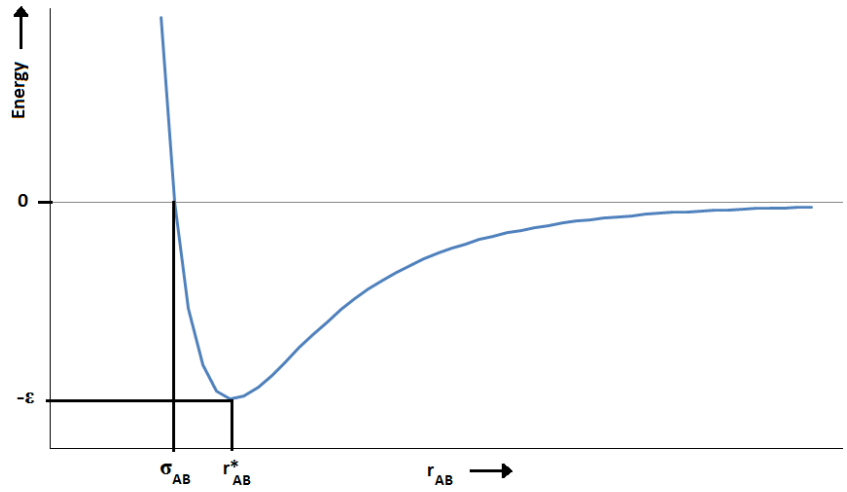

Figure A2. 3 Diagram of the Lennard-Jones potential showing variation of energy with interatomic distance. The parameters of well-depth, ε, the interaction distance where attraction and repulsions cancel, σ_{AB}, and the interatomic distance at the minimum of the well, r^*_{AB}, have been marked.

The electrostatic interaction between atoms A and B can be modelled simply by Coulomb's law, by modelling each atom as a fixed charge, giving the interaction energy as: [2]

$$E_{AB} = \frac{q_A q_B}{\varepsilon_{AB} r_{AB}} \qquad [A2.8]$$

Where ε_{AB} is the dielectric constant, r_{AB} is the distance between the two charges, and q_A and q_B are the partial charges of A and B respectively. There are other alternative schemes that consider the interaction of charge polarisation of different chemical bonds, in terms of their dipole moments, but the above is by far the simplest way to calculate the electrostatic interaction energy: [2]

Most force-fields consider non-bonding interactions in the way laid out above. However, the special case of hydrogen-bonding, which results from the extreme difference in electronegativities of a H atom and a heteroatom, often needs to be considered separately, either by taking it into account in the non-bonding functions, or by adding in a separate 10-12 Lennard-Jones type potential. This potential has a much more rapid decay term in the attractive region, consequently forcing the hydrogen bond to vary little from its equilibrium position.[2]

A2.1.6 *Parameterisation*

The many different constants included in the functions above are derived from fits to experimental data. This process is known as parameterisation, and is carried out by collecting all of the experimental data (structural data, energetic data, molecular moment data etc.) and all of the predicted data using the above functions, and then measuring the deviation between the two by use of a "penalty function". A typical function considers observables such as bond angle and length, and inputs the data for all of them derived from experiment and calculation. It is also able to take into account varying quantities of data for each observable, and is able to set specific tolerances by which each observable can vary between experiment and calculation. The goal is then to minimise the penalty function for a given force-field.[2]

Force-fields typically require a large number of parameters, however, as every single different atom can give rise to a large number of different parameters in combination with any other atom for just one of the functions described above, the sum over all functions can give rise to an extremely large number of parameters (100,000,000 + for a force-field generalised over the 1st 100 elements of the periodic table). The scope of

this problem can be reduced by making force-fields that are specific to a small sub-set of atoms that cover a particular range of molecules (for example, one could design an organic force-field only containing the elements H, C, N, O, F, Si, P, Cl, Br). However, many atoms will have parameters associated with different types of bonding or hybridizations. For example, a C-C bond can be single, double or triple, giving rise to different equilibrium bond lengths, and a C centre can be sp, sp^2 or sp^3 hybridized, giving rise to multiple equilibrium angles. This creates a myriad of additional parameters which can be accounted for by using atom types instead of parameters based on atomic number.[2]

Thus, a given force field may contain a very large number of atom types (MM3 and MMFF have 153 and 99 atom types respectively, and OPLS, which is a general bio-molecular force-field, has 41 types). The number of atom types can be decreased by making certain parameters depend on more than one atom, which are in turn dependent on single, atom-specific parameters. For example, the Lennard-Jones potential that describes the energy of interaction between atoms A and B has two parameters that are specific to the AB system. If, however, these two parameters are made to be functions of both A and B separately, then the problem is simplified by reducing N(N+1)/2 diatomic parameters to N atomic parameters: [2]

$$\sigma_{AB} = \sigma_A + \sigma_B \qquad \text{[A2.9]}$$

$$\varepsilon_{AB} = (\varepsilon_A \varepsilon_B)^{1/2} \qquad \text{[A2.10]}$$

Force-fields that can treat a larger range of molecules make use of strategies such as this by deriving a large range of diatomic parameters from a much smaller sub-set of

atomic parameters. Organic force-fields such as MM3 and MM2, make use of a similar strategy by which constants are made to depend on computed π bond orders between atoms.[2]

Clearly, force-fields optimised for certain ranges of molecules are useful alternatives to more generalised force-fields that are not designed to reproduce experimental results as precisely. MM2 is good for organics but has been largely replaced by MM3 which is good for organics and biomolecules. The MM4 force-field is an even more specialist organic force-field which has been optimised for use with hydrocarbons. In this work, the MMFF94 force-field has been used as implemented through SPARTAN, and is a general use force-field that is good for organics and biomolecules.[2] Several features distinguish it from MM3. The first is that it is tailored to reproduce MM3 accuracy for small molecules, yet still be applicable to much larger systems such as proteins. The second is that the core portion of the force-field has been parameterised from high-level computational data. Finally, it has been parameterised for chemical systems that contain common combinations of functional groups of interest to organic and medicinal chemists.[3]

A2.1.7 *Conformational Sampling using Molecular Dynamics*

Much importance in Chemistry is attached to the location of the global minimum. Unfortunately, analysing the whole potential energy surface of a molecule in a systematic manner would take an unfeasibly long time. A simple example is the consideration of the alkane, $CH_3(CH_2)_{n+1}CH_3$, where each C-C bond is allowed to rotate by 120 degree increments. Butane, where n = 1, has 3^1 = 3 different conformations, however, and even moderately larger values for n increase the number

of conformers to an unmanageable sum (e.g. n = 5 gives 3^5 = 243, n = 10 gives 3^{10} = 59 049, and n = 15 gives 3^{15} = 14 348 907 different combinations). Conformationally complex molecules, with 3 degrees of freedom per atom, leading to 3N-6 degrees of freedom for a molecule comprising N atoms (-3 for translation and -3 for rotation), give rise to a multi-dimensional problem, necessitating the use of an algorithm. The conformational search function in SPARTAN is based on an algorithm that tries to find the global minimum, whilst avoiding letting the system become trapped in a local minimum. It is set-up in such a way that it saves all of the low-energy structures that it samples during this optimization process, which makes it an effective tool for producing a series of low energy conformers that can be further optimized.[4]

For such an energy minimisation algorithm using molecular dynamics, the force-field is used to determine the energy of a given conformation. On applying a "random kick" to a set of torsion rotations, a new conformation is generated, and the energy is recalculated. The "random kick" that SPARTAN uses is a basic set of torsion rotations for non-ring containing structures. A comparison of the two results reveals which is lower in potential energy, and therefore whether the step taken has lowered or increased the potential energy of the molecule. A "greedy algorithm" always decides to accept moves that lower the energy, but such a strategy limits itself to locating only local minima, and in addition, will sample very little of the conformational space.[5]

Instead of this, SPARTAN uses a Monte Carlo method known as "simulated annealing". Simulated annealing works by utilising a simulated temperature ramp, which is a function dependent on the number of iterations and where the Temperature,

T, is related to the probability that the move is accepted.[6] The temperature ramp is calculated as follows: [5]

$$T = T_{final} + \left(T - T_{final}\right)\left(1 - \frac{I}{I_{max}}\right)^3 \qquad \text{[A2.11]}$$

Where T is the temperature, T_{final} is the final temperature, I is the number of iterations, and I_{max} is the maximum number of iterations. The default in Spartan is to start at 5000K and end at 300K.[7] High temperatures favour downhill moves, as well as uphill moves, which allows the algorithm to escape local minima and sample more of the conformational space. This is useful if the starting structure is very different to the actual global minimum. However, as the temperature lowers, downhill moves are increasingly favoured over uphill moves. The end stages of the algorithm are very much like a "greedy algorithm", whereby most moves are only accepted if they lower the energy. The algorithm then "halts" when the set number of maximum iterations is reached.[6] The net-effect is that when the number of iterations are low, the algorithm is allowed to take moves that raise the energy, in order to avoid local minima, but as the number of iterations increases, the probability of acceptance of a worsening move decreases according to the function above, such that in the end-stages of the algorithm only moves that lower the energy are accepted. An example of this type of conformational search conducted in Spartan is given in Figure 4.4: [8]

```
SPARTAN '08 CONFORMATION SEARCH: PC/x86       132


 Initializing 4 threads

  Using Monte-Carlo algorithm.

Using rotatable bonds from rule normal set.

 Conf     Energy     Temp.
(6561)     kJ/mol    (K)  Local%     Remark
-----------------------------------------------
    1    -540.248    5000  100.00*
    2    -537.368    4998  100.00*
    3    -540.248    4996  100.00* Prev. Min.
    4    -517.645    4994  100.00*
    5    -510.604    4991  100.00*
    6    -536.627    4989  100.00*
    7    -534.865    4987  100.00*
    8    -549.955    4985  100.00* !New Best!
    9    -531.676    4983   88.89
   10    -561.699    4981   90.00* !New Best!
   11    -552.764    4979   90.91*
   12    -552.871    4976   91.67*
   13    -552.871    4974   84.62  Same Min.
   14    -552.871    4972   78.57  Same Min.
   15    -548.100    4970   80.00*
   16    -548.100    4968   75.00  Same Min.
   17    -552.871    4966   76.47* Prev. Min.

SPARTAN '08 CONFORMATION SEARCH: PC/x86       132
```

Figure A2. 4 An example of a SPARTAN output file for a conformational search, starting from an initial temperature of 5000 K. The maximum number of iterations is in brackets at the top (6561), and is calculated according to the degree of conformational flexibility available to the molecule. As can be seen, the algorithm is permitted to take worsening moves at high temperature.

In SPARTAN, the conformer distribution results from taking all conformations that have an energy window that is less than 10 kcal mol^{-1} from the current lowest energy conformer. If the number of conformers with acceptable energies exceeds the specified maximum number of conformers, the program discards some with a view to keeping as diverse a group as possible. The default in SPARTAN is to retain the 100 lowest energy conformations.[7]

SPARTAN actually, combines this varying probability algorithm with a Boltzmann weighting factor for the conformation, as a way of finding chemically relevant structures. This is known as the Metropolis method,[9] and it works as follows: Starting from some coordinates, q_1, the energy is calculated. This value, q_1, is then randomly changed to give a new set of coordinates, q_2. At constant temperature, the probability that the point q_2 is accepted is then: [2]

$$p = min\left[1, \frac{exp(-E_2/k_BT)}{exp(-E_1/k_BT)}\right] \qquad [A2.12]$$

This means that if the energy of point q_2 is lower than q_1, it is always accepted. However, if the energy of q_2 is higher than q_1, then p is compared to a random number between 0 and 1, z, and the move is accepted if $p \geq z$. Accepting the point means that a new value for the energy is calculated, and the process is then repeated with the new coordinates. If the point is not accepted, then a new random perturbation is applied to the old set of coordinates, q_1.[2]

The algorithm within SPARTAN combines simulated annealing with a Metropolis-type method to generate its conformer distribution. The conformer distribution function within SPARTAN was used in this work as a way of producing a range of chemically relevant conformers that could be further optimized.[5]

A2.2 Quantum Mechanics

A2.2.1 *Basis of Quantum Mechanics Calculations*

The non-relativistic form of the Schrodinger equation is given as: [10]

$$\hat{H}\Psi = E\Psi \qquad [A2.13]$$

Where $\hat{H}$ is the Hamiltonian (total energy operator), Ψ is the wavefunction and E is the total energy of the system. The Hamiltonian operator takes into account 5 different contributions to the total energy of the system: the kinetic energy of the electrons and nuclei, electrons-nucleus attraction, electron-electron repulsion and nucleus-nucleus repulsion, represented as: [2]

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}} \qquad [A2.14]$$

Where i and j describe electrons, k and l describe nuclei, and ∇^2 is the Laplacian operator. In 3D Cartesian coordinates, the Laplacian operator is given as: [2]

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \qquad [A2.15]$$

Because electrons move around nuclei that are not themselves fixed, all of the terms in the above become correlated, that is, no particle moves independently of another one. However, as nuclei move significantly slower than electrons, one can take the nuclei as being fixed, and only consider the movement of electrons. This

simplification of the above expression for the Hamiltonian is known as the Born-Oppenheimer approximation, and results in an electronic Hamiltonian. This electronic Hamiltonian only then contains terms for the kinetic energy of the electrons, the electron-nucleus attraction, and the electron-electron repulsion. The electronic Schrodinger equation is given as: [10]

$$\widehat{H}_{el}\Psi_{el} = E_{el}\Psi_{\text{el}} \qquad \text{[A2.16]}$$

Adding the nucleus-repulsion energy, which is now a constant, gives: [10]

$$E_{tot} = E_{el} + \sum_{k<l}\frac{e^2 Z_k Z_l}{r_{kl}} \qquad \text{[A2.17]}$$

A2.2.2 *Linear Combination of Atomic Orbitals Basis-sets*

According to the variational principle, the best wavefunction will be the one lowest in energy, as the lowest-energy state is the one that is closest to the true ground-state energy. This means that the trial wavefunction does not have to be a linear combination of orthonormal orbitals (wavefunctions), but can be constructed from any arbitrary set of orbitals: the quality of the guess is entirely dependent on how low the energy is.[2] Trial wavefunctions can be constructed from linear combinations of simpler exact hydrogenic atomic orbitals, termed the basis set: [2]

$$\phi = \sum_{i=1}^{N} a_i \varphi_i \qquad \text{[A2.18]}$$

Thus, a trial wavefunction can be made up over a set of N functions that are each weighted by a coefficient. Minimal basis sets can be built from simple Gaussian Orbital (GO) functions, and comprise exactly the number of functions required to accommodate all of the electrons on the atom. In the STO-KG formalism, however, more accurate Slater Orbital (SO) functions are approximated as expansions of simple GOs. Further to this, these SOs can then be used in split-valence basis sets that contain the usual number of functions for inner-shell electrons (Single-zeta), but each function is replaced by two functions with different exponential factors in the valence region (Double-zeta). The K-K'K"G split valence basis sets were suggested by J. A. Pople and co-workers, and won the Nobel prize in 1998: [10]

Inner shell: $\tilde{\chi}_{nl}(r) = \sum_{k=1}^{K} d_{nl,k}\chi_{GO,l}(\alpha_{nk}, r)$

Contracted valence shell: $\tilde{\chi}'_{nl}(r) = \sum_{k=1}^{K'} d'_{nl,k}\chi_{GO,l}(\alpha'_{nk}, r)$

Diffuse valence shell: $\tilde{\chi}''_{nl}(r) = \sum_{k=1}^{K''} d''_{nl,k}\chi_{GO,l}(\alpha''_{nk}, r)$ [A2.19]

The coefficients, d, and the exponential factors, α, are determined through unrestricted HF calculations (UHF) on isolated atoms. As an example, the 6-31G split valence basis set (K=6, K'=3 and K"=1),[11] has 6 GOs for 1s, 2s, 2p and 3p orbitals, and a split valence pair of 3 and 1 primitives for valence orbitals, which are 4s, 4p and 3d. This is an improvement on a minimal basis set, such as STO-3G, which comprises 3 GOs per orbital, and the smaller 3-21G or 4-31G split valence sets which have fewer GOs for inner shell electrons.[10]

These basis sets can be further improved by adding in polarization (*) and diffuse functions (+). The various additions of polarization functions correspond to additional d-type functions on heavy atoms and p-type functions on hydrogens, whereas the diffuse functions correspond to various additional functions that have small Gaussian exponents. In this work, the basis set 6-31+G** was used for the calculations on arginine. This corresponds to the 6-31 spilt valence basis set augmented with one set of 6-d orbitals on 2nd row atoms and 3p orbitals on hydrogen for the polarisation functions, and one set of diffuse s and p type functions on second row atoms for the diffuse functions. Additional polarisation functions are useful to allow for small displacements of electronic charge away from the nucleus, whereas additional diffuse functions are good for significant shifts of electron density away from atomic centres, particularly in H-bonded systems and anions.[10]

A2.2.3 *Many-electron Hartree-product Wavefunctions*

The many-electron Hamiltonian is separable when the only terms in the one-electron Hamiltonians are for kinetic energy and nuclear attraction, which allows it to be expressed as a sum of one-electron Hamiltonians. The eigenfunctions of each of these one-electron Hamiltonians must each satisfy their respective one-electron Schrodinger equations: [2]

$$H = \sum_{i=1}^{N} h_i \qquad \text{[A2.20]}$$

$$h_i \psi_i = \varepsilon_i \psi_i \qquad \text{[A2.21]}$$

As a consequence, its many-electron eigenfunctions can each be expressed as a product of one-electron eigenfunctions, and thus the eigenvalue of a many-electron eigenfunction can be expressed as a sum of one-electron eigenvalues: [2]

$$H\Psi_{HP} = H\psi_1\psi_2 \dots \psi_N$$

$$= \sum_{i=1}^{N} h_i\psi_1\psi_2 \dots \psi_N$$

$$= (h_1\psi_1)\psi_2 \dots \psi_N + \psi_1(h_2\psi_2) \dots \psi_N + \cdots + \psi_1\psi_2 \dots (h_N\psi_N)$$

$$= (\varepsilon_1\psi_1)\psi_2 \dots \psi_N + \psi_1(\varepsilon_2\psi_2) \dots \psi_N + \cdots + \psi_1\psi_2 \dots (\varepsilon_N\psi_N)$$

$$= \sum_{i=1}^{N} \varepsilon_i\psi_1\psi_2 \dots \psi_N$$

$$= (\textstyle\sum_{i=1}^{N} \varepsilon_i)\Psi_{HP} \qquad \text{[A2.22]}$$

This result provides a means by which the molecular orbital energies can be calculated when the correct one-electron Hamiltonians in terms of the one-electron kinetic energy, nuclear attraction, and the interelectronic repulsion energy are used. The latter is dependent on the charge or probability density of the electron, which is equal to the square of the wavefunction. This presents a paradox, as the point of solving the equation was to find the wavefunction, and so the equations must be solved in a self-consistent manner: the initial wavefunctions are guessed at, the Schrodinger equation is solved for each, and a set of new wavefunctions is formed from which new improved Hamiltonians are calculated and applied to the new set of wavefunctions, generating yet a new set of wavefunctions. The result is that the energy continues to improve until convergence is reached, which is when the energy

between cycles is not significantly improved to within a certain threshold criterion. If the threshold is reduced, it takes more cycles till convergence.[2]

A2.2.4 *Electron Spin, Antisymmetry and Slater Determinants*

Each electron in the total wavefunction can be expressed in terms of a spatial coordinate (r) and spin coordinate (ω). However, the anti-symmetry requirement of the Pauli-Exclusion principle states that the total wavefunction must be anti-symmetric with respect to the interchange of the coordinates of any two electrons. For example, two electrons of the same spin occupying different orbitals could have a Hartree-product wavefunction such as: [1]

$$\Psi_{HP} = \psi_a(1)\alpha(1)\psi_b(2)\alpha(2) \qquad \text{[A2.23]}$$

However, this does not satisfy the Antisymmetry requirement of the Pauli-Exclusion principle, because swapping the electrons around yields a new wavefunction which is not the negative of the old one i.e. $\psi_b(1)\alpha(1)\psi_a(2)\alpha(2) \neq -\psi_a(1)\alpha(1)\psi_b(2)\alpha(2)$. The problem is solved by taking an appropriate combination of orbitals and normalising, which can be expressed using a Slater Determinant (SD) that is anti-symmetric with respect to the interchange of particles: [1]

$$\Psi_{SD} = \frac{1}{\sqrt{2}}[\psi_a(1)\alpha(1)\psi_b(2)\alpha(2) - \psi_a(2)\alpha(2)\psi_b(1)\alpha(1)]$$

$$\Psi_{SD} = \frac{1}{\sqrt{2}}\begin{vmatrix}\psi_a(1)\alpha(1) & \psi_b(1)\alpha(1)\\ \psi_a(2)\alpha(2) & \psi_b(2)\alpha(2)\end{vmatrix} \qquad \text{[A2.24]}$$

Slater determinants can be expanded to N by N determinants of any number of spin-orbitals (products of space and spin orbitals), and express the indistinguishability of quantum particles, which is not accounted for in the Hartree-product wavefunctions. Further to this, evaluation of the energy for interelectronic repulsion between SDs results in the reduction of the classical coulomb repulsion energy (J_{ab}) between the two electrons by an exchange integral (K_{ab}): [2]

$$\int \Psi_{SD} \frac{1}{r_{12}} \Psi_{SD} dr_1 \, d\omega_1 dr_2 d\omega_2 = J_{ab} - K_{ab} \qquad \text{[A2.25]}$$

Thus, there is a reduced probability of finding two electrons of the same spin in the same space. The same evaluation carried out on the determinant for electrons of different spins results in J_{ab} alone, and the exchange integral vanishes.[2]

A2.2.5 *Hartree-Fock Self-consistent Field Method*

The Hartree-Fock (HF) method unites the Hartree-product wavefunctions with Slater Determinants, allowing for the inclusion of the above exchange effects. Thus, it uses a trial wavefunction in the form of a Slater determinant. For a closed-shell system, the HF equations are: [10]

$$\hat{F}\varphi_k = \varepsilon_k \varphi_k \qquad \hat{F} = \hat{h} + \sum_{j=1}^{n} \left(2\hat{J}_j - \hat{K}_j\right) \qquad \text{[A2.26]}$$

Where the Fock operator is the sum of the one-electron Hamiltonians augmented by the HF potential, $2J_i$-K_i, a combination of the exchange and Coulomb operators, which are themselves defined so as to compute the coulomb and exchange integrals

described above.[2] Each MO is expanded as a user chosen basis set of AO-like functions ($\varphi_i = \sum_{\mu=1}^{M} \chi_\mu c_{\mu i}$ where $\{\chi_\mu | \mu = 1,2,\dots,M\}$), and substitution of this expansion into the HF equations allows the derivation of the Hartree-Fock-Roothaan equations in matrix form: [10]

$$\boldsymbol{FC} = \boldsymbol{SCe} \qquad \text{[A2.27]}$$

Where **F** is the Fock matrix, **C** is a matrix of orbital coefficients (where each column represents an MO (eigenvector)), **S** is an overlap matrix, and **e** is a diagonal matrix of eigenvalues. Thus the eigenvalues (the non-zero elements along the main diagonal of **e**) and the eigenvectors (the columns in **C**) can be calculated. However, the elements of the Fock matrix, **F**, depend on the elements of the one-electron density matrix, **D**: [10]

$$F_{\mu\nu} = \langle \chi_\mu | \hat{F} | \chi_\nu \rangle = \langle \chi_\mu | \hat{h} + \sum_{j=1}^{n} \left(2\hat{J}_j - \hat{K}_j\right) | \chi_\nu \rangle$$

$$= h_{\mu\nu} + \sum_{j=1}^{n} \left(2\langle \chi_\mu | \hat{J}_j | \chi_\nu \rangle - \langle \chi_\mu | \hat{K}_j | \chi_\nu \rangle\right)$$

$$= h_{\mu\nu} + \sum_{\rho=1}^{M}\sum_{\tau=1}^{M}\left(\sum_{j=1}^{n} c_{\rho j}^{*} c_{\tau j}\right)\left(2\langle \chi_\mu \chi_\rho | \chi_\nu \chi_\tau \rangle - \langle \chi_\mu \chi_\rho | \chi_\tau \chi_\nu \rangle\right)$$

$$= h_{\mu\nu} + \sum_{\rho=1}^{M}\sum_{\tau=1}^{M} D_{\rho\tau}\left(\langle \chi_\mu \chi_\rho | \chi_\nu \chi_\tau \rangle - \tfrac{1}{2}\langle \chi_\mu \chi_\rho | \chi_\nu \chi_\tau \rangle\right)$$

$$D_{\mu\nu} = 2\sum_{j=1}^{n} c_{\rho j}^{*} c_{\tau j}$$

$$E = \tfrac{1}{2}\sum_{\mu=1}^{M}\sum_{\nu=1}^{M} D_{\mu\nu}\left(h_{\mu\nu} + F_{\mu\nu}\right) \qquad \text{[A2.28]}$$

As can be seen from the derivation, the elements of **F** depend on the elements of **D** and the one- and two-electron integrals. The elements of **D**, however, depend on the orbital coefficients, which were the target quantities of the calculation. As a result of this, the HF equations need to be calculated in a self-consistent manner (as with Hartree-product wavefunctions), known as Self Consistent Field (SCF). The method by which this is usually carried out is given in Figure A2.5.[10]

The chief disadvantage of HF theory is that it does not take into account any form of electron correlation, apart from exchange. The fundamental assumption of HF theory is that each electron sees all of the others as an average field. Thus, most post HF methods have been devised in order to take this into account. The two post HF methods used in this work are MPT and DFT. MPT introduces perturbations to the Hamiltonian of the Fock operator, whilst DFT is based on the Hohenberg-Kohn theorem, which states that the ground-state energy is determined completely by the electronic density.[2]

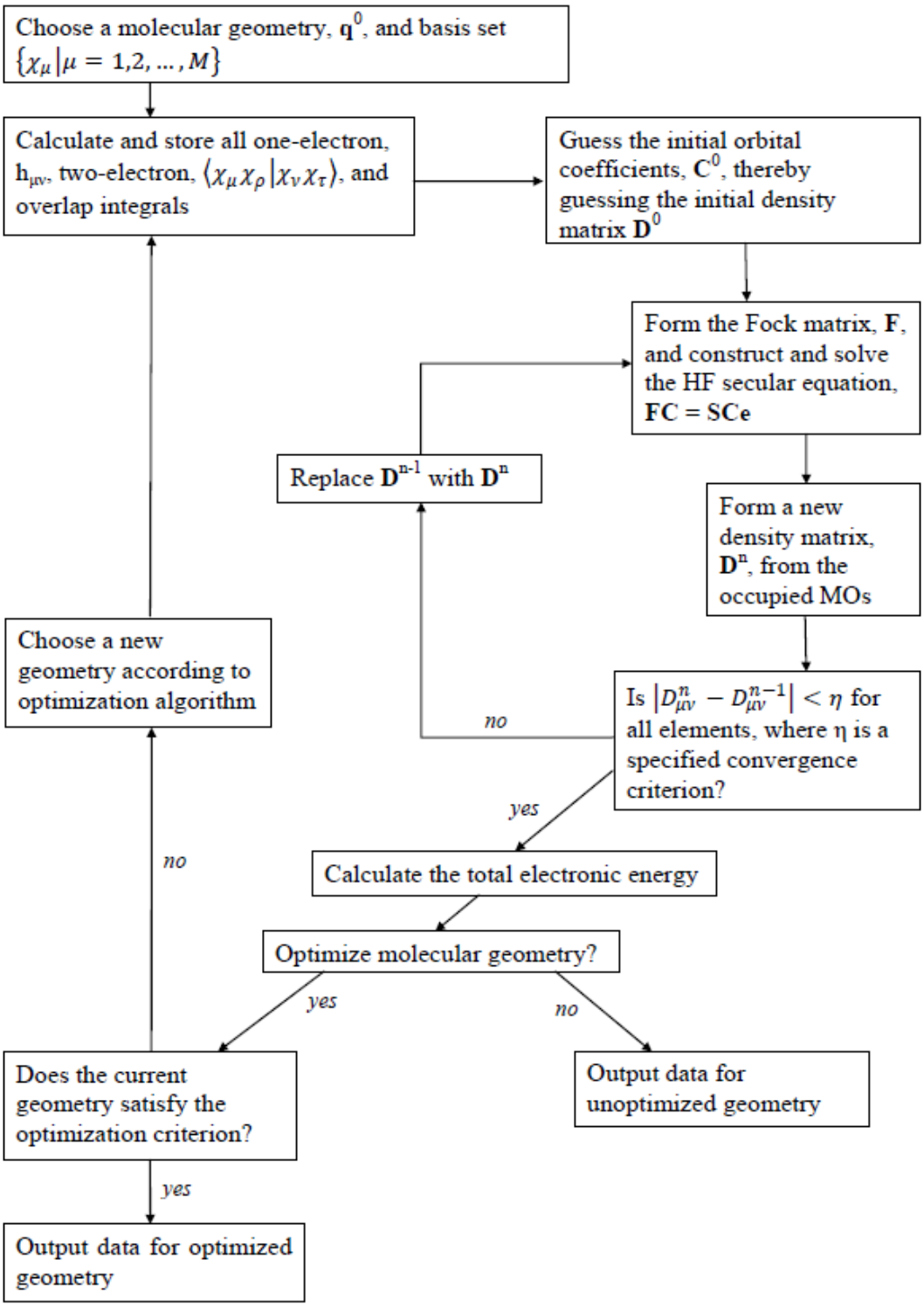

Figure A2. 5 Flow chart for the HF SCF procedure. The fact that the density matrix which is used to form the Fock matrix is itself dependent on the orbital coefficients means that the calculation must be solved iteratively. When the geometry is left unoptimized, this is referred to as a single point calculation (SPE). Flow chart adapted from references [2] and [10].

A2.2.6 *Møller-Plesset Perturbation Theory*

The difference between the true energy and the HF energy is often termed the correlation energy, $E_{corr} = E_{exact} - E_{HF}$. Møller-Plesset (MP) Perturbation Theory is a post HF method that takes into account electron correlation. The theory starts from the basis of Many-Body Perturbation Theory (MBPT), by using a much simpler version of the many-electron Schrödinger equation, involving a simpler 0^{th} order Hamiltonian, $\hat{H}_0$, and treating any difference to the true hamiltonian, $\hat{H} - \hat{H}_0$, as a 'perturbation' to $\hat{H}_0$. This difference is given as $\lambda\hat{V}$, where λ is some small number and $\hat{V}$ the perturbation operator, and so the many-electron Schrödinger equation becomes: [10]

$$\left(\hat{H}_0 + \lambda\hat{V}\right)\Psi_i = E_i\Psi_i \qquad \text{[A2.29]}$$

If $\lambda\hat{V}$ is small, then the wavefunctions and eigenvalues of this equation can be given as perturbation expansions: [10]

$$\Psi_i = \underbrace{\Psi_i^{(0)}}_{\substack{0^{th}order\\term}} + \underbrace{\lambda\Psi_i^{(1)}}_{\substack{1^{st}order\\correction}} + \underbrace{\lambda^2\Psi_i^{(2)}}_{\substack{2^{nd}order\\correction}} + \ldots$$

$$\mathrm{E}_i = \underbrace{E_i^{(0)}}_{\substack{0^{th}order\\term}} + \underbrace{\lambda\mathrm{E}_i^{(1)}}_{\substack{1^{st}order\\correction}} + \underbrace{\lambda^2\mathrm{E}_i^{(2)}}_{\substack{2^{nd}order\\correction}} + \ldots \qquad \text{[A2.30]}$$

Where:

$$E_i^{(0)} = \left\langle \Psi_i^{(0)} \middle| \hat{H}_0 \middle| \Psi_i^{(0)} \right\rangle$$

$$E_i^{(1)} = \left\langle \Psi_i^{(0)} \middle| \hat{V} \middle| \Psi_i^{(0)} \right\rangle$$

$$E_i^{(2)} = \left\langle \Psi_i^{(0)} \middle| \hat{V} \middle| \Psi_i^{(1)} \right\rangle = \sum_{j \neq i} \frac{\left| \left\langle \Psi_i^{(0)} \middle| \hat{V} \middle| \Psi_j^{(0)} \right\rangle \right|^2}{E_i^{(0)} - E_j^{(0)}}$$

In MP Perturbation Theory, the 0[th] order Hamiltonian is chosen to be the sum of one-electron Fock operators for all electrons, and so the eigenvalue for $\hat{H}_0$ is equal to the sum of one-electron eigenvalues: [10]

$$\hat{H}_0 = \sum_{i=1}^{N} \hat{F}(\boldsymbol{x}_i)$$

$$\hat{H}_0 \Psi_0^{(0)} = E_0^{(0)} \Psi_0^{(0)} = \left(\sum_{i=1}^{N} \varepsilon_i\right) \Psi_0^{(0)} \qquad \text{[A2.31]}$$

However, this sum is not equal to the HF energy, and contains an error whereby each electron-electron repulsion is counted twice. The difference between this energy and the HF energy is accounted for in the 1[st] order correction, which returns the HF energy (assuming λ = 1): [2]

$$E_0^{(1)} = \left\langle \Psi_0^{(0)} \middle| \hat{V} \middle| \Psi_0^{(0)} \right\rangle = \left\langle \Psi_0^{(0)} \middle| \hat{H} - \hat{H}_0 \middle| \Psi_0^{(0)} \right\rangle = E_{HF} - \sum_{i=1}^{N} \varepsilon_i$$

$$E_0^{(0)} + E_0^{(1)} = E_{HF} \qquad \text{[A2.32]}$$

It is only with the 2[nd] order correction that the perturbation can lower the energy. This is given by the second-order energy, MP2: [10]

$$E_0^{(2)} = \sum_i \sum_{i<j} \sum_p \sum_{p<q} \frac{\left(\langle \psi_p \psi_q | g | \psi_i \psi_j \rangle - \langle \psi_p \psi_q | g | \psi_j \psi_i \rangle\right)^2}{\varepsilon_i + \varepsilon_j - \varepsilon_p - \varepsilon_q}$$

$$E_{MP2} = E_0^{(0)} + E_0^{(1)} + E_0^{(2)} = E_{HF} + E_0^{(2)} \qquad \text{[A2.33]}$$

The MP2 energy contains a significant amount of correlation energy, which can sometimes be more than necessary as the perturbations do not obey the variation principle, and therefore the total energy can sometimes overshoot the exact ground state energy. Higher accuracy can be achieved by using higher order correction terms (MP3 and MP4), though in this work, MP2 was deemed sufficient, giving an appropriate balance of computational cost to accuracy.[10]

A2.2.7 *Density Functional Theory*

Density Functional Theory (DFT) was developed as an alternative to the HF formalism, and is based on the Hohenberg-Kohn theorem, which states that the ground-state electronic energy is determined completely by the electronic density, and so there exists a unique functional (a prescription for producing a number from a function) which determines the ground-state energy and electronic density exactly. The simplest way of doing this is using the Local Density Approximation (LDA), which is based on the exact exchange energy for a uniform electron gas. The approximation uses Kohn-Sham orbitals, which allows the representation of the wavefunction as a Slater determinant. The total energy is given as: [10]

$$E_{DFT}[\rho] = T_s[\rho] + E_{en}[\rho] + J[\rho] + E_{xc}[\rho] \qquad [A2.34]$$

Where:

$$T_s[\rho] = \sum_{i=1}^{N} \left\langle \psi_i \middle| -\tfrac{1}{2}\nabla^2 \middle| \psi_i \right\rangle$$

$$E_{en}[\rho] = -\sum_{I=1}^{M} \int \frac{Z_I \rho(\boldsymbol{r})}{|\boldsymbol{R}_I - \boldsymbol{r}|} dr$$

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\boldsymbol{r}_1)\rho(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} d\boldsymbol{r}_1 d\boldsymbol{r}_2$$

$$E_{xc}[\rho] = (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho])$$

Each of these terms depends on the electron density of the system: the kinetic energy of the system as given by a Slater determinant is $T_s[\rho]$, the potential energy of attraction between the electrons and nuclei is $E_{en}[\rho]$, the potential energy of repulsion between the electrons is $J[\rho]$, and the final term is the exchange correlation, which is a term that compensates for the differences between the exact kinetic energy and the kinetic energy calculated from a Slater determinant (1st bracket), and between the exact electron-electron repulsion energy and the classical Coulomb repulsion of the electron density (2nd bracket). As discussed previously with HF, the Kohn-Sham equations are solved in a self-consistent manner, though the exchange functional has to be calculated numerically to a high degree of accuracy.[10]

In LDA, E_{xc} depends entirely on the electron density. However, there exists a modification to this, in the gradient-corrected or Generalised Gradient Approximation (GGA), where E_{xc} is both dependent on the electron density, and the gradient to the electron density. The B3LYP functional used in this work is a hybrid of many types of both functionals, chosen so as to give the best result: [10]

$$E_{xc}^{B3LYP}[\rho] = E_x^{LDA}[\rho] + c_0(E_x^{HF}[\rho] - E_x^{LDA}[\rho]) + c_x E_x^{Becke88}[\rho] + E_c^{VWN3}[\rho] + c_c(E_c^{LYP}[\rho] - E_c^{VWN3}[\rho]) \quad \text{[A2.35]}$$

A2.2.8 *Calculation of Molecular Properties – Geometry Optimization*

A molecule of N atoms can be described using 3N coordinates that give the positions of each atom. However, in non-linear molecules, there are 3N-6 coordinates, therefore any arbitrary arrangement of atoms can be described by 3N-6 internal coordinates, $Q_1 ... Q_{3N-6}$, and the ground-state energy hypersurface is a function of each of these coordinates ($E = E(Q_1, Q_2, \dots, Q_{3N-6})$). This surface contains minima, maxima and saddle points which are points where the 1st derivatives with respect to each of these coordinates are zero. These first derivatives can be collected together to form the gradient, **g**: [10]

$$\boldsymbol{g} = \begin{pmatrix} \frac{dE}{dQ_1} \\ \frac{dE}{dQ_2} \\ \vdots \\ \frac{dE}{dQ_{3N-6}} \end{pmatrix}$$

However, it is necessary for minima to be distinguished from maxima and saddle points by the use of a matrix of second derivative elements, known as the Hessian: [10]

$$\boldsymbol{H} = \begin{bmatrix} \frac{d^2E}{dQ_1^2} & \frac{d^2E}{dQ_1 dQ_2} & \cdots & \frac{d^2E}{dQ_1 dQ_{3N-6}} \\ \frac{d^2E}{dQ_2 dQ_1} & \frac{d^2E}{dQ_2^2} & \cdots & \frac{d^2E}{dQ_2 dQ_{3N-6}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2E}{dQ_{3N-6} dQ_1} & \frac{d^2E}{dQ_{3N-6} dQ_2} & \cdots & \frac{d^2E}{dQ_{3N-6}^2} \end{bmatrix}$$

The hessian is symmetric and can be diagonalized by multiplication by an appropriate factor (all off-diagonal elements are zero), giving diagonal elements corresponding to second derivatives in each of the coordinates. The condition of a minimum is then that each of these elements is positive. The geometry optimization process makes use of the gradient and Hessian, and is carried out by the Newton-Raphson method, in which the energy function is approximated as a second order Taylor expansion: [10]

$$E(\boldsymbol{Q}) = E(\boldsymbol{Q}_0) + \tilde{\boldsymbol{q}}\boldsymbol{g}(\boldsymbol{Q}_0) + \frac{1}{2}\tilde{\boldsymbol{q}}\boldsymbol{H}(\boldsymbol{Q}_0)\boldsymbol{q} + \cdots \qquad [A2.36]$$

Where **g** is the gradient, **Q** is a column vector for each of the new (**Q**) and old ($\mathbf{Q}_0$) coordinates, **q** is the difference between these column vectors, and **H** is the Hessian matrix. The new coordinates are then: [10]

$$\boldsymbol{Q} = \boldsymbol{Q}_0 - \boldsymbol{H}^{-1}(\boldsymbol{Q}_0)\boldsymbol{g}(\boldsymbol{Q}_0) \qquad [A2.37]$$

This equation is often solved approximately, but to a high degree of accuracy using a method such as conjugate gradient. There also exist various quasi Newton-Raphson methods where an approximation to the Hessian is used. [10]

References

[1] Grant, G. H.; Richards, W. G. *Oxford Chemistry Primers: Computational Chemistry*; Oxford University Press: New York, NY, 1995

[2] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons, LTD: Chichester, UK, 2002

[3] Halgren, T. A. *J. Comp. Chem.* **1995**, *17, 490*

[4] Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester, UK, 1999

[5] Klunzinger, P. *FAQ web pages for SPARTAN^{TM} customer support*; Wavefunction, INC. [updated 3 Jun 2006; cited 15 Nov 2009]. Available from http://www.wavefun.com/support/sp_compfaq/support_compFAQ.html

[6] Michalewicz, Z.; Fogel, D. B. *How to solve It: Modern Heuristics*; 2nd ed. Springer: Berlin, 2000

[7] *Spartan '08 windows: Tutorial and Users Guide*; Wavefunction, INC.: Irvine, 2006-2009.

[8] Screenshot from the SPARTAN^{TM} 08 program, Wavefunction, INC.: Irvine, 1999-2009

[9] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J Chem. Phys.* **1953**, *21*, 1087

[10] Karadakov, P. B. *Quantum Chemistry* [unpublished lecture notes], University of York, UK

[11] Rassalov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *J. Comp. Chem.* **2001**, *22*, 976

Abbreviations:

AMBER, AMBER94/99	Biomolecules force-fields of P. Kollman et al.
BIRD	Blackbody infrared dissociation
B3LYP	Beck 3 parameter exchange functional with LYP
CAD	Collision activated dissociation
cc-pVDZ	Correlation consistent polarised valence double zeta basis set of Dunning et al.
cc-pVTZ	Correlation consistent polarised valence triple zeta basis set of Dunning et al.
CCSD(T)	A coupled cluster method
CHARMM27	Biomolecules force-field of Brooks et al.
CID	Collision induced dissociation
CRM	Charged-residue model
DFT	Density functional theory
ESI	Electrospray ionisation
FTICR	Fourier transform ion cyclotron resonance
GPA	Gas-phase acidity
GPB	Gas-phase basicity
GGA	Generalised gradient approximation
GO	Gaussian orbital
GnRH	Gonadotropin-releasing hormone
HF	Hartree Fock
ICR	Ion cyclotron resonance
IEM	Ion evaporation model
IR	Infrared
IRMPD	Infrared multi-photon dissociation
LACVP	Hybrid basis set of Hay and Wadt, with an effective core-potential basis set for K, Rb and Cs atoms, and 6-31G** for all other atoms
LDA	Local density approximation
LIF	Laser-induced fluorescence
MALDI	Matrix-assisted laser desorption ionisation
MBPT	Many-body perturbation theory
MCA	Multiply-charged anion
MM	Molecular mechanics
MM2	Organics force-field of N. L. Allinger et al.
MM3	Organics and biomolecules force-field of N. L. Allinger et al.
MMFF94	Organics and biomolecules force-field of T. A. Halgren
MP2	Moller plesset perturbation theory to the second order
MP3	Moller plesset perturbation theory to the third order
MP4	Moller plesset perturbation theory to the fourth order
MPT	Moller plesset perturbation theory
MS	Mass Spectrometry
OPLS	Organics and biomolecules force-field of Jorgensen et al.
PA	Proton affinity

PES	Potential energy surface or photon-electron spectroscopy
QIT	Quadrupole ion trap
QM	Quantum mechanics
RCB	Repulsive coulomb barrier
REMPI	Resonance-enhanced multi-photon ionisation
RNA	Ribose nucleic acid
SCF	Self-consistent field
SD	Slater determinant
SORI	Sustained off-resonant infrared
SP	Single-point
SPE	Single-point energy
STO	Slater-type Orbital
STO-KG	Minimal basis set
STO-3G	STO-KG minimal basis set where K = 3
TBA-F	Tetrabutyl ammonium fluoride
TOF	Time-of-flight
UHF	Unrestricted Hartree-Fock
UV	Ultraviolet
VDE	Vertical detachment energy
3-21G	Split valence double zeta basis set of Pople et al.
4-21G	Split valence double zeta basis set of Pople et al.
6-31G	Split valence double zeta basis set of Pople et al.
6-31+G**	Split valence double zeta basis set of Pople et al., augmented with 2 polarisation functions and 1 diffuse function
6-31+G*	Split valence double zeta basis set of Pople et al., augmented with 1 polarisation function and 1 diffuse function