Analysis of Non–Covalent Systems through Mass Spectrometry Techniques

A thesis submitted to the University of Manchester for the degree of Doctor in Philosophy in the Faculty of Science and Engineering

2016

Chris Nortcliffe

School of Chemistry

Contents

List of figures

List of tables

List of equations

Word Count 56,182

Abstract

Non–covalent interactions represent a major area of interest across the scientific community, with the interactions of biomolecules being an interesting and influential field. Regulating protein function by influencing binding of co–factors and other proteins is at the forefront of many research directions. However, non–covalent complexes can present challenging targets for analysis as many techniques are unable to differentiate between species in a heterogeneous population. Mass spectrometry (MS) experiments provide the unique ability to isolate individual components in a mixture prior to analysis, even when they exist in a dynamic equilibrium. When this is combined with hybrid techniques such as ion mobility–MS, along with fragmentation, the differences incorporated within a complex can be identified structurally and related back to the binding interface. The work presented in this thesis demonstrates how MS and related techniques are suitable methods to address the challenge of analysing non–covalent complexes

The first system in chapter two is a DNA aptamer that has been designed to bind an antibiotic kanamycin. Aptamers represent an emerging field in biopharmaceuticals due to their ability to provide specific antigen binding analogous to antibodies whilst providing advantages in chemical synthesis and immune response. The interaction between these two molecules was assessed using a variety of MS techniques to elucidate the nature of the interface, if the structure of the molecule is pre–organised for binding, and the conformational changes that binding induced. Additional sequence constructs were compared for their binding and structural similarities.

The systems under scrutiny in chapter three belong to a class of anion sensing molecules designed to detect pyrophosphate *via* displacement of a dye molecule, with a high specificity over similar cationic molecules. These unique sensors have the potential to monitor cellular respiration by detecting pyrophosphate released from ATP. The sensors consist of a circular peptide scaffold with two flexible arms each leading to bound Zn (II) ions, with a variety of permutations of length and size examined. Gas–phase kinetic studies were performed to assess efficacy and selectivity of binding pyrophosphate. Subsequently, ion mobility and computational data were used to determine structure–function relationships.

Chapters four and five consider the structural properties of the transcription factor c–MYC and its binding partner MAX. Deregulation of c–MYC is implicated in human cancer progression, therefore making it an attractive pharmaceutical target. However, the leucine zipper binding region of the two proteins is intrinsically disordered, undergoing a disorder to order transition upon binding which presents difficulties for characterisation using traditional structural techniques. These chapters explore the effects of a series of small molecules that have been shown to inhibit dimerisation of c–MYC and MAX through a postulated stabilisation of these disordered states. Peptides derived from the proteins are investigated in the work presented in chapter four, whilst in chapter five extended sequences are interrogated. Further work in chapter five also considers the ability of the dimers to bind DNA in the presence and absence of ligands.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations)

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Perdita Barran for her exemplary support both professionally and personally throughout the many years I have spent under her tutelage. She remains a source of inspiration and a fountain of encouragement.

Furthermore, I would like to thank all my collaborators I had the pleasure of working with. Dr Dave Clarke has been a mentor throughout many of the projects I have presented in particular the Ky2 project. I am eternally grateful to Stacy Panova and Nikita Vekaria for gifting me c–MYC protein along with the entire Jon Waltho group. The project on anion sensing would not be possible without the molecules created by Kate Jolliffe and all her lab in Sydney.

I have been truly blessed by an abundance of friends during my tenure in the Barran group who I have shared many trials and tribulations with. (Not to mention a pint or two). So, in order I can best remember I would like to thank: Martin, Roland, Jude, Yana, Hattie, Ewa, Kamila, Max, Jason, Sophie, Albert, Jakub, Bex, Ashley, Ellie, Callum, Dave, Chris, Lukasz, Alina, Rosie, Ben, Ellie 2, Bruno, Emily, and Aidan. Lukasz deserves an extra special mention for his limitless knowledge of computational chemistry which he has helped me with. Some of these people deserve extra mentions for the hours they spent with me outside the lab fighting dragons, trading sheep, and battling over small grey wood mice.

I would like to thank Tom for being a constant source of alcohol and amusement outside the office. Hannah for being a force of chaos and laughter whenever I came home from a long day in the lab. Wendy and Jamie, and Dave and Clem for many evenings of fun filled conversation. Ruth has been my oldest friend and a source of excellent and witty conversations since many moons ago when we were young. I owe a lot to Caroline for her support, and for everything she has done for me over the years.

Julie has been a spectacular and phenomenal help during the final stages of my thesis for proof reading, advice, and generally keeping me sane and happy. May she continue to tolerate me.

I would like to thank my wonderful parents Chris and Jayne Horton for listening to me talk about what I do with blank looks on their faces. They have given me so much support over the years which I will never be able to repay. I would like to thank my grandparents; in particular my Grandad Rowland to whom this thesis is dedicated. He remains who I aspire to be.

Finally, I would like to give an extra special thanks to everyone who helped me through my personal challenges I faced during the years of my PhD. I have so many people who helped me, which I am eternally grateful for.

Preface

This thesis outlines some of the research I have carried out during my PhD on the use of mass spectrometry techniques to study the interactions of non–covalent complexes and the effects of ligand binding on biomolecular structure. The introduction provides a stepping off point into the work from basic terminology and equipment through to current and past work. It begins with a discussion of non–covalent interactions and complexes in a general sense, and then goes on to describe them in more specific and application–focused directions.

The main body consists of four first author papers, one of which has been published at the time of submission. These have been formatted in a way to properly display them in this thesis format. Supplementary information is provided in the appendices.

All four papers follow a similar trend, in that they utilise a variety of mass spectrometry techniques to study non–covalent interactions within biological systems. The first paper discusses the binding of kanamycin to a DNA aptamer and seeks to identify the orientation of the interaction between these two molecules. The second paper interrogates a macromolecular construct designed to bind ATP specifically over other anions. This paper compares varying structures and ligands to relate structural features to ATP specificity. Papers three and four both focus on the oncogenic system c–MYC and its binding partner MAX on a peptide and protein level respectively. They look at a series of small molecules designed to inhibit their interactions, and how these affect their gas–phase structures.

Abbreviations

ATD	Arrival time distribution
ATP	Adenosine triphosphate
CCS	Collisional cross section
CCSD	Collisional cross section distribution
$^{DT}CCS_{He}$	$^{\text{Drift tube}}$ collisional cross section $_{\text{helium}}$
CEM	Chain ejection model
CI	Chemical ionisation
CID	Collision induced dissociation
CRM	Charged residue model
CSD	Charge state distribution
Da	Dalton
DC	Direct current
DNA	De–oxy ribo nucleic acid
DMSO	Dimethyl Sulfoxide
DT–IM–MS	Drift tube ion mobility–mass spectrometry
ECD	Electron capture dissociation
EDD	Electron detachment dissociation
EHSS	Exact hard sphere scattering
EI	Electron ionisation
ESI	Electrospray ionisation
FAB	Fast atom bombardment
FAIMS	Field asymmetric ion mobility spectrometry

FD	Field desorption
FRET	Fluorescence resonance energy transfer
FTIR	Fourier transform infrared spectroscopy
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IEM	Ion ejection model
IM–MS	Ion mobility–mass spectrometry
IMS	Ion mobility spectrometry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRMPD	Infrared multiphoton dissociation
ITC	Isothermal calorimetry
m/z	Mass–to–charge ratio
MALDI	Matrix–assisted laser desorption ionisation
MCP	Microchannel plate
MC	Monte Carlo
MD	Molecular Dynamics
MOBCAL	A computational program to generate molecular cross sections
MoQToF	Mobility quadrupole time–of–flight mass spectrometer
MS	Mass spectrometry
n–ESI	Nano–electrospray ionisation
NMR	Nuclear magnetic resonance
NOE	Nuclear overhauser effect

PA	Projection approximation
PDB	Protein data bank
PSA	Projection superposition approximation
PTM	Post–translational modification
Q–ToF	Quadrupole time–of–flight mass spectrometer
RF	Radiofrequency
RNA	Ribo nucleic acid
TM	Trajectory method
TOF	Time–of–flight
TWIM–MS	Travelling–wave ion mobility–mass spectrometry
UV	Ultraviolet
UVPD	Ultraviolet photo–dissociation

Greet every morning with a smile.

That way it won't know what you're planning to do to it.

This thesis is dedicated to Rowland Nortcliffe,

for instilling in me autodidactic tendencies…

1.

Introduction

1.1. Non–covalent systems

1.1.1. *Non–covalent interactions*

Non–covalent bonds or interactions occur either within or between macromolecules *via* a variety of electromagnetic interactions. Non–covalent bonds differ from covalent bonds in that they do not involve the sharing of a pair of electrons [1]. A single non–covalent bond is of much lower energy than a covalent one, typically between 1–5 kcal/mol, and due to this most molecules will be able to break non–covalent bonds at room temperature (25°C) [2]. Multiple non–covalent bonds together can form highly stable and specific structures or interactions. These non–covalent bonds are as important to biological structures as the sequence of amino acids or DNA bases. Types of non–covalent interactions that are prevalent in protein chemistry and molecular biology include: ionic interactions, hydrophobic interactions, transient dipoles, van der Waals forces, and finally hydrogen bonding. Multiple molecules connected by non–covalent bonds can be referred to as a non–covalent complex. Non–covalent bonds and complexes can often be difficult to analyse due to their weak and often dynamic nature.

Ionic interactions are a type of non–covalent bond that exist between charged ions of opposite polarity. At cellular pH, there exist many charged groups on bio–molecules, along with a large number of cations and anions (Figure 1.1). In aqueous environments, ions exist surrounded by water in a hydration shell interacting with waters inherent dipole. This hydration is highly energetically favoured and often overcomes the lattice energy of solid ionic crystals [3–5]. Ionic interactions commonly stabilise macromolecular structures, with metal ions often found interacting with proteins and phosphate groups [6, 7]. Salt bridges are an example of ionic interactions and can be formed within proteins where anionic and cationic residues lie within a distance 3–5 Å of each other [8]. Salt bridge interactions commonly have both ionic and hydrogen bonding interactions.

Whereas ionic bonds in a solid salt crystal are very strong (in the region of 100 kcal/mol), in solution they become much weaker due to the hydration of water with a reduced strength of around 1–4 kcal/mol. However, when buried within a protein core with an absence of water, these bonds can become stronger and more influential. For a cationic and anionic pair of amino acid residues (aspartic acid and arginine for example) within a protein core in the complete absence of water, ionic bonding strength could be as high as 20 kcal/mol [9].

Hydrophobic interactions commonly occur in areas rich in C–H bonds which do not interact with water (Figure 1.1). These non–polar regions cluster together as clustering is more energetically favoured than the generation of a non–interacting water cage around the hydrophobic molecules. Singular non–polar groups force the water to rigidly hydrogen bond with itself which is entropically unfavourable [10]. The positive entropy change, due to clustering effects of hydrophobic regions, contributes to a negative change in the Gibbs free energy of protein folding thereby having a stabilising effect. For example, a singular side chain of the amino acid leucine presents 100 Å^2 of surface area interactions with water. The removal of this hydrocarbon chain can lead to an energy change of –2 kcal/mol in the Gibbs free energy [11]. For some globular proteins, the energy stabilisation from grouping of non–polar interactions could account for nearly 50% of the energy of the stable protein fold [11].

Hydrophobic bonding is especially important when generating structural information through mass spectrometry as the gas–phase is a highly hydrophobic environment relative to solution and could therefore influence protein structures [12–13]. Ionic interactions however, can be retained more strongly as the effect of solvent is reduced.

Electron movement around molecules is constantly in flux and as two molecules approach each other they can affect each other (through repulsion of charges) leading to a transient electric dipole between the two (Figure 1.1). This generates a weak non–specific attractive force termed a van der Waals interaction [14]. Molecules with a permanent electric dipole can also induce a transient dipole in others [14]. Van der Waals forces are particularly important for non–polar liquids and solids which cannot form ionic or hydrogen bonds.

Van der Waals interactions require atoms to be in close proximity (1–2 Å) and become effectively zero past 5–6 Å. Additionally, the energies involved are only around 0.02–0.2 kcal/mol [15, 16]. These factors together mean that van der Waals interactions could be seen to be inconsequential on the scale of energies in protein folding. However, as the interiors of proteins are densely packed, along with broad flat interactions between some proteins there can be several hundred van der

Waals interactions occurring at one time both within and between proteins. The sum of these interactions leads to effects on proteins that are disperse but highly influential [15]. Van der Waals interactions are retained in the gas–phase like ionic interactions.

Hydrogen bonding is the most common non–covalent bond in biological systems and is crucial to the tertiary fold of proteins and DNA, along with their interactions with ligands and substrates [17]. The hydrogen bond involves hydrogen atoms bound to a highly electronegative atom (commonly oxygen or nitrogen) which results in an uneven distribution of electrons generating a partial positive charge δ^+ upon the hydrogen (Figure 1.1). This partial charge can interact with an acceptor electronegative atom with a pair of nonbonding electrons in its outer shell. Water is the most common example of hydrogen bonding forming networks of interfacing molecules. Due to the prevalence of water within cellular environments many hydrogen bonds within cells have a dynamic nature, with hydrogen bonds between amino acids in exchange with separate interactions with water.

Hydrogen bonds commonly have a distance of 2.5–3 Å between the two electronegative atoms, with the length of the bond inversely proportional to its strength ranging from 0.5–7 kcal/mol [18]. Alignment is critical in hydrogen bonding with weakening of the interaction if the donor, hydrogen and acceptor molecules are not linearly aligned. Hydrogen bonding is electronic in nature similar to ionic interactions, and is therefore retained in the gas–phase.

A further interaction which should be mentioned is that of metal coordination (Figure 1.1). Metal:ligand binding occurs through the overlap of electron orbitals leading to a transfer or sharing of electrons. Though metal coordination bonds have some covalent character and are often much stronger and more stable other non–covalent interactions, some coordinative bonds can be weak and dynamic [19, 20]. The strengths of metal:ligand bonds vary with: the metal ion involved, its oxidation state, the particular ligand involved and its orientation. There are several biological metal coordinations that are commonly found such as the heme group in proteins such as cytochromes and myoglobin, and zinc ions in proteins such as zinc fingers. The free energy of myoglobin stabilisation by the heme group was calculated as –18.5 kcal/mol through binding to two histidine residues [21]. The heme group in myoglobin has an oxygen atom coordinated with the metal centre through which one of the histidine residues binds, the removal of this leads to a change in the energy of ±3.4 kcal/mol [21]. This type of interaction is observed in chapter 3 between the zinc ions and the dipicolylamine groups.

The final non–covalent interaction type which will be described is the interactions of delocalised π systems (Figure 1.1). Aromatic ring systems containing delocalised electrons exhibit a strong quadrupole moment with a positive potential on the atoms within the ring and a corresponding negative potential above and below. Delocalised systems of different electron densities can stack together to stabilise structures, which is observed biologically in DNA base stacking and is present in small molecules interacting with enzymes. π systems have also been observed interacting with C–H bonds, metal ions, and in uncommon cases halogens; depending upon the electron withdrawing of donating groups around benzene rings [22]. π–π aromatic stacking interactions are generally weak (<2 kcal/mol) but can be strengthened with mono or bivalent cations (6 kcal/mol and 22 kcal/mol respectively) [23].

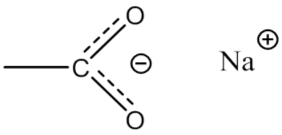
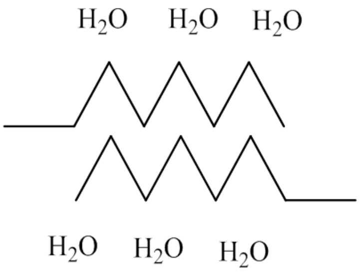
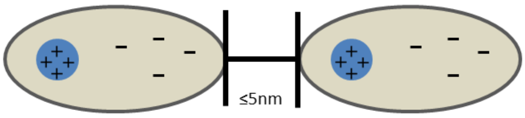
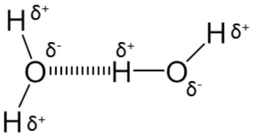
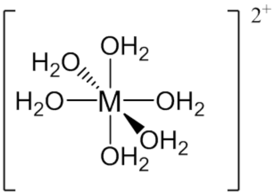
Bonding	**Diagram**	**Energies**
Ionic Bonding		1-20 kcal/mol
Hydrophobic		2 kcal/mol per hydrocarbon group
Van der Waals		0.02-0.2 kcal/mol
Hydrogen Bonding		0.5-7 kcal/mol
Metal Coordination		15-25 kcal/mol
π Stacking		2-22 kcal/mol

Figure 1.1: Examples of non–covalent interactions across a variety of molecular systems with a range of their energies.

1.1.2. Non–covalent complexes of biomolecules

Life simply could not exist without non–covalent complexes and they are ubiquitous across nature. This section will cover several non–covalent complexes and interactions: firstly, intramolecular interactions (protein folding), intermolecular interactions (DNA double helix, enzymes, protein complexes, and antibodies), aggregating systems (lipid bilayers, aggregating proteins), and finally with drug:biomolecule interactions. These complexes and interactions contain many non–covalent bonds of differing characteristics as discussed in section 1.1.1. No process or example uses only a single kind of non–covalent interaction, many work together through different parts of the molecule.

Protein structure lies at the very heart of protein function, with the secondary sequence undergoing complex folding to create pockets and surfaces where chemical reactions may take place [25]. Non–covalent interactions are a small factor when looking at intramolecular structure, with covalent disulfide bridges being a larger factor determining three–dimensional structure of many proteins. Disulphide bridges typically have a bond strength of 60 kcal/mol with a length of around 2 Å, therefore they occur over a similar distance to non–covalent interactions, though the bridges are considerably stronger [24]. Another important feature of disulphide bonds is that they are redox active and therefore are not stable in all cellular environments.

There are five polar amino acid side chains which carry a charge at biological pH (arginine, lysine, histidine, aspartic acid and glutamic acid), which can undergo ionic bonding, or these charges can be used to sequester metal ions, or cofactors essential for structure and function [25]. As the cell exists as mainly an aqueous environment, hydrophobic amino acids such as leucine and valine often become buried within the centre of a protein structure forcing more polar amino acids to the surface. This effect can be reversed in membrane–bound proteins which will be discussed later with lipid bilayers. Amino acids with delocalised ring systems such as phenylalanine, tryptophan and tyrosine can undergo π–π stacking when near each other, and these residues are often found near active sites, acting to orientate substrate molecules with catalytic residues [26]. Hydrogen bonding can occur throughout the protein, however most polar groups readily form bonds with water, making them commonly found on the surface of proteins. Techniques that study protein structure are important as they allow us to uncover how biological proteins perform their function, opening up new avenues for drug targets.

DNA is a polymer molecule where each individual strand is formed of covalent bonds; these two strands come together in a massive hydrogen bonded structure [27]. The vast network of hydrogen bonds not only holds these expansive macromolecular chains together; it is highly specific, leading to complementary binding with mismatches easily recognised. The hydrogen bond exchange in DNA duplex formation is exothermic, partly due to the increase in hydrogen bonds formed. The highly directional nature of the hydrogen bonding is enhanced by the highly polar nature of the phosphate backbone of DNA, thus having a strong interaction with the aqueous cell environment [27, 28]. The other internal non–covalent contributor to DNA structure is base stacking between nucleotides, however π–stacking occurs both inter– and intra–strand. As expected, the degree of stabilisation depends upon the DNA sequence, with the nearest neighbours determining the DNA stability. The negative charges on the phosphodiester backbone of DNA have a strong repulsive effect upon one another which is tempered with the presence of cationic metals. Duplex stability increases with increasing salt concentration, with divalent cations such as Mg^{2+} being more stabilising than Na^+ or K^+ [29]. Techniques that study DNA structure are important as they help understand the expression and activation of genes allowing greater understanding of how the genome is processed.

Enzymes are biological catalysts that bind substrates in an active site, lowering the activation energy for reactions. The active site's three–dimensional geometry is crucial to the operation of the enzyme, with many weak non–covalent interactions being present to align the reactants in the specific orientation required to stabilise transitional states. The first idea proposed for enzyme:substrate interactions was the 'lock–and–key' model (Figure 1.2) developed by Emil Fischer in 1894 [30]. This model states that the active site of the enzyme is perfect for the arrival of the target/specific substrate molecule. The 'lock–and–key' hypothesis was found to be lacking, as enzymes are stabilised by substrate binding and undergo a conformational change. In 1958 Daniel Koshland modified the theory to suggest that as the enzyme is flexible, the active site is reshaped by its interactions with the substrate. This theory is known as 'induced–fit' and is viewed as more comprehensive than 'lock–and–key' (Figure 1.2) [31].

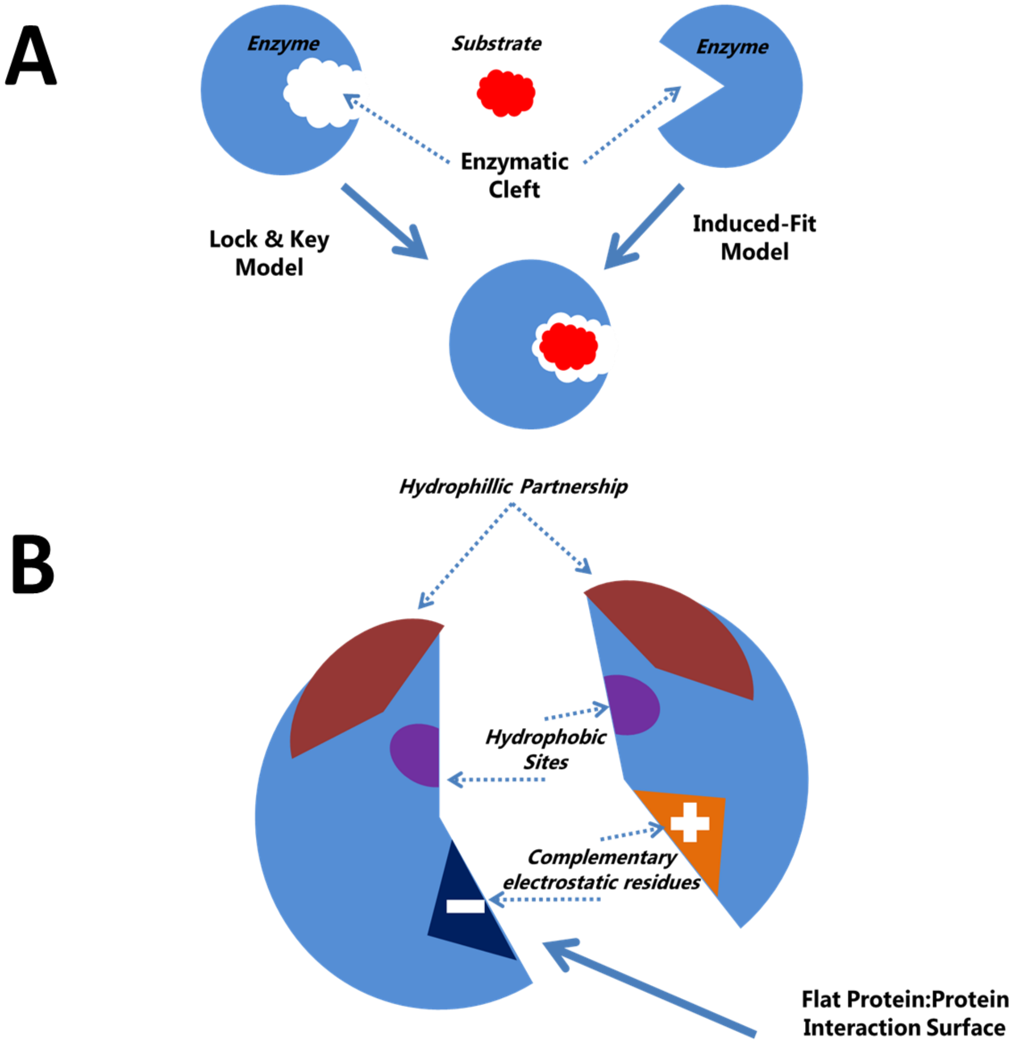


Figure 1.2 A) Visual description of the lock and key model and induced fit model. The lock and key model proposes that the binding site is already preconfigured for the substrate to bind. In the induced fit model, the ligand binding induces a structural change in the enzyme, bringing catalytic groups together. B) Shows a cartoon depiction of flat protein:protein interaction surfaces. Hydrophilic interactions could include polar amino acid residues (*e.g.* serine), or post translational modifications which increase the hydrophilicity of the protein. Hydrophobic patches are commonly localised amino acids high in CH chains (valine, leucine, and isoleucine). Electrostatic interactions can be formed by complementary positive (lysine, arginine, and histidine) and negative (aspartate, glutamate) residues.

Enzymatic active sites often have several potential non–covalent interactions, due to their need to interact strongly with the desired substrate. Catalytic sites are often composed of clefts or pockets and commonly exist at subunit or domain interfaces [32]. They usually have more hydrophobic exposed surfaces than an average protein surface; this can lead to the displacement of water being a driving force in ligand binding [32–34]. One example of how these interactions assist with enzyme function is with enzyme aldose reductase. Aldose reductase contains a very hydrophobic pocket which provides a favourable interaction with linear alkyl chains (up to 3.2 kcal/mol) through hydrophobic bonding [35]. Additionally, there is a key hydrogen bonding interaction between the enzyme and the aldehyde hydroxyl group which contributes 4 kcal/mol to the enzyme:substrate binding energy [35]. Another example comes from the enzymes trypsin and chymotrypsin. Trypsin and chymotrypsin are very similar serine proteases; however, trypsin has a negatively charged aspartic acid in its binding pocket whereas chymotrypsin has a serine. This difference results in chymotrypsin cleaving after large hydrophobic amino acids whereas trypsin specifically cleaves after positive residues. It has been shown that replacing this amino acid in trypsin partially converts it to chymotrypsin activity [36]. By understanding enzyme functions and interactions these can be controlled for either pharmacological effect or *in vitro* use of the enzyme.

Protein–protein interfaces are held together by the same forces discussed previously in chapter 1.1.1, however, whereas enzyme:substrate interactions take place in a small highly focused cleft, protein–protein interfaces are large and flat. The distinct sites of interaction between two proteins are modified and defined by how the two surfaces align against each other. There have been a variety of computational approaches used to map and compare the landscapes of protein–protein interfaces to determine how they align [37–41]. There is some debate over which forces play the biggest role in these interfaces, with differing research favouring electrostatics [42–44], hydrophobic [45, 46], or hydrophilic partnership [47]. It is generally accepted that there exists residue ‘hot–spots’ that play a more important role in these interfaces than others [48–52].

The lipid bilayer of cells is one of the largest macromolecular structures found in nature and is essential to eukaryotic life. It forms through a strong hydrophobic interaction between the long carbon chains of phospholipids, combined with the strong hydrophilic interactions of the polar phosphate head group [55]. These two opposing forces cause alignment of adjacent phospholipids to maximise these competing conditions. On a small–scale, phospholipids will form micelles in aqueous media with a hydrophobic internal space, though with increasing lipid concentration larger structures begin to form, such as bilayers. Many proteins associated with the lipid membrane have specific amino acid orientations to maximise their positioning and interactions [54]. Proteins which interact with only one side of the membrane often have a hydrophobic domain or surface on one side of the protein which will self–associate with the lipid interior. Trans–membrane proteins flank a hydrophobic core with hydrophilic surfaces so as to become placed across the membrane [54]. The gramicidin ion channel is an example of a membrane bound protein that has been studied by molecular dynamics. It was shown that one of the ester groups of the phospholipid formed a specific hydrogen bond with Trp–15 of gramicidin [55]. Additionally, the choline head group of the lipid underwent electrostatic interactions with gramicidin which aligned the hydrophobic groups [55]. Transport across the lipid bilayer is important for many infectious diseases as well as delivery of drug molecules.

Protein aggregation is the root of many neuro–degenerative disorders including Parkinson’s and Alzheimer’s; it is also the disease mechanism of prion proteins. When proteins are produced, they are usually chaperoned to fold into their proper three–dimensional conformation. However, occasionally the protein is misfolded resulting in hydrophobic residues which are usually buried within the protein appearing on the protein surface [56, 57]. Normally these proteins are tagged for degradation, though under certain conditions these hydrophobic areas can combine with similar areas on other proteins. Intrinsically disordered proteins (those without a defined three–dimensional structure) and peptides such as α–synuclein and the amyloid–β peptides are especially susceptible to aggregation [58, 59].

Drug design is heavily influenced by non–covalent interactions with the Lipinski rule of five still being a starting point for small molecule ligands. The rule of five states that there should be no more than five hydrogen bond donors and ten bond acceptors, and details how hydrophobic the drug molecule should be through its octanol–water partition coefficient [60, 61]. Additions to this rule of thumb have stated that there should be fewer than three rotatable bonds which affect how these bonding sites can orientate. Drug molecules are often targeted at existing active sites, therefore structural data from X–ray crystallography and NMR spectroscopy can identify where likely non–covalent interactions are, allowing drugs to be intelligently designed to maximise interactions.

Acetylcholinesterase (AChE) is an enzyme which converts acetylcholine into choline and acetic acid in the synapse, and has been targeted by several small molecule drugs. The structure of the enzyme was determined by X–ray crystallography, which showed the enzyme had its catalytic site at the base of an aromatic cleft 20 Å deep [62]. The top of the cleft also contained an anionic site which would interact with the ammonium moiety of acetylcholine. A wide variety of different drugs have been designed to target this site, from medicinal compounds to insecticides and nerve agents. The ligand molecules take advantage of the non–covalent interactions within molecules; with a cationic group on one end to interact with the top of the anionic gorge, and residues to interact with the enzymatic site at the opposing end. A greater understanding of how drug molecules interact with proteins on a specific interaction level allows better maximisation of functional activity.

1.1.3. Techniques to analyse non–covalent complexes

Many analytical techniques are employed in the study of non–covalent interactions and complexes. Differing techniques offer various kinds and level of information which are often complementary. The following section covers the techniques: NMR spectroscopy, X–ray crystallography, isothermal calorimetry, spectroscopic approaches, and computational modelling.

Nuclear magnetic resonance spectroscopy (NMR spectroscopy) is a technique used to determine the structure of molecules utilising the relaxation of specifically excited isotopes in a magnetic field. The isotopes that undergo nuclear magnetic resonance have an intrinsic magnetic moment and angular momentum, giving them a nonzero spin. The chemical environment of the atomic nuclei affects their chemical shift, altering how they relax in the magnetic field and therefore giving different signals. Most standard techniques only observe atomic nuclei environments through covalent bonding; however as non–covalent bonds are formed the chemical shifts will also become altered. The magnitude and direction of the shift is determined by the degree of electron density that is added or subtracted to the molecule, and is especially visible on protons adjacent to the site of the non–covalent bond, or heteroatoms in an aromatic ring [63, 64]. Metal atoms have a strong effect on the adjacent chemical shifts which is especially useful in supramolecular chemistry to determine nearby nuclei [65, 66].

Protein NMR spectroscopy utilises multidimensional experiments to compensate for many thousands of overlapping 1 dimensional signals from nuclei in similar environments. For NMR spectroscopy of particularly large systems, enrichment for certain isotopes (^{13}C and ^{15}N) or the use of NMR active tags is required. A two–dimensional heteronuclear single quantum correlation experiment of ^{15}N (^{15}N–HSQC) is commonly used as a signal is provided for every amine in the protein backbone (with the exception of proline), along with signals from nitrogen–containing side chains [67].

An example of the use of NMR spectroscopy is the visualisation/analysis of interactions of protein kinase G with myosin light chain phosphatase. Through the use of chemical shift perturbation, certain hydrophobic and charged residues of one domain of myosin light chain phosphatase were shown to interact with protein kinase G [68].

The dipicolylamine complex in chapter 3 has been previously studied by ^{31}P NMR spectroscopy which was used to monitor the shifts on ATP molecules in the presence of the anion sensor. The phosphorus nuclei in ATP interacting with the zinc ions had a different chemical environment than free ATP and therefore had a different shift. This was used to determine which of the nuclei were interacting and therefore potential modes of binding could be determined.

X–ray crystallography looks at the specific diffraction of x–rays by molecular crystals. The angles and intensities of these diffracted beams can be used to generate a three–dimensional image of electron density (and therefore atomic positioning) within the crystals. Intra–molecular non–covalent interactions can often be uncovered by identifying which atoms are near each other in space and likely to form bonds. Co–crystallisation of proteins with ligands often uncovers the site of binding between the two molecules however, forming these co–crystals can be difficult [69–74].

MLI is a protein found in mistletoe, consisting of two protein chains linked by a disulphide. The protein was known to have a specific sugar binding site that triggers internalisation across cell membranes [75]. Through crystallisation of the protein with several sugar molecules one particular asparagine residue (when glycosylated) led to a stabilisation of the molecule. Through the study of the electron density changes with the various sugars, molecular interactions were attributed to to sugar affinity [75].

A structure of the peptides of c–MYC and MAX used in chapter 4 has been generated with X–ray crystallography in their ordered dimer state [76]. This highlights the drawback of the technique as the proteins are captured in a single conformation whereas mass spectrometry and ion mobility–mass spectrometry allows the analysis of the system in a dynamic fashion.

Isothermal calorimetry is a quantitative technique used to measure changes in enthalpy of a solution when binding occurs. A sample is held in a highly efficient thermal conductive jacket with a reference so that both are held at exactly the same temperature. As ligand is added to the sample any slight exothermic reaction is measured against the reference. It is highly useful in second stage drug development as it provides affinity and thermodynamic data upon ligand binding [77–79].

Spectroscopic methods can be fast and efficient ways of measuring ligand protein interaction through direct or indirect means [80]. There are many differing procedures: FTIR [81], UV–Vis absorbance [82, 83], fluorescence depolarisation [84–86], and FRET [87, 88] which all fall under the bracket of spectroscopy. All these techniques rely upon the optical properties of the host or guest and how these properties change upon binding of ligands or structural changes. These spectroscopic techniques have more effective uses on systems which have a strong chromophore such as heme [89] as they display a strong signal without the need of additional labels which may influence the interactions being measured.

The Bradford protein assay is a spectroscopic method for determining protein concentration which is used daily in many labs. The dye Coomassie brilliant blue G–250 undergoes a shift from 465 nm to 595 nm when interacting with hydrophobic areas of proteins through non–covalent interactions [90]. This shift is easily detectable and the degree of shift is related back to protein concentration.

Computational chemistry seeks to use properties of atomic and molecular structure such as atomic positioning, electron density, dipole moments, and bond vibrational frequencies to calculate energy minimised structures of molecules and favourable interactions. It has been shown to be useful in some non–covalent systems through determination of electron densities to uncover possible sites for non–covalent interactions, along with docking studies to determine how two structures can come together to form the lowest energy structure [91]. Different theoretical approaches have been shown to model different interactions with varying degrees of accuracy. The Hartree–Fock method, for instance, describes hydrogen bonding well; however, it does not take dispersion forces (including van der Waals) into consideration. A drawback of these modelling approaches is striking a balance between accuracy in the analysis *versus* the time and cost of computing [91].

Chapter 3 utilises computational techniques to produce possible structures of the anion sensors studied which are then compared to the experimental data. Additionally, the peptides of c–MYC and MAX in chapter 4 have previously been analysed with similar techniques.

1.1.4. Current challenges

Though the techniques outlined in the previous section have been used to great success to study non–covalent complexes they are not without their problems. Most of these techniques require a large amount of sample, which is often destroyed during the experiment (X–ray crystallography, ITC). This is especially problematic for proteins or chemicals which are difficult to express or synthesize. A further issue is a product of the nature of non–covalent complexes as a chemical species, in which they are often transient and short lived. This is especially problematic for intrinsically disordered systems which may form complexes on a dynamic time scale too rapid for some of these methods to detect. Some techniques measure an average signal which may not accurately represent the bound species separately (NMR spectroscopy). Computational docking models can be very useful and effective; however, the accuracy of the model is heavily affected by the size and chemistry of the system [92].

Some of the systems mentioned (ITC, spectroscopic methods) are very sensitive to when a ligand binds, however they don't provide any structural data on the location of binding, inversely the highly structural techniques are not dynamic and are unable to observe active systems. As mentioned previously, X–ray crystallography though providing high resolution images, is limited by the ability to produce high resolution crystals. When crystals are formed they only capture a single solid phase structure which may vary from solution phase, and may not show interactions such as aggregations. It has been suggested that some of these traditional techniques have presented a bias against proteins that exist in multiple conformations [91].

1.2. Use of mass spectrometry for biomolecules

The following section outlines how mass spectrometry has been utilised in the study and characterisation of biomolecules. The use of mass spectrometry for systems with specific non–covalent character is covered in section 1.3. The use of mass spectrometry for these systems has grown greatly since the development of soft ionisation techniques, such as electrospray ionisation (ESI) and matrix–assisted laser desorption/ionisation (MALDI) that allow the transmission of these large fragile molecules into the gas–phase. With the implementation of these new soft techniques, MS approaches began to be utilised to study a variety of previously inaccessible systems. With the widespread implementation of genomics, mass spectrometry took on a further role bridging the gap between known genetic sequences and biological proteomes.

1.2.1. Proteins

The amino acid sequence of proteins is at the centre of biological sciences, and was commonly determined through the use of Edman degradation [93]. Chain fragmentation mass spectrometry briefly enjoyed a period of use for protein sequencing, however with the widespread rise of genomics the role of mass spectrometry changed. Due to the highly reproducible fragmentation nature of proteins, combined with the high accuracy and throughput of mass spectrometry, biological samples could be rapidly identified if the gene was known. This led to the birth of the field of proteomics, which deviated from the field of single protein analysis. Proteomics deals with the identification of proteins and peptides leading up to total proteome analysis, whereas single protein analysis deals more with protein structure, kinetics and binding.

Before discussing further protein structure experiments, a note should be made upon the utlisiation of buffers and denaturing solvents during MS experiments. Following on from the development of soft ionisation methods, proteins could now be transmitted into the gas–phase without fragmentation. However, common mass spectrometry buffers used mainly highly polar solvents such as methanol and acetonitrile which are known to denature proteins. These biomolecules were therefore transmitted into the gas–phase intact but not in their biologically folded conformation. Other techniques kept samples in their correct structure through the addition of high salt concentrations to mimic the conditions found in cells. However, due to the nature of mass spectrometry, these high salt conditions led to a large degree of non–specific aggregation from cationic metals such as sodium and potassium making experiments difficult. Work was done to replace these solutions with volatile salts that will keep proteins in ‘native–like’ conformations, whilst being volatile enough to be lost during ionisation [94–97]. This led to the use of volatile ammonium salts such as acetate, and bicarbonate.

Top–down proteomics is a technique that bridges the gap between bottom–up proteomics and single protein analysis. In a top–down experiment the protein is introduced into the mass spectrometer intact and then subsequently fragmented rather than being proteolytically digested pre–analysis (bottom–up). Top down experiments are highly effective at localisation of post translational modification, as well as providing structural data [98–101].

Top–down experiments and single protein experiments can provide a wealth of structural information through both native and denaturing experiments. Studies have shown that the gas–phase structures of proteins are strongly related to those in solution provided they are ionised from 'native–like' solution conditions [102–104]. The number of charges a protein accumulates during desolvation is dependent upon the number of ionisable sites upon the surface of the protein in droplets formed during electrospray ionisation. Spectra acquired from 'native–like' conditions are in a folded state and therefore present fewer ionisable residues. Spectra acquired generally present narrow charge state distributions with relatively little charge. Spectra acquired from solutions with a high concentration of organic solvent or low pH are expected to be unfolded and present broad charge state distributions and higher charges, as more ionisable sites are now available in this denatured state [105].

One protein system which has been interrogated by top down proteomics by several groups is histone proteins and their modifications. Histones are extensively modified through acetylation and methylation, and separation of the various forms can prove challenging [106]. Top–down MS allows the individual protein isoforms to be interrogated separately, providing more information on the location of these post translational modifications.

Chapter 2 applies top–down proteomics strategies to the DNA aptamer Ky2 through fragmentation of the bound complex in the gas–phase to determine the nature and interface of binding.

1.2.1.1. Disordered systems and regions

A note should be made on protein and peptide systems that have no defined structure, or have regions of disorder. These intrinsically disordered proteins (IDP's) are difficult to study with many of the traditional techniques outlined due to their lack of structure on most experimental time scales. Early work believed these regions to be rare, but they have now been found to be not only abundant but highly biologically relevant across cell signalling, transcription and transduction [107].

Though the exact nature of what makes a protein disordered is not entirely known, low mean hydrophobicity, along with high net charge and enrichment in polar and charged amino acids are common features of disordered systems [105]. Mass spectrometry has proved highly useful in analysis of IDP's and their structure due to its fast time scales and low sample requirements [105].

Casein is an example of one of the early protein families found to have disordered regions. Experiments in the 1950's showed that the protein isoforms of casein were highly resistant to thermal or chemical denaturation along with being highly sensitive to protease digestion [107]. Additionally, optical studies showed casein structurally resembled denatured proteins [107].

The proteins c–MYC and MAX are examples of proteins with disordered regions and are studied in chapters 4 and 5.

1.2.2. Ion mobility–mass spectrometry

Ion mobility–mass spectrometry (IM–MS) is a hybrid technique that allows increased information to be determined than during a standard MS experiment. The details of instrument methodology will be discussed later; this section will discuss applications. Ion mobility spectrometry is a gas–phase separation technique; ions travel at different speeds due to their mobility (related to the ions rotationally averaged collisional cross section and charge) in a buffer gas. It is used extensively for rapid detection of volatile small molecules such as explosives and narcotics. It was coupled with mass spectrometry in the early 1960s [109], which led to a variety of MS systems culminating in the publication of the text 'Transport Properties of Ions in Gases' [110].

Due to the relation of ion mobility to collisional cross section (CCS) IM–MS has found extensive use in studying biomolecules, as experimentally derived data can be compared to models generated from other techniques (NMR spectroscopy, X–ray crystalisation, and computational modelling [111–113]). Clemmer *et al.* showed the conformations of a protein using IM–MS to separate out folded from unfolded species [114]. Since this experiment, a large number of proteins and complexes have been studied through IM–MS, some species such as ubiquitin extensively so [115–117].

The field of proteomics has benefited from IM–MS through the extra dimension of separation that can be gathered by utilising mobility. Often, tryptic peptides from complex systems occur at coincident *m/z* making analysis complicated. IM–MS can then separate these peaks clearly [109, 110]. IM–MS also allows trends of charge–state lines to be used to identify related peptides and isomers [118–120].

IM–MS has also been applied to a variety of non–biological systems such as silicon ion clusters [121], transition metals [122] and carbon clusters [123], as well as extended organic molecules such as fullerenes and nanotubes [124, 125].

1.3. Use of mass spectrometry and ion mobility–mass spectrometry to study non–covalent systems

The types of non–covalent bonds, their importance in biological systems and a variety of techniques that exist to study them have so far been discussed. This section will now specifically look at those complexes studied by mass spectrometry and how the technique adds to the field.

1.3.1. Advantages of mass spectrometry

Mass spectrometry has a variety of advantages over other analytical techniques. Sample quantity is a key area of advantage *versus* other methods with only nanograms of analyte being required. The sensitivities of MS techniques are especially useful for non–covalent complexes as these often are not the most abundant species observed. High sensitivity and dynamic range allows low abundance species to be identified and then isolated. Specificity is a major asset to MS analysis as specific species can be viewed in unison to provide binding information as well as affinity.

Analysis of quaternary structure and stoichiometry is particularly effective, as MS does not measure an average value, rather each species is analysed separately. This can be very useful when species form a complex series such as dimers, trimers or higher multimers. The analysis of systems with multiple subunits has always been a strength of MS, and with advances in the field of surface induced dissociation, the various affinities and interactions of these interfaces can be studied in depth.

1.3.2. Proteins

Proteins engage in non–covalent interactions with other proteins as well as with ligands and co–factors. Examples of homo–dimers studied by MS come from early work on avidin and concanavalin A, which were shown to have tetrameric structures which supported data gathered from other techniques [126]. Importantly neither trimeric nor pentameric species were observed, showing that solution phase interactions were preserved rather than artificially generated or destroyed. One example of a large hetero–multimeric non–covalent complex which has been extensively studied through MS is the GroEL chaperone complex. Not only has the full multicomponent complex been transferred to the gas–phase [127], it has also been observed with substrates bound within the cavity, all with non–covalent interactions [128]. The analysis of large protein complexes using MS has been shown to provide better stoichiometric information than some other techniques. For example, the protein TRAP was believed to be undecameric until studied with MS, where it was revealed to exist as a dodecameric form and as a double–ring 24–mer in the presence of tryptophan [129]. Similarly, serum amyloid P component (SAP) was shown to assemble into 20/30–mers in the presence of dAMP whereas only penta– and decamers had been previously observed [130]. Recent studies using ion mobility, which have further investigated these SAP multimers, recommend caution when studying these structures, as some conformations may be generated as artefacts through the ionisation process [131].

Disruption of intermolecular interactions is the basis of many novel new drug targets. A recent study into the t–RNA modifying enzyme Tgt has shown that particular ‘hot–spots’ between homodimers contain key non–covalent interactions which when removed or disrupted prevent dimer formation [132].

Binding of small molecule drugs to proteins can be employed as a method of high throughput screening [133]. One example comes from Loo *et al.* studying a variety of Src SH2 peptide inhibitors [134]. The observed binding strengths of the protein:peptide distributions, were consistent with the observed solution–phase binding constants. This highlights how data acquired in the gas–phase is still relevant to solution phase interactions.

An interesting study by Zenobi *et al.* [135] outlines how the affinity of a weak non–covalent interaction was measured using mass spectrometry. NCoA–1 PAS–B binds to STAT6 weakly in cells, and is involved in recruiting transcription factors. The affinity was measured by the displacement of a known standard peptide by an inhibitor to act as a reference. Through this method the direct weak interaction was not measured, rather its effect on an interaction which would have been retained more strongly through desolvation. This further study highlights how even weak interactions can be quantified in the gas–phase using mass spectrometry.

1.3.3. Host–guest systems

Whereas aggregates and adducts are examples of non–specific interactions, host–guest systems have a highly organised, often preformed, binding pocket or site where a reliable and repeatable interaction occurs. Good examples of this which have been studied extensively are heme binding proteins such as myoglobin [136, 137] and cytochrome–C [138]. It has been shown that the protein complex can be ionised, retaining the specific interaction with the heme group. Another example of MS analysis involved determining the stoichiometry of the Ca^{2+} binding protein calbindin D_{28K} [139]. Though calbindin D_{28K} had six calcium binding domains it had been suggested that between three and six ions were bound to the protein. Additionally, several studies could not provide an exact integral value for number of ions bound [140–142]. MS observed a mass increase in the presence of high excess of calcium acetate equal to four Ca^{2+} ions confirming the stoichiometry. Matrilysin is a similar example of a metal binding protein, which binds both zinc and calcium ions [143]. The stochiometry and ratio of metal ion binding couldn't be distinguished *via* other techniques. However, it was shown to bind two of each of the ions as the relative mass shifts were easily distinguishable via MS.

1.3.4. DNA

Oligonucleotides binding to proteins are an area of interest for looking at gene expression and transcription. These complexes have high levels of electrostatic interactions which are strengthened in the gas–phase leading to very stable complexes [144–146].

DNA is the target of many small molecule drugs. Though many of these involve permanent covalent bond formation with DNA such as cisplatin [147], some of these drugs function by intercalating between base pairs in a non–covalent manner [148], or bind to the major or minor grooves [149]. Mass spectrometry is an excellent technique for measuring structural changes upon binding of DNA. For example, a recent study followed the dimerization of the retinoic acid receptor upon DNA binding, following all species simultaneously as the reaction progressed [150].

Inter– and intra–molecular structures of G–quadruplexes are an interesting system held together by non–covalent interactions which can be studied *via* mass spectrometry. Guanine rich areas of DNA or RNA bind together with Hoogsteen base–pairing with a cation centralised to offset the degree of negative charge. These interesting structures have become research targets due to their potential involvement in telomerase inhibition along with gene expression. Various studies have invoked MS to study these systems including: formation [151], drug stabilisation [152], and protein interaction [153, 154]. They have also been used by Gabelica *et al.* as measures of instrument conditions drawing attention to how various non–covalent systems behave under different source layouts [155].

1.3.5. Supramolecular complexes

Supramolecular complexes have been studied by MS to a lesser degree than equivalent proteins. MS and particularly IM–MS have however been used to great effect in several studies. One example from Schalley *et al.* [156, 157] describes the use of CID fragmentation (in a mass spectrometry experiment) to show the structural evidence for free movement of the ring of a rotaxane. Similar strategies were used to show the structural formation of catenanes, and trefoil knots [158].

Mass spectrometry can be useful for determining encapsulation events in supramolecular structures. Pyrogallarenes without a host cation form nonspecific clusters. When small guests are added the signal shifts to a purely dimer orientation, and large guests produce a spread from trimers to octamers. When a guest of the correct size is added, however, a cube like hexamer structure is generated which is supported by crystal structures [159–163].

Highly symmetrical molecules such as Stang–type squares, composed of four metal complexes linked (historically) by 4, 4’–Dipyridyl ligands, represent challenging molecules to study by classical analytical techniques. After synthesis, complete squares and half structures are routinely observed, which often have similar NMR signals [164, 65]. Under MS conditions these half structures can also overlap by having the same mass to charge ratio as a full structure of double charge. For some species, these can be determined by the spacing of ^{13}C isotopes to determine the charge state, and if species overlay the isotope pattern can be shifted at alternate spacings. However, for highly charged species or those containing elements with more complicated isotope structures (such as metal ions) separating these patterns can be difficult. Isolation experiments along with ion mobility can easily separate out these multiple structures [165]. These types of structures are also examples of N–pyridyl ligands bound to metal. Lusby *et al.* used pyridyl ligands and palladium ions to form self–assembling supramolecular complexes [165]. These compounds had been difficult to fully characterise *via* NMR due to the similar chemical environments of the various incomplete structures [165]. Metal binding pyridyl ligands are also used in chapter 3 when binding to zinc ions through dipicolylamine groups.

1.3.6. Drawbacks

The major drawback of studying non–covalent complexes through MS is transferring the intact complexes from the solution to the gas–phase. Due to the energies involved with the ionisation and desolvation process and the low affinities of many of these interactions, not every solution–phase complex will be seen at the same intensity or at all. However, in some protein systems, despite the complex not being visible, structural changes that were induced can still be retained and observed [166]. Another problem is that once complexes enter into the gas–phase, hydrophobic interactions will be weakened and electrostatic ones strengthened, leading to possible structural changes [167].

One example where solution and gas–phase results differ heavily is with the binding preferences of acyl–CoA with differing lengths of fatty acid, to the acyl–CoA binding protein. In solution, different interaction potentials were observed across chain lengths of zero to sixteen carbons. The greater carbon chains had less favourable interactions and therefore less bound complexes will be present. However, in the gas–phase, the amounts of the various complexes remained the same despite changing chain length [168]. Therefore, the carbon chain had no effect on the degree of overall binding. Acyl–CoA consists of a highly polar group containing several phosphates connected to a fatty acid chain of varying length. The acyl–CoA binding protein has several conserved residues to interact with these phosphates. Upon binding in solution, the hydrophobic fatty acid is exposed, resulting in an increasingly unfavourable interaction as chain length increases. However, in the gas–phase (which is inherently hydrophobic) the exposure of this hydrophobic region is no longer unfavourable resulting in similar binding strengths for each of the ligands.

A common question asked of mass spectrometry structural biology methods is “How relevant is a gas–phase experiment for measuring an interaction in solution?” These types of experiment involve the removal of the solvent shell around a protein and exposure to strong vacuum conditions, and therefore the question arises: are the structures observed the same as in solution or are they an artefact of the gas–phase? The first case to be made for mass spectrometry is determining protein stoichiometry. Mass spectrometry is an excellent tool for determining the stoichiometry of complexes and in most cases the results match with solution. Particularly high protein concentrations do show nonspecific protein aggregation; however, this can usually be avoided through decreasing the analyte concentration [169].

Another argument for studying structural biology with mass spectrometry is the use of ‘soft landing’ experiments. In these experiments, the mass spectrometer is used as both an analytical and preparative tool whereby ions that have travelled through the mass spectrometer are collected and studied using other techniques. This allows the separation of species that would be difficult to distinguish by other means. Robinson *et al.* used this technique to study the cage protein ferritin using the mass spectrometer to separate the bound and unbound forms [170]. The collected complexes were then analysed through transmission electron microscopy which revealed that characteristic structural features had been retained.

1.4. Mass spectrometry instrumentation

The mass spectrometer has seen vast improvements in speed, accuracy, precision and sensitivity since its first conception. Its roots can be traced back to 1886 when Eugen Goldstein first observed rays in gas discharges within an electric field. Wilhelm Wien built on Goldstein’s work using electric and magnetic fields to separate these positive rays by their mass to charge ratio. JJ Thomson furthered this work to produce the mass spectrograph [171, 172]. These early instruments were able to separate ions and isotopes but as the technology developed the range of samples increased and now covers a broad range of potential samples.

The following sections will now discuss the development various parts of the MS instrumentation and how they function.

1.4.1. Ionisation

The first developed ionisation technique was electron ionisation (originally referred to as electron impact or EI). During EI, an electric current is passed through a wire filament which is then heated leading to thermal emission of electrons. These electrons are accelerated and concentrated into a perpendicular beam. Impact of these high energy electrons with the introduced volatile sample then leads to production of a radical cation along with high degrees of fragmentation. These fragments could be exploited for discerning structural features of smaller molecules, but often the entire precursor was dissociated and spectra were complex. The extensive fragmentation of the precursor caused by EI led to the development of softer methods, such as chemical ionisation (CI) [173]. In CI, analyte vapor is introduced into chamber with a vast excess of a reagent gas (such as methane). Electrons from a heated filament cause the gas to form a plasma of radical cations similar to EI. Secondary gas–phase reactions with neutral reagent gas occur, forming a range of reactive gas–phase cations. These cations then react with the analyte by a number of mechanisms (*e.g.* proton transfer, charge exchange and adduct formation), forming analyte gas–phase ions. Typically, less fragmentation was observed in CI compared to EI and some of the precursor was often retained making molecular weight determination easier. Though softer than EI, CI still led to fragmentation and was not applicable for all analytes.

Fast atom bombardment (FAB) was a technique, developed later, that led to less fragmentation compared to EI and CI [174]. FAB uses a beam of high energy atoms aimed at a surface to generate ions. The sample to be analysed is usually suspended in a matrix, and the atoms are commonly inert gases such as argon and xenon. Field desorption (FD) ionisation was also developed in which a high potential filament is used to ionise gaseous analytes. This technique also produces little fragmentation and the major charged species are molecular radical cations [175]. Techniques such as electrospray ionisation (ESI) (given the frequency of use of this ionisation technique within this thesis, further detail is provided in the following sections: 1.4.1.1 and 1.4.1.2.) and matrix–assisted laser desorption/ionisation (MALDI) [176] cause very little to no fragmentation of samples and are often referred to as ‘soft’ ionisation techniques. MALDI uses a laser to ionise a sample suspended in an organic matrix. The matrix absorbs the light much more readily than the sample and the ionisation is transferred to the analyte. There are no set rules for matrix choices, however the common design includes: low molecular weight (for easy vaporisation), acidic (to act as proton source during positive ionisation), strong optical absorption in UV or IR, and containing a chromophore [176]. MALDI generates mainly singly charged ions and is therefore commonly combined with time–of–flight detection systems. The pulsed nature of MALDI also lends itself to the pushes in time–of–flight detectors.

The development of these ‘soft’ ionisation strategies allowed MS to be used for previously inaccessible samples. Now fragile biological and non–covalently bonded samples could be transferred into the gas–phase without fragmentation.

1.4.1.1. Electrospray ionisation

The first paper on the behaviour of electrospray can be traced to John Zeleny in 1914 who published work describing the behaviour of droplets leaving glass capillaries [177]. The use of ESI for MS was published in 1968, and the Nobel Prize was awarded to John Fenn in 2002 for his work advancing and developing the field [178].

The operation of ESI occurs as following, and is shown graphically in Figure 1.3. During electrospray ionisation, a potential of several kV is applied to the volatile analyte solution leading to the production of a Taylor cone and charged droplets containing the analyte [179]. The droplets travel along an electric gradient whilst being reduced in size *via* a drying gas. As the droplet decreases in size it approaches the Rayleigh limit; the maximum size a droplet of a certain charge can access [180]. At the Rayleigh limit the droplet becomes deformed due to the repulsion of like charges. At a certain point, the repulsion becomes stronger than the surface tension of the droplet and it undergoes Coulombic fission resulting in a dispersion of smaller droplets. There are several theories for how the final stages of desolvation occur resulting in bare ions. The three major theories are the ion ejection model (IEM), the charge residue model (CRM), and the chain ejection model (CEM) [181, 182]. None of the models are believed to be exclusively correct, rather differing molecules will ionise through different or multiple processes.

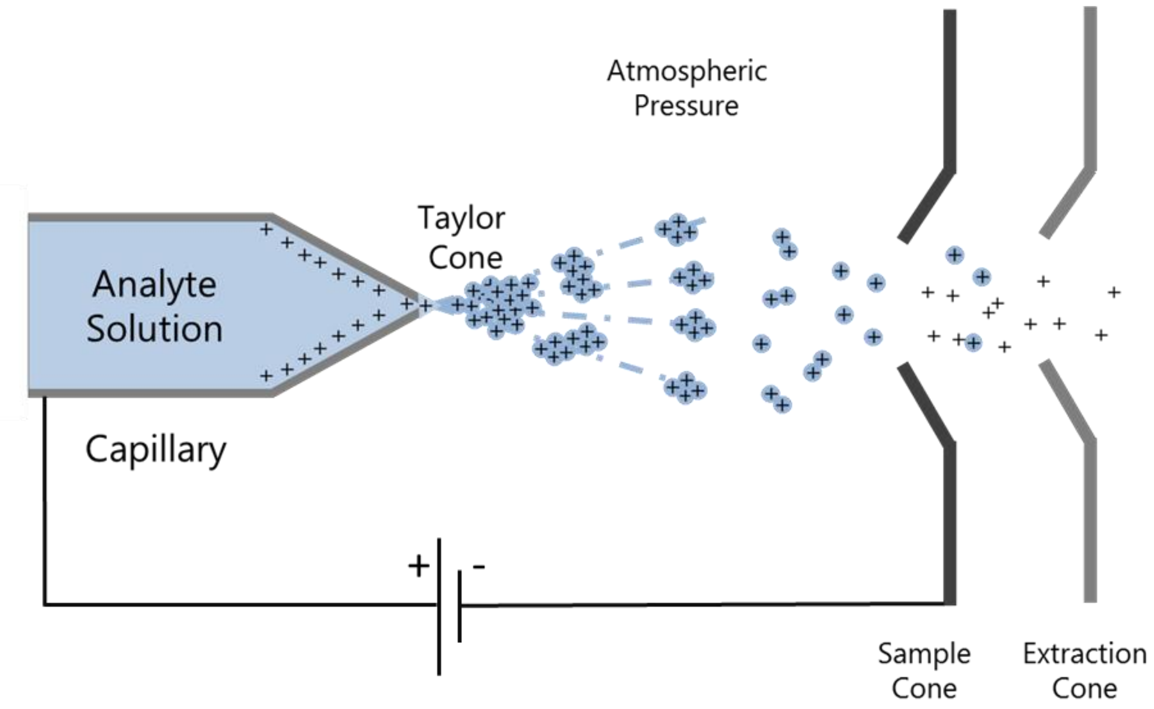


Figure 1.3: Schematic representation of electrospray ionisation. Features include Taylor cone formation, and solvent removal from ions through various vacuum stages.

The IEM, proposed by Iribane and Thomson [183], is believed to be favoured for low molecular weight ions whereby the electric field in a droplet at the Rayleigh limit is suitably high to cause the ejection of these small solvated ions from the surface. The CRM, originally described by Dole *et al.* [184], states Rayeligh limited droplets which contain a single analyte will lose solvent leading to a bare ion. The charge present on the surface of the droplet is transferred to the analyte when the final solvation shell is lost. This model is believed to be favoured by globular species such as natively folded proteins [185, 186]. The final model, CEM, has been postulated for disordered or unfolded protein systems along with polymers, whereby due to the exposed hydrophobic surfaces, the interior of the ESI droplet is energetically unfavourable. This leads to the analyte migrating to the surface of the droplet and a tail being ejected from the hydration shell. As the tail is sequentially ejected a large degree of charge is transferred to the analyte.

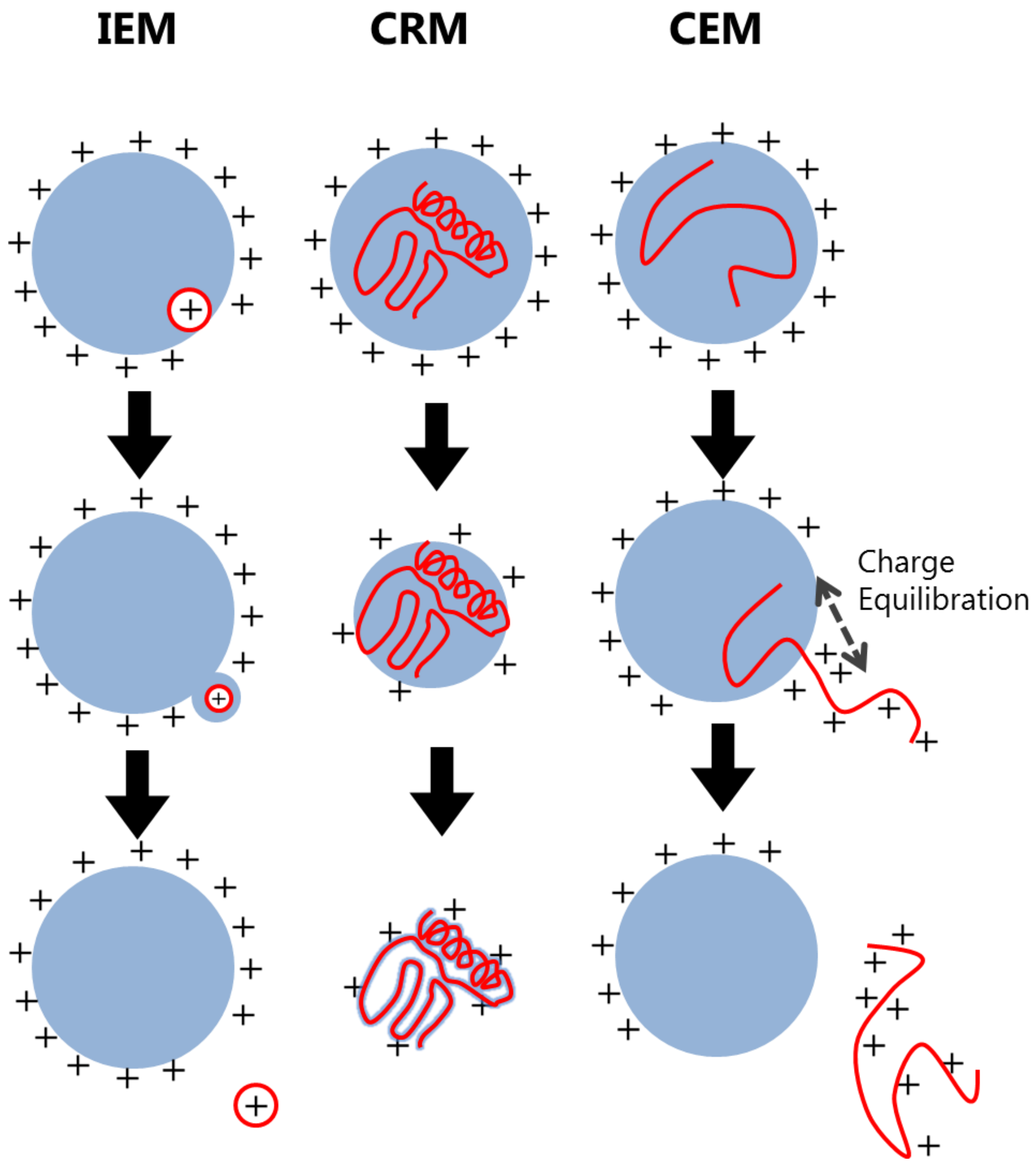


Figure 1.4: Showing various mechanisms in the ESI process: the ion ejection model (IEM), charge residue model (CRM), and the chain ejection model (CEM). Modified from reference [182].

During positive electrospray ionisation, the analyte commonly becomes multiply charged through the addition of several protons in the form $[M + xH]^{+x}$. As mass spectrometry measures the mass to charge ratio of ions rather than the molecular weight, this means species can be ionised in multiple charge states. Large species such as proteins can present in a charge envelope of varying charges. This charge state envelope presents in a normal distribution, with the majority of the species in the middle of the distribution and fewer populated states towards higher and lower charges. Analytes can vary in charge state distribution depending upon the buffers in which they are ionised from. Proteins analysed under highly organic conditions become unfolded and can therefore expose a high number of ionisable residues, presenting broader charge state distributions at lower *m/z,* due to a high number of charges. When the same species are analysed from conditions designed to buffer folded conformations (such as ammonium acetate) they present fewer ionisable residues and therefore appear in much narrower conformations with fewer charges at higher *m/z*. Proteins with disordered regions or multiple possible conformations can present in two charge state envelopes, as the different conformations accrue a different number of protons.

1.4.1.2. Nano–electrospray ionisation

The development of ESI has led to the development of a low flow version of ESI termed nano–ESI (n–ESI). During n–ESI samples of as little as 10 µl are placed inside a glass capillary with an internal diameter of ~0.1 mm. The potential applied is less than standard ESI and is done through metal coating, or insertion of inert wire down the capillary to the solution. Due to the smaller capillary, the droplets formed are smaller for n–ESI resulting in a more stable, gentle spray, with reduced signal to noise ratio [187]. The lower sample volumes also allow for lower concentrations, which are particularly desirable for biological systems which may be difficult to express [188, 189].

1.4.2. Mass analysers

There are many kinds of mass analyser which are often combined to form hybrid instruments [190]. These hybrid instruments are often very flexible in their operation due to the combined strengths of their various operations. All mass analysers work on the principle of separating ions *via* their *m/z* to allow resolved detection.

1.4.2.1. Quadrupole

The quadrupole analyser consists of four metal rods in a parallel alignment (Figure 1.5). Each parallel pair is connected electronically and an RF voltage is applied with a DC voltage offset relative to the other pair of rods. This generates the oscillating electric field described by equation 1.1 [190].

$$\pm\boldsymbol{\phi_0} = \pm(\boldsymbol{U} - \boldsymbol{V}\ \boldsymbol{cos\omega t})$$

Equation 1. 1

Where ϕ_0 is the potential applied to the rods, *U* the DC voltage, *V* the zero–peak amplitude of the RF potential and *ω* it;s angular frequency with *t* as time.

The movement of the ion through the quadrupole is the mode of separation and is perturbed by the potentials (*U* and *V*) and the frequency (*ω)* employed. One of two things can occur at this point; either the ion collides with the rod and is neutralised, or the potential will switch before collision leading to it becoming attracted to a different rod. Ions will therefore travel in an oscillating trajectory which can be modelled with the help of the Mathieu differential equation (equation 1.2). The values of *a* and *q* determine whether the trajectory of the ion will be stable (allowing transmission) or unstable [191]. The RF amplitude (V) can be scanned to allow ions of different *m/z* to pass through the instrument or the amplitude can be fixed to allow only ions with a specific *m/z* value to pass through with a stable trajectory; all other ions will collide with the rods and be lost. This mode allows the quadrupole to act as a mass filter allowing a single species through for further analysis.

$$\frac{\boldsymbol{d^2Y}}{\boldsymbol{dX^2}} + [\boldsymbol{a} - \boldsymbol{2q}\ \boldsymbol{cos}\ (\boldsymbol{2x})]\boldsymbol{Y} = \boldsymbol{0}$$

Equation 1.2

Equation 1. 2

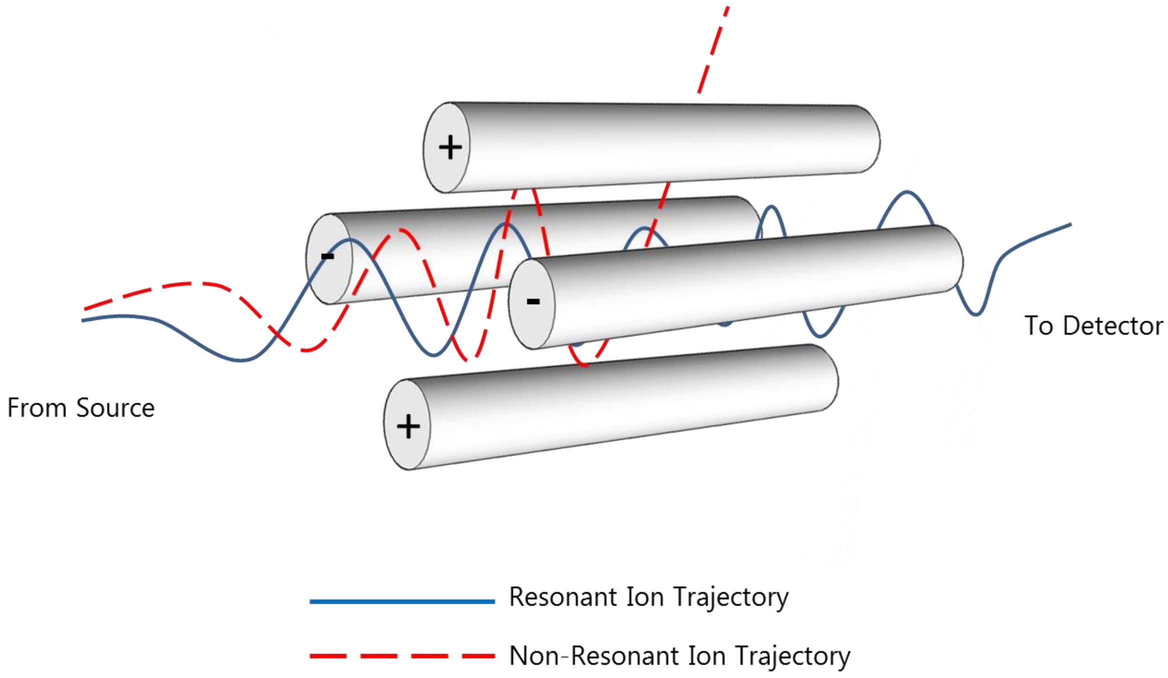


Figure 1.5: Schematic representation of quadrupole mass analyser. Ions resonant with the RF field travel through on a stable trajectory whereas those with an unstable trajectory collide with the rods.

1.4.2.2. Time–of–flight

Time–of–flight (ToF) is a method used to determine an ion's *m/z via* the time it spends travelling through a field free region of known distance (Figure 1.6). In the ToF tube, ions are pushed in a packet as they are all accelerated to the same kinetic energy. Ions of the same charge will have the same kinetic energy; therefore, their velocity will be dependant only upon their masses. Ions of a low mass will reach the detector earlier than ions with a higher mass for a specific charge. The movement of these ions can be described by the following equations:

$$E_K = \frac{mv^2}{2} = zeV$$

Equation 1. 3

$$t = \frac{d}{v}$$

Equation 1. 4

$$t^2 = \frac{m}{z} \left(\frac{d^2}{2eV}\right)$$

Equation 1. 5

Where E_K is the kinetic energy of the ion and *m* it's mass mass. *v* is the velocity of the ion upon entering the ToF region, *ze* the total charge of the ion, *V* the accelerator voltage, *t* is the time the ion takes to travel and *d* is the flight tube length.

Taken together, these equations state that the time taken for an ion to travel the drift region is given by equation 1.4 which when combined with equation 1.3 leads to the relationship between the time taken in the field free region and the *m/z* of the ion (equation 1.5).

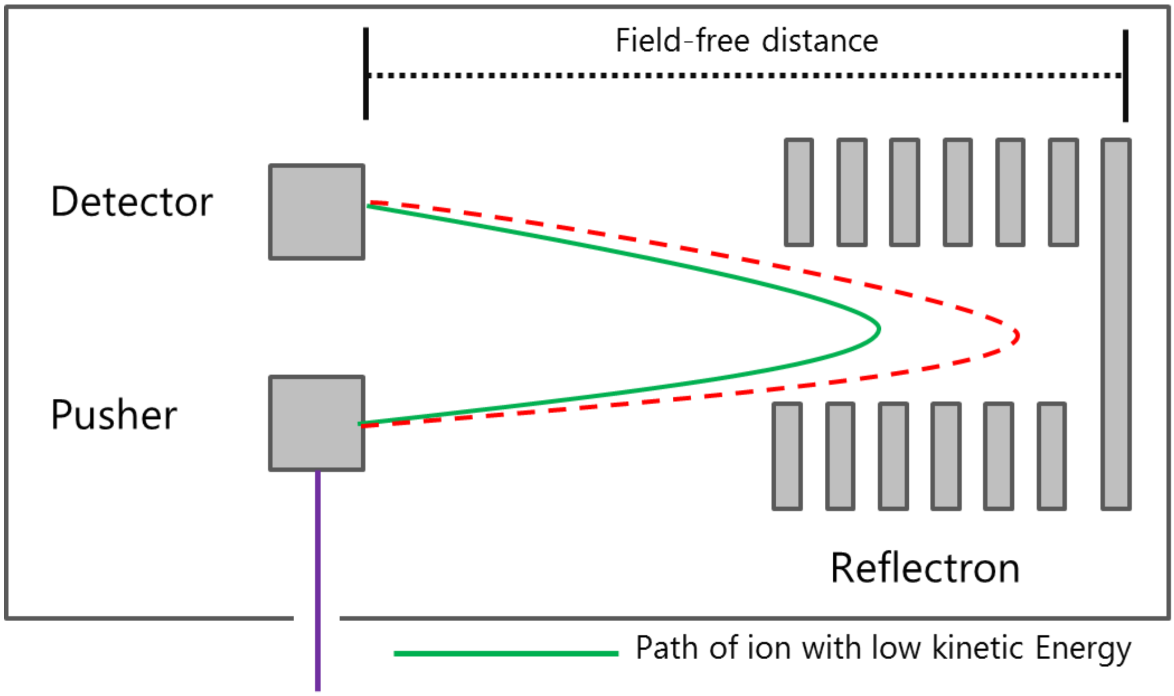


Figure 1.6: Schematic representation of time–of–flight mass analyser with reflectron. Ions with lower kinetic energy do not penetrate as deeply into the reflectron.

The power of the ToF instrument was greatly improved with the development of the reflectron by Mamyrin in 1973 [192]. The reflectron is a series of grids and ring electrodes which reflect ions back up the flight tube, with ions of differing kinetic energies penetrating different depths. Not only does the reflectron effectively double the drift length available in an instrument, it also corrects for slight changes in kinetic energy which led to a loss of resolution. Ions with lower kinetic energy of the same *m/z* will not penetrate the reflectron as deeply as those with higher kinetic energy and therefore spend less time within it. As a result, both fast (higher energy) and slow (lower energy) ions will reach the detector at the same time.

1.4.3. Microchannel plate detectors

Microchannel plates (MCPs) are a common detector during MS analysis after separation *via m/z.* MCP’s consist of parallel channels of semiconductors held within a plate. Ions impact upon the inner surface which is held at a potential of ~ –1 kV, this impact triggers the release of an electron cascade down the potential gradient of the channel with each secondary electron triggering a further release of electrons down the channel. Each microchannel is a continuous–dynode electron multiplier under the presence of a strong electric field (~ 2 kV). Amplification of the original ion impact is of the order of up to 10^4 with several plates being used in conjunction to generate up to a 10^8 increase [190]. The final burst of electrons leads to the generation of current which can be measured. MCP detectors are often used in ToF instrumentation as they generate rapid signals which lead to narrow pulses and precise arrival times.

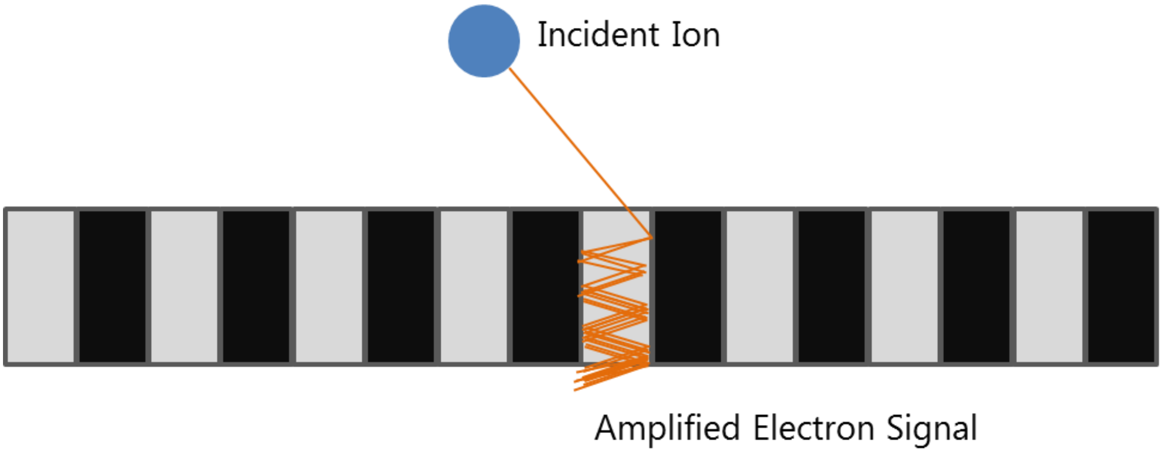


Figure 1.7: Schematic representation of the electron cascade that occurs in microchannel plates. The incident ion enters a channel triggering the release of an electron. This electron then triggers a cascade which is detected at the exit of the channel providing a signal in the mass spectrometer.

1.5. Mass spectrometry techniques

The field of mass spectrometry has been hybridised with a variety of gas–phase ion chemistries to add additional orthogonal structural information to *m/z* measurements. The two broad techniques discussed further are ion mobility and fragmentation methods.

1.5.1. Ion mobility–mass spectrometry

Ion mobility spectrometry (IMS) is a technique used to separate ions by their mobility in the gas–phase whilst travelling through a buffer gas. It has been commonly employed in the security industry for the detection of drugs and explosives. Due to both IMS and MS being gas–phase ion techniques, the coupling of the two was synergistic. McDaniel first coupled a low field mobility cell to a sector mass spectrometer [99], followed by IMS–ToF instrumentation by McAfee and Edelson [193]. Various other layouts were developed leading up to Mason and McDaniel publishing their seminal text 'Transport properties of ions in gases' in 1988 [110].

The first commercial ion mobility–mass spectrometry (IM–MS) instrument was produced by Waters Corporation in 2006. The Synapt is a Q–ToF instrument with a post quadrupole travelling wave separation cell [194]. Recently, Agilent Technologies have released a drift tube IM–MS Q–ToF instrument [195]. The release of these commercial instruments is bringing the field to the forefront of structural biology for separation and cross section analysis.

The nomenclature for ion mobility–mass spectrometry uses the term CCS for collisional cross sections with a preceding superscript for the type of ion mobility and a preceeding subscript for the gas used where appropriate *e.g.* ${}^{DT}CCS_{He}$.

1.5.1.1. *Drift tube*

Drift tube ion mobility spectrometry is the simplest of the IMS techniques and relies on direct application of ion movement from theory. Ions are injected into a drift cell filled with an inert buffer gas. Helium is the preferred buffer gas for these systems; however commercial instruments often used molecular nitrogen as higher voltages can be used without electrical gas breakdown. Carbon dioxide is also used if high separation is required, although the use of this gas means that cross sections cannot be modelled [196, 197]. The drift cell has an electric field gradient which draws ions through, whilst they are being impeded by collisions with the gas. These collisions are at a much lower energy than during collision induced dissociation and therefore do not result in fragmentation. The speed at which the ions travel through the cell is due to their ion mobility, which is directly proportional to the rotationally averaged collisional cross sections (CCS, Ω, $Å^2$) along with their charge. Larger ions will have a larger CCS and will therefore have more collisions with the buffer gas, impeding the ion's movement, and will therefore arrive later. Higher charged species will be more affected by the electric field and therefore arrive quicker. In its simplest functions, IM–MS can separate dimers which appear at coincident *m/z* values. The process of DT–IM–MS is outlined in Figure 1.8.

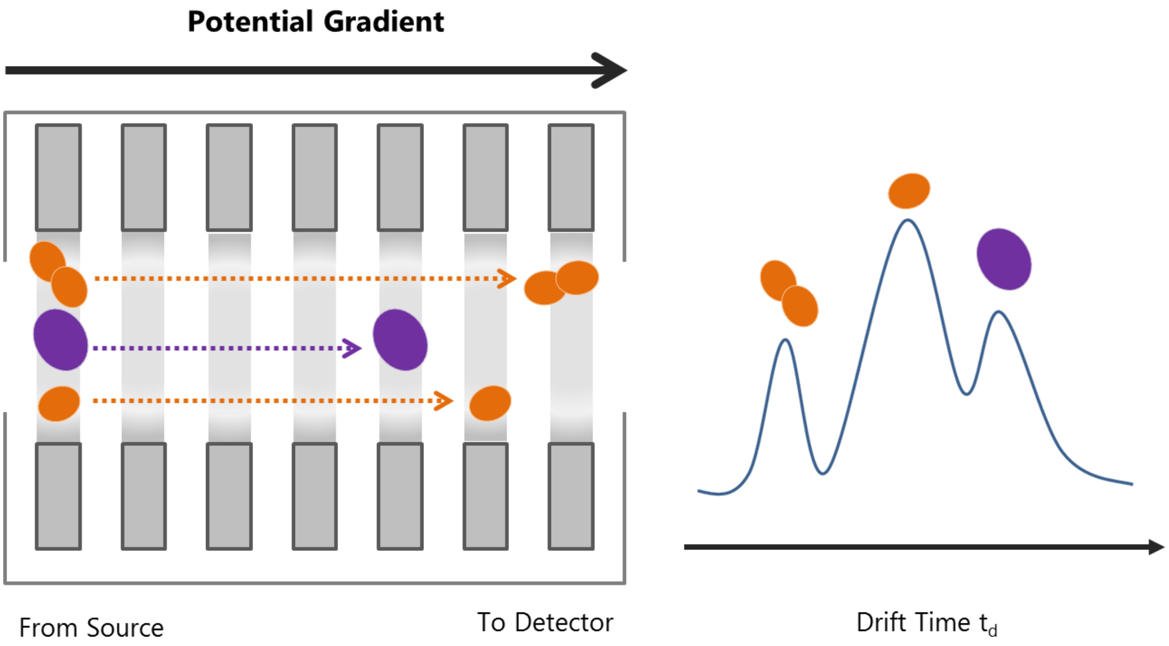


Figure 1.8: Schematic representation of a drift tube ion mobility cell with the resulting arrival time distribution. The orange circle represents an ion of a certain charge, whilst the double circle is a dimer of this ion. The purple represents a different larger ion. The higher charged dimer arrives first as ion mobility is affected more by charge than size. Next to arrive is the orange monomer as it represents a smaller cross section than the purple monomer.

The mobility of an ion under a weak linear electric field can be derived as its velocity in travelling a drift cell of known length over time [110]:

$$\frac{L}{t_d} = v_d = KE$$

Equation 1. 6

Where *L* is the cell drift length, t_d is the drift time time, v_d is velocity of the ion, *K* is the mobility and *E* the electric field.

The electric field can be considered weak as long as the average ion energy acquired from it is small compared to the thermal energy of the buffer gas. The field is measured in Townsend (Td) and is a ratio of the field strength over buffer gas density *N* (number of molecules per unit volume). A weak field has values of *E/N* ≤ 4 Td.

To decouple the mobility of experimental results from variables such as temperature and pressure, these can be normalised to standards (760 Torr/273.5 K) to yield a reduced mobility.

$$\boldsymbol{K_0 = K\,\frac{T_0 P}{T P_0}}$$

Equation 1. 7

Experimental values can now be inserted into the Mason–Schamp equation to produce rotational averaged collisional cross sections (CCS, Ω, $Å^2$).

$$\boldsymbol{\Omega = \frac{3ze}{16N}\sqrt{\frac{2\pi}{\mu k_B T}}\,\frac{1}{K_0}}$$

Equation 1. 8

Where K_0 is the reduced mobility (measured mobility corrected to temperature and pressure) *z* is the ion charge state, e is the elementary charge state, *N* is the number gas density, *μ* is the reduced mass of the ion–pair, k_B is the Boltzmann constant and *T* is the buffer gas temperature.

1.5.1.2. Travelling wave

Travelling wave ion mobility–mass spectrometry (TWIM–MS) was developed at Waters Corporation for their commercial ion mobility–mass spectrometer to provide an extra dimension of separation with higher transmission than traditional DT–IM–MS. As in DT–IM–MS, ions are stored in an ion guide region before being injected into the ion mobility cell. The ion mobility cell uses nitrogen as its buffer gas and has a series of planar electrodes which apply an RF field to radially confine the ion beam. The travelling wave is then imposed on the ions through the superimposition of transient DC voltages on the RF field [194, 198]. These DC voltages travel through the cell in a series of waves interacting with the ions. Ions with high mobility are carried by the waves and are transmitted through the cell more quickly, whereas lower mobility ions are crested over the waves and take longer to travel. By altering the wave height and velocity the degree of separation can be tuned for individual systems. This feature is particularly useful to separate conformers with high structural similarity. Unfortunately, the CCS of ions cannot be directly measured through TWIM–MS unlike DT–IM–MS; however, it is proportional to various instrumental settings. Therefore, CCS can be determined by comparing arrival times to a set of appropriate standard calibrants with CCS predetermined through DT–IMS [199–201].

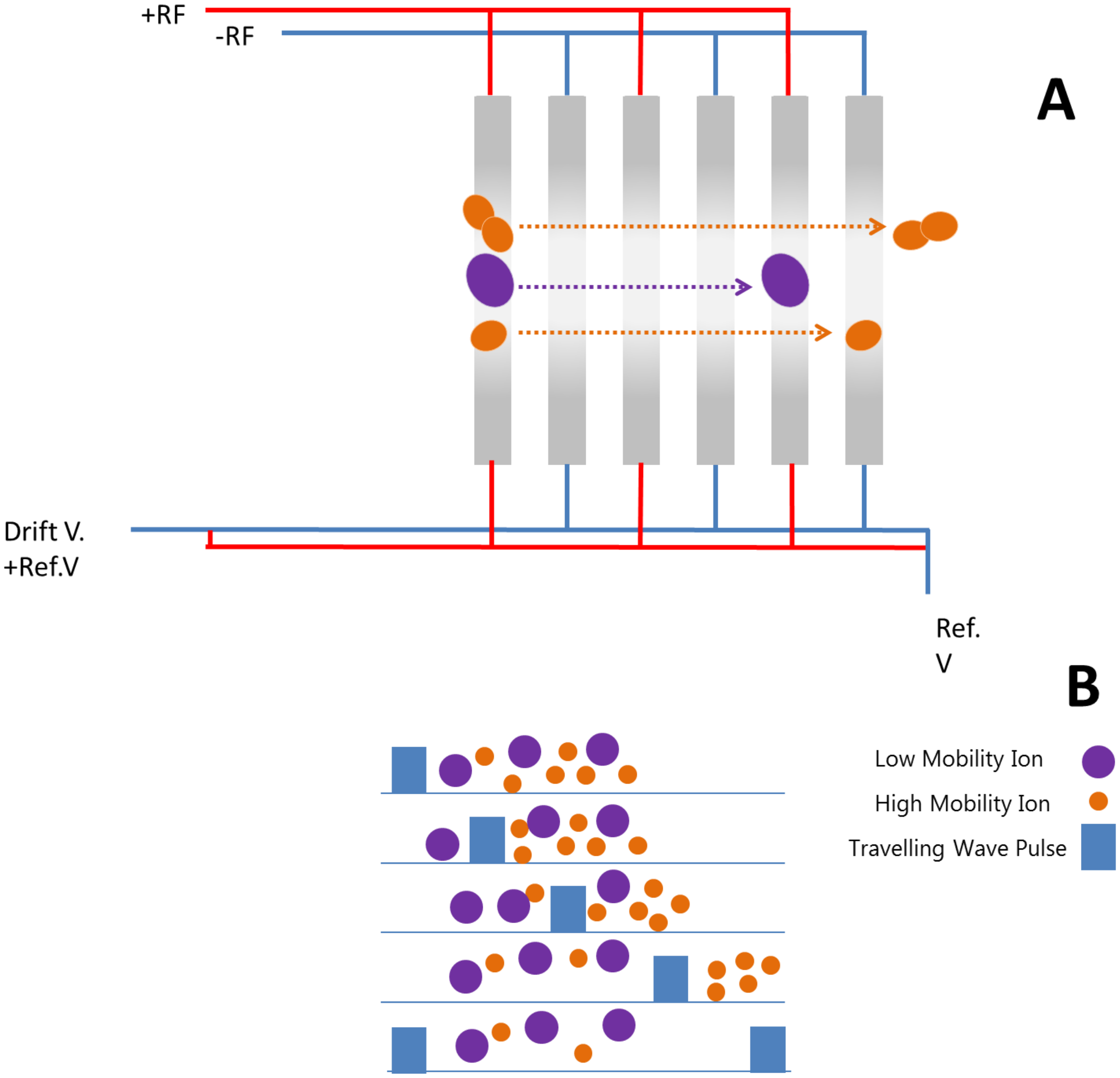


Figure 1.9: Schematic representation of travelling wave cell (A) and separation

of ions (B). Voltage is applied to alternate electrodes creating a "wave" of voltage. In B, ions of a low mobility are crested over the wave whereas high mobility ions are carried forwards. In A, standard separation ions are affected by multiple waves.

Though the introduction of commercial ion mobility instrumentation has allowed a greater degree of access to the field, care must be taken with the use of TWIM–MS as studies have shown a degree of heating introduced to the ions through separation which may induce conformational changes in proteins [202].

1.5.1.3. Field asymmetric ion mobility

Field asymmetric ion mobility (FAIMS) or differential mobility (DMS) represents a shift from DT–IM–MS and TWIM–MS, as separation is usually done at the start of the instrument, rather than post quadrupole, because separation is achieved at atmospheric pressure and temperature. FAIMS was first reported by Purves *et al.* who utilised the mobility of an ion between a pair of electrodes with a gas flow [203, 204]. During FAIMS the lower electrode is held at a constant ground potential whilst the upper electrode is switched between a short high voltage component followed by a longer low voltage of the opposite polarity. Ions are strongly attracted for a short time followed by a lesser repulsion for a long time till their eventual collision with one of the electrodes or transmission. To allow transmission a DC 'compensation voltage' (CV) is applied in relation to the high and low field mobility (Figure 1.10). This CV can be viewed like a quadrupole in that it can be scanned to allow transmission of many ions or fixed to only allow a single mobility; however, during scanning mode ions will be separated. Unfortunately, CCS cannot be determined by FAIMS as ion mobility is not proportional to the field and is instead dependant on the voltages applied. Also, a high degree of heating is applied to the ion making it unsuitable to study molecular conformation. The heating of ions results in denaturation of the molecular folding; therefore, the transmitted conformation is not the same as that in solution.

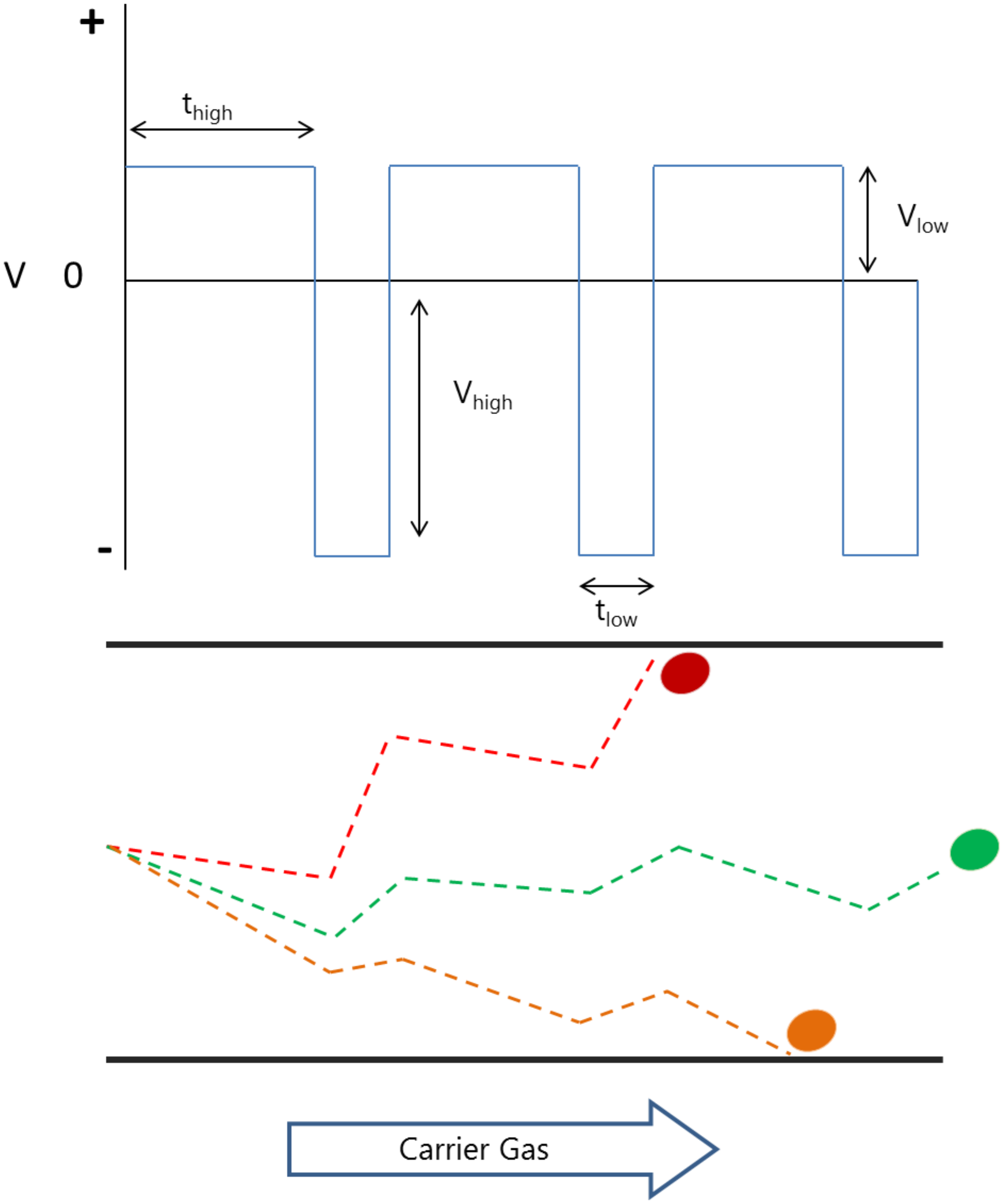


Figure 1.10: Schematic representation of the voltage changes during a FAIMS experiment and their effects on ion trajectory. By altering the compensation voltage only ions of a single mobility will be transmitted.

1.5.2. Fragmentation

Fragmentation is a highly useful gas–phase technique to determine molecular structure or sequence. As ions fragment in specific pathways the production of these product ions can be predicted and used to confirm identities. There are a variety of techniques that can be used for fragmentation ranging from physical, to electronic, to radiation induced.

1.5.2.1. Collision–induced dissociation

Collison–induced dissociation (CID) also known as collisionally activated dissociation (CAD) is the most commonly available fragmentation technique and is widely available across a range of instrumentation. Ions are accelerated into a collision cell and then collide with neutral gas molecules (often nitrogen or argon) [190]. With a series of collisions, the internal energy of the ion is increased, eventually resulting in the breaking of an internal bond. CID usually results in the breaking of the weakest bond first, and is therefore poor for studying non–covalent complexes and post translational modifications. Additionally, as the fragmentation is the result of a series of low energy collisions rather than a single high energy one, there is often molecular unfolding or heating. Each small collision results in the transmission of energy from the fragmentation gas to the ion; as the energy of the ion increases (below the threshold that fragmentation occurs), bonds may rotate, resulting in different conformations than would exist in solution. This molecular unfolding can be utilised through collision induced unfolding (CIU) experiments to observe internal molecular stability [205].

A standard nomenclature is used to detail the fragmentation of proteins and peptides based upon where the fragmentation occurs. CID produces mainly fragmentation between the carbon and nitrogen atoms of the peptide bond. These are referred to as b (N terminal) and y (C terminal) type ions (Figure 1.11) [206].

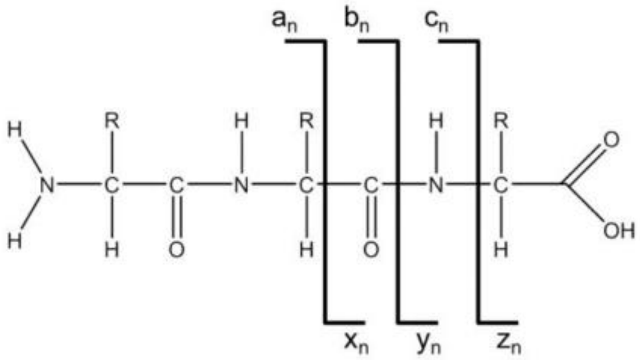


Figure 1.11: Nomenclature for protein and peptide fragmentation [206].

1.5.2.2. Laser induced fragmentation

Laser fragmentation methods involve the excitement of analyte molecules by photons. These fragmentation methods commonly require ions to be trapped during fragmentation and have mainly been utilised on FT and ion trap instruments. Infrared (IR) lasers have been utilised for IR multi–photon dissociation (IRMPD) for some time and more recently ultraviolet (UV) lasers for UV photo–dissociation. Use of IRMPD for proteins was first conducted in 1994 by McLafferty [207]. Due to the low energies of the IR lasers, multiple photon excitations are required until the analyte has passed to higher energy vibrational states leading to bond fragmentation. More recently UV lasers (200–300 nm) have been used which can led to single photon excitation [208–210].

1.5.2.3. Electronic dissociation

Electron capture dissociation (ECD) was developed as a low energy fragmentation option. It involves the interaction of the gas–phase cation with a free electron generating a radical ion which then fragments faster than typical bond vibration [211–213]. The fragments produced are different to those of CID (primarily c and z as shown in Figure 1.11 [214]), and the bonds broken are no longer necessarily those of lowest energy, meaning labile PTMs may remain covalently linked to their appropriate residue. Because of the instrumental requirements, ECD has only been applied in ICR and ion trap instruments. A method was developed to utilise the low energy fragmentation pathway through transfer of electrons through an intermediary molecule in electron transfer dissociation (ETD) [215, 216]. ETD generates radical anions separately to the analyte cation which are then allowed to interact in a controlled manner within the instrument. The radical anion transfers its additional electron to the analyte cation causing it to dissociate *via* pathways similar to those used in ECD. While ECD is only applicable to cationic species, the negative mode equivalent, electron detachment dissociation (EDD), can be used for anion fragmentation. EDD utilises higher energy electrons, leading to ejection of electrons from the analyte anion and subsequent fragmentation [217–219].

1.6. Molecular modelling

Molecular modelling methods have been widely adopted within the field of mass spectrometry, particularly ion mobility, to predict structures analytes may take up in solution or gas–phase. These computational operations can be compared to experimental distributions, charge states, or conformations to corroborate the results. These methods break down into two broad categories based on whether they only account for physical interactions, or if they also include electron potentials.

Molecular mechanics describes the physical motions of atoms from the principles of classical mechanics. Interactions between neighbouring atoms are described by van der Waals interactions and chemical bonds represented as spring interactions [220]. Atomic velocities are determined by their molecular mass in relation to the temperature of the system and their starting coordinates. Starting coordinates are often derived from structures that have already been produced by NMR spectroscopy or X–ray crystallography in the protein data bank (PDB) [221, 222]. These parameters, as well as bond length, partial charges and force constants are known as a force field. This force field is applied to the system to generate a potential function related to the system's internal energy. These force fields try to find the lowest energy of the systems by rotating bonds to their lowest energy state in a local energy minimum. These forms of computational molecular modelling seek to generate potential structures which can be further compared to data collected by other techniques such as ion mobility–mass spectrometry, discussed in section 1.6.1.

The drawback of these simple force fields is they fail to consider the effects of electronic orbitals on structure. Additionally, not all elements have been characterised for these calculations, particularly uncommon metal atoms. An alternative to these force fields are more advanced quantum methods, which attempt to approximate the Schrodinger equation. Examples of these include the Hartree–Fock method [223] and density functional theory [224–226]. These methods seek to compute orientation of atoms and bonds based upon electronic orbital theory and bonding along with general assumptions and simplifications. Some of these simplifications include boundary conditions (to limit the size of the simulation), the Born–Oppenheimer simplification (which assumes nuclei remain in place), or the simplification of potential energy surfaces to exclude electronic effects. Additionally, the Monte Carlo method [227] is based on statistical mechanics of Boltzman distributions which can be used to model potential conformations.

1.6.1. Calculation of theoretical collisional cross section

A large part of IM–MS is comparison of experimentally derived cross sections to those determined from other methods. The prediction of CCS comes from molecular dynamic strategies that predict how the ions will move through space and interact with buffer gas. There are three major methods used to calculate theoretical CCS: the projection approximation (PA), exact hard sphere scattering (EHSS), and the trajectory method (TM). PA models the ions by treating them as a collection of overlapping hard spheres with each sphere radius determined by its collision distance [228]. This method works well for convex systems; however, it fails to accurately predict concave systems [229]. Concave systems may experience fewer collisions due to shielding, or greater collisions due to reflections of buffer gas within a cavity. EHSS is an evolution of PA with an added term to account for scattering angles between the buffer gas on both departure and arrival [230]. This is then used to generate an orientation–averaged momentum transfer to the CCS. TM considers the analyte as a collection of atoms, allowing the consideration of long–range interactions as well as close collisions [231]. An offshoot of PA has been developed, termed the projected superposition approximation (PSA) [232–235]. PSA calculations include shape and size terms to provide more accurate CCS values. TM is considered the most reliable method; however, computations can be very slow for large systems; therefore, the faster PSA is becoming more widespread.

1.7. References

[1] B. Silvi, A. Savin, Classification of chemical bonds based on topological analysis of electron localization functions, *Nature.* **371** (2006) 683–686.

[2] A. E. Reed, L. A. Curtiss, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **88** (1988) 899-926.

[3] N. E. Zhou, C. M. Kay, R. S. Hodges, The role of interhelical ionic interactions in controlling protein folding and stability. De novo designed synthetic two–stranded alpha–helical coiled–coils, *J. Mol. Biol.* **237** (1994) 500–512.

[4] T. Webb, The free energy of hydration of ions and the electrostriction of the solvent, *J. Am. Chem. Soc.* **48** (1926) 2589–2603.

[5] R. H. Stokes, R. A. Robinson, Ionic hydration and activity in electrolyte solutions, *J. Am. Chem. Soc.* **70** (1948) 1870–1878.

[6] A. L. Lehninger, Role of metal ions in enzyme systems, *Physiol. Rev.* **30** (1950) 393–429.

[7] K. B. Jacobson, J. E. Turner, The interaction of cadmium and certain other metal ions with proteins and nucleic acids, *Toxicology.* **16** (1980) 1-37.

[8] B. H. Honig, W. L. Hubbell, Stability of "salt bridges" in membrane proteins, *Proc. Natl. Acad. Sci.* **81** (1984) 5412–5416.

[9] A. Horovitz, L. Serrano, B. Avron, M. Bycroft, A. R. Fersht, Strength and co–operativity of contributions of surface salt bridges to protein stability, *J. Mol. Biol.* **216** (1990) 1031–1044.

[10] R. B. Hermann, Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area, *J. Phys. Chem.* **76** (1972) 2754–2759.

[11] A. A. Pakula, R. T. Sauer, Reverse hydrophobic effects relieved by amino–acid substitutions at a protein surface, *Nature*. **344** (1990) 363–364.

[12] C. Chotia, Hydrophobic bonding and accessible surface area in proteins, *Natur*e. **248** (1974) 338–339.

[13] G. Némethy, H. A. Scheraga, Structure of water and hydrophobic bonding in proteins. II. A Model for the thermodynamic properties of aqueous solutions of hydrocarbons, *J. Chem. Phys.* **36** (1962) 3382–3400.

[14] C. J. Oss, D. R. Absolom, A. W. Neumann, Applications of net repulsive van der Waals forces between different particles, macromolecules, or biological cells in liquids, *Colloids Surf.* **1** (1980) 45–56.

[15] C. J. Oss, R. J. Good, M. K. Chaudhury, The role of van der Waals forces and hydrogen bonds in “hydrophobic interactions” between biopolymers and low energy surfaces, *J. Colloid Interface Sci.* **111** (1986) 378–390.

[16] T. Steiner, The hydrogen bond in the solid state, *Angew. Chemie. Int. Ed. Engl.* **41** (2002) 48–76.

[17] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Danneberg P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci, D. J. Nesbit, Definition of the hydrogen bond (IUPAC Recommendations 2011), *Pure Appl. Chem*. **83** (2011) 1637–1641.

[18] I. K. McDonald, J. M. Thornton, Satisfying hydrogen bonding potential in proteins, *J. Mol. Biol.* **238** (1994) 777–793.

[19] E. R. Johnson, S. Keinan, P. Mori–Sanchez, J. Contreras–Garcia, A. J. Cohen, W. Yang, Revealing noncovalent interactions, *J. Am. Chem. Soc.* **132** (2010) 6498-6506.

[20] F. Wang, J. Zhang, X. Ding, S. Dong, M. Liu, B. Zheng, S. Li, L. Wu, T. Yu, H. W. Gibson, F. Huang*,* Metal coordination mediated reversible conversion between linear and cross–linked supramolecular polymers, *Angew. Chemie. Int. Ed. Engl.* **122** (2010) 1108–1112.

[21] B. A. Springer, K. D. Egeberg, S. G. Sligar, R. J. Rohlfs, A. J. Mathews, J. S. Olson. Discrimination between oxygen and carbon monoxide and inhibition of auto–oxidation by myoglobin. Site–directed mutagenesis of the distal histidine, *J. Biol. Chem.* **264** (1989) 3057–3060.

[22] Y. X. Lu, J. W. Zou, Y. H. Wang, Q. S. Yu, Substituent effects on noncovalent halogen/π interactions: theoretical study, *Int. J. Quantum Chem.* **107** (2007) 1479–1486.

[23] D. A. Dougherty, Cation–pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp, *Science.* **271** (1996) 163-168.

[24] S. F. Betz, Disulfide bonds and the stability of globular proteins, *Protein Sci.* **2** (1993) 1551–1558.

[25] N. Ul, Isoelectric points and conformation of proteins: II. Isoelectric focusing of α–chymotrypsin and its inactive derivative, *Biochim. Biophys. Acta.* **229** (1971) 582–589.

[26] M. M. Flocco, S. L. Mowbray, Planar stacking interactions of arginine and aromatic side–chains in proteins, *J. Mol. Biol.* **235** (1994) 709–717.

[27] L. Strekowski, B. Wilson, Noncovalent interactions with DNA: an overview, *Mutat. Res. Mol. Mech. Mutagen.* **623** (2007) 3–13.

[28] E. T. Kool, Hydrogen bonding, base stacking, and steric effects in DNA replication, *Annu. Rev. Biophys. Biomol. Struct*. **30** (2001) 1–22.

[29] J. Černý, P. Hobza, Non–covalent interactions in biomacromolecules, *Phys. Chem. Chem. Phys.* **9** (2007) 5291.

[30] T. J. Matray, E. T. Kool, A specific partner for abasic damage in DNA, *Nature.* **399** (1999) 704-708.

[31] H. R. Bosshard, Molecular recognition by induced fit: how fit is the concept?, *News*

Physio. Sci. **16** (2001) 171-173.

[32] L. Serrano, A. Matouschek, A. R. Fersht, The folding of an enzyme: VI. The folding pathway of barnase: Comparison with theoretical models, *J. Mol. Biol.* **224** (1992) 847–859.

[33] P. L. Privalov, G. I. Makhatadze, Contribution of hydration and non–covalent interactions to the heat capacity effect on protein unfolding, *J. Mol. Biol.* **224** (1992) 715–723.

[34] R. R. Knowles, E. N. Jacobsen, Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts, *Proc. Natl. Acad. Sci.* **107** (2010) 20678–20685.

[35] A. Urzhumtsev, F. Tete–Favier, A. Mitschler, J. Barbanton, P. Barth, L. Urzhumtseva, J. F. Biellmann, A. D. Podjarny, D. Moras, A 'specificity'pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil, *Structure.* **5** (1997) 601–612.

[36] L. Hedstrom, J. Perona, W. J. Rutter, Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant, *Biochemistry.* **33** (1994) 8757–8763.

[37] Z. Deng, C. Chuaqui, J. Singh, Structural interaction fingerprint (SIFt): a novel method for analyzing three–dimensional protein–ligand binding interactions, *J. Med. Chem.* **47** (2004) 337–344.

[38] B. K. Biswal, M. M. Cherney, M. Wang, L. Chan, C. G. Yannopoulos, D. Bilimoria, N. Nicolas, J. Bedard, M. N. James, Crystal structures of the RNA–dependent RNA polymerase genotype 2a of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non–nucleoside inhibitors, *J. Biol. Chem.* **280** (2005) 18202–18210.

[39] L. Swint–Kruse, C. S. Brown, Resmap: automated representation of macromolecular interfaces as two–dimensional networks, *Bioinformatics.* **21** (2005) 3327–3328.

[40] L. Swint–Kruse, Using networks to identify fine structural differences between functionally distinct protein states, *Biochemistry.* **43** (2004) 10886–10895.

[41] H. Zhu, I. Sommer, T. Lengauer, F. S. Domingues, Alignment of non–covalent interactions at protein–protein interfaces, *PLoS One.* **3** (2008) e1926.

[42] A. Heifetz, E. Katchalski–Katzir, M. Eisenstein, Electrostatics in protein–protein docking, *Protein Sci.* **11** (2002) 571–587.

[43] F. Sheinerman, Electrostatic aspects of protein–protein interactions, *Curr. Opin. Struct. Biol*. **10** (2000) 153–159.

[44] C. L. Vizcarra, S. L. Mayo, Electrostatics in computational protein design, *Curr. Opin. Chem. Biol.* **9** (2005) 622–626.

[45] L. Young, R. L. Jernigan, D. G. Covell, A role for surface hydrophobicity in protein–protein recognition, *Protein Sci.* **3** (1994) 717–729.

[46] A. Berchanski, B. Shapira, M. Eisenstein, Hydrophobic complementarity in protein–

protein docking, *Proteins.* **56** (2004) 130–142.

[47] P. Krapivsky, E. Ben–Naim, Shattering transitions in collision–induced fragmentation, *Phys. Rev. E.* **68** (2003) 021102.

[48] U. Ogmen, O. Keskin, A. S. Aytuna, R. Nussinov, A. Gursoy, PRISM: protein interactions by structural matching, *Nucleic Acids Res.* **33** (2005) 331–336.

[49] O. Keskin, I. Bahar, A. Y. Badretdinov, O. B. Ptitsyn, R. L. Jernigan, Empirical solvent–mediated potentials hold for both intra–molecular and inter–molecular inter–residue interactions, *Protein Sci.* **7** (1998) 2578–2586.

[50] S. Miyazawa, R. L. Jernigan, Estimation of effective interresidue contact energies from protein crystal structures: quasi–chemical approximation, *Macromolecules.* **18** (1985) 534–552.

[51] Y. Ofran, B. Rost, Analysing six types of protein–protein interfaces, *J. Mol. Biol.* **325** (2003) 377–387.

[52] F. Glaser, D. M. Steinberg, I. A. Vakser, N. Ben–Tal, Residue frequencies and pairing preferences at protein–protein interfaces, *Proteins.* **43** (2001) 89–102.

[53] E. Frieden, Non–covalent interactions: key to biological flexibility and specificity, *J. Chem. Educ*. **52** (1975) 754–761.

[54] A. Lee, Lipid–protein interactions in biological membranes: a structural perspective, *Biochim. Biophys. Acta.* **1612** (2003) 1–40.

[55] D. A. Kelkar, A. Chattopadhyay, The gramicidin ion channel: a model membrane protein, *Biochim. Biophys. Acta.* **1768** (2007) 2011–2025.

[56] J. P. Zbilut, A. Colosimo, F. Conti, M. Colafranceschi, C. Manetti, M. Valerio, C. L. Webber, A. Giuliani, Protein aggregation/folding: the role of deterministic singularities of sequence hydrophobicity as determined by nonlinear signal analysis of acylphosphatase and Aβ (1–40), *Biophys. J.* **85** (2003) 3544–3557.

[57] F. Heinkel, J. Gsponer, Determination of protein folding intermediate structures consistent with data from oxidative footprinting mass spectrometry, *J. Mol. Biol.* **428** (2015) 365–371.

[58] R. Beveridge, A. S. Phillips, L. Denbigh, H. M. Saleem, C. E. MacPhee, P. E. Barran, Relating gas–phase to solution conformations: lessons from disordered proteins, *Proteomics.* **15** (2015) 2872–2883.

[59] A. S. Phillips, A. F. Gomes, J. M. D. Kalapothakis, J. E. Gillam, J. Gasparavicius, F. C. Gozzo, T. Kunath, C. MacPhee, P. E. Barran, Conformational dynamics of α–synuclein: insights from mass spectrometry, *Analyst.* **140** (2015) 3070–3081.

[60] P. Leeson, Drug discovery: chemical beauty contest, *Nature.* **481** (2012) 455–456.

[61] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.* **46** (2001) 3–26.

[62] M. Harel, D. M Quinn, H. K. Nair, I. Silman, J. L Sussman, The X–ray structure of a transition state analog complex reveals the molecular origins of the catalytic power and substrate specificity of acetylcholinesterase, *J. Am. Chem. Soc*. **118** (1996) 2340–2346.

[63] R. Cabot, C. A. Hunter, L. M. Varley, Hydrogen bonding properties of non–polar solvents, *Org. Biomol. Chem.* **8** (2010) 1455–1462.

[64] P. Mazzei, A. Piccolo, Quantitative evaluation of noncovalent interactions between glyphosate and dissolved humic substances by NMR spectroscopy, *Environ. Sci. Technol*. **46** (2012) 5939–5946

[65] M. Fuss, H. U. Siehl, B. Olenyuk, P. J. Stang, Dynamics of noncovalent supramolecular complexes. NMR study of the rotational barriers in chiral BINAP palladium (II) and platinum (II) bis (phosphane) complexes that resemble the minimal subunits of chiral polygons and polyhedra, *Organometallics.* **18** (1999) 758–769.

[66] Q. Yi, J. E. Erman, J. D. Satterlee, Proton NMR studies of noncovalent complexes of Cytochrome C peroxidase–cyanide with horse and yeast ferricytochromes C, *Biochemistry.* **32** (1993) 10988–10994.

[67] A. D. Bax, S. Grzesiek. Methodological advances in protein NMR, *Acc. Chem. Res.* **26** (1993) 131–138.

[68] A. K. Sharma, G. P. Zhou, J. Kupferman, H. K. Surks, E. N. Christensen, J. L. Chou, M. E. Mendelsohn, A. C. Rigby, Probing the interaction between the coiled coil leucine zipper of cGMP–dependent protein kinase Iα and the C terminus of the myosin binding subunit of the myosin light chain phosphatase, *J. Biol. Chem.* **283** (2008) 32860–32869.

[69] R. Balakrishnan, N. Ramasubbu, K. I. Karaughese, R. Parthasarathy, Crystal structures of the copper and nickel complexes of RNase A: metal–induced interprotein interactions and identification of a novel copper binding motif, *Proc. Natl. Acad. Sci.* **94** (1997) 9620–9625.

[70] R. Kubota, S. Tashiro, M. Shiro, M. Shionoya, *In situ* X–ray snapshot analysis of transient molecular adsorption in a crystalline channel, *Nat. Chem.* **6** (2014) 913–918.

[71] M. Mirzaei, H. Eshtiagh–Hosseini, Z. Bolouri, Z. Rahmati, A. Esmaeilzadeh, A. Hassanpoor, A. Bauza, P. Ballester, M. Barcelo–Oliver, J. Y. Mague, B. Notash, A. Frontera, Rationalization of noncovalent interactions within six new M^{II} /8–aminoquinoline supramolecular complexes (M^{II} = Mn, Cu, and Cd): a combined experimental and theoretical DFT study, *Cryst. Growth Des.* **15** (2015) 1351–1361.

[72] J. B. Thoden, H. M. Holden, Z. Zhuang, D. Dunaway–Mariano, X–ray crystallographic analyses of inhibitor and substrate complexes of wild–type and mutant 4–hydroxybenzoyl–CoA thioesterase, *J. Biol. Chem.* **277** (2002) 27468–27476.

[73] M. E. Goldberg, F. Schaeffer, Y. Guillou, L. Djavadi–Ohaniance, Pseudo–native motifs in the noncovalent heme–apocytochrome, *Biochemistry.* **274** (1999) 16052–16061.

[74] P. Dinis, D. L. M. Suess, S. J. Fox, J. E. Harmer, R. C. Driesener, L. De La Paz, J. R. Swartz, J. W. Essex, R. D. Britt, P. L. Roach, X–ray crystallographic and EPR spectroscopic analysis of HydG, a maturase in [FeFe]–hydrogenase H–cluster assembly, *Proc. Natl. Acad. Sci.* **112** (2015) 1362–1367.

[75] H. Niwa, A. G. Tonevitsky, I. I. Agapov, S. Saward, U. Pfüller, R. A. Palmer, Crystal structure at 3 Å of mistletoe lectin I, a dimeric type-II ribosome-inactivating protein, complexed with galactose, *Eur. J. Biochem.* **270** (2003) 2739–2749.

[76] A. R. Ferré–D'Amaré, G. C. Prendergast, E. B. Ziff, S. K. Burley, Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain, *Nature.* **363** (1993) 38–45.

[77] V. Ball, C. Maechling, Isothermal microcalorimetry to investigate non–specific interactions in biophysical chemistry, *Int. J. Mol. Sci.* **10** (2009) 3283–3315.

[78] I. Jelesarov, H. R. Bosshard, Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition*, J. Mol. Recognit.* **12** (1999) 3–18.

[79] M. M. Pierce, C. S. Raman, B. T. Nall, Isothermal titration calorimetry of protein–protein interactions, *Methods.* **19** (1999) 213–221.

[80] J. Piehler, New methodologies for measuring protein interactions *in vivo* and *in vitro*, *Curr. Opin. Struct. Biol.* **15** (2005) 4-14.

[81] M. Johnson, E. Garand, M. Z. Kamrath, Determination of non–covalent docking by IR spectroscopy of cold gas–phase complexes, *Science.* **335** (2012) 694–698.

[82] K. Suntharalingam, O. Mendoza, A. A. Duarte, D. J. Mann, R. Vilar, A platinum complex that binds non–covalently to DNA and induces cell death *via* a different mechanism than cisplatin, *Metallomics.* **5** (2013) 514–523.

[83] M. Sirajuddin, S. Ali, A. Badshah, Drug–DNA interactions and their study by UV–Visible, fluorescence spectroscopies and cyclic voltametry, *J. Photochem. Photobiol.* **124** (2013) 1–19.

[84] X. Michalet, S. Weiss, M. Jäger, Single–molecule fluorescence studies of protein folding and conformational dynamics, *Chem. Rev.* **106** (2006) 1785–1813.

[85] V. DeRocco, T. Anderson, J. Piehler, D. Erie, K. Weninger, Four–color single–molecule fluorescence with noncovalent dye labeling to monitor dynamic multimolecular complexes, *Biotechniques.* **49** (2010) 807–816.

[86] R. A. Agbaria, D. Gill, Non-covalent polymers of oxadiazole derivatives induced by γ-cyclodextrin in aqueous solutions–flourescence study, *J. Photochem. Photobiol*. **78** (1994) 161-167.

[87] J. Valanciunaite, A. S. Klymchenko, A. Skripka, L. Richert, S. Steponkiene, G. Streckyte, Y. Mely, R. Rotomskis, A non–covalent complex of quantum dots and chlorine: efficient energy transfer and remarkable stability in living cells revealed by FLIM, *RSC Adv*. **4** (2014) 52270–52278.

[88] I. L. Medintz, A. R. Clapp, H. Mattoussi, E. R. Goldman, B. Fisher, J. M. Mauro, Self–assembled nanoscale biosensors based on quantum dot FRET donors, *Nat. Mater.* **2**

(2003) 630–638.

[89] D. W. Urry, The heme chromophore in the ultraviolet, *J. Biol. Chem.* **242** (1967) 4441–4448.

[90] J. H. Morrissey, Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity, *Anal. Biochem.* **117** (1981) 307–310.

[91] A. G. Murzin, Biochemistry. Metamorphic proteins, *Science.* **320** (2008) 1725–1726.

[92] A. Oda, K. Tsuchida, T. Takakura, N. Yamaotsu, S. Hirono, Comparison of consensus scoring strategies for evaluating computational models of protein−ligand complexes, *J. Chem. Inf. Model.* **46** (2006) 380–391.

[93] P. Edman, Method for determination of the amino acid sequence in peptides, *Acta Chem. Scand.* **4** (1950) 283–293.

[94] M. C. Kuprowski, L. Konermann, Signal response of coexisting protein conformers in electrospray mass spectrometry, *Anal. Chem.* **79** (2007) 2499–2506.

[95] S. K. Chowdhury, V. Katta, B. T. Chait, Probing conformational–changes in proteins by mass–spectrometry, *J. Am. Chem. Soc.* **112** (1990) 9012–9013.

[96] A. Dobo, I. A. Kaltashov, Detection of multiple protein conformational ensembles in solution *via* deconvolution of charge–state distributions in ESI MS, *Anal. Chem.* **73** (2001) 4763–4773.

[97] B. Domon, R. Aebersold, Mass spectrometry and protein analysis, *Science.* **312** (2006) 212–217.

[98] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, D. E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, *Annu. Rev. Biophys*. **41** (2012) 429–452.

[99] A. Bensimon, A. J. R. Heck, R. Aebersold, Mass spectrometry–based proteomics and network biology, *Annu. Rev. Biochem.* **81** (2012) 379–405.

[100] J. F. Kellie, J. C. Tran, J. E. Lee, D. R. Ahlf, H. M. Thomas, I. Ntai, A. D. Catheman, K. R. Durbin, L. Zamdborg, A. Vellaichamy, P. M. Thomas, N. L. Kelleher, The emerging process of top down mass spectrometry for protein analysis: biomarkers, protein–therapeutics, and achieving high throughput, *Mol. Biosyst.* **6** (2010) 1532–1539.

[101] B. A. Garcia, What does the future hold for top down mass spectrometry, *J. Am. Soc. Mass Spectrom.* **212** (2010) 193-202.

[102] M. C. Kuprowski, L. Konermann, Signal response of coexisting protein conformers in electrospray mass spectrometry, *Anal. Chem.* **79** (2007) 2499–2506.

[103] S. K. Chowdhury, V. Katta, B. T. Chait, Probing conformational–changes in proteins by mass–spectrometry, *J. Am. Chem. Soc.* **112** (1990) 9012–9013.

[104] A. Dobo, I. A. Kaltashov, Detection of multiple protein conformational ensembles in solution *via* deconvolution of charge–state distributions in ESI MS, *Anal. Chem*. **73**

(2001) 4763–4773.

[105] R. Beveridge, S. Covill, K. J. Pacholarz, J. M. D. Kalapothakis, C. E. Macphee, P. E. Barran, A mass–spectrometry–based framework to define the extent of disorder in proteins, *Anal. Chem.* **86** (2014) 10979–10991.

[106] A. Moradian, A. Kalli, M. J. Sweredoski, S. Hess, The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications, *Proteomics.* **14** (2014) 489–497.

[107] V. N. Uversky, Intrinsically disordered proteins from A to Z, *Int. J. Biochem. Cell Biol.* **43** (2011) 1090–1103.

[108] S. Sammak, G. Zinzalla, Targeting protein–protein interactions (PPIs) of transcription factors: Challenges of intrinsically disordered proteins (IDPs) and regions (IDRs), *Prog. Biophys. Mol. Biol.* **119** (2015) 41–46.

[109] E. W. McDaniel, D. W. Martin, W. S. Barnes, Drift tube–mass spectrometer for studies of low–energy ion–molecule reactions, *Rev. Sci. Instrum.* **33** (1962) 2–7.

[110] E. A. Mason, E. W. McDaniel, *Transport properties of ions in gases*, Wiley–VCH Verlag GmbH and Co. KGaA, Weinheim. 1st Edn. (1988).

[111] A. Politis, A. Y. Park, S. J. Hyung, D. Barsky, B. T. Ruotolo, C. V. Robinson, Integrating ion mobility–mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes, *PLoS One.* **5** (2010) e12080

[112] A. Politis, A. Y. Park, Z. Hall, B. T. Ruotolo, C. V. Robinson, Integrative modelling coupled with ion mobility–mass spectrometry reveals structural features of the clamp loader in complex with single–stranded DNA binding protein, *J. Mol. Biol.* **425** (2013) 4790–4801.

[113] A. Politis, F. Stengel, Z. Hall, H. Hernández, A. Leitner, T. Walzthoeni, C. V. Robinson, R. Abersold, A mass spectrometry–based hybrid method for structural modelling of protein complexes, *Nat. Methods.* **11** (2014) 403–406.

[114] D. E. Clemmer, R. R. Hudgins, M. F. Jarrold, Naked protein conformations: cytochrome C in the gas–phase, *J. Am. Chem. Soc.* **117** (1995) 10141–10142.

[115] H. Shi, N. Atlasevich, S. I. Merenbloom, D. E. Clemmer, Solution dependence of the collisional activation of ubiquitin $[M + 7H]^{7+}$ ions, *J. Am. Soc. Mass Spectrom.* **25** (2014) 21–26.

[116] H. Shi, D. E. Clemmer, Evidence for two new solution states of ubiquitin by IMS–MS analysis, *J. Phys. Chem. B.* **118** (2014) 3498–3506.

[117] C. Graf, M. Stankiewicz, R. Nikolay, M. P. Mayer, Insights into the conformational dynamics of the E3 ubiquitin ligase CHIP in complex with chaperones and E2 enzymes, *Biochemistry.* **49** (2010) 2121–2129.

[118] K. Venne, E. Bonneil, K. Eng, P. Thibault, Improvement in peptide detection for proteomics analyses using nano LC–MS and high–field asymmetry waveform ion mobility–mass spectrometry, *Anal. Chem.* **77** (2005) 2176–2186.

[119] J. A. McLean, B. T. Ruotolo, K. J. Gillig, D. H. Russell, Ion mobility–mass spectrometry: a new paradigm for proteomics, *Int. J. Mass Spectrom.* **240** (2005) 301–315.

[120] B. Srebalus, A. E. Hilderbrand, S. J. Valentine, D. E. Clemmer, Resolving isomeric peptide mixtures: A combined HPLC/ion mobility–TOFMS analysis of a 4000–component combinatorial library, *Anal. Chem.* **74** (2002) 26–36.

[121] C. S. Hoaglund–Hyzer, Y. J. Lee, A. E. Counterman, D. E. Clemmer, Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures, *Anal. Chem.* **74** (2002) 992–1006.

[122] B. T. Ruotolo, J. A. McLean, K. J. Gillig, D. H. Russell, The influence and utility of varying field strength for the separation of tryptic peptides by ion mobility–mass spectrometry, *J. Am. Soc. Mass Spectrom.* **16** (2005) 158–165.

[123] M. F. Jarrold, V. A. Constant, Silicon cluster ions: evidence for a structural transition, *Phys. Rev. Lett.* **67** (1991) 2994–2997.

[124] M. T. Bowers, P. Koppen, P. R. Kemper, Reactions of state–selected Co^+ with C_3H_8, *J. Am. Chem. Soc.* **114** (1992) 1083–1084.

[125] G. von Helden, M. T. Hsu, N. Gotts, M. T. Bowers, Carbon cluster cations with up to 84 atoms: structures, formation mechanism, and reactivity, *J. Phys. Chem.* **97** (1993) 8182–8192.

[126] K. J. Light–Wahl, B. L. Schwartz, R. D. Smith, Observation of the noncovalent quaternary associations of proteins by electrospray ionization mass spectrometry, *J. Am. Chem. Soc.* **116** (1994) 5271–5278.

[127] A. A. Rostom, C. V. Robinson, Detection of the intact GroEL chaperonin assembly by mass spectrometry, *J. Am. Chem. Soc.* **121** (1999) 4718–4719.

[128] E. van Duijn, D. A. Simmons, R. H. H. Heuvel, P. J. Bakkes, H. van Heerikhuizen, R. M. A. Heeren, C. V. Roninson, S. M. van der Vies, A. J. Heck, Tandem mass spectrometry of intact GroEL–substrate complexes reveals substrate–specific conformational changes in the trans ring, *J. Am. Chem. Soc.* **128** (2006) 4694–4702.

[129] M. G. McCammon, H. Hernández, F. Sobott, C. V. Robinson, Tandem mass spectrometry defines the stoichiometry and quaternary structural arrangement of tryptophan molecules in the multiprotein complex TRAP, *J. Am. Chem. Soc*. **126** (2004) 5950–5951.

[130] J. A. Aquilina, C. V. Robinson, Investigating interactions of the pentraxins serum amyloid P component and C–reactive protein by mass spectrometry, *Biochem. J.* **375** (2003) 323–328.

[131] L. Han, B. T. Ruotolo, Ion mobility–mass spectrometry differentiates protein quaternary structures formed in solution and in electrospray droplets, *Anal. Chem.* **87** (2015) 6808–6813.

[132] S. Jakobi, P. T. X. Nguyen, F. Debaene, S. Cianférani, K. Reuter, G. Klebe, What glues a homodimer together: analysis of the stabilizing effect of an aromatic hot spot in the protein–protein interface of the tRNA–modifying enzyme Tgt, *ACS Chem. Biol.*

10 (2015) 1897–1907.

[133] N. Jonker, J. Kool, H. Irth, W. M. A. Niessen, Recent developments in protein–ligand affinity mass spectrometry, *Anal. Bioanal. Chem.* **399** (2011) 2669–2681.

[134] J. A. Loo, P. F. Hu, P. McConnell, W. T. Mueller, T. K. Sawyer, V. Thanabal, A study of Src SH2 domain protein–phosphopeptide binding interactions by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.* **8** (1997) 234–243.

[135] P. Wolff, C. Viega, E. Ennifar, G. Bec, G. Guichard, P. Dumas, Native ESI mass spectrometry can help to avoid qrong interpretations from isothermal titration calorimetry in difficult situations, *J. Am. Soc. Mass Spectrom.* **28** (2017) 347–357.

[136] B. Ganem, Detection of noncovalent receptor–ligand complexes by mass spectrometry, *J. Am. Chem. Soc.* **113** (1991) 6294–6296.

[137] V. Katta, B. T. Chait, Observation of the heme–globin complex in native myoglobin by electrospray–ionization mass spectrometry, *J. Am. Chem. Soc.* **113** (1991) 8534–8535.

[138] Y. T. Li, Y. L. Hsieh, J. D. Henion, B. Ganem, Studies on heme binding in myoglobin, hemoglobin, and cytochrome C by ion spray mass spectrometry, *J. Am. Soc. Mass Spectrom.* **4** (1993) 631–637.

[139] T. D. Veenstra, K. L. Johnson, A. J. Tomlinson, S. Naylor, R. Kumar, Determination of calcium–binding sites in rat brain calbindin D 28K by electrospray ionization mass spectrometry, *Biochemistry.* **36** (1997) 3535–3542.

[140] D. Touboul, L. Maillard, A. Grässlin, R. Moumne, M. Seitz, J. Robinson, R. Zenobi, How to deal with weak interactions in noncovalent complexes analyzed by electrospray mass spectrometry: cyclopeptidic inhibitors of the nuclear receptor coactivator 1–STAT6, *J. Am. Soc. Mass Spectrom.* **20** (2009) 303–311.

[141] M. D. Gross, G. L. Nelsestuen, R. Kumar, Observations on the binding of lanthanides and calcium to vitamin D–dependent chick intestinal calcium–binding protein. Implications regarding calcium–binding protein function, *J. Biol. Chem.* **262** (1987) 6539–6545.

[142] T. D. Veenstra, M. D. Gross, W. Hunziker, R. Kumar, Identification of metal–binding sites in rat brain calcium–binding protein, *J. Biol. Chem.* **270** (1995) 30353–30358.

[143] R. Feng, A. L. Castelhano, R. Billedeau, Z. Yuan, Study of noncovalent enzyme—inhibitor complexes and metal binding stoichiometry of matrilysin by electrospray ionization mass spectrometry, *J. Am. Soc. Mass Spectrom.* **6** (1995) 1105–1111.

[144] X. Cheng, P. E. Morin, A. C. Harms, J. E. Bruce, Y. Ben–David, R. D. Smith, Mass spectrometric characterization of sequence–specific complexes of DNA and transcription factor PU.1 DNA binding domain, *Anal. Biochem.* **239** (1996) 35–40.

[145] X. Cheng, A. C. Harms, P. N. Goudreau, T. C. Terwilliger, R. D. Smith, Direct measurement of oligonucleotide binding stoichiometry of gene V protein by mass spectrometry, *Proc. Natl. Acad. Sci.* **93** (1996) 7022–7027.

[146] S. Hofstadler, R. H. Griffey, Analysis of noncovalent complexes of DNA and RNA by

mass spectrometry, *Chem. Rev.* **101** (2001) 377–390.

[147] J. Butour, K. Jankowski, J. Macquet, Mass Spectrometry study of DNA–cisplatin complexes: perturbation of guanine–cytosine base–pairs, *Biochem. Biophys. Res. Commun.* **92** (1980) 68–74.

[148] R. M. Suárez, P. Bosch, D. Sucunza, A. M. Cuadro, A. Domingo, F. Mendicuti, J. J. Vaquero, Targeting DNA with small molecules: a comparative study of a library of azonia aromatic chromophores, *Org. Biomol. Chem.* **13** (2015) 527–538.

[149] A. Triolo, F. M. Arcamone, A. Raffaelli, P. Salvadori, Non–covalent complexes between DNA–binding drugs and double–stranded deoxyoligonucleotides: a study by ion spray mass spectrometry, *J. Mass Spectrom.* **32** (1997) 1186–1194.

[150] N. T. Nguyen–Huynh, J. Osz, C. Peluso–Iltis, N. Rochel, N. Potier, E. Leize–Wagner, Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry, *Biophys. Chem.* **210** (2015). 2–8.

[151] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw, V. Gabelica, G–quadruplex DNA assemblies: loop length, cation identity, and multimer formation, *J. Am. Chem. Soc.* **130** (2008) 10208–10216.

[152] M. Islam, S. Fujii, S. Sato, T. Okauchi, S. Takenaka, A Selective G–quadruplex DNA–stabilizing ligand based on a cyclic naphthalene diimide derivative, *Molecules.* **20** (2015) 10963–10979.

[153] L. P. Bai, J. Liu, L. Han, H. M. Ho, R. Wang, Z. H. Jiang, Mass spectrometric studies on effects of counter ions of TMPyP4 on binding to human telomeric DNA and RNA G–quadruplexes, *Anal. Bioanal. Chem.* **406** (2014) 5455–5463.

[154] V. Brázda, L. Hároníková, J. C. C. Liao, M. Fojta, DNA and RNA quadruplex–binding proteins, *Int. J. Mol. Sci.* **15** (2014) 17493–17517.

[155] F. Balthasart, J. Plavec, V. Gabelica, Ammonium ion binding to DNA G–quadruplexes: do electrospray mass spectra faithfully reflect the solution–phase species?, *J. Am. Soc. Mass Spectrom.* **24** (2013) 1–8.

[156] B. Baytekin, H. T. Baytekin, C. A. Schalley, Mass spectrometric studies of non–covalent compounds: why supramolecular chemistry in the gas–phase?, *Org. Biomol. Chem.* **4** (2006) 2825–2841.

[157] C. A. Schalley, Supramolecular chemistry goes gas phase: the mass spectrometric examination of noncovalent interactions in host–guest chemistry and molecular recognition, *Int. J. Mass Spectrom.* **194** (2000) 11-39.

[158] P. E. Barran, H. L. Cole, S. M. Goldup, D. A. Leigh, P. R. McGonigal, M. D. Symes, J. Wu, M. Zengerle, Active–metal template synthesis of a molecular trefoil knot, *Angew. Chemie. Int. Ed. Engl.* **123** (2011) 12280–12284.

[159] L. C. Palmer, J. Rebek, Hydrocarbon binding inside a hexameric pyrogallol[4]arene capsule, *Org. Lett.* **7** (2005) 787–789.

[160] L. Avram, Y. Cohen, Molecules at close range: encapsulated solvent molecules in pyrogallol[4]arene hexameric capsules, *Org. Lett.* **8** (2006) 219–222.

[161] L. Avram, Y. Cohen, Hexameric capsules of lipophilic pyrogallolarene and resorcinarene in solutions as probed by diffusion NMR: one hydroxyl makes the difference, *Org. Lett.* **5** (2003) 3329–3332.

[162] L. Avram, Y. Cohen, Discrimination of guests encapsulation in large hexameric molecular capsules in solution: pyrogallol[4]arene *versus* resorcin[4]arene capsules, *J. Am. Chem. Soc.* 1**25** (2003) 16180–16181.

[163] J. L. Atwood, L. R. MacGillivray, A chiral spherical molecular assembly held together by 60 hydrogen bonds, *Nature.* **389** (1997) 469–472.

[164] S. R. Seidel, P. J. Stang, High–symmetry coordination cages *via* self–assembly, *Acc. Chem. Res.* **35** (2002) 972–983.

[165] J. Ujma, M. De Cecco, O. Chepelin, H. Levene, C. Moffat, S. J. Pike, P. J. Lusby, P. E. Barran, Shapes of supramolecular cages by ion mobility–mass spectrometry, *Chem. Commun.* **48** (2012) 4423-4425.

[166] S. R. Harvey, M. Porrini, C. Stachl, D. Macmillan, G. Zinzalla, P. E. Barran, Small–molecule inhibition of c–MYC:MAX leucine zipper formation Is revealed by ion mobility–mass spectrometry, *J. Am. Chem. Soc.* **134** (2012) 19384–19392

[167] J. A. Loo, Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom. Rev.* **16** (1997) 1–23.

[168] C. V. Robinson, E. W. Chung, B. B. Kragelund, J. Knudsen, R. T. Aplin, F. M. Poulsen, C. M. Dobson, Probing the nature of non–covalent protein ligand interactions by mass spectrometry, *J. Am. Chem. Soc.* **118** (1996) 8646–8653.

[169] S. J. Shire, Z. Shahrokh, J. Liu, Challenges in the development of high protein concentration formulations, *J. Pharma. Sci.* **93** (2004) 1390–1402.

[170] V. A. Mikhailov, T. H. Mize, J. L. Benesch, C. V. Robinson, Mass–selective soft–landing of protein assemblies with controlled landing energies, *Anal. Chem*. **86** (2014) 8321–8328.

[171] G. Squires, Francis Aston and the mass spectrograph, *J. Chem. Soc. Dalt. Trans.* **23** (1998) 3893–3900.

[172] J. J. Thomson, Bakerian lecture: rays of positive electricity, *Proc. R. Soc. A. Math. Phys. Eng. Sci.* **89** (1913) 1–20.

[173] F. Field, M. Munson, Chemical ionisation mass spectrometry, *J. Am. Chem. Soc.* **88** (1966) 2621–2630.

[174] H. R. Morris, M. Panico, M. Barber, R. S. Bordoli, R. D. Sedgwick, A. Tyler, Fast atom bombardment: A new mass spectrometric method for peptide sequence analysis, *Biochem. Biophys. Res. Commun.* **101** (1981) 623–631.

[175] R. P. Lattimer, H. R. Schuten, Field ionisation and field desorption mass spectrometry: past present and future, *Anal. Chem.* **61** (1989) 1201–1215.

[176] M. Karas, D. Bachmann, F. Hillenkamp, Influence of the wavelength in high–irradiance ultraviolet laser desorption mass spectrometry of organic molecules, *Anal.*

Chem. **57** (1985) 2935–2939.

[177] J. Zeleny, The electrical discharge from liquid points, *Phys. Rev.* **3** (1914) 69–91.

[178] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science.* **246** (1989) 64–71.

[179] P. Kebarle, L. Tang, From ions in solution to ions in the gas–phase – the mechanism of electrospray mass spectrometry, *Anal. Chem.* **65** (1993) 972–986.

[180] L. Rayleigh, On the equilibrium of liquid conducting masses charged with electricity, *Philos. Mag. Ser.* **5** (1882) 184–186.

[181] X. Yue, S. Vahidi, L. Konermann, Insights into the mechanism of protein electrospray ionization from salt adduction measurements, *J. Am. Soc. Mass Spectrom.* **25** (2014) 1322–1331.

[182] L. Konermann, E. Ahadi, A. D. Rodriguez, S. Vahidi, Unraveling the Mechanism of Electrospray Ionization, *Anal. Chem.* **85** (2012) 2–9.

[183] J. V. Iribarne, On the evaporation of small ions from charged droplets, *J. Chem. Phys.* **64** (1976) 2287–2294.

[184] M. Dole, Molecular beams of macroions, *J. Chem. Phys.* **49** (1968) 2240-2249.

[185] E. Ahadi, L. Konermann, Modelling the behavior of coarse–grained polymer chains in charged water droplets: Implications for the mechanism of electrospray ionization, *J. Phys. Chem. B.* **116** (2012) 104–112.

[186] L. Konermann, A. D. Rodriguez, J. Liu, On the formation of highly charged gaseous ions from unfolded proteins by electrospray ionization, *Anal. Chem.* **84** (2012) 6798–6804.

[187] R. Juraschek, T. Dülcks, M. Karas, Nanoelectrospray – more than just a minimized–flow electrospray ionization source, *J. Am. Soc. Mass Spectrom.* **10** (1999) 300–308.

[188] M. S. Wilm, M. Mann, Electrospray and Taylor–Cone theory, Dole's beam of macromolecules at last?, *Int. J. Mass Spectrom.* **136** (1994) 167–180.

[189] M. Wilm, M. Mann, Analytical properties of the nanoelectrospray ion source, *Anal. Chem.* **68** (1996) 1–8.

[190] J. L. Wiza, Microchannel plate detectors. *Nucl. Instrum. Methods.* **162** (1979) 587-601.

[191] R. E. March, An introduction to quadrupole ion trap mass spectrometry, *J. Mass Spectrom.* **32** (1997) 351–369.

[192] B. A. Mamyrin, V. I. Karataev, D. V Shmikk, V. A. Zagulin, The mass–reflectron, a new nonmagnetic time–of–flight mass spectrometer with high resolution, *Sov. Phys. JETP.* **37** (1973) 45–48.

[193] K. McAfee, D. Edelson, Identification and mobility of ions in a townsend discharge by time–resolved mass spectrometry, *Pro. Phys. Soc.* **81** (1962) 382–384.

[194] S. D. Pringle, K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinos, R. H. Bateman, M. T. Bowers, G. H. Scrivens, An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa–ToF instrument, *Int. J. Mass Spectrom.* **261** (2007) 1–12.

[195] M. E. Belov, M. A. Buschbach, D. C. Prior, K. Tang, R. D. Smith, Multiplexed ion mobility spectrometry–orthogonal time–of–flight mass spectrometry. *Anal. Chem.* **79** (2007) 2451 –2462.

[196] S. Rokushika, H. Hatano, H. H. Hill, Ion mobility spectrometry in carbon dioxide, *Anal. Chem.* **58** (1986) 361–365.

[197] G. R. Asbury, H. H. Hill, Using different drift gases to change separation factors (α) in ion mobility spectrometry, *Anal. Chem.* **72** (2000) 580–584.

[198] K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, R. H. Bateman, Applications of a travelling wave–based radio–frequency–only stacked ring ion guide, *Rapid Commun. Mass Spectrom.* **18** (2004) 2401–2414.

[199] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, Collision cross sections of proteins and their complexes: A calibration framework and database for gas–phase structural biology, *Anal. Chem.* **82** (2010) 9557–9565.

[200] C. B. Lietz, Q. Yu, L. Li, Large–scale collision cross–section profiling on a traveling wave ion mobility mass spectrometer, *J. Am. Soc. Mass Spectrom*. **25** (2014) 2009–2019.

[201] D. P. Smith, T. W. Knapman, I. Campuzano, R. W. Malham, J. T. Berryman, S. E. Radford, A. E. Ashcroft, Deciphering drift time measurements from travelling wave ion mobility spectrometry–mass spectrometry studies, *Eur. J. Mass Spectrom*. **15** (2009) 113–130.

[202] S. I. Merenbloom, T. G. Flick, E. R. Williams, How hot are your ions in TWAVE ion mobility spectrometry?, *J. Am. Soc. Mass Spectrom.* **23** (2012) 553–562.

[203] R. W. Purves, R. Guevremont, S. Day, C. W. Pipich, M. S. Matyjaszczyk, Mass spectrometric characterization of a high–field asymmetric waveform ion mobility spectrometer, *Rev. Sci. Instrum.* **69** (1998) 4094–4104.

[204] R. W. Purves, R. Guevremont, Electrospray ionization high–field asymmetric waveform ion mobility spectrometry–mass spectrometry, *Anal. Chem.* **71** (1999) 2346–2357.

[205] Y. Zhong, L. Han, B. T. Ruotolo, Collisional and coulombic unfolding of gas–phase proteins: high correlation to their domain structures in solution, *Angew. Chemie. Int. Ed. Engl.* **53** (2014) 1–5.

[206] P. Roepstorff, J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11** (1984) 601.

[207] D. P. Little, J. P. Speir, M. W. Senko, P. B. O'Connor, F. W. McLafferty, Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing, *Anal. Chem.* **66** (1994) 2809–2815.

[208] B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I. Compagnon, P. E. Barran, UV photodissociation of trapped ions following ion mobility separation in a Q–ToF mass spectrometer, *Analyst.* **139** (2014) 6348–6351.

[209] B. Bellina, I. Compagnon, L. Joly, F. Albrieux, A. R. Allouche, F. Bertorelle, J. Lemoine, R, Antoine, P. Dugord, UV spectroscopy of entire proteins in the gas–phase, *Int. J. Mass Spectrom.* **297** (2010) 36–40.

[210] V. Gabelica, F. Rosu, E. De Pauw, R. Antoine, T. Tabarin, M. Broyer, P. Dugord, Electron photodetachment dissociation of DNA anions with covalently or noncovalently bound chromophores, *J. Am. Soc. Mass Spectrom.* **18** (2007) 1990–2000.

[211] R. Zubarev, N. L. Kelleher, F. W. McLafferty, Electron capture dissociation of multiply charged protein cations. *J. Am. Chem. Soc.* **120** (1998) 3265–3266.

[212] R. A. Zubarev, D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter, F. W. McLafferty, Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal. Chem.* **72** (2000) 563–573.

[213] K. Breuker, H. B. Oh, C. Lin, B. K. Carpenter, F. W. McLafferty, Nonergodic and conformational control of the electron capture dissociation of protein cations, *Proc. Natl. Acad. Sci.* **101** (2004) 14011–14016.

[214] H. J. Cooper, Investigation of the presence of b ions in electron capture dissociation mass spectra, *J. Am. Soc. Mass Spectrom.* **16** (2005) 1932–1940.

[215] K. O. Zhurov, L. Fornelli, M. D. Wodrich, U. A. Laskay, Y. O. Tsybin, Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis, *Chem. Soc. Rev.* **42** (2013) 5014–5030.

[216] F. Lermyte, T. Verschueren, J. M. Brown, J. P. Williams, D. Valkenborg, F. Sobott, Characterization of top–down ETD in a travelling–wave ion guide, *Methods.* **89** (2015) 22–29.

[217] I. Anusiewicz, M. Jasionowski, P. Skurski, J. Simons, Backbone and side–chain cleavages in electron detachment dissociation (EDD), *J. Phys. Chem. A.* **109** (2005) 11322–11337.

[218] M. Taucher, K. Breuker, Top–down mass spectrometry for sequencing of larger (up to 61 nt) RNA by CAD and EDD, *J. Am. Soc. Mass Spectrom*. **21** (2010) 918–929.

[219] J. Yang, K. Håkansson, Fragmentation of oligoribonucleotides from gas–phase ion–electron reactions, *J. Am. Soc. Mass Spectrom.* **17** (2006) 1369–1375.

[220] V. G. S. Box, The molecular mechanics of quantized valence bonds, *J. Mol. Med*. **3** (1997) 124–141.

[221] A. D. Mackerell, Empirical force fields for biological macromolecules: Overview and issues, *J. Comput. Chem.* **25** (2004) 1584–1604.

[222] J. Parsons, J. B. Holmes, J. M. Rojas, J. Tsai, C. E. M. Strauss, Practical conversion from torsion space to Cartesian space for *in silico* protein synthesis, *J. Comput. Chem*. **26** (2005) 1063–1068.

[223] C. Froese–Fischer, General Hartree–Fock program, *Comput. Phys. Commun*. **43** (1987) 355–365.

[224] S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections, *J. Comput. Chem.* **25** (2004) 1463–1473.

[225] K. Burke, L. O. Wagner, DFT in a nutshell, *Int. J. Quantum Chem.* **113** (2013) 96–101.

[226] O. A. Lilienfeld, I. Tavernelli, U. Rothlisberger, D. Sebastiani, Optimization of effective atom centered potentials for london dispersion forces in density functional theory, *Phys. Rev. Lett.* **93** (2004) 153001–153004.

[227] D. P. Kroese, T. Brereton, T. Taimre, Z. I. Botev, Why the Monte Carlo method is so important today, *WIREs Comp. Stat.* **6** (2014) 386–392.

[228] T. Wyttenbach, G. von Helden, J. J. Batka, D. Carlat, M. T. Bowers, Effect of the long–range potential on ion mobility measurements, *J. Am. Soc. Mass Spectrom.* **8** (1997) 275–282.

[229] D. E. Clemmer, M. F. Jarrold, Ion mobility measurements and their applications to clusters and biomolecules, *J. Mass Spectrom.* **32** (1997) 577–592.

[230] A. A. Shvartsburg, M. F. Jarrold, An exact hard–spheres scattering model for the mobilities of polyatomic ions, *Chem. Phys. Lett.* **261** (1996) 86–91.

[231] M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, M. F. Jarrold, Structural information from ion mobility measurements: effects of the long–range potential, *J. Phys. Chem.* **100** (1996) 16082–16086.

[232] C. Bleiholder, T. Wyttenbach, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method, *Int. J. Mass Spectrom.* **308** (2011) 1–10.

[233] C. Bleiholder, S. Contreras, T. D. Do, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (II). Model parameterization and definition of empirical shape factors for proteins, *Int. J. Mass Spectrom.* **345** (2013) 89–96.

[234] S. E. Anderson, C. Bleiholder, E. R. Brocker, P. J. Stang, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (III): application to supramolecular coordination–driven assemblies with complex shapes, *Int. J. Mass Spectrom.* **330** (2012) 78–84.

[235] C. Bleiholder, S. Contreras, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (IV). Application to polypeptides, *Int. J. Mass Spectrom.* **354** (2013) 275–280.

2.

Novel application of mass spectrometry techniques to characterise a DNA aptamer ligand binding site

2. Declaration

This chapter consists of one draft experimental paper awaiting submission: C. Nortcliffe, A. Theisen, B. Bellina, D. Clarke, P. E. Barran.

Novel application of mass spectrometry techniques to characterise a DNA aptamer ligand binding site

As first author on this publication I performed all experimental work including: MS, ligand binding and IM–MS experiments. Alina Theisen assisted with instrument and technical expertise in the UVPD experiments.

Novel application of mass spectrometry techniques to characterise a DNA aptamer ligand binding site

C. Nortcliffe, A. Theisen, B. Bellina, D. Clarke, P. E. Barran.

Supporting Information

Figures detailing DT–IM–MS experiments for alternate DNA structures and solvent conditions can be found in Appendix 1.

2.1. Abstract

The use of aptamers in biochemical research represents an emerging field, allowing for molecular recognition through a mechanism analogous to that of antibodies. There are currently few biophysical techniques available to characterise the structures of aptamers and their interactions with targets. The work presented here applies mass spectrometry (MS) techniques to probe the interaction between a DNA aptamer (Ky2) and its small molecule target, the aminoglycoside antibiotic kanamycin. Several MS–based fragmentation techniques reveal the loop region of the DNA hairpin to be the site of kanamycin binding. Alterations to the DNA sequence are made in order to establish the impact of base orientation on the ligand binding site. Ion mobility coupled to mass spectrometry reveals reduced flexibility (25% reduction in FWHM, 15% reduction in ${}^{DT}CCS_{He}$) of the DNA aptamer upon kanamycin binding, indicating that the small molecule induces a stabilising effect. This multi–technique approach highlights structural features that are highly important for target interactions and may assist in the future design of aptamers with higher selectivity and specificity for their targets.

2.2. Introduction

Nucleic acid sequences are able to form many more tertiary structures than just the traditional double helix macromolecular fold. These structures include G–quadruplexes, i–motif DNA, hairpins, pseudo knots, double–stranded crucioforms as well as many others [1–4]. It is this ability to form a diverse range of folded forms that makes oligonucleotides an attractive possibility in the use of targeted intermolecular interactions [5–10]. Aptamers are small molecules designed to recognise a specific target molecule through a three–dimensional fold or binding pocket; analogous to monoclonal antibodies, although aptamers are typically of much lower molecular weight. Aptamers can be formed of peptides, DNA or RNA. Large libraries of aptamers can be generated by a process known as systematic evolution of ligands by exponential enrichment (SELEX), creating up to 106 randomised sequences with a variety of three dimensional structures [11]. Although the SELEX process is best suited to systems with large surface areas, such as proteins with potential pockets and interfaces, it has been applied to a variety of low molecular weight targets including metal ions and small organic molecules including cholic acid, cocaine and oxytetracycline [12].

Aptamers have been shown to have a wide variety of potential uses in biosensing, drug development and probing potential drug targets [13–17]. They can fulfil a similar role to antibodies, but with the advantages of generating a lower immune response and being much easier to chemically synthesize and modify [18].

The first aptamer characterised for a molecular target was the thrombin–binding aptamer (TBA) in 1992 [20]. This DNA 15–mer forms a G–quadruplex and was generated through the SELEX process where a library of 10^{13} DNA sequences was screened against α–thrombin binding. Several variants of TBA have been developed since its discovery and the X–ray crystal structure of a thrombin–modified TBA complex was solved in 2011 [5]. Since then, many aptamers have been developed for a variety of systems which can be roughly divided into three classes: small molecules, biomolecules and whole cells [12, 13, 16]. Small species can be targeted by the SELEX process even down to inorganic ions as Zn^{2+} or Ni^{2+} [21, 22]. A challenge in the targeting of these small molecules is the need for chemical immobilisation of the substrate on a solid support, which may reduce the specificity of aptamers due to the inaccessibility of the target molecules to the aptamer [12]. Biomolecules are the main targets of aptamer studies, with many varied uses including affinity purification, bio–sensing, imaging, drug delivery and medical applications as well as discovery of new molecules or targets. The SELEX process can be applied to whole cells or tissues; however, a new technique has been developed called cell–SELEX [23] for these larger targets. Issues remain, however, as membrane bound proteins and cell surface markers are prolific, therefore it is difficult to generate specific aptamers.

The analysis of intact DNA by MS is possible due to the development of 'soft' ionisation techniques such as ESI [24] and MALDI [25] which allow the gentle transfer of ions into the gas–phase. The first reported ESI spectrum of an oligonucleotide came in 1988 by Covey *et al.* [26] who observed pieces of DNA of up to 14 bases long when operating their instrument in negative ionisation mode. The majority of current work utilising mass spectrometry with DNA falls into two major categories: analysis of DNA modifications (including methylation and acetylation) [27–36], and DNA binding to proteins [37–43].

Although some aptamers have been structurally characterised by classical techniques such as NMR spectroscopy [5] and X–ray crystallography [44], the use of ion mobility–mass spectrometry (IM–MS) has not been applied. IM–MS is a gas–phase electrophoretic technique which separates ions based on their mass, charge and size/shape, giving a coarse view of an ion's conformation in the form of a rotationally averaged collisional cross section (CCS, Ω, $Å^2$). This allows changes in conformation upon ligand binding to be measured, and when combined with computational modelling, can be used to help assign binding sites. IM–MS analysis of oligonucleotides was first reported in 1997 [45] when a ten–base thymidine oligomer was analysed in a solution of water and acetonitrile with ammonium hydroxide. CCS values were compared to calculated models, with the most likely candidate structures found to have the majority of the charge sitting on the phosphodiester backbone. Following on from this work, several studies examined the retention of oligonucleotide structure in solvent–free environments [46–50]. This approach was followed by analysis of more specific tertiary structures such as G–quadruplexes, cruciforms, and hairpins [50–55]. The next key evolution of DNA ion mobility experiments came from the analysis of an organometallic anti–cancer drug bound to an oligonucleotide target [56]. G–quadruplexes have since been the subject of a great deal of interest due to their complex 3D structure and ability to retain cations [54, 57].

The process of dirft time (DT) IM–MS involves sending a pulse of ions through a drift cell filled with an inert gas which is held along a potential. Ions drift along the potential field, and are slowed *via* collisions with buffer gas molecules, with larger molecules undergoing more collisions than smaller ones. The speed at which ions travel the drift cell region is related to not just the size but the charge upon the ion which has a measurable mobility (K), which is a constant related to drift velocity (v_d) and the electric field (E). Using equation 1.8 it is possible to convert the arrival time of ions into a $^{DT}CCS_{He}$ (Ω, $Å^2$).

$$\boldsymbol{\Omega} = \frac{3ze}{16N}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{1}{K_0}$$

Equation 1. 8

Where K_0 is the reduced mobility (measured mobility corrected to temperature and pressure) z is the ion charge state, e is the elementary charge state, N is the number gas density, μ is the reduced mass of the ion–pair, k_B is the Boltzmann constant and T is the buffer gas temperature.

2.3. Aims

The binding mode of a previously reported single stranded DNA (ssDNA) aptamer which has been shown to specifically bind kanamycin [19] is analysed here through mass spectrometry. Kanamycin is an aminoglycoside antibiotic commonly used to treat Gram–negative bacteria (Figure 2.1 A). The binding affinity is analysed using mass spectrometry, and the conformations of the bound forms are elucidated using ion mobility–mass spectrometry and other mass spectrometry techniques. Previous work by Song *et al.* [19] used this aptamer (hereby referred to as Ky2) to detect kanamycin in milk *via* an indicative colour change caused by a bound gold nano particle. This study seeks to further this work by probing the structure of the DNA:antibiotic complex and mapping the binding site. Determination of the nature of the complex by mass spectrometry will pave the way for improved aptamer design for use in other systems, and uncover information which may not be available from other techniques.

Additional variations of the Ky2 aptamer were tested (shown in Figure 2.1 C) including the removal of the tail residues (Ky2–L), the removal of the loop region (Ky2–T), an inversion of the loop residues (Ky2–I), and replacement of the loop region with a control sequence (Ky2–C). These alternate sequences were employed to help uncover facets of the aptamer:ligand interface.

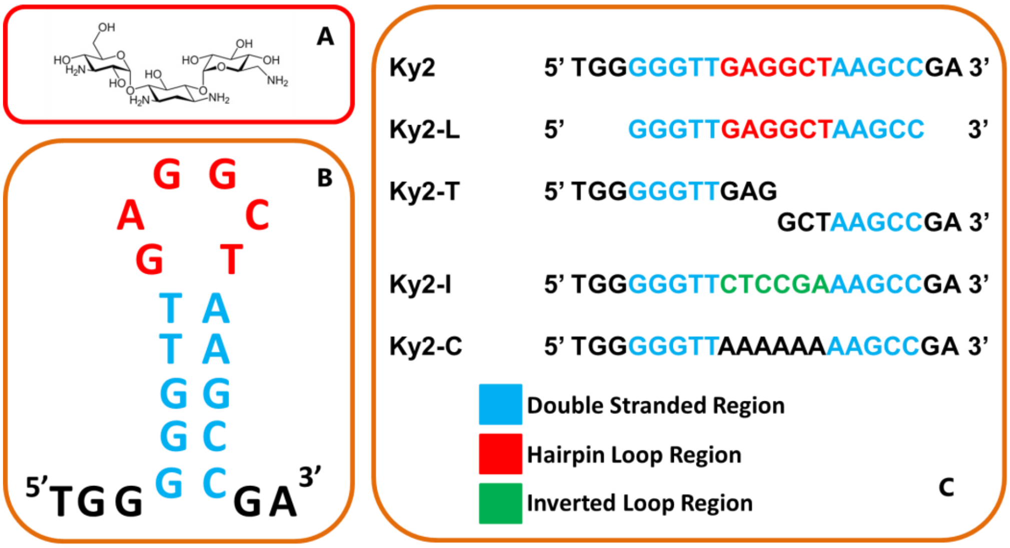


Figure 2.1: A) Molecular structure of the antibiotic kanamycin. B) Structural representation of the Ky2 aptamer as a hairpin. C) Alternate aptamer sequences used in this study. Double stranded regions are shown in blue, hairpin loops are shown in red and inverted loops are shown in green.

2.4. Experimental details

Single stranded DNA aptamers were purchased from Sigma Aldrich (Dorset, UK) at HPLC purification grade. LC–MS grade solvents and ammonium acetate were purchased from Fisher Scientific (Loughborough, UK). DNA was prepared in water to a stock concentration of 200 µM and diluted to a concentration of 20 µM for all experiments. Two solution conditions were used in these experiments, defined as 'buffered' and 'denaturing'. Buffered DNA was prepared in 100 mM ammonium acetate with 5% methanol, whereas denatured DNA samples were prepared in a solution of 50:50 methanol:H_2O, in accordance with previously published work [58, 59].

Kanamycin sulfate was obtained from Sigma Aldritch as a mixture of kanamycin A (major component >99%), kanamycin B and C. A stock solution of concentration 2 mM was prepared and diluted to 100 µM for each experiment. DNA and ligand were incubated together for 20 minutes before each experiment at room temperature.

Nano–ESI tips were pulled from 0.5 mm glass capillaries using a Flaming/Brown micropipette puller (Sutter Instrument Company, Novato, CA, USA). Platinum wire (0.125 mm Goodfellow) was inserted into the pulled glass tip to provide the ionisation potential to the solution.

2.4.1. n–ESI mass spectrometry

Mass spectrometry experiments were performed using a Synapt G2 (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer operated in negative ionisation mode. Instrument parameters were optimised to retain the non–covalent aptamer:ligand complex and were maintained as closely as possible across experiments. Typical source conditions: capillary voltage ~ –1.6 kV, sampling cone ~ 20 V, extraction cone ~2 V.

2.4.2. Ion mobility–mass spectrometry

DT–IM–MS experiments were performed on a Q–ToF 1 (Waters, Micromass, Manchester, UK) quadrupole time–of–flight mass spectrometer that has been modified in–house to contain a 5.1 cm drift cell [60]. The cell is pressurised to ~4 Torr with helium. The potential across the drift cell was stepped from –50 V to –20 V in discrete acquisition experiments where temperature and pressure were monitored. Rotationally averaged collisional cross sections (${}^{DT}CCS_{He}$) were obtained by plotting experimental arrival times against *P/V* for a number of injections at a specified drift voltage (V). Experiments were performed with three technical repeats; values reported are means with standard deviation.

2.4.3. Collision induced unfolding

The Ky2 aptamer and the Ky2–I aptamer were subjected to increasing energy of collisions to encourage unfolding of tertiary structures, following electrospray from both solutions. Bound and unbound $[M - 4H]^{4-}$ and $[M - 5H]^{5-}$ complexes were isolated in the quadrupole region of a Synapt G2 (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer and then excited in the trap region for 200 scans with steps of 5 V until the precursor ion was no longer visible (maximum CE used 40 V). These were then plotted using Python CIU suite software [61] to create heat maps depicting the change in drift time of the ions at increasing collision energy, allowing visualisation of molecular unfolding.

2.4.4. CID fragmentation

CID fragmentation experiments were conducted using a Synapt G2 (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer. Bound and unbound $[M - 5H]^{5-}$ species were isolated using the quadrupole followed by excitation from collisions with argon in the trap cell up to energies of 35 V.

Fragments were then assigned using the McLuckey nomenclature outlined in Figure S1 [62].

2.4.5. UV photo–dissociation

Ultraviolet photo–dissociation (UVPD) is a new and emerging technique using lasers to induce rapid electronic fragmentation of ions. It has been utilised for the fragmentation of DNA along with DNA:chromophore complexes [63, 64]. UVPD experiments were performed using a modified Synapt G2–S (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer which has been fitted with a CaF_2 window allowing a laser beam to enter the instrument and irradiate ions trapped in the transfer cell (described previously [65]). Ions for the $[M - 5H]^{5-}$ unbound and bound forms of the aptamer were mass isolated in the quadrupole and accumulated in the transfer region of the instrument for 10 seconds before being exposed to a Q–switched Nd:YAG Continuum Minilite II 266 nm laser for 1 second. The average pulse energy was estimated to be 0.4–1 mJ. After irradiation, ions were activated using a low travelling wave voltage. The laser pulse causes excitation of the ions leading to electron detachment of the molecule, which results in the $[M - 5H]^{4-\bullet}$ radical species, along with some low abundance fragments. The post activation, with travelling wave voltage, leads to further fragmentation analogous to that of CID in other studies [63].

2.5. Results

2.5.1. n–ESI MS

The DNA aptamer was analysed at 20 µM in 100 mM ammonium acetate, and presents in the charge states $[M - 3H]^{3-}$ through $[M - 6H]^{6-}$, with the $[M - 4H]^{4-}$ (*m/z* 1640) being most dominant at about 50% of the total ion signal intensity. When analysed from denatured conditions, the charge state distribution (CSD) widens, with the $[M - 5H]^{5-}$ / $[M - 6H]^{6-}$ charge states now most dominant, with up to the $[M - 11H]^{11-}$ charge state being observed. DNA peaks show a marked degree of salt retention with adducts of masses corresponding to both sodium and potassium observed; attributed to the high affinity of DNA for cations. Upon addition of kanamycin (100 µM), peaks attributable to the Ky2:kanamycin complex at *m/z* 1761 and *m/z* 1409 were observable and are assigned as $[M + \text{ligand} - 4H]^{4-}$ and $[M + \text{ligand} - 5H]^{5-}$ (Figure 2.2A). Upon binding, there is no large shift in CSD and there are always peaks attributable to the unbound aptamer, even with the addition of 400 µM kanamycin. This is in disagreement with the work of Song *et al.* [19] who reported K_d <100 nM, suggesting that the complex is dissociating during transmission into the gas–phase.

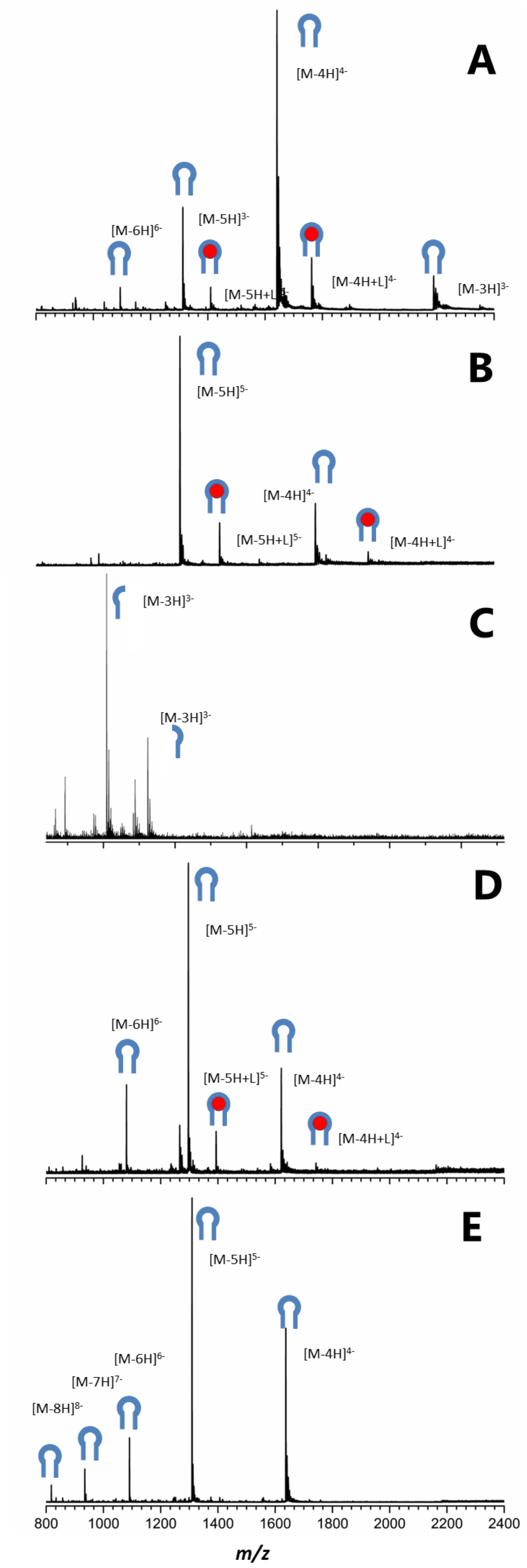


Figure 2.2: Mass spectra of all aptamers employed. Blue hairpin represents the unbound aptamer and a red circle the ligand. A) Mass spectrum of Ky2 aptamer (20 µM), kanamycin (100 µM), 100 mM ammonium acetate, 5% methanol. B) Mass spectrum of Ky2-L aptamer (20 µM), kanamycin (100 µM), 100 mM ammonium acetate, 5% methanol. C) Mass spectrum of Ky2-T aptamer (both strands 20 µM annealed), kanamycin (100 µM), 100 mM ammonium acetate, 5% methanol. D) Mass spectrum of Ky2-I aptamer (20 µM), kanamycin (100 µM), 100 mM ammonium acetate, 5% methanol. E) Mass spectrum of Ky2-C aptamer (20 µM), kanamycin (100 µM), 100 mM ammonium acetate, 5% methanol.

2.5.2. Solvent conditions

Under methanol:H_2O conditions there is an increase in the intensity of the higher charge states both in the presence and absence of the ligand, though the $[M - 4H]^{4-}$ and $[M - 5H]^{5-}$ charge states remain the most intense. The ratio of bound to unbound forms is larger under denaturing methanol:H_2O conditions than under 'native like' ammonium acetate conditions. A screen was carried out with various concentrations of methanol, from 50% to 0% (vol %) in 100mM ammonium acetate (Figure S2.2). In conditions without buffer, bound peaks account for ~20% of total intensity, dropping to 5–10% with a mix of denaturing and buffering solvent, and finally dropping to less than 5% in fully buffered conditions. There are several possible explanations for this behaviour: one possibility is that since desolvation occurs more readily in the presence of methanol, more complex is carried into the gas–phase or fewer complexes dissociate in the source region. An alternative explanation is that methanol induces the formation of a structure which can interact more favourably with the ligand. It has previously been reported that the addition of up to 20% methanol has little or no effect on the charge state distribution from oligonucleotides [59]; however, there have been fewer investigations of its effects on structure.

2.5.3. Alternative structures

To further investigate the binding of the ligand to the DNA aptamer, several alternative and truncated sequences were obtained, to investigate the different facets of the ligand interface. The first sequence consists of residues 4–19, leaving the loop region and the structural double stranded region intact, but removing the excess tail residues, denoted Ky2–Loop (Ky2–L). The Ky2–L displayed a narrow mass spectrum with two dominant charge states at $[M - 5H]^{5-}$ (*m/z* 1240) and $[M - 4H]^{4-}$ *m/z* 1653) (Figure 2.2 B). Ligand bound complexes were also observed, again mainly $[M - 5H]^{5-}$ (*m/z* 1361) and $[M - 4H]^{4-}$ (*m/z* 1816) ions presented, with a similar relative intensity of bound to unbound as Ky2 (15:1) (Figure 2.2 B). From this data, it can be inferred that the removed residues have little effect on ligand binding. The second analysed construct consisted of two separate strands of DNA comprised of residues 1–11 and 12–21, providing the tail region without the loop structure (denoted Ky2 Tail/Ky2–T). The strands were analysed both with and without annealing and in the presence and absence of the ligand. Each individual strand presented as a $[M - 3H]^{3-}$ ion at *m/z* of 1157 and 1013 respectively, the double strand was observed, although it was only as a low abundance species (Figure 2.2 C). There is no evidence of ligand binding to either strand, or a ligand stabilised double strand. This strongly suggests that the structure of the loop region is essential to ligand binding. Due to the low abundance of the double stranded DNA, this may prevent a suitable conformation for kanamycin binding. Therefore, two alternate forms were used: one where the loop region had a sequence inversion (G»C, C»G, A»T, T»A) (denoted Ky2–Inverse/ Ky2–I); and a control with residues 9–14 replaced with adenosine (denoted Ky2–Adenosine/ Ky2–A). Ky2–I presented a similar mass spectrum to Ky2, though with a shift in distribution towards higher charge states (Figure 2.2 D). Kanamycin binding was observed nearly as strongly as with Ky2 with peaks observed at *m/z* 1394 and *m/z* 1742*,* attributable to the $[M - 4H]^{4-}$ and $[M - 5H]^{5-}$ complexes respectively. Ky2–I has the same hydrogen bonding structure as Ky2 but with inverted base positions (see Figure 2.1 C); the difference is the presence or absence of the 5–membered rings in the nucleobase. As a complex is still observed, this indicates that the hydrogen bonding is specific and the backbone is flexible enough to compensate for this change in positioning. Ky2–A represents a negative control, as the structure is the same as Ky2; however, the interface chemistry is different in relation to the hydrogen bonds of the bases. Ky2–A is observed in a wider charge state distribution from $[M - 4H]^{4-}$ to $[M - 8H]^{8-}$, though the majority of the intensity still lies in the $[M - 4H]^{4-}$ and $[M - 5H]^{5-}$ peaks (Figure 2.2 E). No binding to kanamycin was observed in the mass spectrum even at 20:1 ligand excess.

2.5.4. Ion mobility–mass spectrometry

Figure 2.3 shows the collisional cross section distributions ($^{DT}CCSD_{He}$) of the Ky2 aptamer across the observed charge states. Results are displayed for the unbound aptamer in the absence of kanamycin (control, Figure 2.3 A), the unbound aptamer in the presence of kanamycin (unbound, Figure 2.3 B), and the aptamer:ligand complex (bound, Figure 2.3 C). Data from methanol:H_2O conditions are shown in Figure S2.3.

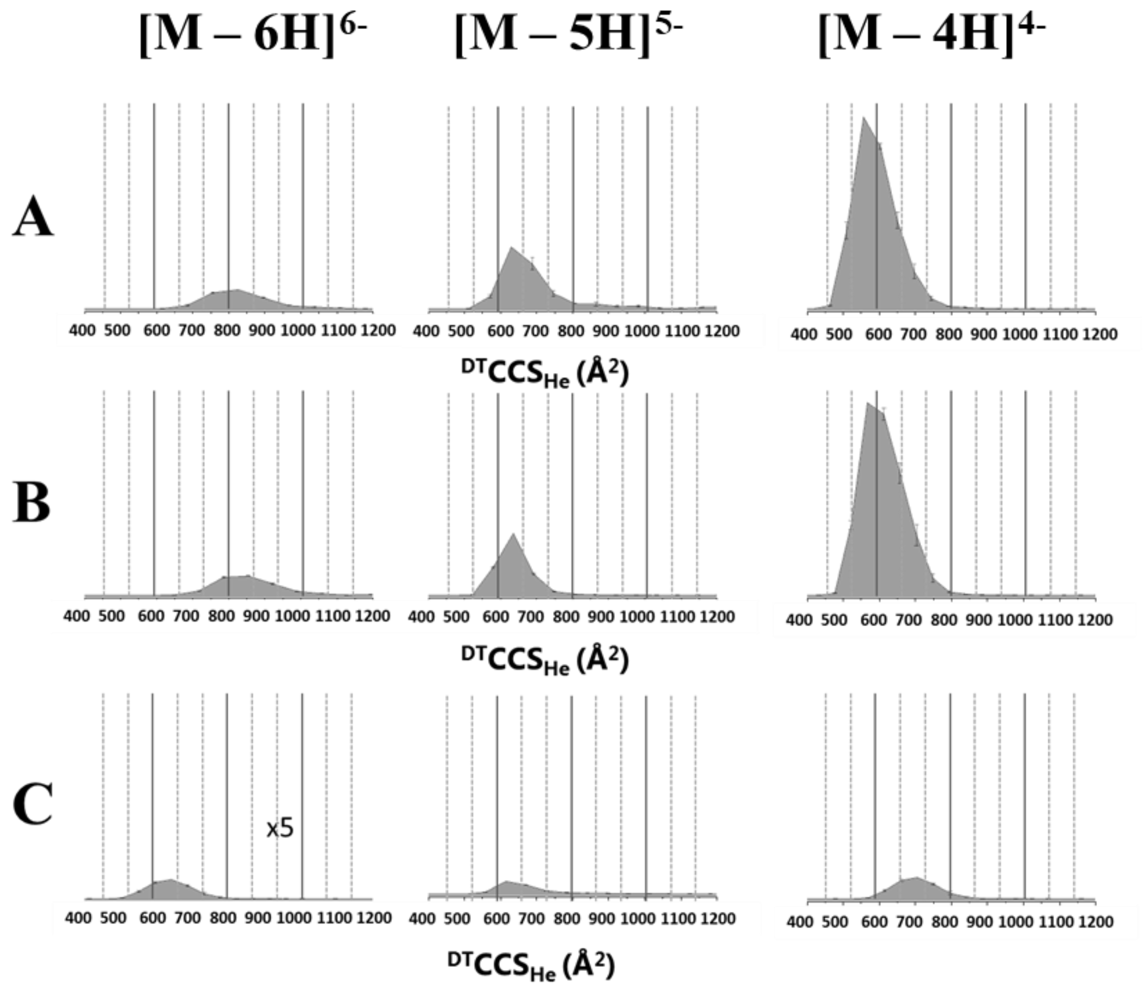


Figure 2.3: $^{DT}CCSD_{He}$ (A^2) for the Ky2 aptamer (20 µM) in the [M – 4H]$^{4-}$ through [M – 6H]$^{6-}$ charge states. Solvent conditions were 100 mM ammonium acetate, 5% methanol. A) Control experiment of Ky2 aptamer with no ligand. B) Unbound Ky2 aptamer in the presence of kanamycin (100 µM kanamycin). C) Bound aptamer:ligand complex (100 µM kanamycin).

The average $^{DT}CCS_{He}$ values and width of the $^{DT}CCSD_{He}$ peak for each charge state of Ky2 in the presence and absence of kanamycin are reported in Table 2.1. The $^{DT}CCSD_{He}$ value is taken from the apex of the distribution, and for a normal distribution this will be the average value. The full width at half maximum (FWHM) for each distribution gives an indication of the flexibility of the molecule. Some width from these peaks is inherent in the system flexibility; however, some of this is an artefact of the system resulting from the original pulse width in addition to ion diffusion whilst travelling in the drift cell. Under buffered conditions, the aptamer:ligand complex (Figure 2.3 C) is present at a much lower intensity than the unbound Ky2 (Figure 2.3 A and B). Upon binding to kanamycin, changes in the

$^{DT}CCS_{He}$ of the complex compared with Ky2 alone were observed (Figure 2.3 A vs Figure 2.3 C). The $[M - 4H]^{4-}$ charge state undergoes an increase in average $^{DT}CCS_{He}$, from 573 Å^2 to 691 Å^2. The $[M - 5H]^{5-}$ charge state shows no change in average $^{DT}CCS_{He}$ upon binding, whilst the $[M - 6H]^{6-}$ charge state undergoes a compaction event to produce a $^{DT}CCS_{He}$ centred on a lower $^{DT}CCS_{He}$ (641 Å^2 compared with 804 Å^2). Additionally, the FWHM is notably smaller for the bound complexes in the $[M - 6H]^{6-}$ and $[M - 5H]^{5-}$ charge states compared with unbound Ky2. It appears that the binding of kanamycin to Ky2 rearranges the aptamer into a single conformation, regardless of the number of charges held on the surface. The average $^{DT}CCS_{He}$ for Ky2 in the presence of kanamycin but not bound is similar to that of Ky2 in the absence of kanamycin, indicating that either a low proportion of kanamycin is able to bind to Ky2 in solution or that the aptamer is able to return to its unbound conformation faster than the timescale of the experiment.

When comparing the results from the methanol:H_2O solution to those from ammonium acetate, it was observed that the denaturing conditions led to larger cross sections throughout. However, the effect of the denaturing conditions is less marked in the $[M - 4H]^{4-}$ and $[M - 5H]^{5-}$ charge states, with only ~10% increase compared to buffered conditions (Figure S2.3, Table 2.1). The $[M - 6H]^{6-}$ charge state has a larger $^{DT}CCS_{He}$ (851 Å^2) with a larger FWHM, suggesting higher flexibility. The $[M - 6H]^{6-}$ charge state of the bound and unbound forms of the complex are of higher intensity under these conditions. It would be expected for a protein that under these conditions denaturing would occur, resulting in much larger conformation. Since it is also observed that the signal owing to the aptamer:ligand complex is of greater intensity under these conditions, these slightly larger structures may be more able to retain the ligand during desolvation.

The $^{DT}CCSD_{He}$ for the bound $[M - 5H]^{5-}$ charge state ion is reduced in width under both conditions, therefore, it can be hypothesized that this charge state provides some optimum distribution, or orientation, of charges across the molecule which improves the strength of kanamycin binding. The $[M - 5H]^{5-}$ charge state also has the highest intensity of ligand bound, again attributable to favourable interactions stabilised at this net charge state.

Table 2.1: $^{DT}CCS_{He}$ of Ky2 aptamer across varying charge state errors taken from three technical replicates

Solvent	Charge state	Bound State	Average $^{DT}CCS_{He}$ (A^2)	FWHM $Å^2$
Ammonium acetate	$[M – 4H]^{4–}$	Control	573 ± 11	145 ± 9
		Unbound	585 ± 6	150 ± 5
		Bound	691 ± 10	140 ± 10
	$[M – 5H]^{5–}$	Control	647 ± 7	140 ± 6
		Unbound	620 ± 10	95 ± 6
		Bound	635 ± 4	120 ± 5
	$[M – 6H]^{6–}$	Control	804 ± 5	195 ± 4
		Unbound	839 ± 3	205 ± 5
		Bound	641 ± 8	150 ± 11
Methanol:H_2O	$[M – 4H]^{4–}$	Control	601 ± 19	115 ± 13
		Unbound	627 ± 6	160 ± 6
		Bound	675 ± 4	150 ± 5
	$[M – 5H]^{5–}$	Control	662 ± 15	140 ± 9
		Unbound	636 ± 15	140 ± 11
		Bound	676 ± 9	135 ± 7
	$[M – 6H]^{6–}$	Control	818 ± 17	200 ± 16
		Unbound	858 ± 7	245 ± 10
		Bound	851 ± 6	265 ± 8

2.5.5. Collision induced unfolding

Collision induced unfolding is used here to further examine the stability of the complexes. Visual heat maps constructed from ion mobility distribution data allow relative ion stabilities under activation conditions to be measured. These experiments considered both the effect of solvent conditions and the sequence of the hairpin region for Ky2 and Ky2–I. The results under ‘buffered’ conditions are displayed in Figure 2.4 and for methanol:H_2O in Figure S2.4. On the heat maps, areas in red represent intensity in a drift time, therefore the most abundant conformations. As collision energy is increased, the molecule unfolds leading to intensity appearing at new drift times due to new structures. For Ky2 (Figure 2.4: A and B, C and D), in all conditions the major species is stable up to 30 V, where it begins to break down into a more extended conformer along with undergoing distinct fragmentation. By 45 V the isolated charge state is entirely lost, though some fragments remain with similar drift times. The bound states require more energy to unfold, indicating that the ligand stabilises the fold, as seen for the $[M - 4H]^{4-}$ and $[M - 5H]^{5-}$ charge states, with less distinct unfolding events. For Ky2–I (Figure 2.4: E and F), unbound forms break down at lower energies than for Ky2. The bound forms (Figure 2.4: G and H) begin unfolding at lower energies with noticeable ligand loss, indicating that the ligand interaction is not as strong and consequently stabilises the fold to a lesser degree.

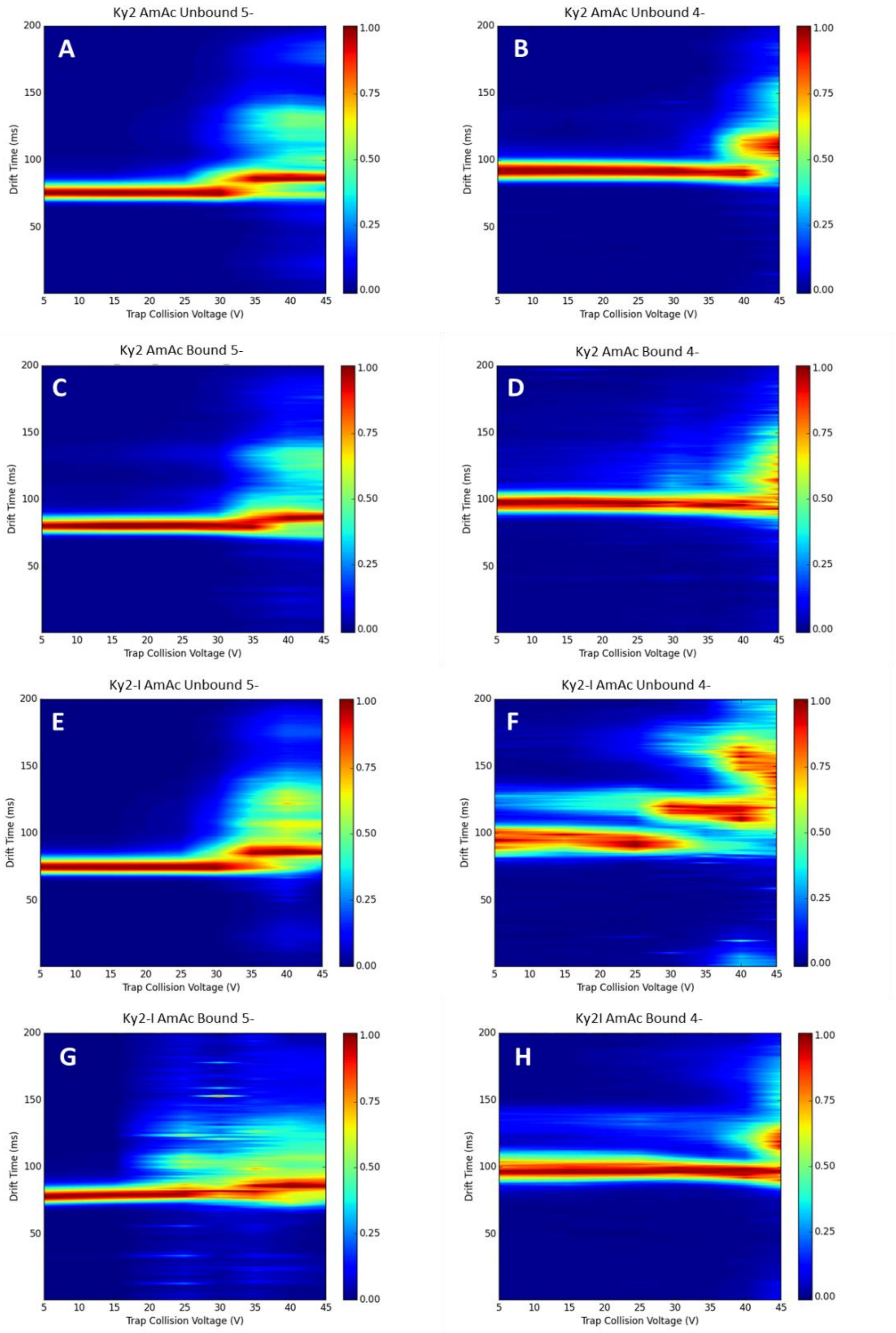


Figure 2.4: Collision induced unfolding heat maps for the [M – 4H]$^{4-}$ and [M – 5H]$^{5-}$ charge state of the unbound Ky2 (A and B) and Ky2-I Aptamers (E and F) along with the bound Ky2 (C and D) and bound Ky2-I (G and H). Red indicates areas of intensity in drift time. Shifting drift times can be seen through pattern changes as energy is increased. DNA aptamer 20 µM, 100 mM, ammonium acetate, 5% methanol, 100 µM kanamycin.

2.5.6. *Fragmentation*

2.5.6.1. CID

CID fragmentation was performed using a Synapt G2 (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer with the $[M - 5H]^{5-}$ unbound or bound aptamer being isolated in the quadrupole followed by fragmentation in the trap region with energies from 5 to 35 V (Figure 2.5). Just as proteins fragment along specific pathways and certain types of fragment are observed depending upon the fragmentation method, DNA follows a similar fragmentation pattern [62]. As expected, mainly w and a–B fragments were observed for both the bound and unbound forms during CID. The unbound form presented nearly a full sequence coverage of w (90%) and a–B (66%) fragments in charge states varying from 1– to 4– (w_{1-19} a–B_{2-15}) with a very small number of low abundance fragments from other possible pathways (a_4, y_2, $d_{1-2,}$ c_{3-4}). When isolating the bound form, a large number of unbound fragments were observed. From the isolated bound form however, a small proportion of w and a–B fragments are observed with the ligand still attached. It is interesting to note that these fragments do not cover the full sequence as observed in the unbound form, but were only observed at $w_{10-14,\ 16}$ and a–$B_{9-13,\ 15}$. This strongly indicates that the ligand is bound within the loop region.

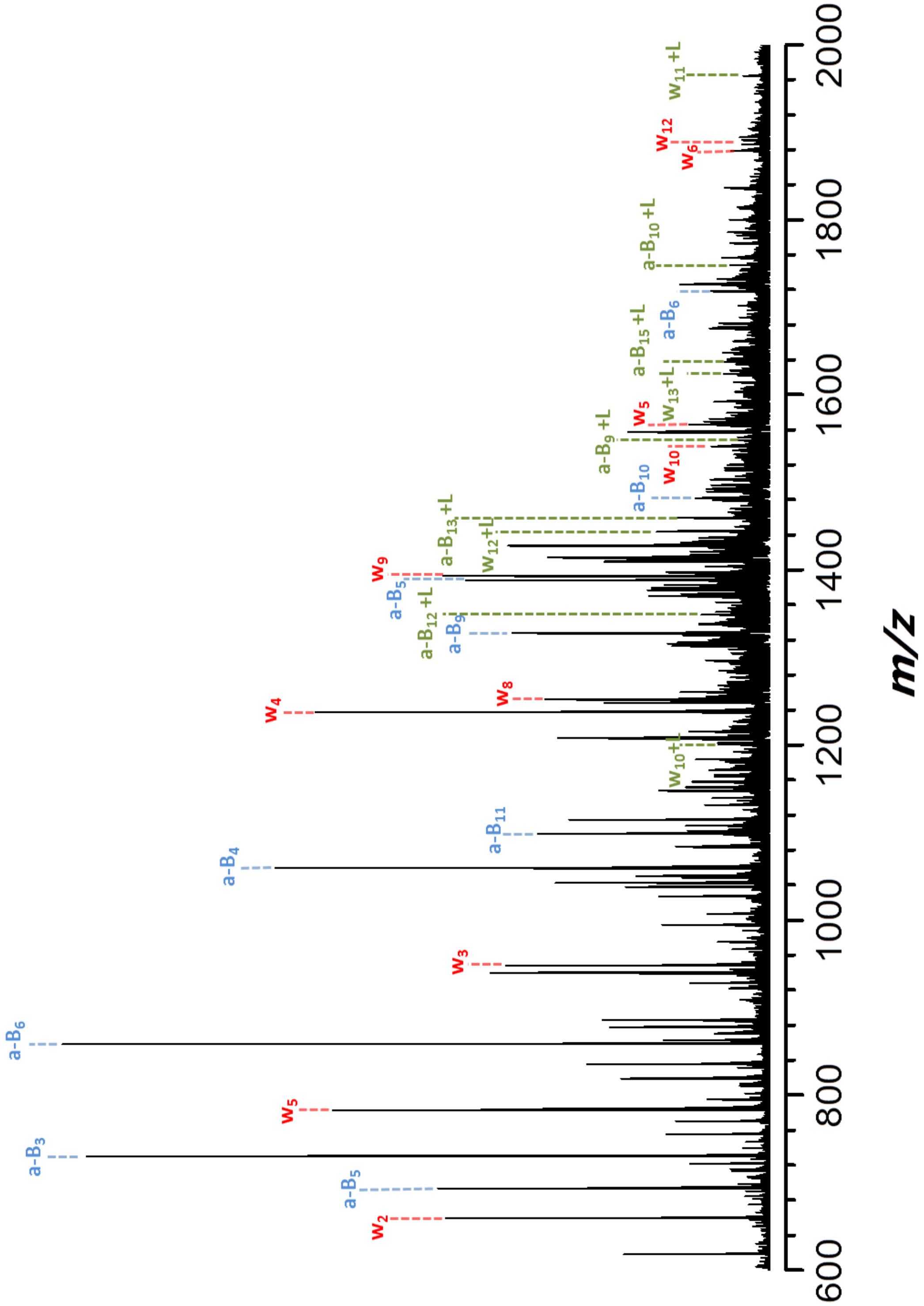


Figure 2.5: n–ESI mass spectrum of Ky2 aptamer:ligand bound complex [M – 5H]$^{5-}$ charge state *m/z* 1409 after isolation and fragmentation at 40 V. w fragments are labelled in red, a–B fragments in blue, and fragments which retain the ligand after fragmentation are labelled in green.

2.5.6.2. UVPD

UVPD was performed on an in–house modified Synapt G2–S mass spectrometer (Waters, Manchester, UK) described previously [65]. For all species analysed, a high degree of electron detachment was observed with the primary loss of a single electron and a lesser loss of a second electron (Figure 2.6). This has been shown previously in other studies, with guanine being the most photo–active nucleotide [64, 68]. Interestingly, irradiation of the bound form led to a greater amount of electron detachment than irradiation of the corresponding unbound form. In addition to charge reduction, fragmentation within the termini of the aptamer was also identified. This was mainly the loss of guanine bases and some chain fragmentation. Fragments were observed for the unbound aptamer for $w_{2–6}$ and $a–B_{2–6}$, some larger fragments (with $a–B_{12}$ the largest fragment) were observed but these did not form a complete series. These fragments were not observed when the bound aptamer was isolated but not irradiated, even though the signal to noise ratio was much lower.

Neither the detachment of electrons nor the loss of bases from the termini of the aptamer led to a large loss of ligand binding, as indicated by the presence of the $[M + L – 5H]^{4–\bullet}$ and $[M + L – 4H]^{3–\bullet}$ bound species. From this, it can be inferred that base and nucleotide loss from the tail region does not interfere with aptamer:ligand binding.

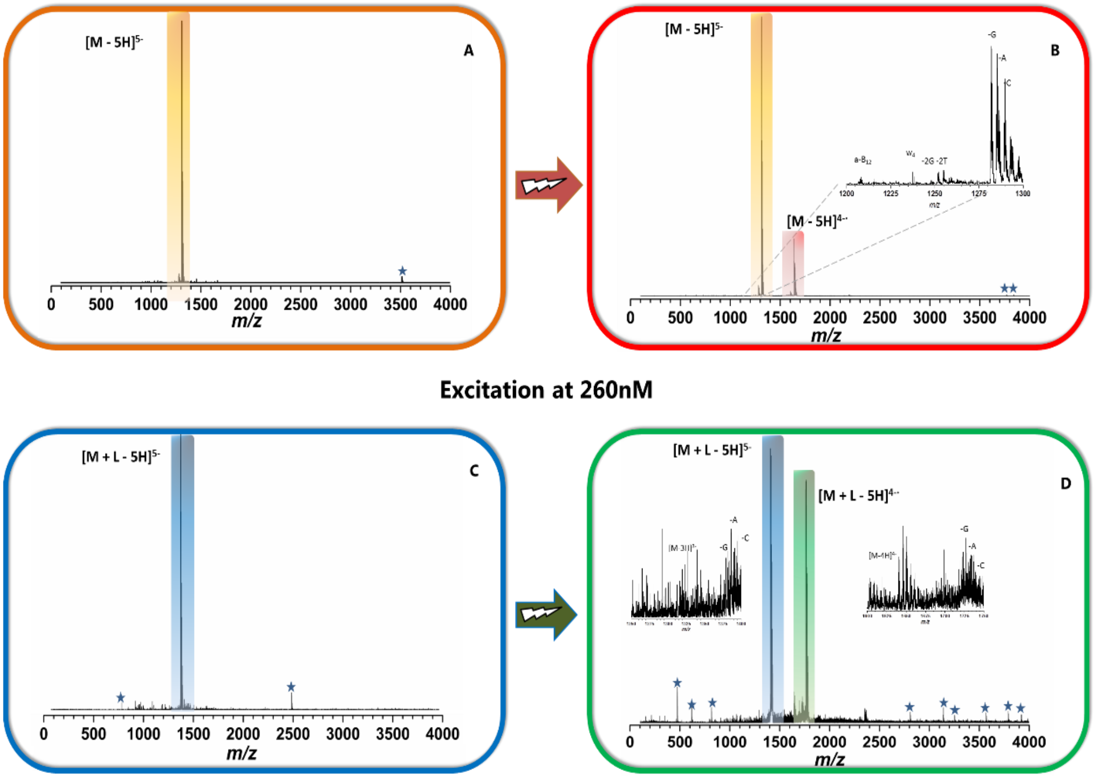


Figure 2.6: A) Isolated [M – 5H]$^{5–}$ unbound aptamer. B) Isolated [M – 5H]$^{5–}$ unbound aptamer after 1 second laser exposure, red bar indicates [M – 5H]$^{4–\cdot}$ charge state caused by single electron detachment. C) Isolated [M + L – 5H]$^{5–}$ bound aptamer:ligand complex. D) Isolated [M + L – 5H]$^{5–}$ bound aptamer:ligand complex after 1 second laser exposure, green bar indicates [M + L – 5H]$^{4–\cdot}$ charge states caused by single electron detachment. Inserts show magnification of major fragments. * represents electronic noise peaks that are not attributable to the analyte.

2.6. Discussion

The purpose of this study was to uncover how a DNA aptamer binds to its target *via* the use of mass spectrometry techniques including: ion mobility, alternative structures, fragmentation, collision induced unfolding and various solvent conditions. It can be inferred from the data with a high degree of confidence, that the loop region is highly structurally important for ligand binding. A paper by Shi *et al.* [69] proposed a G–quadruplex structure for this aptamer which was not observed in these experiments. A G–quadruplex would present with a much higher collisional cross section and drift time during the IM–MS experiments, as well as a different charge state distribution. No evidence of a structure of this nature was observed in these experiments.

A range of Ky2 sequences were created which isolated both the hairpin loop (Ky2–L) and tail regions (Ky2–T), allowing determination of where in the structure kanamycin was binding. Using native mass spectrometry, no binding of kanamycin was observed to occur in the tail region, whereas binding was retained in the loop. The control Ky2 (KY2–C) sequence also did not bind the ligand, indicating that the correct sequence was required; rather than the formation of a loop of any structure. Interestingly, binding was observed using a Ky2 construct with a G↔C A↔T sequence inversion (Ky2–I). In the sequence inversion, the hydrogen bonding chemistry at the bases remained the same. This shows that there is some flexibility allowed in the loop region whilst retaining the ligand.

IM–MS was employed to determine the conformational changes Ky2 undergoes upon kanamycin binding. An extension of the low charge state and a compaction of the high charge was observed when kanamycin was bound, indicating that the binding of kanamycin stabilises the inherently flexible loop region into a single conformation. Collision induced unfolding was used to show that the binding of kanamycin to Ky2 confers structural stability in comparison with the unbound form. Additionally, comparing Ky2–I to Ky2 highlighted that in the absence of the ligand there was little difference in the CIU plots, however, in the presence of the ligand, Ky2–I began unfolding at lower energies, suggesting that the altered construct is unable to form as stable a complex.

Fragmentation provides an excellent insight into how molecules interact; in this set of experiments both CID and UVPD were used to determine how the aptamer:ligand complex is formed. Aptamer:ligand fragments were observed only on residues $w_{10–14,\ 16}$ and $a–B_{9–13,\ 15}$, strengthening the theory that the loop region is the site of ligand binding. These fragments however, were only observed at very low intensity, therefore, other fragments could also be present below the limit of detection. UVPD was employed as a fast fragmentation technique to try and preserve the aptamer:ligand complex during the dissociation process. UVPD led to minor fragmentation in the terminal residues with loss of several bases, along with some nucleotides. The fragmentation was minimal and mainly contained to the terminal residues, and the ligand was retained after these losses, indicating binding in the loop region.

The second major feature of this work was assessing the stability of the system under what, in protein chemistry, would be referred to as 'denaturing' conditions. Under high concentrations of methanol, with identical concentrations of ligand and aptamer, the intensity of signals from bound complexes was higher than under buffered conditions, although It is unclear whether these patterns are a solution or desolvation effect. DT–IM–MS and CIU show increased flexibility under high methanol conditions, though less than would be expected for proteins. This suggests that the DNA aptamers could be more stable than proteins under these denaturing conditions. The collision induced unfolding data also display similar distributions for the unbound $[M - 5H]^{5-}$ charge state in both 'buffered' and 'denaturing' conditions, which could suggest that the aptamer was not perturbed by the altered solvent conditions.

2.7. Conclusions

Mass spectrometry techniques have been applied to the concept of aptamer structure. It has been shown that for the Ky2 aptamer, the loop region of the aptamer is the location of the ligand binding, with the double stranded region providing structure to the system. It has also been revelaed that DNA structure is resilient to methanol mediated unfolding, particularly at low charge states. This work demonstrates the strength of mass spectrometry to structurally characterise non–covalent complexes and map binding sites.

2.8. Acknowledgments

Thanks, are expressed to the DTC in Cell and Proteomic Technologies EP/F500424/1, and BBSRC for awards BB/L015048/1 and BB/H013636/1. Additional thanks, to the British Mass Spectrometry Society for a grant that allowed the purchase of a nanospray tip pipette puller.

2.9. References

[1] O. Y. Fedoroff, A. Rangan, V. V. Chemeris, L. H. Hurley, Cationic porphyrins promote the formation of i–motif DNA and bind peripherally by a nonintercalative mechanism, *Biochemistry.* **39** (2000) 15083–15090.

[2] I. M. Pedroso, L. F. Duarte, G. Yanez, K. Burkewitz, T. M. Fletcher, Sequence specificity of inter– and intramolecular G–quadruplex formation by human telomeric DNA, *Biopolymers.* **87** (2007) 74–84.

[3] C. K. Singleton, Effects of salts, temperature, and stem length on supercoil–induced formation of cruciforms, *J. Biol. Chem.* **258** (1983) 7661–7668.

[4] K. M. Sullivan, D. M. Lilley, Influence of cation size and charge on the extrusion of a salt–dependent cruciform, *J. Mol. Biol.* **193** (1987) 397–404.

[5] I. Russo-Krauss, A. Merlino, C. Giancola, A. Randazzo, L. Mazzarella, F. Sica, Thrombin–aptamer recognition: a revealed ambiguity, *Nucleic Acids Res.* **39** (2011) 7858–7867.

[6] R. B. Kotia, L. Li, L. B. McGown, Separation of nontarget compounds by DNA aptamers, *Anal. Chem.* **72** (2000) 827–831.

[7] H. Ulrich, A. B. Martins, J. B. Pesquero, RNA and DNA aptamers in cytomics analysis, *Cytometry.* **59** (2004) 220–231.

[8] N. Savory, K. Abe, K. Sode, K. Ikebukuro, Selection of DNA aptamer against prostate specific antigen using a genetic algorithm and application to sensing, *Biosens. Bioelectron.* **26** (2010) 1386–1391.

[9] M. Famulok, Oligonucleotide aptamers that recognize small molecules, *Curr. Opin. Struct. Biol.* **9** (1999) 324–329.

[10] S. Jeong, S. R. Han, Y. J. Lee, S. W. Lee, Selection of RNA aptamers specific to active prostate–specific antigen, *Biotechnol. Lett.* **32** (2010) 379–385.

[11] M. Darmostuk, S. Rimpelova, H. Gbelcova, T. Ruml, Current approaches in SELEX: An update to aptamer selection technology, *Biotechnol. Adv.* **33** (2015) 1141–1161.

[12] Y. Kim, M. Gu, Advances in aptamer screening and small molecule aptasensors, *Adv. Biochem. Eng. Biotechnol.* **140** (2014) 29–67.

[13] A. B. Iliuk, L. Hu, W. A. Tao, Aptamer in bioanalytical application, *Anal. Chem.* **83** (2011) 4440–4452.

[14] F. Li, H. Zhang, Z. Wang, A. Newbigging, M. S. Reid, X. F. Li, X. C. Li, Aptamers facilitate amplified detection of biomolecules, *Anal. Chem.* **87** (2015) 274–292.

[15] M. C. Mercier, M. Dontenwill, L. Choulier, Selection of nucleic acid aptamers targeting tumor cell–surface protein biomarkers, *Cancers* **9** (2017) 69–102.

[16] C. L. Hamula, H. Zhang, L. L. Guan, X. Li, X. C. Le, Selection of aptamers against live bacterial cells, *Anal. Chem.* **80** (2008) 7812–7819.

[17] Z. Tang, P. Parekh, P. Turner, R. W. Moyer, W. Tan, Generating aptamers for recognition of virus-infected cells, *Clin. Chem.* **55** (2009) 813–822.

[18] G. Mayer, The chemical biology of aptamers, *Angew. Chemie. Int. Ed. Engl.* **48** (2009) 2672–2689.

[19] K. M. Song, M. Cho, H. Jo, K. Min, S. H. Jeon, T. Kim, M. S. Han, J. K. Ku, C. Ban,

Gold nanoparticle–based colorimetric detection of kanamycin using a DNA aptamer, *Anal. Biochem.* **415** (2011) 175–181.

[20] L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas, J. J. Toole, Selection of single–stranded DNA molecules that bind and inhibit human thrombin, *Nature.* **355** (1992) 564–566.

[21] J. Ciesiolka, J. Gorski, M. Yarus, Selection of an RNA domain that binds Zn^{2+}, *RNA.* **1** (1995) 538–550.

[22] H. P. Hofmann, S. Limmer, V. Hornung, M. Sprinzl, Ni^{2+}–binding RNA motifs with an asymmetric purine–rich internal loop and a G–A base pair, *RNA.* **3** (1997) 1289–1300.

[23] D. Shangguan, Y. Li, Z. Tang, Z. C. Cao, H. W. Chen, P. Mallikaratchy, K. Sefah, C. J. Yang, W. Tan, Aptamers evolved from live cells as effective molecular probes for cancer study, *Proc. Natl. Acad. Sci.* **103** (2006) 11838–11843.

[24] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science.* **246** (1989) 64–71.

[25] R. Wang, B. Chait, High–accuracy mass measurement as a tool for studying proteins, *Curr. Opin. Biotechnol.* **5** (1994) 77–84.

[26] T. R. Covey, R. F. Bonner, B. I. Shushan, J. Henion, The determination of protein, oligonucleotide and peptide molecular weights by ion–spray mass spectrometry, *Rapid Commun. Mass Spectrom.* **2** (1988) 249–256.

[27] S. Frelon, T. Douki, J. L. Ravanat, J. P. Pouget, C. Tornabene, J. Cadet, High–performance liquid chromatography–tandem mass spectrometry measurement of radiation–induced base damage to isolated and cellular DNA, *Chem. Res. Toxicol.* **13** (2000) 1002–1010.

[28] L. A. Haff, I. P. Smirnov, Single–nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI–TOF mass spectrometry, *Genome Res.* **7** (1997) 378–388.

[29] X. Liu, M. A. Lovell, B. C. Lynn, Development of a method for quantification of acrolein–deoxyguanosine adducts in DNA using isotope dilution–capillary LC/MS/MS and its application to human brain tissue, *Anal. Chem.* **77** (2005) 5982–5989.

[30] A. P. M. Loureiro, S. A. Marques, C. C. M. Garcia, P. Di Mascio, M. H. G. Medeiros, Development of an on–line liquid chromatography–electrospray tandem mass spectrometry assay to quantitatively determine 1, *N*(2)–etheno–2'–deoxyguanosine in DNA, *Chem. Res. Toxicol.* **15** (2002) 1302–1308.

[31] R. Singh, P. B. Farmer, Liquid chromatography–electrospray ionization–mass spectrometry: the future of DNA adduct detection, *Carcinogenesis.* **27** (2006) 178–196.

[32] H. Oberacher, P. J. Oefner, W. Parson, C. G. Huber, On–line liquid chromatography mass spectrometry: A useful tool for the detection of DNA sequence variation, *Angew. Chemie. Int. Ed. Engl.* **40** (2001) 3828–3830.

[33] L. Song, S. R. James, L. Kazim, A. R. Karpf, Specific method for the determination of genomic DNA methylation by liquid chromatography–electrospray ionization tandem mass spectrometry DNA methylation that utilizes liquid chromatography, *Anal. Chem.* **77** (2005) 504–510.

[34] C. Siethoff, I. Feldmann, N. Jakubowski, M. Linscheid, Quantitative determination of DNA adducts using liquid chromatography/electrospray ionization mass spectrometry and liquid chromatography/high–resolution inductively coupled plasma mass spectrometry, *J. Mass Spectrom.* **34** (1999) 421–426.

[35] M. Ehrich, M. R. Nelson, P. Stanssens, M. Zabeau, T. Liloglou, G. Xinarianos, C. R. Cantor, J. K. Field, D. van der Boom, Quantitative high–throughput analysis of DNA methylation patterns by base–specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci.* **102** (2005) 15785–15790.

[36] Y. T. Chang, C. H. Lin, C. T. Lee, M. W. Lin, L. Y. Liu, C. C. Chen, D. D. Lee, H. N. Lin, S. F. Tasi, I. Matsuura, Detection of common mutations in sporadic primary localized cutaneous amyloidosis by DNA mass spectrometry, *Br. J. Dermatol.* **170** (2014) 974–976.

[37] N. T. Nguyen–Huynh, J. Osz, C. Peluso–Iltis, N. Rochel, N. Potier, E. Leize–Wagner, Monitoring of the retinoic acid receptor–retinoid X receptor dimerization upon DNA binding by native mass spectrometry, *Biophys. Chem.* **210** (2015) 2–8.

[38] A. Politis, A. Y. Park, Z. Hall, B. T. Ruotolo, C. V Robinson, Integrative modelling coupled with ion mobility–mass spectrometry reveals structural features of the clamp loader in complex with single–stranded DNA binding protein. *J. Mol. Biol.* **425** (2013) 4790–4801.

[39] B. Ganem, J. D. Henion, Y. T. Li, Y. L. Hsieh, Studies of non–covalent interactions of actinomycin D with single–stranded oligodeoxynucleotides by ion spray mass spectrometry and tandem mass spectrometry, *Biol. Mass Spectrom.* **23** (1994) 272–276.

[40] N. Nagesh, G. Raju, R. Srinivas, P. Ramesh, M. D. Reddy, C. R. Reddy, A dihydroindolizino indole derivative selectively stabilizes G–quadruplex DNA and down–regulates c–MYC expression in human cancer cells, *Biochim. Biophys. Acta.* **1850** (2015) 129–140.

[41] X. Ma, S. Shah, M. Zhou, C. K. Park, V. H. Wysocki, N. C. Horton, Structural analysis of activated SgrAI–DNA oligomers using ion mobility–mass spectrometry, *Biochemistry.* **52** (2013) 4373–4381.

[42] A. Triolo, F. M. Arcamone, A. Raffaelli, P. Salvadori, Non–covalent complexes between DNA–binding drugs and double–stranded deoxyoligonucleotides: A study by ionspray mass spectrometry, *J. Mass Spectrom.* **32** (1997) 1186–1194.

[43] X. Cheng, A. C. Harms, P. N. Goudreau, T. C. Terwilliger, R. D. Smith, Direct measurement of oligonucleotide binding stoichiometry of gene V protein by mass spectrometry, *Proc. Natl. Acad. Sci.* **93** (1996) 7022–7027.

[44] S. B. Long, M. B. Long, R. R. White, B. A. Sullenger, Crystal structure of an RNA aptamer bound to thrombin, *RNA.* **14** (2008) 2504–2512.

[45] C. S. Hoaglund, Y. Liu, A. D. Ellington, M. Pagel, D. E. Clemmer, Gas–phase DNA: oligothymidine ion conformers, *J. Am. Chem. Soc.* **119** (1997) 9051–9052.

[46] J. Gidden, A. Ferzoco, E. S. Baker, M. T. Bowers, Duplex formation and the onset of helicity in poly d$(CG)_n$ oligonucleotides in a solvent–free environment, *J. Am. Chem. Soc* **126** (2004) 15132–15140.

[47] E. S. Baker, M. T. Bowers, B–DNA helix stability in a solvent–free environment, *J. Am. Soc. Mass Spectrom.* **18** (2007) 1188–1195.

[48] A. Burmistrova, V. Gabelica, A. S. Duwez, E. De Pauw, Ion mobility spectrometry reveals duplex DNA dissociation intermediates, *J. Am. Soc. Mass Spectrom.* **24** (2013) 1777–1186.

[49] J. Gidden, M. T. Bowers, Gas–phase conformations of deprotonated and protonated mononucleotides determined by ion mobility and theoretical modelling, *J. Phys. Chem. B.* **107** (2003) 12829–12837.

[50] J. Gidden, E. S. Baker, A. Ferzoco, M. T. Bowers, Structural motifs of DNA complexes in the gas–phase, *Int. J. Mass Spectrom.* **240** (2005) 183–193.

[51] E. S. Baker, S. L. Bernstein, M. T. Bowers, Structural characterization of G–quadruplexes in deoxyguanosine clusters using ion mobility–mass spectrometry*, J. Am. Soc. Mass Spectrom.* **16** (2005) 989–997.

[52] E. S. Baker, N. F. Dupuis, M. T. Bowers, DNA hairpin, pseudoknot, and cruciform stability in a solvent–free environment, *J. Phys. Chem. B.* **113** (2009) 1722–1727.

[53] E. D. Leriche, M. Hubert–Roux, C. Afonso, C. M. Lange, M. C. Grossel, F. Maire, C. Loutelier–Bourhis, Investigation of dendriplexes by ion mobility–mass spectrometry, *Molecules.* **19** (2014) 20731–20750.

[54] F. Balthasart, J. Plavec, V. Gabelica, Ammonium ion binding to DNA G–quadruplexes: do electrospray mass spectra faithfully reflect the solution–phase species?, *J. Am. Soc. Mass Spectrom.* **24** (2013) 1–8.

[55] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw, V. Gabelica, G–quadruplex DNA assemblies: loop length, cation identity, and multimer formation, *J. Am. Chem. Soc.* **130** (2008) 10208–10216.

[56] J. P. Williams, J. A. Lough, I. Campuzano, K. Richardson, P. J. Sadler, Use of ion mobility–mass spectrometry and a collision cross–section algorithm to study an organometallic ruthenium anticancer complex and its adducts with a DNA oligonucleotide, *Rapid Commun. Mass Spectrom.* **23** (2009) 3563–3569.

[57] J. Zhou, F. Rosu, S. Amrane, D. N. Korkut, V. Gabelica, J. L. Mergny, Assembly of chemically modified G–rich sequences into tetramolecular DNA G–quadruplexes and higher order structures, *Methods.* **67** (2014) 159–168.

[58] A. Marchand, V. Gabelica, Native electrospray mass spectrometry of DNA G–quadruplexes in potassium solution, *J. Am. Soc. Mass Spectrom.* **25** (2014) 1146-1154.

[59] F. Rosu, E. De Pauw, V. Gabelica, Electrospray mass spectrometry to study drug–nucleic acids interactions, *Biochimie.* **90** (2008) 1074–1087.

[60] B. J. McCullough, J. Kalapothakis, H. Eastwood, P. Kemper, D. MacMillan, K. Taylor, J. Dorin, P. E. Barran, Development of an ion mobility quadrupole time of flight mass spectrometer, *Anal. Chem.* **80** (2008) 6336–6344.

[61] J. D. Eschweiler, J. N. Rabuck–Gibbons, Y. Tian, B. T. Ruotolo CIUSuite: A quantitative analysis package for collision induced unfolding measurements of gas–phase protein ions, *Anal. Chem.* **87** (2015) 11516–11522.

[62] J. Wu, S. A. McLuckey, Gas–phase fragmentation of oligonucleotide ions, *Int. J. Mass Spectrom.* **237** (2004) 197–241.

[63] V. Gabelica, T. Tabarin, R. Antoine, F. Rosu, I. Compagnon, M. Broyer, E. De Pauw, P. Dugord, Electron photodetachment dissociation of DNA polyanions in a

quadrupole ion trap mass spectrometer, *Anal. Chem.* **78** (2006) 6564–6572.

[64] S. I. Smith, J. S. Brodbelt, Characterization of oligodeoxynucleotides and modifications by 193 nm photodissociation and electron photodetachment dissociation, *Anal. Chem.* **82** (2010) 7218-7226.

[65] B. Bellina, J. M. Brown, J. Ujma, P. Murray, K. Giles, M. Morris, I, Compagnon, P. E. Barran, UV photodissociation of trapped ions following ion mobility separation in a Q–ToF mass spectrometer, *Analyst.* **139** (2014) 6348–6351.

[66] M. Busman, A. L. Rockwood, R. D. Smith, Activation energies for gas–phase dissociations of multiply charged ions from electrospray ionization mass spectrometry, *J. Phys. Chem.* **96** (1992) 2397–2400.

[67] J. A. Taraszka, J. Li, D. E. Clemmer, Metal–mediated peptide ion conformations in the gas–phase, *J. Phys. Chem. B.* **104** (2000) 4545–4551.

[68] V. Gabelica, F. Rosu, E. De Pauw, R. Antoine, T. Tabarin, M. Broyer, P. Dugourd, Electron photodetachment dissociation of DNA anions with covalently or noncovalently bound chromophores, *J. Am. Soc. Mass Spectrom.* **18** (2007) 1990–2000.

[69] Y. P. Xing, C. Liu, X. H. Zhou, H. C. Shi, Label–free detection of kanamycin based on a G–quadruplex DNA aptamer–based fluorescent intercalator displacement assay, *Sci. Rep.* **5** (2015) 8125–8132.

3.

The potential of ion mobility–mass spectrometry for tuning synthetic host guest systems: A case study using novel zinc (II) dipicolylamine anion sensors

3. Declaration

This chapter consists of one published experimental paper: C. Nortcliffe, L. G. Migas, X. Liu, H. T. Ngo, K. A. Jolliffe, P. E. Barran.

The potential of ion mobility–mass spectrometry for tuning synthetic host guest systems: A case study using novel zinc (II) dipicolylamine anion sensors

International Journal of Mass Spectrometry, Volume 391, 30 November 2015, Pages 62–70.

This article has been reproduced in an unchanged format except for minor adjustments to incorporate them in to this thesis.

As first author on this publication I performed all experimental work including: MS, Ligand Binding and DT–IM–MS experiments, and drafted the manuscript. I jointly contributed to the theoretical CCS calculations with Lukasz Migas.

The potential of ion mobility–mass spectrometry for tuning synthetic host guest systems: A case study using novel zinc (II) dipicolylamine anion sensors

Chris Nortcliffe[a], Lukasz G. Migas[a], Xuejian Liu[b], Huy Tien Ngo[b], Katrina A. Jolliffe[b], Perdita E. Barran[a]

[a] *The Michael Barber Centre for Collaborative Mass Spectrometry, The School of Chemistry, Manchester Institute of Biotechnology, The University of Manchester, perdita.barran@manchester.ac.uk*
[b] *School of Chemistry (F11), The University of Sydney, NSW 2006, Australia, kate.jolliffe@sydney.edu.au; Fax:+61 2 9351 3329; Tel:+61 2 9351 2297*

Supporting Information

Additional mass spectra are included in appendix 2. Additional computational settings and outputs are included.

3.1. Abstract

Synthetic approaches to produce scaffolds that will act as efficient biosensors have been the subject of considerable research efforts in recent years. The preparation of host complexes which will specifically bind a given guest necessitates structural analysis of the potential complexes formed, as well as an assessment of the strength of the binding interaction. A challenge to the down–stream analytical scientist is to develop a rapid method that can probe the strength of binding, and the architecture of the complex. In this work, nano–ESI mass spectrometry and drift tube ion mobility–mass spectrometry (DT–IM–MS) are employed to examine host guest interactions. Data is presented on 5 cyclic peptides bis–(Zn^{II}) dipicolylamines which have been systematically designed as potential molecular biosensors for pyrophosphate (PPi). The competition of the macrocycles for PPi compared with other small anions were investigated. Mass spectrometry provides the stoichiometry of the complex, and the proportion of complex preserved into the mass spectrometer correlates with solution phase data. It was found that the macrocycles have a higher affinity for PPi than for ATP or pyrocatechol violet (PV). The absolute scale of affinity from mass spectrometry is less (4–5 log K_a) than that found in solution (4–9 log K_a), but within the series of macrocycles trends in relative between the macrocycles affinities are comparable. DT–IM–MS shows that the collisional cross sections ($^{DT}CCS_{He}$) of the macrocycles are unchanged on binding PPi, indicating that the anion is within the designed binding pocket. It was also observed that binding has tightened the conformation of the macrocycle, as revealed from analysis of the collisional cross section distributions ($^{DT}CCSD_{He}$), where for 4/5 of the macrocycles there is a decrease in cross sectional width. Molecular mechanics (MMFF94) and semi empirical methods (ONIOM PM6:B3LYP/aug–cc–pvDZ) were used to generate candidate geometries for **4** to support our assertions on the nature of the binding pocket in the uncomplexed host, and to provide a route for improvement in the design.

Keywords

Ion mobility–MS

Binding studies

Non–covalent complexes

Pyrophosphate

Anion sensing

3.2. Introduction

Mass spectrometry (MS) is widely used for studying many different molecular classes and with the development of soft ionisation, it is now possible to routinely analyse non–covalently bound complexes in the gas–phase [1, 2]. As well as providing *m/z* and hence stoichiometric information on the formed complex, mass spectrometry can also be used to measure the strength of binding interactions, yielding dissociation constants within minutes using only nanogram amounts of sample [3]. Furthermore, using the more recently developed hybrid technique of ion mobility–mass spectrometry (IM–MS), rotationally averaged collisional cross sections can be measured yielding information on conformational changes occurring during complex formation [4–7].

3.2.1. Anion sensing

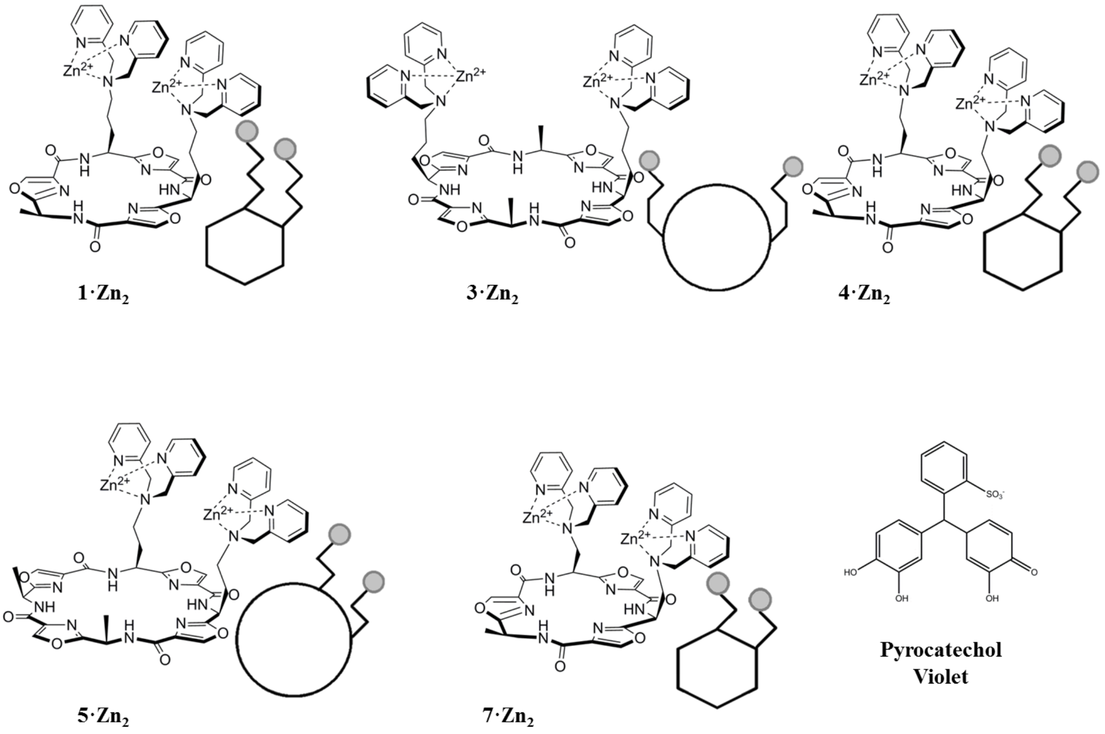
Low molecular weight anions (<500 Da.) have many key biological functions [8–10]. They are often found in metabolomic [11] and cell signalling processes [12, 13], and are also critical to enzyme turnover [14, 15]. Cellular phosphate is an example: kinase enzymes add and remove phosphates to control protein activity and folding [16, 17]; phosphates are also a major component in the DNA and RNA backbones of cells [18–20]. Pyrophosphate (PPi, $P_2O_7^{4-}$), is a by–product of many enzymatic reactions, such as those carried out by DNA transferases and polymerases [21], so the ability to directly detect PPi concentrations would allow real–time analysis of biochemical reactions such as DNA amplification [22]. Furthermore, since the elevation of PPi is implicated in several illnesses (osteoarthritis, uremia, and acromegaly) a developed assay for PPi could be applied for medical benefit [23–25].

One of the first selective inorganic phosphate sensor was designed over 100 years ago, with the development of a molybdenum(IV) assay by Taylor and Miller in 1914 [26]. Although this method has been improved over the years [27–32] it still has several flaws including the generation of toxic metal waste. It is also highly time consuming and suffers from interference from other anions, which can lead to false positives in the results. Other methods have involved chemical addition of chromophores to molecular receptors, however these require special handling, high cost, and complicated synthetic procedures [33].

There has been a great deal of interest in the generation of selective anion sensors for use in aqueous environments [33–49], or which could be used directly upon biological samples; however, this is a challenge due to the high hydration energies of the ions. The large size of the phosphate anion, along with its high hydrophilicity, places it near the bottom of the Hofmeister selectivity series [50]. Further complications are caused by the acid–base properties of phosphate anions, and the large concentration of other biological anions such as chloride, carboxylates, and hydroxide.

Dipicolylamine (DPA) complexes have been proposed as a bio–compatible tool for anion sensing providing a tridentate ligand specific to Zn^{2+} as opposed to other common biological cations. Previous work has generated a series of macrocyclic peptides with two Zn(II)–DPA substituted side arms for the purpose of binding pyrophosphate [51–53] (Figure 3.1). The receptors herein possess similar structural features with an oxazole–modified cyclic peptide of varying sizes with two

DPA arms attached by varying length of linkers; this is shown in Figure 3.1 as full chemical structures and simplified cartoons. Using indicator displacement assays, previous work has revealed that inclusion of small methyl sidechains on the opposing side of the macrocycles to the linking arms favours the specificity of the binding pocket for PPi over ATP. However, differences in peptide binding affinity to the indicators (predominantly pyrocatechol violet) were also observed and the indicator displacement approach does not allow a direct measurement of binding affinity to the anions, providing only apparent K_a values. Synthesis and design of these molecules have been detailed in previously published work [51]. The five molecules (**1**, **3**–**5**, and **7,** Figure 3.1) can be separated into two categories: those with three (**1**, **4, 7**) or four (**3**, **5**) oxazole units within the cyclic scaffold. Within each series differences include modifications of side arm length and position. Solution phase studies of all 5 of the molecules shown in Figure 3.1 showed full indicator displacement upon addition of 1 equivalent of PPi, suggesting a 1:1 complex. It was also found that PPi bound more tightly than all other anions examined. The hypothesized binding mode for PPi–macrocycle, (supported through ^{31}P NMR experiments) involves PPi bridging between the two metal ions [54].



1·Zn_2 3·Zn_2 4·Zn_2

5·Zn_2 7·Zn_2 Pyrocatechol Violet

Figure 3.1: Chemical structures of five macrocyclic peptides designed to bind pyrophosphates that have been investigated here. Molecules 1, 4, and 7 vary through the length of the hydrocarbon linker between the oxazole ring and the dipicolylamine groups. For 1 it is $(CH_2)_3$, 4 is $(CH_2)_2$ and 7 is (CH_2). 3 and 5 include a larger ring with differing placement of the linker arms around the ring these species have the same length of linker arm which is $(CH_2)_2$. Simplified cartoons are presented next to each macrocycle to visualise trends.

3.2.2. IM–MS

Ion mobility–mass spectrometry (IM–MS) is a hybrid technique that separates ions based on mass, charge and conformation. Ion mobility allows the elucidation of structural isomers and ligand interactions by adding an extra dimension to a standard mass spectrometry experiment. Interesting structural features such as different binding modes can be tested against experimental data through the use of molecular modelling strategies to determine which model best fit with observed results. Many supramolecular systems have been studied using IM–

MS as this technique allows assembly mechanisms and intermediate structures to be investigated [55–60].

In this work drift tube IM–MS (DT–IM–MS) was used to study molecular conformations. In a typical DT–IM–MS experiment, ions are stored and pulsed into a cell of known length filled with inert buffer gas across which is a weak (5–40 Vcm^{-1}) electric field. The velocity at which the ion drift through the cell is proportional to the electric field applied, with the constant of proportionality being the mobility of the ion (K) and is dependent on the nature of the drift gas ion interaction and on the temperature of the drift gas. It is common to report K in terms of the reduced mobility K_0 which is normalised for drift gas temperature and pressure. K_0 is inversely proportional to the ions rotationally averaged collisional cross section ($^{DT}CCS_{He}$, Ω, $Å^2$) as described by equation 1.8:

$$\boldsymbol{\Omega} = \frac{\mathbf{3}\boldsymbol{ze}}{\mathbf{16}\boldsymbol{N}}\sqrt{\frac{\mathbf{2}\boldsymbol{\pi}}{\boldsymbol{\mu k_B T}}}\,\frac{\mathbf{1}}{\boldsymbol{K}_0} \qquad \textit{Equation 1. 8}$$

Where K_0 is the reduced mobility (measured mobility corrected to temperature and pressure) z is the ion charge state, e is the elementary charge state, N is the number gas density, μ is the reduced mass of the ion–pair, k_B is the Boltzmann constant and T is the buffer gas temperature. It is possible to generate theoretical CCS from atomistically resolved structures obtained *via* crystallography, NMR or from calculations [4, 61–63]. In order to do this, a number of algorithms which evaluate the CCS for a given set of molecular coordinates are utilised. This synergistic use of experiment and theory to obtain CCS values for comparison can be very powerful. It can be used to suggest candidate geometries to fit experimental findings and hence to understand the relation between structure, function and specificity [64–68].

MS and DT–IM–MS are here employed along with computation to examine five macrocyclic peptides and their binding to pyrocatechol violet, PPi and ATP.

3.3. Experimental details

3.3.1. Materials

All the macrocycles analysed here were synthesized as described previously [51–53, 69]. Aqueous stock solutions of each macrocycle at 500 µM were prepared from lyophilised powder with 5 mM ammonium acetate (Fisher Scientific, UK AR grade), and diluted to 50 µM, 5 mM ammonium acetate with 5% propan–2–ol (Fisher Scientific, UK AR grade) and the pH adjusted to 6.8. It was found that in solution conditions more commonly used for small molecule analysis (50:50 methanol:H_2O 0.1% formic acid) the Zn^{2+} ions bound to the dipicolylamines dissociate and the cyclic peptides were mainly observed with the Zn^{2+} ions unbound (data not shown). This was not the case in the aqueous ammonium acetate conditions, which also are more akin to those used for the solution phase assays [51], where the bound zinc ions could be retained. Although it was necessary to use propan–2–ol to improve the spray quality, especially with solutions containing the phosphate anion. Ligands were prepared at stock concentrations of 1 mM in 5 mM ammonium acetate and then serially diluted to values between 0–400 µM. Ligands used were pyrocatechol violet (Sigma Aldrich, UK), tetra–sodium pyrophosphate (Sigma Aldrich, UK) and adenosine 5′–triphosphate disodium salt hydrate (Sigma Aldritch, UK). Each ligand was made up to a concentration of 50 µM and was incubated with a macrocycle at an equivalent concentration for a stoichiometry of 1:1, for 20 minutes. No increase in binding was observed with longer incubation. For the K_d assay (see below) experiments were repeated at a range of anion concentrations (0–400 µM).

Nano–electrospray tips were made using a Flaming/Brown micropipette puller (Sutter Instrument Company, Novato, CA, USA) using 0.5 mm walled glass capillaries. Ionisation potentials were applied through platinum wire (0.125 mm Goodfellow) inserted into the pulled glass tips.

3.3.2. n–ESI–mass spectrometry

A Q–ToF 2 (Waters, Micromass, Manchester, UK) quadrupole time–of–flight mass spectrometer with an orthogonal nano–electrospray source was used to obtain all mass spectra. Instrument parameters in the source region were optimised to retain non–covalent complexes and maintained across experiments. Particular care was taken to minimise the source cone voltage since this correlated with optimum successful transfer of species that were complexes. Typical instrument source operating settings involved: capillary voltage ~1.4 kV, cone voltage ~25 V, source temperature 80°C. Experiments were performed with three technical repeats; values reported are mean with standard deviation.

3.3.3. Determination of dissociation constants from n–ESI–MS data

Measurement of K_d from the mass spectra was performed based upon the method of Sannes–Lowery, Griffey and Hofstader [3] with titrations of the peptide against solutions containing one ligand, along with two–ligand competitive binding experiments. n–ESI–MS allows the relative stoichiometries and abundances of both the *holo* and *apo* complexes to be detected simultaneously *via* their monoisotopic peak intensities. The intensities of each species including adduct peaks was summed and normalised to total macrocycle concentration. This is shown in equation 3.1 where R is the free macrocycle concentration, L is the free ligand, and RL is the 1:1 macrocycle– ligand complex.

$$\boldsymbol{K_d} = \frac{[\boldsymbol{R}][\boldsymbol{L}]}{[\boldsymbol{RL}]} \qquad \text{Equation 3.1}$$

The amount of bound and unbound species was then compared and converted into the values of concentration of free ligand and ratio of molar concentration of bound ligand to total molar concentration of macrocycle. The total peaks owing to the macrocycle should be equivalent in intensity to the total concentration of macrocycle added. K_ds were obtained by plotting these values against each other as [RL]/[R] vs [RL] yields a slope of $1/K_d$.

Several assumptions were made in the fitting of the K_d measurements which may have influenced the values obtained. Firstly, it was assumed that all the species behaved similarly in the source of the instrument with respect to ionisation and retention of bound complexes. Next, creating the baseline of the unbound species, all species owing to the macrocycle were utilised including those that did not have Zn^{2+} and those which had Na^+ and K^+ bound. However, the absence or presence of these metal ions will have had an effect on the affinity of the macrocycle for the ligands, altering the results. Peak intensity was measured using the monoisotopic peak rather than the integration of peak area however previous work has shown this has little effect on K_d measurements when the mass of the complex and the uncomplexed ions are relatively similar [70, 71]. Finally, the starting concentrations were created from stock solutions and are therefore subject to errors present in balances and pipettes which may have altered final solutions. K_d measurements are in µM and K_a are in M. Data is reported in K_a when making comparisons to previous work (where this was the form of the equilibrium constant used) [51]. When comparing within the work presented here, K_d is used for clearer comparison of trends within data

3.3.4. Anion competition assay

n–ESI MS was used to determine the competition for binding of PPi and PV to selected macrocycles (**3** and **5**) because they both showed strong binding to each anion. Macrocycles were incubated with PV (100 µM) for 20 minutes leading to the macrocycle:PV complex accounting for ~100% of the spectra. PPi was thereafter added to solutions in increasing concentrations and spectra were obtained measuring the appearance of the macrocycle:PPi complex.

3.3.5. Ion mobility–mass spectrometry

IM–MS experiments were performed using an in–house modified Q–ToF 1 (Waters, Micromass, Manchester, UK) quadrupole time–of–flight mass spectrometer that includes a 5.1 cm drift cell (pressurised to ~4 Torr with helium) located before the quadrupole as described previously [7]. The voltage across the drift cell was altered from ~60 – 20 V in seven discrete steps whilst temperature and pressure were recorded at each step and averaged over the acquisition time. Data were acquired for 5 minutes at each voltage and summed with all arrival–time distributions recorded synchronously with each injection of ions through the cell. Rotationally averaged collisional cross sections ($^{DT}CCS_{He}$) were obtained by plotting experimental arrival times against *P/V* for each drift voltage (V) as described previously [7, 72, 73]. Experiments were performed with three technical repeats; values reported are mean with standard deviation.

3.3.6. Calculations

Molecular structures were built using ChemDraw [74] with previous ^{31}P NMR restraints in place [51]. The structures were pre–optimised using AVOGADRO and the MMFF94 level of theory. This was followed by ONIOM PM6:B3LYP/aug–cc–pvDZ calculations in Gaussian09 [75]. In previous studies, ^{31}P NMR shifts have shown that PPi binds to Zn^{2+} *via* all four negatively charged O–P oxygen atoms in a symmetrical fashion. This binding mode was considered when generating potential structures. The structures were then used for collisional cross section calculations using the trajectory method in MOBCAL [76, 77] and projection superposition approximation (PSA) [78–81]; the calculated values were then compared to drift–time ion mobility measurements [82]. Full details are provided in supplementary information.

3.4. Results

3.4.1. n–ESI MS to provide anion affinity

Figure 3.2 shows n–ESI spectra in 5 mM ammonium acetate for **3** with each of the anions examined here. In the full spectra (supporting information Figure S3.1–5) all five macrocycles present mainly in two charge states with the majority for each as a $[M - 2\ H \cdot 2\ Zn]^{2+}$ ion which is shown for **3** in an expanded region of the mass spectrum in Figure 3.2 (a–d). The assignment of the charge and elemental composition of each ion is supported by fitting theoretical isotopic distributions to the abundant ions (see inserts in Figure 3.2 (a–d)), and by the excellent mass agreement; for example, **3** +2 species a measured mass of *m/z* 564.145 *versus* a predicted of *m/z* 564.142. Though the $[M - H \cdot 2\ Zn]^{3+}$ ion is observed at low abundance, there is no evidence of the $[M \cdot 2\ Zn]^{4+}$ species. For each macrocycle, the +2 ion is most abundant and represents the macrocycle with the loss of two H^+ ions with two Zn^{2+} ions resulting in an overall charge of +2 (see Figure 3.2 and supporting Figure S3.1–5). For the bare macrocycle (Figure 3.2 (a) and S3.1–4 (a)) some ions are observed due to the loss of one or both Zn^{2+} ions, this effect is most pronounced for **7**, which indicates that the binding of zinc ions is not as favoured for **7** as for the other macrocycles, a point which will be discussed later. The side chains are shortest on **7** affording it less conformational flexibility and possibly some steric hindrance for binding of the Zn^{2+} ions. Some binding of acetate was also observed in the unbound species, likely to the highly positively charged Zn^{2+} ions, and highlighting the softness of our n–ESI source.

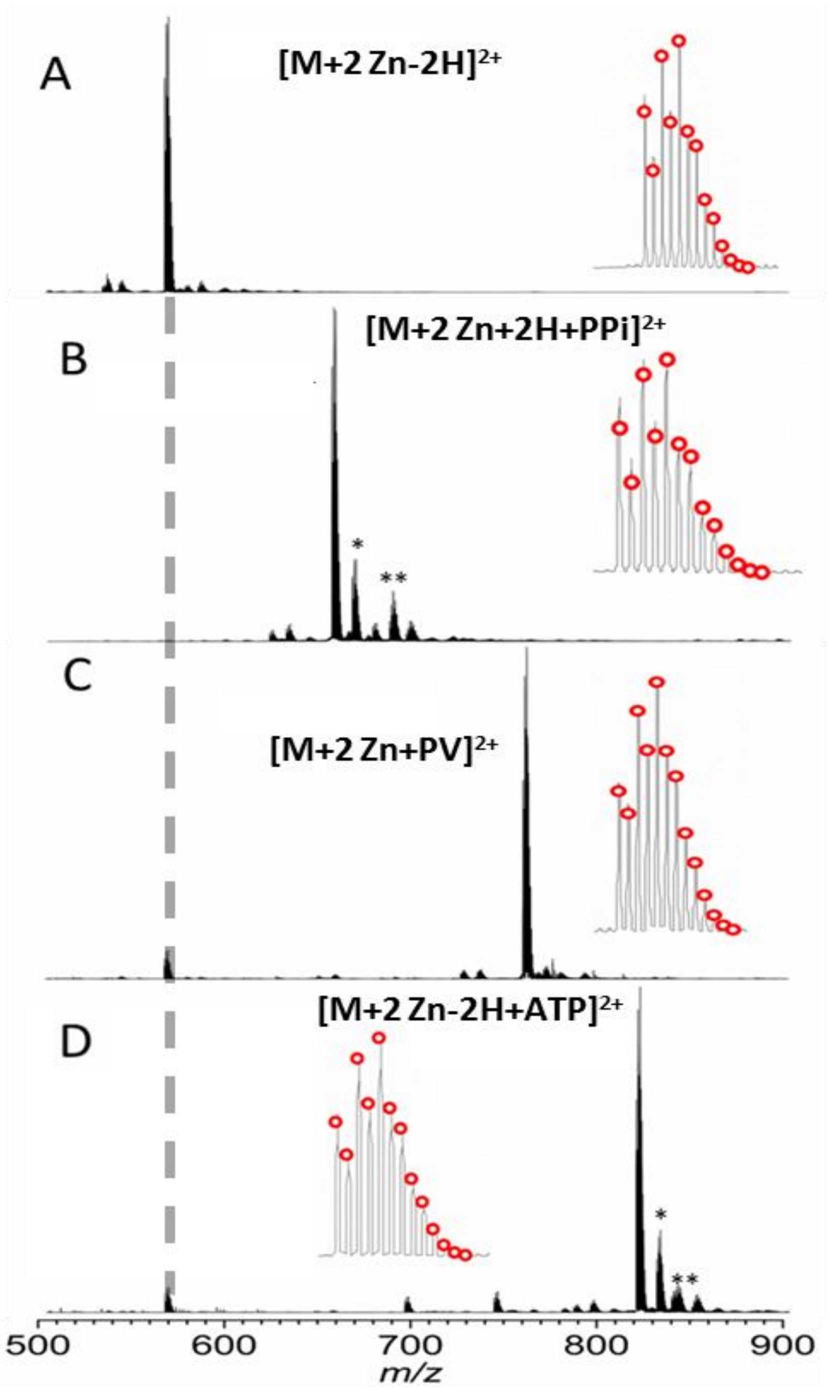


Figure 3.2: n–ESI spectra of the 2+ charge state of 3 for (A) the *Apo* form, and with:(B) PPi^{4-}, (C) PV^{1-}, (D) ATP. Insets show theoretical isotope fitting of labelled species superimposed on experimental isotope distribution. Conditions: 50 μM 3, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 μM sodium pyrophosphate, (C) 200 μM pyrocatechol violet, (D) 100 μM adenosine 5′–triphosphate (ATP) disodium salt hydrate. Impurity peaks are observed and annotated as sodium (*) and potassium (**) adducts.

Ligand bound species are observed at both charge states, with the ratio between bound and unbound being highly comparable for each charge state. In addition, upon ligand binding there is no shift in the relative intensities of the unbound forms (Figure 3.2 B–D and Figure S3.1–4 B–D). It is interesting to note that upon ligand binding, all macrocycle:anion complex peaks are always observed with two Zn^{2+} ions bound, and no zinc loss species; this strongly suggests that in each case the ligand stabilises the dipicolylamine–zinc interactions in the gas–phase and that the ligand binds preferentially to the di–zinc form of the macrocycle, as has previously been suggested by Hamachi and co–workers for other bis–(Zn^{II}) dipicolylamine complexes [83]. **7** exhibits the lowest affinity for PPi (Figure 3.4 B) which can be attributed to steric hindrance resulting from the short linker arms between the scaffold and dipicolylamine group. The two macrocycles with the larger scaffolds **3** and **5** display similar ratios between the intensities of the bound and unbound species despite the linker arms being located at various positions on the macrocycle, implying that there is sufficient flexibility between them that the mode of binding is unaffected by this orientation.

Table 3.1: K_d values with associated errors for macrocycle binding experiments with a series of anions

	PPi	PV	ATP
$1 \cdot Zn_2$	4.9 ± 0.7	16.2 ± 1.4	
$3 \cdot Zn_2$	12.3 ± 2	25.8 ± 1.8	16 ± 0.5
$4 \cdot Zn_2$	3.5 ± 0.3	5.5 ± 0.5	
$5 \cdot Zn_2$	13.7 ± 2.3	18 ± 1.9	17.9 ± 2.6
$7 \cdot Zn_2$	14.9 ± 1.7	40.5 ± 8.9	

Errors quoted are associated with standard deviation of three biological repeats and accuracy of fitting.

To quantify the effects of anion binding, the intensities of the monoisotopic peaks from each species were used to determine binding affinities for each ligand following a method adapted from that previously described [3]. Table 3.1 reports the K_d values obtained from the n–ESI–MS titration experiments, which can be compared with those obtained previously (Table S3.1) There is generally good agreement between the n–ESI titration data and the colorimetric/UV binding assay; with both approaches the macrocycles showed higher affinity for PPi than PV, however a greater affinity for ATP than PV was observed, for both **3** and **5**, which differs from solution findings (Table 3.1 *cf.* Table S1). With the addition of a 1:1 equivalent of PPi to macrocycle, the unbound peak had almost totally disappeared in favour of the bound peak. The K_d values are displayed in Figure 3.3, illustrating the relative affinity across macrocycles and ligands. All the trends match well with solution phase data with **4** showing the tightest binding to PPi and **7** showing the greatest discrimination between ligands. Lower dissociation constants are observed using MS than in the previous indicator displacement assays which can be attributed to several factors including: loss of anion during ionisation, the limit to µM concentrations causing inaccuracy for strong binding, and related issues with the dynamic range of the mass spectrometer as a detector. Solution phase data was measured in log K_a therefore this will be used to compare the data in this format. Looking specifically at **5** the PPi binding constants were >9 in solution and 4.8 through mass spectrometry whilst ATP binding constants were 6.7 and 4.7 respectively. Mass spectrometry consistently shows weaker binding than found by UV. The strongest binding observed through MS is for **4** at 5.5 log K_a which is several orders of magnitude lower than solution. As shown in the original work differing buffers led to different binding strengths, therefore it is safe to assume that the ammonium acetate values will differ from HEPES and Krebs buffers, however the retention of the trends points to the value of the techniques employed.

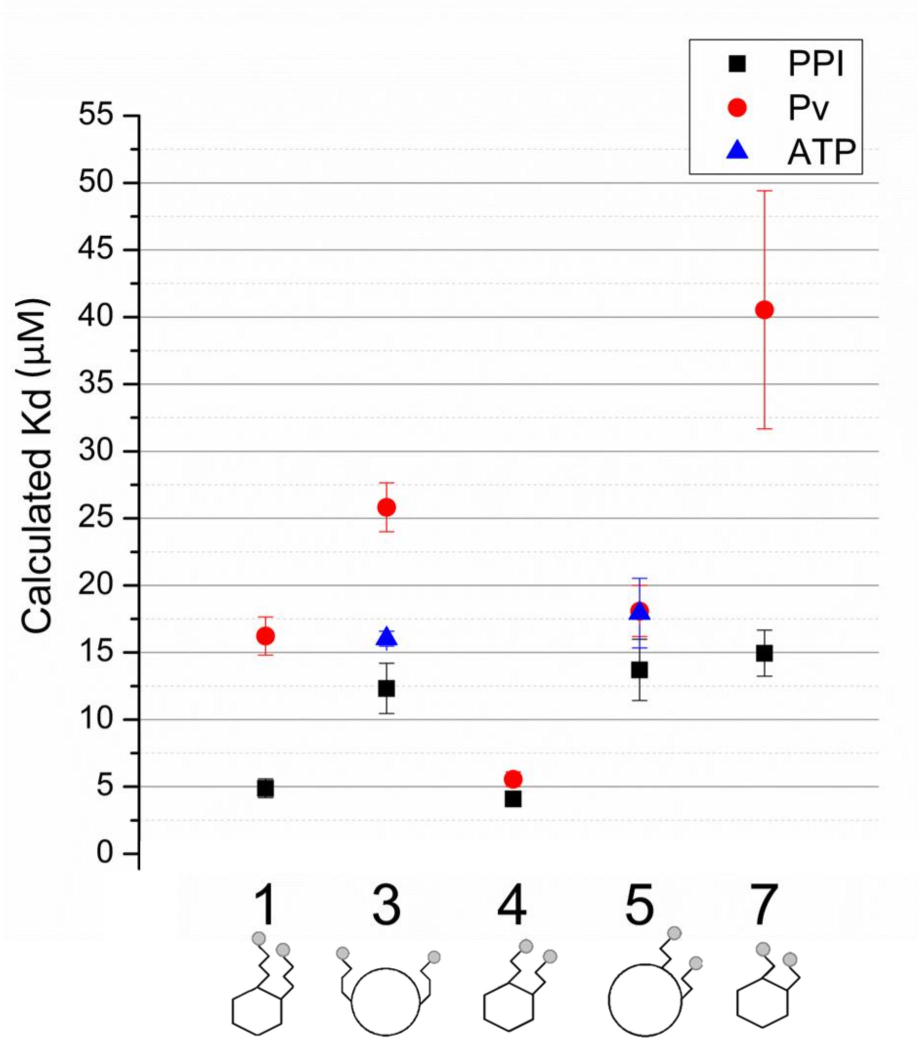


Figure 3.3: K_d (μM) measurements for all anions and macrocycle molecules displayed with error bars to observe trends. Simplified cartoons of structures are shown beneath.

3.4.2. Competition assay

Experiments were carried out to measure how easily PPi could displace PV from a macrocycle:PV complex. This experiment was designed to accurately recreate the solution work and to show the effect of PV displacement on PPi dissociation constant. It was found that PPi easily displaced PV with a similar ratio of bound vs. unbound PPi observed (see Figure 3.4 for data at stoichiometric ratios). This supports the solution phase results and the design rationale whereby the macrocycle bound to the dye is easily displaced by PPi leading to a visual colour change that can be measured.

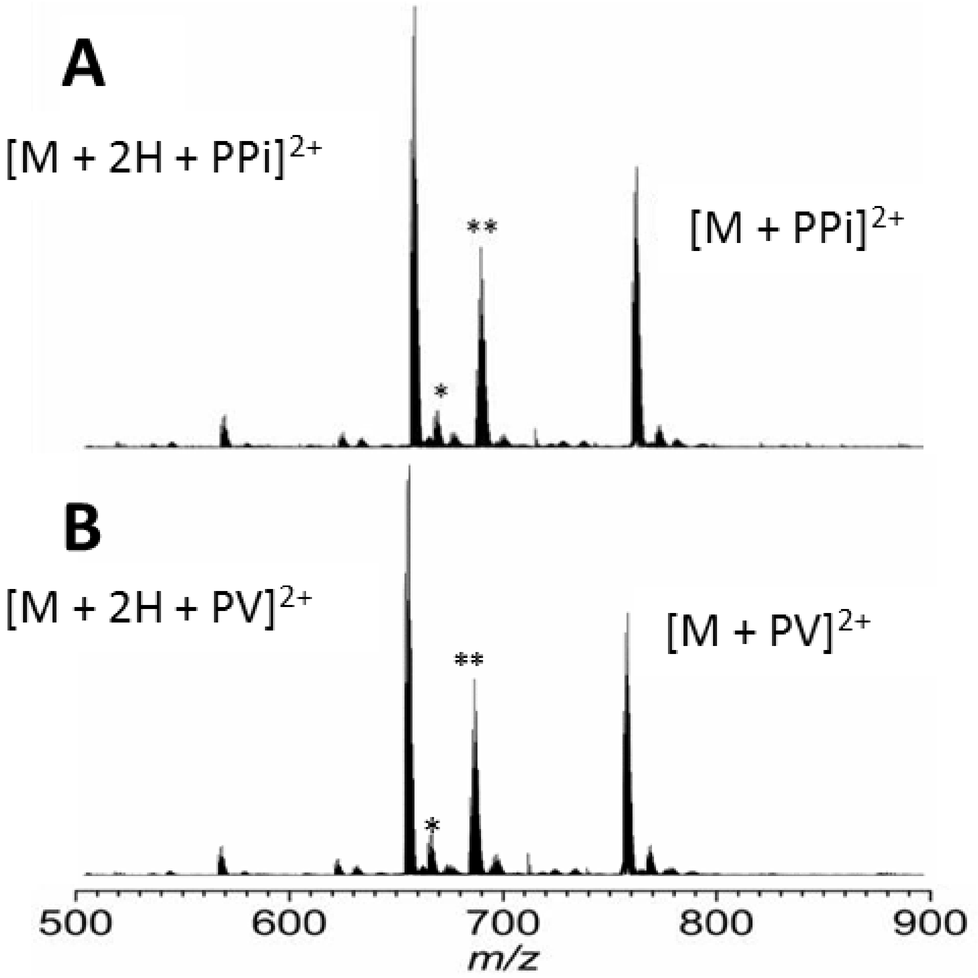


Figure 3.4: n–ESI spectra of the 2+ charge state for (A) competition experiment between 3 with 50 µM PV, and 50 µM PPi. (B) Competition experiment between 5 with 50 µM PV, and 50 µM PPi impurity peaks are observed and annotated as sodium (*) and potassium (**) adducts.

3.4.3. IM–MS

IM–MS experiments were carried out on each *apo* macrocycle along with each of the three phosphate ligands. Collisional cross sections were obtained through measurement of arrival time distributions (ATDs) across a range of drift voltages which were converted to $^{DT}CCS_{He}$ through the use of equation 1.8. This data is presented in Table 3.2 and Figure S 3.5 and allows us to correlate binding and structural information.

Table 3.2: Rotationally averaged collisional cross sections (Å^2) ($^{DT}CCS_{He}$) for all macrocycles and ligands

	Unbound	PPi	PV	ATP
1 · Zn_2	218.9 ±1.8	218.6 ± 0.6	275.5 ± 5.4	256.52 ± 3.8
3 · Zn_2	240.7 ± 3.7	264.2 ± 2.2 /282.9 ± 3.4	289.7 ± 3.9	276.7 ± 3.2
4 · Zn_2	207.9 ± 3.2	206.7 ± 2.1	261.1 ± 6.4	245.6 ± 5.2
5 · Zn_2	217.4 ± 1.1	217.9 ± 2.9	277.5 ±1.9	265.9 ± 2.5
7 · Zn_2	201.6 ± 2.3	215.8 ± 5.2	263 ± 3.5	252.8 ± 9.4

Errors quoted are associated with standard deviation of three biological repeats and accuracy of fitting.

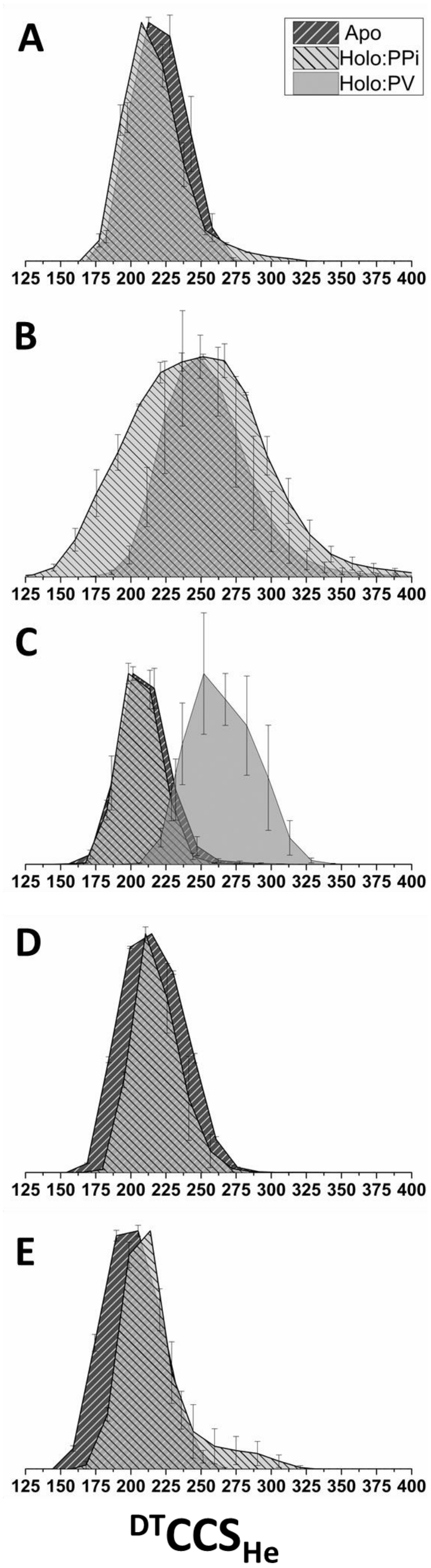


Figure 3.5: Collison cross section distributions ($^{DT}CCS_{He}$) for apo and PPi:*holo* complex of all macrocycles [(A)1 , (B) 3, (C) 4, (D) 5, and (E) 7]. All the macrocycles show evidence of conformational tightening in the presence of the ligand (except 3). 4 has been shown with the $^{DT}CCSD_{He}$ of the PPi:*holo* complex to show that upon PV binding there is a large shift in $^{DT}CCS_{He}$ and an expansion of the structure.

In Figure 3.5 similar macrocycles are compared to obtain trends from the ion mobility data. The image presents cross section intensity intensity on the Y axis *versus* $^{DT}CCS_{He}$ (Å^2) on the X axis. For **1**, **4** and **7** with each further reduction in linker arm length results in a corresponding reduction in cross section in the *apo* form of the complex from 219 Å^2 to 201 Å^2. However, when binding ligands between **1** and **4** there is a reduction in $^{DT}CCS_{He}$ as expected from 218 Å^2 to 207 Å^2, however **7** shows a larger $^{DT}CCS_{He}$ for all *holo* complexes than expected (216 Å^2). **7** has the shortest linker arms of any macrocycle which may prohibit anions binding as tightly as in the other systems. Macrocycles **3** and **5** have equal length linker arms located at various positions on the ring. **5** shows a smaller cross section than **3** for the all complexes though the differences are more pronounced for the *apo* and PPi:*holo* complex. Complexes **1, 4,** and **5** all show a reduction in $^{DT}CCS_{He}$ upon binding of PPi which indicates conformational tightening from a hollow complex to a filled one [81]. This tightening is best realised by a consideration of the widths of the collisional cross section distributions ($^{DT}CCSD_{He}$) where it can be clearly observed that the PPi:*holo* complex is narrower than the *apo* macrocycles (Figure 3.5). **7** shows an increased $^{DT}CCS_{He}$ upon PPi binding however the width of the $^{DT}CCSD_{He}$ is reduced. **3** PPi:*holo* presents a wide $^{DT}CCSD_{He}$ (Figure 3.5 B) with two major conformations observable in some ATD (data not shown). The $^{DT}CCSD_{He}$ ranges from a form that is smaller and more compact than the **3** *apo* form, towards a more extended conformer. The *apo* form of **3** already has the broadest distribution from ~175 Å^2 up to ~350 Å^2. These broad distributions are likely due to the alternate placement of the linker arms.

3.5. Calculation results

In vacuo and solvated molecular dynamics calculations were carried out on **4**; the MD trajectories were analysed for the radius of gyration to determine the compactness of the modelled structures as well as root mean square deviation (RMSD) to determine the similarity of the model to the lowest energy structure (Figure S3.6). Approximately 50 structures were extracted from each trajectory and their CCS_{He} were calculated in MOBCAL [76, 77] and PSA [78–81].

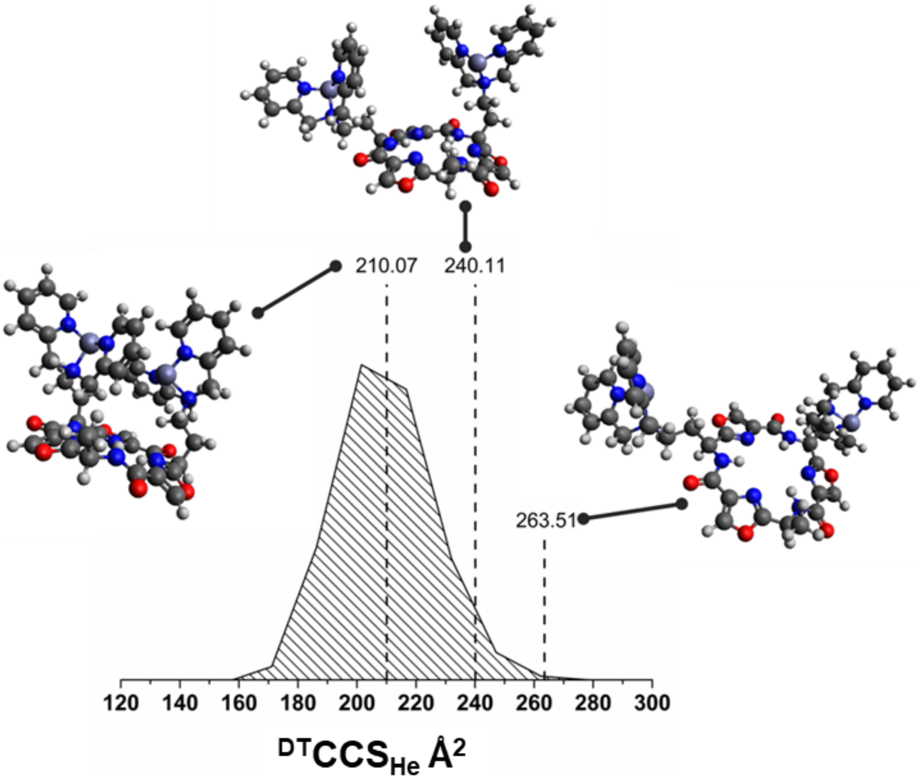


Figure 3.6: Experimental ($^{DT}CCSD_{He}$ Å^2) collisional cross section distribution for 4. Annotations denote computationally generated structures with CCS_{He} predicted using the trajectory method.

The structures fell broadly into two distributions; one more compact and one extended (supplementary data Figures S3.6 and S3.7). *In–vacuo* the compact structures range between ~210 to ~225 Å^2, whilst the extended group range from ~230 to ~270 Å^2. Structures as extended as this, which were not observed during the experiment, have a common feature of the linker arms spreading out from the macrocycle forming more flattened structures where Zn^{2+} ions are as separated as possible. The compact cluster retains a similar structure to that of the starting model with the linker arms extending above the macrocyclic ring. Figure 3.6 shows the experimental collisional cross section distribution for the unbound macrocycle, which has been converted to a $^{DT}CCS_{He}$ and superimposed upon this, is an example low energy structure for each of the compact, extended geometries observed. A lower populated geometry, is classified 'intermediate' and is also shown.

3.6. Discussion

n–ESI mass spectrometry has been used to examine the structure of complexes formed between a series of peptide based macrocycles and biologically relevant anions, and to measure directly the strength of the binding interaction. The same principles have been followed that are common in the field of native mass spectrometry, where the synthetic complex is incubated in an aqueous volatile buffered solution and directly infused with nano–ESI conditions that are kept as gentle as possible. All macrocycles have affinities of 15 µM or less for PPi, and each complex is readily observed, which supports the design criteria and the direct infusion approach taken. Macrocycles **1** and **4** have the highest affinity for PPi indicating that the optimum linker arm length for this anion is 2–3 carbon atoms. When the linker arm was reduced to a single carbon as in **7** the affinity for PPi is reduced, and the discrimination between PPi and PV was increased. For **3** and **5** where the arm location has been altered, the binding is also weaker than for **1** and **4**, implying that both arm length and location are critical for binding. Arms that are too long provide too much flexibility lowering the strength of binding, and those too short led to steric hindrance. Additionally, arms in the wrong location do not provide as effective geometry for binding to occur, or, it is weaker. The mass spectrometry method indicates less discrimination between the macrocycles for each anion (Table 3.1 *cf.* Table S3.1) although values found in the solution assays were taken with higher buffer strengths and were buffer dependent suggesting some competing effects that have not been examined here with the low salt concentration (5 mM) of the volatile ammonium acetate.

$^{DT}CCS_{He}$ of PPi and four of the macrocycle complexes (**1, 3, 4,** and **5**) change little on binding PPi suggesting the pocket is pre–configured for binding this anion, again an indication of good design. This is strongly highlighted in Figure 3.5 whereby a narrowing of the collisional cross distribution is observed indicating a reduction of flexibility. This reduction in width of the $^{DT}CCSD_{He}$ is in the order of ~10 to ~50% (FWHM) between macrocycles. Given the mass increase on binding PPi, the equivalence of the $^{DT}CCS_{He}$ values for the bound and unbound forms implies conformational tightening in the complex. By manually docking the anion into our low energy structure and using the trajectory method [66, 67] a pseudo *holo* CCS_{He} of ~250 Å^2 is obtained, which is larger than the *apo* form by 42 Å^2. This highlights how well the anion fits into the designed binding pocket to provide a complex with the experimental $^{DT}CCS_{He}$ of 207 Å^2. By contrast the **7**:PPi complex is 7% larger than *apo* **7**, suggesting a less well configured binding site. For each macrocycle, the complexes formed with PV and ATP are larger than found for the *apo* and PPi bound forms, with the PV complexes consistently the largest (Table 3.2 and Figure 3.5); this combined with reduced affinity measured (Table 3.1) suggests that the binding pocket is not as well tuned for PV. It can also be seen in Figure 3.5 (C) of the **4**:PV *holo* complex that there is a broadening of the distribution further providing evidence that the binding pocket is not configured for this ligand. The larger complexes observed for ATP and PV may be due to steric hindrance from the non–binding side chains on the cyclic scaffold. PPi binds completely between the two Zn ions and therefore does not come in contact with these methyl groups. This evidence of conformational tightening only in the presence of the targeted anion supports the receptor design process and viability.

Given that these macrocycles have been designed to bind PPi, it is interesting to consider PPi binding and $^{DT}CCS_{He}$ trends as a function of decreasing arm length (**1**, **4**, **7)** along with increasing the ring size (**4** *cf*. **5)**. Between **1** and **4** there is a small decrease in cross section as would be expected but no great change in pyrophosphate binding, although, there is less specificity for PPi over PV. This could be due to an optimum flexibility in the arms that facilitates the most pre–organised binding site; the further reduction of arm length in **7** has reduced affinity for both anions as well an increase in cross section upon ligand binding. This is likely due to the shorter arms making ligand access more restricted forcing a greater conformational change. When comparing PPi complexes for macrocycles **4** and **5,** where the arm length is the same and the ring size and position of the steric groups alter, the $^{DT}CCS_{He}$ values are very close and unaltered from the *apo* forms.

3 has the ring size of **5** and the arm length of **4** however with a differing placement of arms. From the IM–MS data, it can be seen that the difference in $^{DT}CCS_{He}$ from **4** to **3** is greater than from **4** to **5**. This increase from the trend may be attributed to the parallel arm placement resulting in a comparatively larger sized complex for **3**. The fact that similar affinities have been observed for each ligand for **3** *cf* **5** indicates that with the optimum arm length (here 2 carbons) the placement of the arms does not alter the affinity any further. The multiple conformations that the PPi bound form presents in may be due to differing binding modes whereby the PPi may bind across the complex with each phosphorus binding to both Zn^{2+} atoms rather than one each.

Candidate structures were generated for **4** using molecular dynamics and found to have their CCS_{He} ranging from ~211 to ~270 Å^2. The larger structures were found to belong to an extended conformation in which the two–linking arms flex away from the macrocycle. The extended conformer is observed in both solution and gas–phase calculations but not well sampled in the IM–MS experiments, suggesting that it collapses on the time scale of our experiment.

3.7. Conclusions

This paper has analysed the binding of a series of novel zinc (II) dipicolylamine sensors for anionic ligands using a variety of mass spectrometry methods more routinely applied with protein ligand complexes. Anion receptors were shown to have greater affinity for PPi over the dye molecule PV thought due to the alignment of four negatively charged O–P bonds with the Zn^{2+} atoms. The reduced specificity to ATP could be attributed to only 3 O–P bonds being involved in binding; the lower charge density on each oxygen atom involved in binding or the steric effect of the nucleoside base upon with the non–binding sidechains. Ion mobility has provided insights as to how macrocycle structure influences strength of ligand binding. It has been shown how IM–MS can provide K_d values as well as insights to the binding mode which can be used to infer a mechanism. This highlights the role for ion mobility as a tool in structural drug design, with greater versatility than solution phase methods.

The high specificity of PPi over other anionic ligands results in a tight complex (**5** shows ~50% tighter cross section upon binding (FWHM)) that is smaller than the unbound macrocycle, which is a testament to the selectivity of these anion receptors and to the success of their design.

3.8. Acknowledgements

Thanks, are expressed to the DTC in Cell and Proteomic Technologies EP/F500424/1, and BBSRC for awards BB/L015048/1 and BB/H013636/1. Additional thanks, to the British Mass Spectrometry Society for a grant that allowed the purchase of a nanospray tip pipette puller.

3.9. References

[1] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, *Science.* **246** (1989) 64–71.

[2] B. Ganem, J. D. Henion, Going gently into flight: analyzing noncovalent interactions by mass spectrometry, *Bioorg. Med. Chem.* **11** (2003) 311–314.

[3] K. A. Sannes–Lowery, R. H. Griffey, S. A. Hofstadler, Measuring dissociation constants of RNA and aminoglycoside antibiotics by electrospray ionization mass spectrometry, *Anal. Biochem.* **280** (2000) 264–271.

[4] C. Uetrecht, R. J. Rose, E. van Duijn, K. Lorenzen, A. J. R. Heck, Ion mobility mass spectrometry of proteins and protein assemblies, *Chem. Soc. Rev.* **39** (2010) 1633–1655.

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

[6] T. Wyttenbach, M. T. Bowers, Intermolecular interactions in biomolecular systems examined by mass spectrometry, *Annu. Rev. Phys. Chem.* **58** (2007) 511–533.

[7] B. J. McCullough, J. Kalapothakis, H. Eastwood, P. Kemper, D. MacMillan, K. Taylor, J. Dorin, P. E. Barran, Development of an ion mobility quadrupole time of flight mass spectrometer, *Anal. Chem.* **80** (2008) 6336–6344.

[8] E. M. Wright, J. M. Diamond, Anion selectivity in biological systems, *Physiol. Rev.* **57** (1977) 109–156.

[9] D. R. Blake, R. E. Allen, J. Lunec, Free radicals in biological systems. A review orientated to inflammatory processes, *Br. Med. Bull.* **43** (1987) 371–385.

[10] A. J. Meijer, K. van Dam, The metabolic significance of anion transport in mitochondria, *Biochim. Biophys. Acta.* **346** (1974) 213–244.

[11] H. A. Krebs, R. L. Heech, Equilibrium relations between pyridine nucleotides and adenine nucleotides and their roles in the regulation of metabolic processes, *Adv. Enzyme Regul.* **7** (1969) 397–413.

[12] E. K. Hoffmann, P. B. Dunham, Membrane mechanisms and intracellular signalling in cell volume regulation, *Int. Rev. Cytol.* **161** (1995) 173–262.

[13] E. A. MacRobbie, Signal transduction and ion channels in guard cells, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **353** (1998) 1475–1488.

[14] K. Sharp, R. Fine, B. Honig, Computer simulations of the diffusion of a substrate to an active site of an enzyme, *Science.* **236** (1987) 1460–1463.

[15] K. Detmer, V. Massey, Effect of monovalent anions on the mechanism of phenol hydroxylase, *J. Biol. Chem.* **259** (1984) 11265–11272.

[16] R. Wierenga, M. De Maeyer, W. Hol, Interaction of pyrophosphate moieties with alpha–helixes in dinucleotide–binding proteins, *Biochemistry.* **24** (1985) 1346–1357.

[17] D. Barford, A. K. Das, M. P. Egloff, The structure and mechanism of protein phosphatases: insights into catalysis and regulation, *Annu. Rev. Biophys. Biomol. Struct.* **27** (1998) 133-164.

[18] D. E. Koshland, The role of flexibility in enzyme action, *Cold Spring Harb. Symp. Quant. Biol.* **28** (1963) 473–482.

[19] S. Mookerjea, T. Coolbear, M. L. Sarkar, Key role of dolichol phosphate in

glycoprotein biosynthesis, *Can. J. Biochem. Cell Biol.* **61** (1983) 1032–1040.

[20] T. Lindahl, Instability and decay of the primary structure of DNA, *Nature.* **362** (1993) 709–715

[21] C. C. Richardson, Enzymes in DNA metabolism, *Annu. Rev. Biochem*. **38** (1969) 795–840

[22] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén, P. Nyrén, Real–time DNA sequencing using detection of pyrophosphate release, *Anal. Biochem.* **242** (1996) 84–89.

[23] S. H. Moochhala, J. A. Sayer, G. Carr, N. L. Simmons, Renal calcium stones: insights from the control of bone mineralization, *Exp. Physiol.* **93** (2008) 43–49.

[24] M. Doherty, C. Belcher, M. Regan, A. Jones, J. Ledingham, Association between synovial fluid levels of inorganic pyrophosphate and short term radiographic outcome of knee osteoarthritis, *Ann. Rheum. Dis*. **55** (1996) 432–436.

[25] M. Camerlain, D. J. McCarty, D. C. Silcox, A. Jung, Inorganic pyrophosphate pool size and turnover rate in arthritic joints, *J. Clin. Invest.* **55** (1975) 1373–1381.

[26] R. Koenig, C. Johnson, Colorimetric determination phosphorus in biological material, *Ind. Eng. Chem. Anal. Ed.* **14** (1942) 155-156.

[27] H. Baadenhuijsen, A. P. Jansen, Continuous–flow determination of serum inorganic phosphate with a single reagent. The vanadomolybdate method re–evaluated, *Clin. Chem.* **23** (1977) 1275–1280.

[28] C. C. Garber, R. C. Miller, Revisions of the 1963 semidine HC1 standard method for inorganic phosphorus, *Clin. Chem.* **29** (1983) 184–188.

[29] H. H. Taussky, E. Shorr, G. Kurzmann, A microcolorimetric method for the determination of inorganic phosphorus, *J. Biol. Chem.* **202** (1953) 675–685.

[30] P. J. Worsfold, L. J. Gimbert, U. Mankasingh, O. N. Omaka, G. Hanrahan, P. C. Gardolinski, P. M. Haygarth, B. L. Turner, M. J. Keith–Roach, I. D. McKelvie, Sampling, sample treatment and quality assurance issues for the determination of phosphorus species in natural waters and soils, *Talanta.* **66** (2005) 273–293.

[31] Y. Subbarow, C. H. Fiske, The colorimetric determination of phosphorus, *J. Biol. Chem.* **66** (1925) 375–400.

[32] C. Chen, L. Lu, Y. Zheng, D. Zhao, F. Yang, X. Yang, A new colorimetric protocol for selective detection of phosphate based on the inhibition of peroxidase–like activity of magnetite nanoparticles, *Anal. Methods.* **7** (2015) 161–167.

[33] A. E. Hargrove, S. Nieto, T. Zhang, J. L. Sessler, E. V. Anslyn, Artificial receptors for the recognition of phosphorylated molecules, *Chem. Rev.* **111** (2011) 6603–6782.

[34] C. Anda, A. Llobet, V. Salvado, J. Reibenspies, R. J. Motekaitis, A. E. Martell, A systematic evaluation of molecular recognition phenomena. 1. Interaction between phosphates and nucleotides with hexaazamacrocyclic ligands containing *m*–xylylic spacers, *Inorg. Chem.* **39** (2000) 2986–2999.

[35] K. Tohda, M. Tange, K. Odashima, Y. Umezawa, H. Furuta, J. L. Sessler, Liquid membrane electrode for guanosine nucleotides using a cytosine–pendant triamine host as the sensory element, *Anal. Chem.* **64** (1992) 960–964.

[36] C. M. Carey, W. B. Riggan, Cyclic polyamine ionophore for use in a dibasic phosphate–selective electrode, *Anal. Chem.* **66** (1994) 3587–3591.

[37] A. Bencini, A. Bianchi, C. Giorgi, P. Paoletti, B. Valtancoli, V. Fusi, E. Garcia–Espana, J. M. Llinares, J. A. Ramirez, Effect of nitrogen methylation on cation and

anion coordination by hexa– and heptaazamacrocycles. Catalytic properties of these ligands in ATP dephosphorylation, *Inorg. Chem.* **35** (1996) 1114–1120.

[38] E. Kimura, A. Watanabe, H. Nihira, Macrocyclic polyamines as calculi solubilizers, *Chem. Pharm. Bull. (Tokyo).* **31** (1983) 3264–3268.

[39] S. Bunce, E. S. Kong, The interactions between nucleic acids and polyamines. I. High resolution carbon–13 and hydrogen–1 nuclear magnetic resonance studies of spermidine and 5'–AMP, *Biophys. Chem.* **8** (1978) 357–368.

[40] C. Nakai, W. Glinsmann, Interactions between polyamines and nucleotides, *Biochemistry.* **16** (1977) 5636–5641.

[41] E. A. Katayev, J. L. Sessler, V. N. Khrustalev, Y. A. Ustynyuk, Synthetic model of the phosphate binding protein: solid–state structure and solution–phase anion binding properties of a large oligopyrrolic macrocycle, *J. Org. Chem.* **72** (2007) 7244–7252.

[42] E. J. Song, H. Kim, I. H. Hwang, K. B. Kim, A. R. Kim, I. Noh, C. Kim, A single fluorescent chemosensor for multiple target ions: Recognition of Zn^{2+} in 100% aqueous solution and F^{-} in organic solvent, *Sens. Actuat. B Chem.* **195** (2014) 36–43.

[43] M. Cametti, K. Rissanen, Highlights on contemporary recognition and sensing of fluoride anion in solution and in the solid state, *Chem. Soc. Rev.* **42** (2013) 2016–2038.

[44] J. M. M. Rodrigues, A. S. F. Farinha, P. V. Muteto, S. M. Woranovicz–Barreira, F. A. Almeida Paz, M. Neves, J. A. S. Cavaleiro, A. C. Tome, M. Teresam S. R. Gomes, J. L. Sessler, J. P. C. Tome, New porphyrin derivatives for phosphate anion sensing in both organic and aqueous media, *Chem. Commun.* **50** (2014) 1359–1361.

[45] H. T. Niu, D. Su, X. Jiang, W. Yang, Z. Yin, J. He, J. P. Cheng, A simple yet highly selective colorimetric sensor for cyanide anion in an aqueous environment, *Org. Biomol. Chem.* **6** (2008) 3038–3040.

[46] W. Gong, D. Na, L. Fang, H. Mehdi, G. Ning, Preassembly–driven ratiometric sensing of H_2PO_4 anions in organic and aqueous environments, *Org. Biomol. Chem.* **13** (2015) 1979–1982.

[47] R. M. Duke, T. McCabe, W. Schmitt, T. Gunnlaugsson, Recognition and sensing of biologically relevant anions in alcohol and mixed alcohol–aqueous solutions using charge neutral cleft–like glycol–derived pyridyl–amidothiourea receptors, *J. Org. Chem.* **77** (2012) 3115–3126.

[48] T. Nishimura, S. Y. Xu, Y. B. Jiang, J. S. Fossey, K. Sakurai, S. D. Bull, T. D. James, A simple visual sensor with the potential for determining the concentration of fluoride in water at environmentally significant levels. *Chem. Commun.* **49** (2013) 478–480.

[49] J. Liu, K. Wu, X. Li, Y. Han, M. Xia, A water soluble fluorescent sensor for the reversible detection of tin(iv) ion and phosphate anion, *RSC Adv.* **3** (2013) 8924–8928.

[50] F. Hofmeister, Zur Lehre von der Wirkung der Salze - Zweite Mittheilung, *Arch. Exptl. Pathol. Pharmakol.* **24** (1888) 247-260.

[51] X. Liu, H. T. Ngo, Z. Ge, S. J. Butler, K. A. Jolliffe, Tuning colourimetric indicator displacement assays for naked–eye sensing of pyrophosphate in aqueous media, *Chem. Sci.* **4** (2013) 1680–1686.

[52] S. J. Butler, K. A. Jolliffe, Selective pyrophosphate recognition by cyclic peptide receptors in physiological saline, *Chem. Asian J.* **7** (2012) 2621–2628.

[53] S. J. Butler, K. A. Jolliffe, Synthesis of a family of cyclic peptide–based anion receptors, *Org. Biomol. Chem.* **9** (2011) 3471–3483.

[54] W. E. Broderick, J. I. Legg, Catecholate coordination to copper: structural characterization of a tetrachloro–o–catecholate–bridged dicopper (II) complex as a model for intermediates in copper–catalyzed oxidation of catechols, *Inorg. Chem.* **24** (1985) 3725–3727.

[55] P. E. Barran, H. L. Cole, S. M. Goldup, D. A. Leigh, P. R. McGonigal, M. D. Symes, J. Wu, M. Zengerle, Active–metal template synthesis of a molecular trefoil knot, *Angew. Chemie. Int. Ed. Engl.* **123** (2011) 12280–12284.

[56] J. Ujma, M. De Cecco, O. Chepelin, H. Levene, C. Moffat, S. J. Pike, P. J. Lusby, P. E. Barran, Shapes of supramolecular cages by ion mobility–mass spectrometry, *Chem. Commun.* **48** (2012) 4423-4425.

[57] E. R. Brocker, S. E. Anderson, B. H. Northrop, P. J. Stang, M. T. Bowers, Structures of metallosupramolecular coordination assemblies can be obtained by ion mobility spectrometry–mass spectrometry, *J. Am. Chem. Soc.* **132** (2010) 13486–13494.

[58] E. F. Wilson, H. Abbas, B. J. Duncombe, C. Strebe, D. L. Long, L. Cronin, Probing the self–assembly of inorganic cluster architectures in solution with cryospray mass spectrometry: growth of polyoxomolybdate clusters and polymers mediated by silver(I) ions, *J. Am. Chem. Soc.* **130** (2008) 13876–13884.

[59] X. Li, Y. T. Chan, G. R. Newkome, C. Wesdemiotis, Gradient tandem mass spectrometry interfaced with ion mobility separation for the characterization of supramolecular architectures, *Anal. Chem.* **83** (2011) 1284–1290.

[60] S. Trimpin, D. E. Clemmer, Ion mobility spectrometry/mass spectrometry snapshots for assessing the molecular compositions of complex polymeric systems, *Anal. Chem.* **80** (2008) 9073–9083.

[61] E. Vandermarliere, E. Stes, K. Gevaert, L. Martens, Resolution of protein structure by mass spectrometry, *Mass Spectrom. Rev.* **35** (2016) 653-665.

[62] E. Karaca, A. M. Bonvin, On the usefulness of ion–mobility mass spectrometry and SAXS data in scoring docking decoys, *Acta. Cryst.* **69** (2013) 683–694.

[63] A. Politis, A. Y. Park, Z. Hall, B. T. Ruotolo, C. V. Robinson, Integrative modelling coupled with ion mobility–mass spectrometry reveals structural features of the clamp loader in complex with single–stranded DNA binding protein, *J. Mol. Biol.* **425** (2013) 4790–4801.

[64] S. H. Lee, J. C. Rasaiah, Molecular dynamics simulation of ion mobility. 2. Alkali metal and halide ions using the SPC/E model for water at 25°C, *J. Phys. Chem.* **100** (1996) 1420–1425.

[65] H. R. Skullerud, Monte-Carlo investigations of the motion of gaseous ions in electrostatic fields, *J. Phys. B: At. Mol. Phys.* **6** (1973) 728–742.

[66] J. Gidden, M. T. Bowers, Gas–phase conformations of deprotonated and protonated mononucleotides determined by ion mobility and theoretical modelling, *J. Phys. Chem. B.* **107** (2003) 12829–12837.

[67] Z. Hall, A. Politis, C. V. Robinson, Structural modelling of heteromeric protein complexes from disassembly pathways and ion mobility–mass spectrometry, *Structure.* **20** (2012) 1596–1609.

[68] A. Politis, A. Y. Park, S.J. Hyung, D. Barsky, B. T. Ruotolo, C. V. Robinson, Integrating ion mobility–mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes, *PLoS One*. **5** (2010) e12080.

[69] M. J. McDonough, A. J. Reynolds, W. Y. G. Lee, K. A. Jolliffe, Selective recognition of pyrophosphate in water using a backbone modified cyclic peptide receptor, *Chem. Commun.* **28** (2006) 2971–2973.

[70] V. Gabelica, N. Galic, F. Rosu, C. Houssier, E. De Pauw, Influence of response factors on determining equilibrium association constants of non–covalent complexes by electrospray ionization mass spectrometry, *J. Mass Spectrom.* **38** (2003) 491–501.

[71] Y. Berezovskaya, M. Porrini, C. Nortcliffe, P. E. Barran, The use of ion mobility–mass spectrometry to assist protein design: A case study on zinc finger fold *versus* coiled coil interactions, *Analyst.* **140** (2015) 2847–2856.

[72] C. S. Hoaglund, Y. Liu, A. D. Ellington, M. Pagel, D. E. Clemmer, Gas–phase DNA: oligothymidine ion conformers, *J. Am. Chem. Soc.* **119** (1997) 9051–9052.

[73] J. T. Moseley, I. R. Gatland, D. W. Martin, E. W. McDaniel, Measurement of transport properties of ions in gases; results for K^+ Ions in N_2, *Phys. Rev.* **178** (1969) 234–239.

[74] N. Mills, ChemDraw Ultra 10.0 CambridgeSoft, *J. Am. Chem. Soc.* **128** (2006) 13649–13650.

[75] A. E. Reed, F. Weinhold, Natural localized molecular arbitals, *J. Chem. Phys*. **83** (1985) 1736–1740

[76] A. A. Shvartsburg, M. F. Jarrold, An exact hard–spheres scattering model for the mobilities of polyatomic ions, *Chem. Phys. Lett.* **261** (1996) 86–91.

[77] M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, M. F. Jarrold, Structural information from ion mobility measurements: effects of the long–range potential, *J. Phys. Chem.* **100** (1996) 16082–16086.

[78] C. Bleiholder, T. Wyttenbach, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). Method, *Int. J. Mass Spectrom.* **308** (2011) 1–10.

[79] C. Bleiholder, S. Contreras, T. D. Do, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (II). Model parameterization and definition of empirical shape factors for proteins, *Int. J. Mass Spectrom.* **345** (2013) 89–96.

[80] S. E. Anderson, C. Bleiholder, E. R. Brocker, P. J. Stang, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (III): application to supramolecular coordination–driven assemblies with complex shapes, *Int. J. Mass Spectrom.* **330** (2012) 78–84.

[81] C. Bleiholder, S. Contreras, M. T. Bowers, A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (IV). Application to polypeptides, *Int. J. Mass Spectrom.* **354** (2013) 275–280.

[82] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminform*. **4** (2012) 17.

[83] A. Ojida, Y. Mito–Oka, K. Sada, I. Hamachi, Molecular recognition and fluorescence sensing of monophosphorylated peptides in aqueous solution by bis(zinc(II)–dipicolylamine)–based artificial receptors, *J. Am. Chem. Soc.* **126** (2004) 2454–2463.

4.

The use of ion mobility–mass spectrometry to probe the effect of small molecule inhibitors on the interaction of c–MYC with MAX

4. Declaration

This chapter consists of one draft experimental paper awaiting submission: C. Nortcliffe, P. E. Barran.

The use of ion mobility–mass spectrometry to probe the effect of small molecule inhibitors on the interaction of c–MYC with MAX.

As first author on this publication I performed all experimental work including: MS, ligand binding and IM–MS experiments.

Ion mobility–mass spectrometry probes the effect of small molecule inhibitors on the interaction of c–MYC with MAX

Chris Nortcliffe, Perdita E. Barran.

Supporting Information

Additional mass spectra and collisional cross section distributions for peptides used in this study are included in appendix 3.

4.1. Abstract

Mass spectrometry can be used as a tool for high throughput screening of new drugs and ligands, and application of this technique towards studying conformation of intrinsically disordered proteins (IDPs) is increasing. It is appropriate to consider how mass spectrometry can be applied to screen putative inhibitors of IDP complexes, where interactions are commonly weak across flat interaction interfaces. This work applies ion mobility–mass spectrometry (IM–MS) to the study of small molecule inhibitors (SMI) designed to target the leucine zipper complex formed between the transcription factor c–MYC and its endogenous partner MAX. Previous work suggested that these peptides can be readily observed as a dimer with two dominant conformations; one akin to the leucine zipper which is destabilised in the presence of a SMI (10058–F4), and the other a more compact globular form [1]. This study extends those findings to examine the effects of four additional ligands: 10074–G5 (**A**)**,** 5530837 (**B**), 5793353 (**C**) and 9082178 (**D**). Incubation with each of the four ligands results in a reduction of the observed heterodimer complex. Additionally, IM–MS relates structural loss of extended conformations with improved inhibition of heterodimer formation.

4.2. Introduction

Traditional drug design methodologies and strategies usually involve the optimisation of structural features of a given active site or pocket to enable effective and targeted binding. Interactions are maximised by cycling through a series of different functional groups to achieve the strongest fit, whilst minimising toxicity and side reactions [2]. However, these methods are cited as less amenable when dealing with potential targets that lack a distinct solvable structure [3]. Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are increasingly becoming attractive drug targets as they constitute about a third of all eukaryotic proteins [4], especially within pathways involved in cellular signalling [6] and regulation of cellular cycles [5]. It is postulated that the high flexibility of IDPs allows a kinetic advantage when binding to multiple partners and often constitutes a disorder–to–order transition [7]. New techniques are being developed to characterise the binding mode of ligands to disordered systems, often with the use of *in silico*–based applications. However, these computational approaches are often expensive, time consuming and impractical for a high–throughput approach [8, 9]. More critically, they cannot rely on a resolved starting geometry. It has been suggested that certain ligands can stabilise a disordered state of an IDP, hence exploiting the entropic gain that the system would receive upon binding to its target [10, 11]. These stabilisations should be able to be observed experimentally.

The high flexibility and dynamic nature of IDPs means that they present many transient conformations, the shortest lived of which will be difficult to characterise with many biophysical approaches [3, 12]. These molecules are frequently difficult to crystallise and, in general, ensemble averaging is needed to best represent the breadth of conformational states that the molecules can present. Whilst NMR spectroscopy can provide good insights into the dominant forms of these conformational states and produces distance restraints to solve an unknown structure, this global view is not so amenable to a drug discovery approach. Most NMR experiments cannot determine rapid interconversion between conformational states, nor sparsely occupied states. There is a growing body of evidence that highlights the benefit of using mass spectrometry (MS) as well as ion mobility–mass spectrometry (IM–MS) for analysis of IDPs [13–15]. The clearest advantage is the ability of IM–MS experiments to separate conformers in the gas–phase; kinetically trapping and hence resolving rapidly interconverting structures that cannot be seen by other techniques. The second distinct advantage is the ability of MS to provide the stoichiometry of binding within complex mixtures. Taken together, the IM–MS experiments can reveal the effect of one or more binding partners upon the intrinsic conformations.

c–MYC is part of the MYC family of oncoproteins, which function as transcription factors highly involved in regulation of pro–survival and pro–apoptotic gene function [16, 17]. The transcriptional activation of these proteins is managed through protein–protein interactions (PPIs). In the case of c–MYC this is the obligatory co–factor protein MAX which forms a heterodimer with c–MYC, controlling the vast majority of its functions [18]. The c–MYC:MAX complex is involved in between 10–15% of all gene expressions in humans *via* the recruitment of co–factors. c–MYC has been found to be deregulated in a large number of human cancers. In addition, elevated expression of c–MYC is regularly associated with aggressive tumours with poor differentiation. Studies have shown that inhibition of c–MYC leads to the rapid regression of established tumours, making c–MYC an attractive cancer target [19]. Hammoudeh *et al.* [10, 20] reported a number of c–MYC small molecule inhibitors (SMI), whose mechanism of action is to bind to the putative PPI sites. Other work went on to suggest that these SMIs limit the interaction between c–MYC and MAX by inducing a more rigid and defined conformation [1, 21]. This target is attracting lots of attention with various groups proposing differing strategies to prevent heterodimer activity [22–25]. The dimerisation of c–MYC and MAX occurs through a basic helix–loop–helix motif in the leucine zipper (bHLHZip) domain. These regions in both c–MYC and MAX are intrinsically disordered before undergoing an essential disorder–to–order transition to form a four–helix bundle of two α–helices separated by loops.

The work in this paper extends previous investigations of the effect the SMI (10058–F4 (**E**)) [10, 20, 26–29] has on peptides containing the leucine zipper domains of both c–MYC and its partner MAX. IM–MS was conducted on these molecules both in the presence and absence of the ligand **E** along with circular dichroism measurements and molecular modelling [1]. Here IM–MS measurements are conducted on the c–MYC:MAX complex with an expanded panel of SMIs (Figure 4.1).

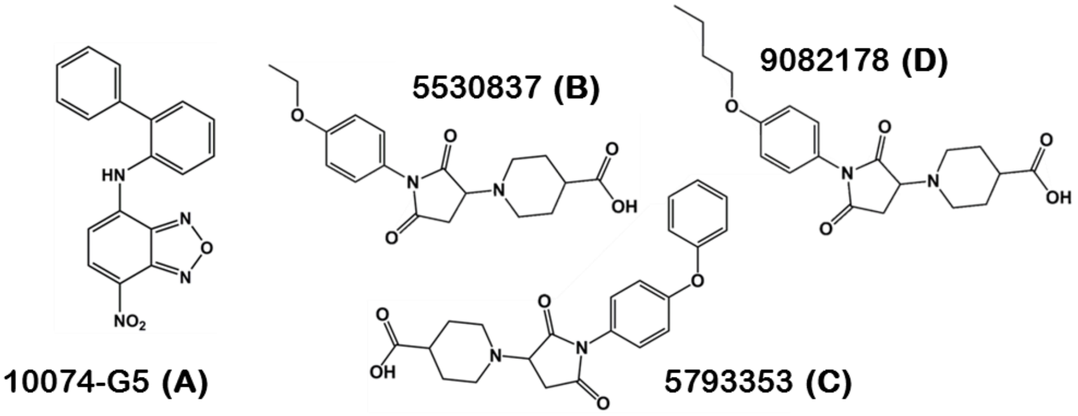


Figure 4.1: Structures of the small molecule inhibitor ligands that were incubated with the c–MYC:MAX complex and studied *via* IM–MS:10074–G5 (A) 5530837 (B), 5793353 (C) and 9082178 (D) [23, 30, 31].

4.3. Experimental details

4.3.1. Peptides

Peptides were purchased from GenScript (USA) as lyophilised powder and prepared to stock concentrations of 1 mg/ml. The sequence of the c–MYC peptide used was: (residues 408 – 437) VQAEEQKLISEEDLLRKRREQLKHKLEQL, and the sequence of the MAX peptide used was: (residues 73 – 102) MRRKNDTHQQDIDDLKRQNALLEQQVRAL. Peptides were analysed at 100 µM in 20 mM ammonium acetate (pH 6.8) (5% DMSO). Ligand **A** was purchased from Sigma Aldrich, (Gillingham, UK). Ligands **B**, **C**, and **D** were purchased from Chembridge (San Diego, CA, USA). All ligands were dissolved in DMSO at 2 mM. Ligand experiments were conducted at 1:1:2 concentrations of 100 µM of c–MYC and MAX in 20 mM ammonium acetate with 200 µM ligand in 5% DMSO. All samples were incubated for 1 hour at 37°C prior to analysis.

4.3.2. Mass spectrometry

A Synapt G2 (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer with an orthogonal nano–electrospray ionisation source was used to obtain mass spectra. Instrument parameters in the source region were retained across all experiments with tuning to transmit non–covalent complexes. Particular care was taken to minimise the source cone voltage since this correlated with optimum successful transfer of complexed species. Typical instrument source operating settings involved: capillary voltage 1.3 kV, cone voltage 25 V, source temperature 80°C. Experiments were performed with three technical repeats; values reported are the mean with standard deviation.

4.3.3. Ion mobility–mass spectrometry

IM–MS experiments were conducted using a Q–ToF 1 (Waters, Micromass, Manchester, UK) quadrupole time–of–flight mass spectrometer that has been modified to include a 5.1 cm copper drift cell (pressurised with helium to ~4 Torr) [32], again with nano–electrospray ionisation. The voltage across the cell was altered from 50 to 20 V in discrete experimental steps with temperature and pressure measured for each acquisition. Data were acquired with synchronous injection into the drift cell at each voltage and summed over a number of injections. Rotationally averaged collisional cross sections ($^{DT}CCS_{He}$) were obtained by plotting experimental arrival times against P/V for each drift voltage (V) as described previously [32–34]. Experiments were performed with three technical repeats; values reported are the mean with standard deviation.

4.4. Results

Peptides were analysed in 20 mM ammonium acetate at pH 6.8 after incubation for one hour at 37°C. As observed previously for this system both peptides presented in a charge state distribution from $[M + 2H]^{2+}$ to $[M + 4H]^{4+}$ with the majority of the intensity in the $[M + 3H]^{3+}$ charge state (*m/z* c–MYC 1196*, m/z* MAX 1178). Homodimers and heterodimers were mainly observed as complexes in the $[M + 5H]^{5+}$ charge state (c–MYC *m/z* 1435*, MAX m/z* 1413*,* heterodimer *m/z* 1424) with some complex in the $[M + 4H]^{4+}$ charge state heterodimer (*m/z* 1777).

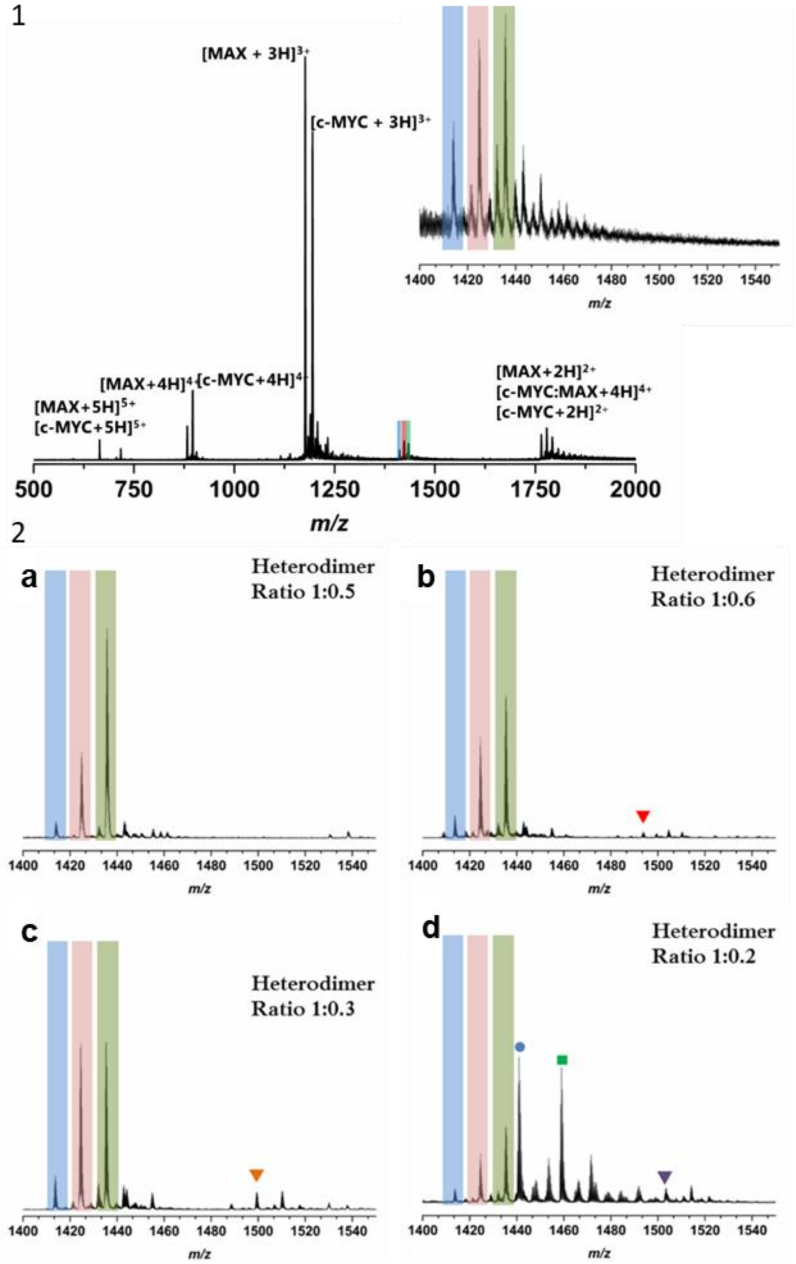


Figure 4.2: n–ESI mass spectrum obtained for 50 µM c–MYC and 50 µM MAX in 20 mM ammonium acetate with 5% DMSO. Inset shows zoom of *m/z* 1400–1550 region containing 5+ heterodimers and homodimers. The coloured bands indicate MAX homodimer (blue), c–MYC homodimer (green) and the heterodimer (pink). Part two shows inset region in the presence of SMI ligands with relative ratio of heterodimer peak *versus* no inhibitor. (a) Mass spectrum with addition of 100 µM ligand A, (b) mass spectrum with addition of 100 µM ligand B, (c) mass spectrum with addition of 100 µM ligand C, (d) mass spectrum with addition of 100 µM ligand D. Red, orange and purple arrows indicate the heterodimer + bound ligands B, C and D, respectively. Blue circles and green squares indicate 3+ c–MYC with ligand D and 3+ MAX with ligand D, respectively.

4.4.1. Effect of ligands on complex formation

Upon the addition of stoichiometric amounts of ligand, no binding of ligand **A** was observed despite a reported micromolar binding affinity, a similar result was seen previously with ligand **E** [1]. However, ligands **B**, **C** and **D** were all observed to bind to monomeric c–MYC (forming complexes seen at +3 charge state ions at *m/z* 1308, 1321 and 1327 respectively), MAX (again forming a +3 complex at *m/z* 1293, 1302 and 1309 respectively) as well as to the heterodimer observed as triply charged complexes with *m/z* 1493, 1499 and 1503. For c–MYC:MAX with ligand **D** the complex was seen with up to three ligands bound (two and three ligand **D** *m/z* 1582 and 1660 respectively), however, these were at very low intensity and may be attributable to ligand clusters forming in solution, rather than to three isolated binding sites. Binding to ligand **A** was never observed even with very 'soft' source conditions (sampling cone 20 V), however, at high cone voltages (above 80 V) complexes observed with the three other SMIs were also lost. This suggests a weak interaction complex and that in all cases some complex may be lost during desolvation. Cone voltages were kept as low as possible for all ion mobility experiments, and capillary voltages were also minimised.

The intensity of the $[M + 5H]^{5+}$ heterodimer was reduced after incubation with each SMI ligand. The intensity ratios of incubated $[M + 5H]^{5+}$ heterodimer relative to the intensity in the absence of an inhibitor are shown in Figure 4.2 A–D relative to the intensity in the absence of any inhibitor. There are negligible differences in intensity for the $[M + 4H]^{4+}$ heterodimer with most of the ligands, although a decrease in intensity was observed when incubated with ligand **A**. This indicates that there is not simply a shift in charge state distribution in the presence of these ligands but an overall shift in complex stability. The calculated ratio of bound peak is inclusive of $[M + 4H]^{4+}$ and $[M + 5H]^{5+}$ charge states but not salt adducts. After incubation with the SMIs (except with ligand **D**) the unbound monomers exhibit a shift to higher charge states with the $[M + 4H]^{4+}$ species increasing in intensity.

4.4.2. Effect of ligands on complex conformations

As reported previously, the $[M + 5H]^{5+}$ heterodimer presents in two conformations: one compact, hereby referred to as C_1 (885 Å^2), and one extended refferd to as E_1 (1010 Å^2) [1]. These have previously been assigned as ordered and disordered based on IM–MS, molecular dynamics and CD data [1]. It was shown in previous work that incubation with ligand **E** lead to a loss of the E_1 conformation [1].

These peptide constructs are not amidated at the C terminus and therefore give cross sectional values that are slightly smaller than previous work (~5%). Figure 4.3 shows the collisional cross section distributions ($^{DT}CCSD_{He}$) of the heterodimer in the presence of all the ligands. The area under the curve is reflective of the intensity of the $[M + 5H]^{5+}$ heterodimer from the mass spectra shown in Figure 4.2. After incubation with ligand **A** (Figure 4.3 B), both conformations are reduced in intensity, and there is proportionately more loss of the extended form E_1, as well as a slight increase in the flexibility of both forms. It can also be observed that both conformations now present in broader distributions than in the absence of the ligand (Figure 4.3). This can be inferred to be resulting from a compaction in the molecular conformations of the heterodimer leading to less abundance in the most extended forms.

After incubation with ligand **B** (Figure 4.3 C), the heterodimer presents similarly to after incubation with ligand **A,** with an increase in flexibility of the extended form. The apex of the extended form is 1035 Å^2, however, there may be a further extended form at 1085 Å^2, labelled E_2. The bound form (Figure 4.3 F) of the complex presents in two conformations (B_1 970 Å^2 and $^BE_2$ 1070 Å^2 (^{B} denotes a conformation only observed in bound species)) that flank E_1. This second conformation exists in the same conformational space as E_2 and is likely to adopt a similar structure. These could be due to ligand binding to both the ordered and the disordered complexes leading to an increase in cross section of both. The increased width of the $^{DT}CCSD_{He}$ for the unbound complex is similar to that for the bound distribution, therefore the 'unbound' conformation may be a result of the bound form having undergone dissociation in the source region and not yet relaxed to the unbound complex conformation. This effect has been shown previously with ligand **E** [1], where no bound complex was directly observed.

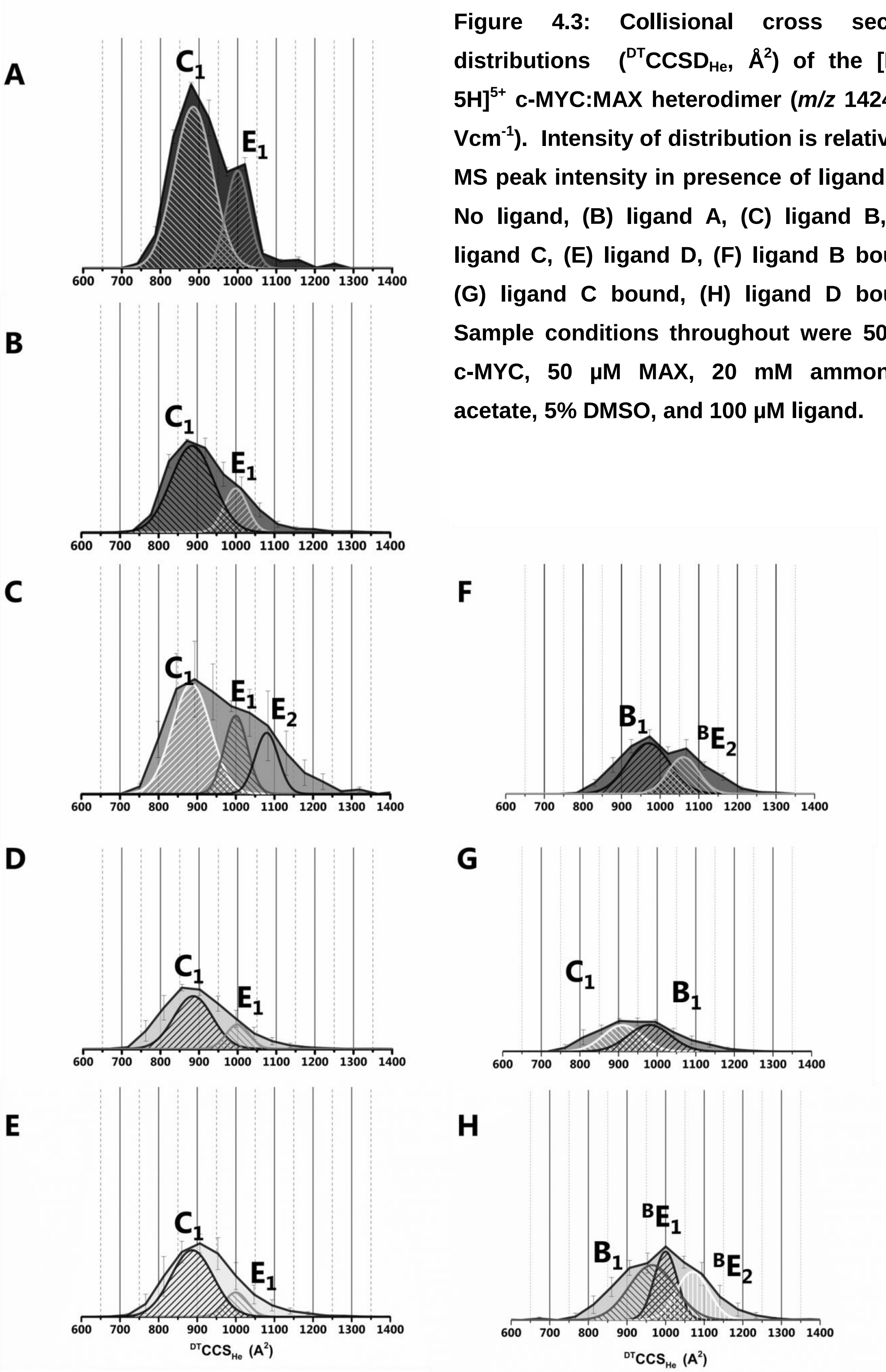


Figure 4.3: Collisional cross section distributions ($^{DT}CCSD_{He}$, Å^2) of the $[M + 5H]^{5+}$ c-MYC:MAX heterodimer (*m/z* 1424) (5 Vcm^{-1}). Intensity of distribution is relative to MS peak intensity in presence of ligand. (A) No ligand, (B) ligand A, (C) ligand B, (D) ligand C, (E) ligand D, (F) ligand B bound, (G) ligand C bound, (H) ligand D bound. Sample conditions throughout were 50 µM c-MYC, 50 µM MAX, 20 mM ammonium acetate, 5% DMSO, and 100 µM ligand.

Incubation of the $[M + 5H]^{5+}$ heterodimer complex in the presence of ligand **C** (Figure 4.3 D) leads to a reduction in complex intensity, and the complete loss of the extended conformation E_1. The bound complex displays a broad $^{DT}CCSD_{He}$ spanning a large area of cross sectional space (250 $Å^2$ at FWHM *versus* 150 $Å^2$ for the compact $[M + 5H]^{5+}$ conformation). This increase in flexibility upon ligand binding (Figure 4.3 G) is unexpected and could be due to an overlapping ensemble of structures rather than a single flexible one. Another possible explanation could be that stabilisation in one region of the disordered complex causes another region to increase in flexibility in order to maintain entropy.

In the presence of ligand **D** there is, again, near complete loss of the extended conformation E_1 (Figure 4.3 E), though there is a slight shift in the maxima of the compact peak and a respective increase in width. The bound complex presents in what could be attributed to a single large conformation or more likely two to three overlapping conformations (Figure 4.3 H). The intensity of this bound peak is much lower than that observed for the other ligands. The conformation of the peak owing to two bound ligands and three bound ligands sees a shift of 80 $Å^2$ to a higher $^{DT}CCS_{He}$ with a similar width of $^{DT}CCSD_{He}$ (350 $Å^2$ FWHM), again in a broad conformation (Figure S4.2). This indicates that despite an increase in mass, the addition of this further ligand has little change on the complex conformation. As three possible binding sites for these ligands have been proposed [10], it is possible that ligand **D** can bind to all three sites, resulting in three peaks each with a distinct conformation owing to a single ligand bound in one or more sites. Binding at any site appears to prevent the disorder–to–order transition, since the unbound complex presents mainly in a single compact conformation in the presence of ligand **D**.

4.5. Discussion

The binding of IDPs to their targets will reduce entropy, which must be accounted for with an overall gain in enthalpy [20, 21]. Stabilising these disordered states therefore exploits this enthalpic gain. A selection of SMI were tested against the leucine zipper region of the peptides c–MYC and MAX. The ligands have been shown to stabilise structures *via* binding at the leucine zipper domain, leading to an increase in disorder (as seen in solution experiments) and inhibition of dimer formation [21].

MS experiments consistently show a reduction in the intensity of the $[M + 5H]^{5+}$ heterodimer after incubation with each of the four SMIs. This reduction in abundance of the heterodimer species indicates that the interaction between c–MYC and MAX is reduced. The $[M + 5H]^{5+}$ heterodimer presents in two conformations, one extended (E_1) and one compact (C_1). Previous work has assigned these to ordered and disordered states, respectively [1]. The SMIs studied here, particularly ligand **A**, have been shown to stabilise certain regions of c–MYC associated with the more disordered form [21]. As the ordered form is required for function of c–MYC, this should reduce its activity and inhibit cancer cells [21].

With the addition of ligands **A**, **C** and **D** there is a reduction in the intensity of the extended state E_1, albeit to varying degrees. Reduction of the extended state is correlated with the greatest loss of heterodimer intensity. ${}^{DT}CCSD_{He}$ widen in the presence of all four ligands, which is also indicative of the conformational changes occurring. Presence of ligand **B** leads to the appearance of a further extended state (E_2) which appears to be retained from the bound form. This would further support the idea of a bound conformation being retained after ligand loss.

Bound complexes of the heterodimer with ligand are observed for three of the SMIs: **B**, **C** and **D.** These bound forms present in an intermediate conformational space (B_1) with a further, more extended form (${}^{B}E_2$). The intensity of the bound peaks however, are very low (less than 5% of heterodimer).

4.6. Conclusion

Four small molecule inhibitors were shown by ion mobility–mass spectrometry to modulate the conformation of the leucine zipper region of c–MYC and its partner MAX. Each ligand has a different effect on the protein complex, indicating the complexity of interactions as well as the strength of IM–MS as a tool to distinguish between binding modes. This work has highlighted the use of IM–MS as a tool for screening ligands that have been shown to stabilise molecular conformations.

4.7. Acknowledgements

Thanks, are expressed for the awards of the DTC in Cell and Proteomic Technologies EP/F500424/1, and BBSRC awards BB/L015048/1 and BB/H013636/1. Additional thanks, to the British Mass Spectrometry Society for a grant that allowed the purchase of a nanospray tip pipette puller.

4.8. References

[1] S. R. Harvey, M. Porrini, C. Stachl, D. Macmillan, G. Zinzalla, P. E. Barran, Small–molecule inhibition of c–MYC:MAX leucine zipper formation is revealed by ion mobility mass spectrometry, *J. Am. Chem. Soc.* **134** (2012) 19384–19392.

[2] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev.* **46** (2001) 3–26.

[3] V. N. Uversky, Intrinsically disordered proteins from A to Z, *Int. J. Biochem. Cell Biol.* **43** (2011) 1090–1103.

[4] A. L. Fink, Natively unfolded proteins, *Curr. Opin. Struct. Biol.* **15** (2005) 35–41.

[5] C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions, *Annu. Rev. Biochem.* **83** (2014) 553–584.

[6] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, Z. Obradović, Intrinsic disorder and protein function, *Biochemistry.* **41** (2002) 6573–6582.

[7] M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, Preformed structural elements feature in partner recognition by intrinsically unstructured proteins, *J. Mol. Biol.* **338** (2004) 1015–1026.

[8] M. Varadi, W. Vranken, M. Guharoy, P. Tompa, Computational approaches for inferring the functions of intrinsically disordered proteins, *Front. Mol. Biosci.* **2** (2015) 45.

[9] V. M. Burger, T. Gurry, C. M. Stultz, Intrinsically disordered proteins: where computation meets experiment, *Polymers.* **6** (2014) 2684–2719.

[10] D. I. Hammoudeh, A. V. Follis, E. V. Prochownik, S. J. Metallo, Multiple independent binding sites for small–molecule inhibitors on the oncoprotein c–Myc, *J. Am. Chem. Soc.* **131** (2009) 7390–7401.

[11] Y. Cheng, T. Legall, C. J. Oldfield, J. P. Mueller, P. Romero, M. S. Cortese, V. N. Uversky, A. K. Dunker, Rational drug design *via* intrinsically disordered protein, *Trends Biotechnol.* **24** (2006). 435–442.

[12] T. Mittag, J. D. Forman–Kay, Atomic–level characterization of disordered protein ensembles, *Curr. Opin. Struct. Biol.* **17** (2007) 3–14.

[13] S. R. Harvey, C. E. MacPhee, P. E. Barran, Ion mobility mass spectrometry for peptide analysis, *Methods.* **54** (2011) 454–461.

[14] Y. Berezovskaya, C. T. Armstrong, A. L. Boyle, M. Porrini, D. N. Woolfson, P. E. Barran, Metal binding to a zinc–finger peptide: a comparison between solution and the gas–phase, *Chem. Commun*. **47** (2011) 412–414.

[15] A. S. Phillips, A. F. Gomes, J. M. Kalapothakis, J. E. Gillam, J. Gasparavicius, F. C. Gozzo, T. Kunath, C. MacPhee, P. E. Barran, Conformational dynamics of α–synuclein: insights from mass spectrometry, *Analyst.* **140** (2015) 3070–3081.

[16] C. E. Nesbit, J. M. Tersak, L. E. Grove, A. Drzal, H. Choi, E. V. Prochownik, Genetic dissection of c–myc apoptotic pathways, *Oncogene.* **19** (2000) 3200–3212.

[17] C. E. Nesbit, J. M. Tersak, E. V. Prochownik, MYC oncogenes and human neoplastic disease, *Oncogene.* **18** (1999) 3004–3016.

[18] L. Kretzner, E. M. Blackwood, R. Eisenman, MYC and MAX proteins possess distinct transcriptional activities, *Nature.* **359** (1992) 426–429.

[19] L. Soucek, J. Whitfield, C. P. Martins, A. J. Finch, D. J. Murphy, N. M. Sodir, A. N. Karnezis, L. B. Swigart, S, Nasi. G. I. Evan, Modelling Myc inhibition as a cancer therapy, *Nature.* **455** (2008) 679–683.

[20] A. V. Follis, D.I. Hammoudeh, A. T. Daab, S. J. Metallo, Small–molecule perturbation of competing interactions between c–Myc and Max, *Bioorg. Med. Chem. Lett.* **19** (2009) 807–810.

[21] A. V. Follis, D. I. Hammoudeh, H. Wang, E. V. Prochownik, S. J. Metallo, Structural Rationale for the coupled binding and unfolding of the c–Myc oncoprotein by small Molecules, *Chem. Biol.* **15** (2008) 1149–1155.

[22] K. Y. Jung, H. Wang, P. Teriete, J. L. Yap, L. Chen, M. E. Lanning, A. Hu, L. J. Lambert, T. Holien, A. Sundan, N. D. Cosford, E. V. Prochownik, S. Fletcher, Perturbation of the c–Myc–Max protein–protein interaction *via* synthetic α–helix mimetics, *J. Med. Chem.* **58** (2015) 3002–3024.

[23] D. Clausen, J. Guo, R. Parise, J. Beumer, M. Egorin, J. S. Lazo, E. V. Prochownik, J. L. Eiseman, *In vitro* cytotoxicity and *in vivo* efficacy, pharmacokinetics, and metabolism of 10074–G5, a novel small–molecule inhibitor of c–Myc/Max dimerization, *J. Pharmacol. Exp. Ther.* **335** (2010) 715–727.

[24] T. Berg, Small–molecule modulators of c–Myc/Max and Max/Max interactions, *Curr. Top. Microbiol. Immunol.* **348** (2010) 139–149.

[25] A. Kiessling, B. Sperl, A. Hollis, D. Eick, T. Berg, Selective inhibition of c–Myc/Max dimerization and DNA binding by small molecules, *Chem. Biol.* **13** (2006) 745–751.

[26] M. Zhang, H. Y. Fan, S. C. Li, Inhibition of c–Myc by 10058–F4 induces growth arrest and chemosensitivity in pancreatic ductal adenocarcinoma, *Biomed. Pharmacother.* **73** (2015) 123–128.

[27] I. Müller, K. Larsson, A. Frenzel, G. Oliynyk, H. Zirath, E. V. Prochownik, N. J. Westwood, M. A. Henriksson, Targeting of the MYCN protein with small molecule c–MYC inhibitors, *PLoS One.* **9** (2014) e97285.

[28] J. Guo, R. A. Parise, E. Joseph, M. J. Egorin, J. S. Lazo, E. V. Prochownik, J. L. Eiseman, Efficacy, pharmacokinetics, tisssue distribution, and metabolism of the Myc–Max disruptor, 10058–F4 [Z, E]–5–[4–ethylbenzylidine]–2–thioxothiazolidin–4–one, in mice, *Cancer Chemother. Pharmacol.* **63** (2009) 615–25.

[29] T. Berg, Inhibition of transcription factors with small organic molecules, *Curr. Opin. Chem. Biol.* **12** (2008) 464–471.

[30] X. Yin, C. Giap, J. S. Lazo, E. V. Prochownik, Low molecular weight inhibitors of Myc–Max interaction and function, *Oncogene.* **22** (2003) 6151–6159.

[31] H. Wang, J. Chauhan, A. Hu, K. Pendleton, J. L. Yap, E. Sabato, J. W. Jones, M. Perri, J. Wu, E. Cione, M. A. Kane, S. Fletcher, E. V. Prochownik, Disruption of Myc–Max heterodimerization with improved cell–penetrating analogs of the small molecule

10074–G5, *Oncotarget.* **4** (2013) 936–947.

[32] B. J. McCullough, J. Kalapothakis, H. Eastwood, P. Kemper, D. MacMillan, K. Taylor, J. Dorin, P. E. Barran, Development of an ion mobility quadrupole time of flight mass spectrometer, *Anal. Chem.* **80** (2008) 6336–6344.

[33] Y. Berezovskaya, M. Porrini, C. Nortcliffe, P. E. Barran, The use of ion mobility–mass spectrometry to assist protein design: a case study on zinc finger fold *versus* coiled coil interactions, *Analyst.* **140** (2015) 2847–2856.

[34] Y. Liu, D. E. Clemmer, Characterising oligosaccharides using injected–ion mobility–mass spectrometry, *Anal. Chem.* **69** (1997) 2504–2509.

5.

The use of ion mobility–mass spectrometry to probe the inhibition of the transcription factor c–MYC:MAX binding to DNA

5. Declaration

This chapter consists of one draft experimental paper awaiting submission: C. Nortcliffe, P. E. Barran.

The use of ion mobility–mass spectrometry to probe the inhibition of the transcription factor c–MYC:MAX binding to DNA

As first author on this publication I performed all experimental work including: MS, ligand binding and IM–MS experiments. Protein expression was performed by collaborators.

The use of ion mobility–mass spectrometry to probe the inhibition of the transcription factor c–MYC:MAX binding to DNA

Chris Nortcliffe, Perdita E. Barran.

Supporting Information

Additional mass spectra for proteins used in this study are included in appendix 4.

5.1. Abstract

Proteins involved in DNA transcription are targeted in drug therapy for a variety of medical applications. Often such proteins lack distinct secondary structure or contain regions of disorder which makes designing specific and selective inhibitors a non–trivial task. Ion mobility–mass spectrometry (IM–MS) is an emerging tool for studying these intrinsically disordered proteins (IDPs). Here a panel of four small molecule inhibitors (SMI) designed to inhibit the interaction between the complex of the transcription factor c–MYC with its endogenous partner MAX are screened, along with their effects on the heterodimer complex binding to DNA. IM–MS shows structural changes induced by these SMI, and a stabilisation of disordered states. In the absence of DNA, low intensity heterodimer peaks from the c–MYC:MAX heterodimer are observed (less than 5% total intensity) in a range of conformations. In the presence of all SMI, extended conformations (above 2000 Å^2) are stabilised in the gas–phase. Upon the addition of DNA, the signal corresponding to the c–MYC:MAX:DNA complex becomes much more intense indicating stabilisation of the heterodimer interaction in the presence of DNA. In the presence of DNA, the SMI stabilise intermediate conformations (2000–2500 Å^2) of the complex with reduction of more extended charge states.

5.2. Introduction

Since first reported, the number of proteins and peptides believed to be intrinsically disordered (IDPs) or to contain regions of disorder (IDRs) has grown, with current estimates that at least a third of all eukaryotic proteins contain disorder to some extent [1]. The inherent increase in flexibility bestowed by these sequences makes them abundant in certain cellular roles such as signalling [2, 3], regulation [4] and interaction with DNA [5, 6]. IDPs have also been highlighted as having an important role in cancer progression; for example, the transcription factor c–MYC has been shown to be over expressed in ~70% of human cancers [7, 8]. IDPs are of great interest as drug targets however, are often described as 'undruggable' due to the lack of resolvable structure, their large binding interfaces and their low affinity interactions. The design of drugs which target IDPs is difficult as the common method of using a solved structure to determine likely candidates *via in silico* screening is not suited to these molecules [9, 10]. Additionally, there are further barriers to be overcome, including stabilisation of disordered states and overcoming large entropic barriers, which are not present for ligands which bind to protein targets with a lock and key model.

c–MYC is member of a family of oncoprotein transcription factors highly involved in cellular regulation, including cell cycling, growth, along with survival and apoptosis [11]. Deregulation of c–MYC has been associated with several human cancers even when it is not the primary oncogenic driver [12–15]. The activity of the MYC family of proteins is highly regulated by protein:protein interactions, with c–MYC in particular being dependent on its cofactor protein MAX.

Structurally c–MYC has a disordered leucine zipper domain however; upon hetero–dimerisation with its partner protein MAX, it undergoes a disorder–to–order transition to form a coiled coil structure through a basic–helix–loop–helix interaction (bHLHZip). This dimerised form of c–MYC–MAX possess an interface of a parallel, left handed, four–helix bundle, with each monomer unit consisting of two α–helicies, interspaced by a disordered loop [16]. There has been extensive research that presents c–MYC inhibition as an attractive therapeutic strategy [13]. Experiments involving both the genetic and pharmacological inhibition of c–MYC (and related MYC proteins) showed regression of tumours [12, 17–20]. Additionally, the down stream effects of c–MYC inhibition appear to be mild and reversible despite c–MYC being required for cellular function in normal cells [21–24]. Due to the difficulties of targeting IDPs there are currently no clinically available drugs that target this system, however Hammoudeh *et al.* [25] have reported a number of small molecules with low micromolar affinities designed to inhibit c–MYC through stabilisation of disordered states. Two inhibitors (10058–F4 and 10074–G5) were shown to bind to separate target regions on the leucine zipper domain of a small number of amino acids.

Biophysical characterisation of IDPs and their interactions is hindered due to their inherently high flexibility and range of conformational states [26, 27]. Ion mobility–mass spectrometry, with its capability to distinguish different conformations in a single experiment [28–30], forms an attractive option for studying these flexible regions. In addition, native MS methods are useful for binding experiments as the *m/z* separation allows both bound and unbound states of non–covalent interaction partners to be interrogated separately [30–33]. There have been several excellent review articles highlighting the use of IM–MS to study intrinsically disordered systems [34–36].

IM–MS instrumentation is becoming more common–place, with the availability of both commercially available drift–time ion mobility (DT–IM–MS), and travelling wave technology. The experiments presented here utilise an in–house modified DT–IM–MS instrument in which ions are pulsed into the drift cell filled with helium gas, whilst temperature and pressure are recorded. Across the cell a weak electric field is applied, where movement of the ions are encouraged by the potential, but impeded by collisions with the helium buffer gas. The mobility of an ion (K) is related to its shape (given as the rotationally averaged collsional cross section CCS, Ω, $Å^2$) along with its mass and charge, as these affect how it interacts with the field and buffer gas. The ratio of the drift velocity (v_d) and the applied electric field (E) can be used to determine the mobility of the ion through the Mason–Schamp equation:

$$\boldsymbol{\Omega} = \frac{\mathbf{3}\boldsymbol{ze}}{\mathbf{16}\boldsymbol{N}} \sqrt{\frac{\mathbf{2}\boldsymbol{\pi}}{\boldsymbol{\mu k_B T}}} \frac{\mathbf{1}}{\boldsymbol{K}_0} \qquad \textit{Equation 1. 8}$$

Where K_0 is the reduced mobility (measured mobility corrected to temperature and pressure) z is the ion charge state, e is the elementary charge state, N is the number gas density, μ is the reduced mass of the ion–pair, k_B is the Boltzmann constant and T is the buffer gas temperature.

Previous work examined the leucine zipper region of c–MYC:MAX using molecular modelling and ion mobility–mass spectrometry [37], and characterised the effects of a single small molecule inhibitor 10058–F4 (**E)** on the structure of the heterodimer. It was shown that upon addition of the ligand, the presence of a smaller compact conformation was increased, which could be attributed to a stabilisation of the disordered structure. This screen was extended to other potential inhibitors designed to stabilise the disordered state of the c–MYC leucine zipper, thereby preventing binding to MAX [25] (manuscript in preparation; chapter 4). Here the ligand screen has been extended, and probes the structural effects upon constructs of c–MYC and MAX containing the leucine zipper region and the DNA binding domain (Figure 5.1 A). The first series of experiments probe the intrinsic structure of c–MYC:MAX heterodimer according to a framework model [38], before studying the effects of four previously studied SMI on their conformations (Figure 5.1 B). The final stage of experiments sought to observe the c–MYC:MAX heterodimer in a bound interaction with dsDNA complex in the gas–phase and then to study the effects of these four ligands once more. This study greatly extends the understanding of the use of IDPs as drug targets and highlights structural changes of the disordered systems in the presence of ligands believed to interact with certain conformational states.

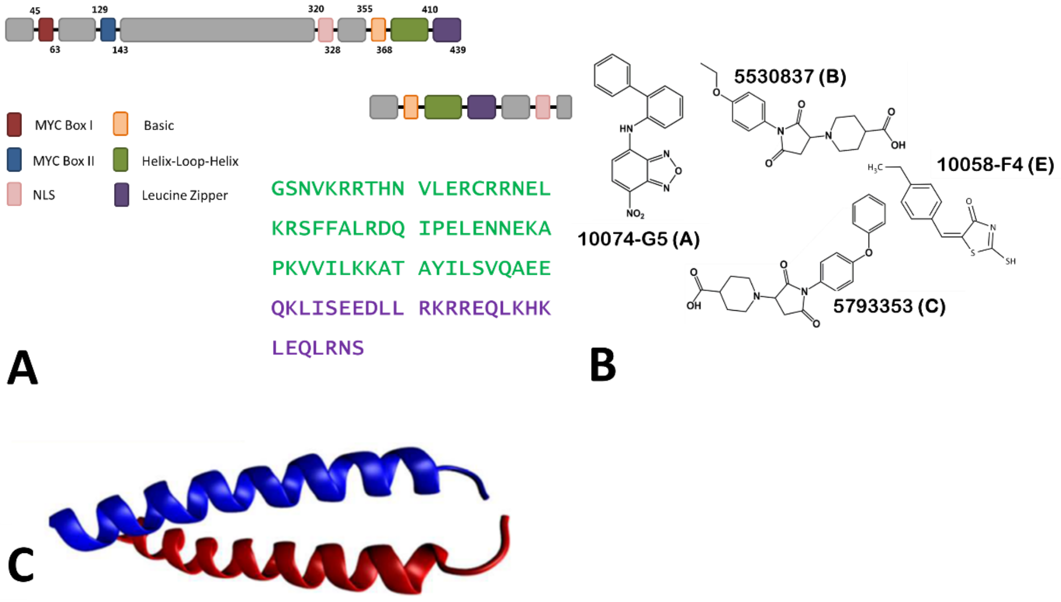


Figure 5.1: A) Domains of full length c–MYC with sequence in purple is the leucine zipper region of c–MYC as examined previously [37]. The sequence in green is the DNA binding region and in purple the leucine zipper domain. These have been shown relative to their positions on the c–MYC and MAX genes. B) Ligands analysed during these experiments: (A) 10074–G5, (B) 5530837, (C) 5793353, (D) 5793353, and (E) 10058–F4 [39]. C) X–ray crystal structure of c–MYC and MAX leucine zipper domains [37].

5.3. Experimental details

5.3.1. *Protein expressions*

5.3.2. *c–MYC expression and purification*

The c–MYC amino acid sequence expressed comprised of residues 351 to 437 (Figure 5.1 A). The pET plasmid encoding bHLHZip and DNA domain of c–MYC was obtained from AstraZeneca. c–MYC was expressed in *Escherichia coli* (*E. coli*) BL21 (DE3), Gold cell lines (Agilent Technologies, Cheadle, UK) at 37°C in LB media (Sigma Aldrich, Gillingham, UK). Protein expression was induced by addition of 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) to bacterial cultures at OD_{600}. Cells were then harvested after 4 h, and frozen at –80°C. Harvest was further re–suspended in 20 mM phosphate buffer with 8 M urea and protease inhibitor tablet (Roche Products Ltd. UK), sonicated and centrifuged for 30 min at 40000 *g*. Supernatant was loaded on a Co^{2+} affinity column, equilibrated with 3 M GuHCl, 20 mM phosphate, 0.1 M NaCl, pH 7.4, washed with 2.5 mM imidazole (Sigma Aldrich, Gillingham, UK) and c–MYC was eluted by addition of 125 mM imidazole. C–terminal His–tag was cleaved by TEV protease and c–MYC was further purified on column for Co^{2+} affinity chromatography to remove His–tag peptides. The purity and identity of the protein was validated by SDS–PAGE. All other chemicals were obtained from Fisher Scientific, UK.

5.3.3. *MAX expression and purification*

The MAX amino acid sequence expressed comprised of residues 19 to 102 (appendix 4). Protein construct (MAX) plasmid was freshly transformed into *E. coli* C41 (DE3) competent cells and selected on LB ampicillin agar plate (100μg/ml). 1 litre Terrific broth (Sigma Aldrich, Gillingham, UK) culture was grown to an OD_{600} and induced by 1 mM IPTG with shaking at 37°C for 20 hrs.

Cells were resuspended in 30 ml of buffer A (20 mM Tris–HCl pH 7.9 and 0.5 M NaCl), with 5 mM imidazole, 5 mM $MgCl_2$, 1% Triton X–100, 1/500 protease inhibitor cocktail (Sigma Aldrich, Gillingham, UK) and 1/5000 benzonase nuclease (Merck, Darmstadt, Germany). The sample was sonicated on ice and centrifuged at 40000 *g* for 30 min at 4°C. The supernatant was then purified on 2 ml nickel resin (Qiagen, UK) with mixing at 4°C for 2 hours, and then washed with imidazole gradient between 5 mM to 100 mM in buffer A with, 0.1 % Triton X–100 and 1/1000 protease inhibitors and MAX was eluted by addition of 500 mM imidazole. C–

terminal His–tag was cleaved by TEV protease and MAX was further purified on nickel resin to remove His–tag peptides. The purity and identity of the protein was validated by SDS–PAGE. All other chemicals were obtained from Fisher Scientific, UK.

5.3.4. Small molecule inhibitors

Ligand **A** and **E** were purchased from Sigma Aldrich, (Gillingham, UK). Ligands **B**, **C**, & **D** were purchased from Chembridge (San Diego, CA, USA). All Ligands were dissolved in methanol at 2 mM. All samples were incubated for 1 hour at 37°C prior to analysis.

5.3.5. Mass spectrometry ionisation

Both MS and IM–MS experiments were conducted with nano–electrospray ionisation (n–ESI). n–ESI tips were prepared in–house using a Flaming/Brown micropipette puller (Sutter Instrument Company, Novato, CA, USA) using thin walled glass capillaries with an internal diameter of 0.5 mm. Potential was applied through the introduction of a platinum wire (0.125 mm, Goodfellow, UK) into the pulled capillary.

5.3.6. Mass spectrometry heterodimer

MS experiments were performed using a Synapt G2 (Waters, Manchester, UK) quadrupole time–of–flight mass spectrometer. Protein samples were prepared at 200 µM stock solutions in 100 mM ammonium acetate before being diluted to 50 µM for analysis, with the exception of the buffer screening experiment. In the buffer screening experiment, c–MYC was buffer exchanged into the appropriate concentration of ammonium acetate and methanol from stock Samples were incubated at 50 µM c–MYC, 50 µM MAX, with potential 100 µM ligand, for one hour at 37°C. Tuning conditions were kept constant across experiments at a capillary voltage of 1.3 kV, cone voltage 35 V, and source temperature 80°C.

5.3.7. Mass spectrometry c–MYC:MAX:dsDNA complex

The DNA consensus sequence used in this experiment was GACCACGTGGTC. DNA was purchased from Sigma Aldritch (Gillingham, UK) as single strands which were dissolved in water at 400 µM. These samples were then incubated at 90°C for 4 minutes followed by slow cooling to room temperature to induce annealing. Samples were incubated at 50 µM c–MYC, 50 µM MAX, 100 µM ds DNA, with potential 100 µM ligand, for one hour at 37°C. Tuning conditions were kept constant across experiments at a capillary voltage of 1.3 kV, cone voltage 35 V, and source temperature 80°C.

5.3.8. Ion mobility–mass spectrometry

IM–MS experiments were performed using an in house modified instrument consisting of a Q–ToF 1 (Waters, Micromass, Manchester, UK) quadrupole time–of–flight mass spectrometer which contains a 5.1 cm copper drift cell pressurised to approximately 4 Torr with helium [40]. Experiments were conducted over at least 5 discrete voltage steps from 50 V to 15 V whilst temperature and pressure were recorded. Rotationally averaged collisional cross sections ($^{DT}CCS_{He}$) were then obtained by plotting the experimental drift arrival time against pressure/voltage. Data is presented as collisional cross sectional distributions ($^{DT}CCSD_{He}$, $Å^2$) to portray conformational ensembles and flexibility. Experiments were performed with 3 technical repeats and values shown in Figures 5.3 and 5.5 are the mean with errors reported as standard deviation. Conformational families have been assigned to the $^{DT}CCSD_{He}$ which were then fitted using Gaussian distributions.

5.4. Results

5.4.1. *Mass spectrometry of the c–MYC:MAX heterodimer*

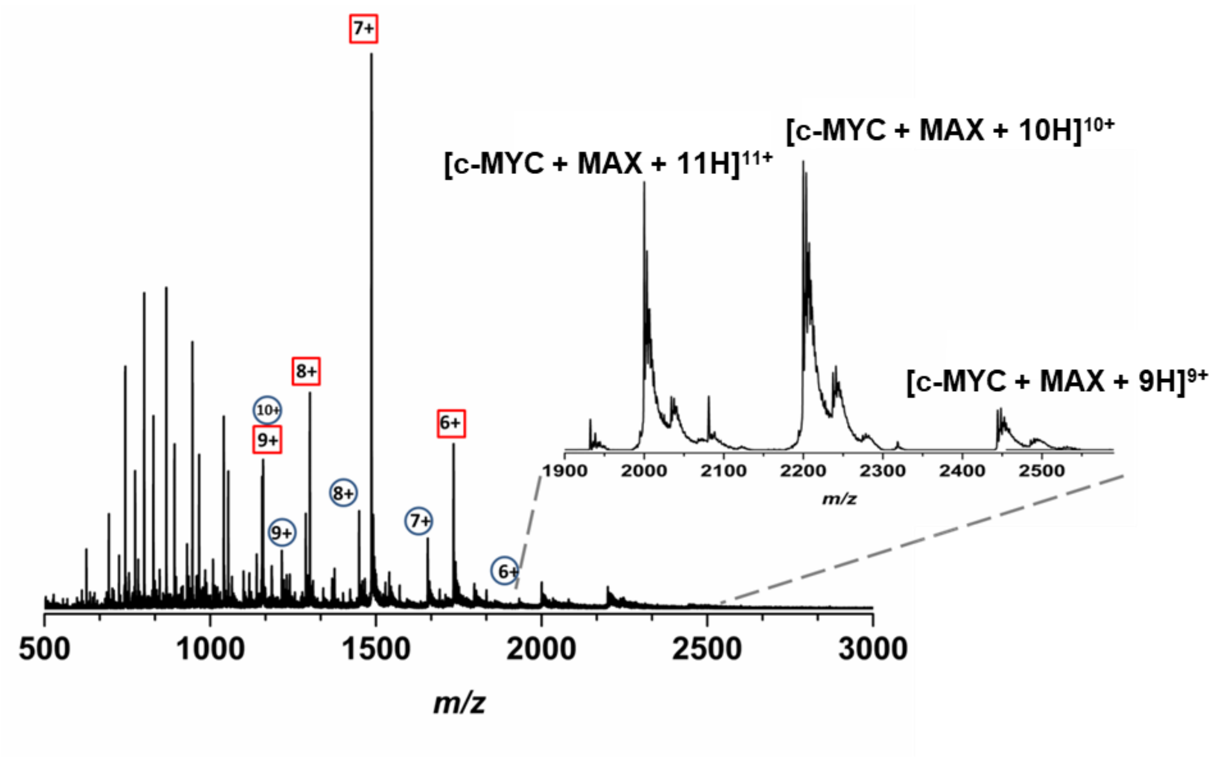


Figure 5.2: n–ESI mass spectrum of c–MYC and MAX incubated together. Red squares denote dominant c–MYC charge states and blue circles MAX charge states. Inset shows the mass spectrum of the heterodimer in the 9+, 10+ and 11+ charge states. Peaks below *m/z* 1000 correspond to higher charge states, but have not been labelled for clarity. Sample conditions were 50 µM c–MYC, and 50 µM MAX, from an aqueous solution with 100 mM ammonium acetate.

The mass spectrum obtained from c–MYC incubated with MAX in equimolar quantities (Figure 5.2), yielded a weak signal that can be attributed to the heterodimer of c–MYC and MAX. The heterodimer presents across three main charge states with the 10+ (*m/z* 2199) and 11+ (*m/z* 1999) being of nearly equal intensity and a lower abundant 9+ (*m/z* 2443) peak. Both proteins are mostly observed as unbound, either due to weak association affinity and/or as monomers produced from dissociation of the heterodimer during ionisation. Conformational families have been assigned to the collisional cross sectional distributions ($^{DT}CCSD_{He}$) of the heterodimer (Figure 5.3 control). Gaussian fitting has been applied to these distributions to generate a series of conformations ranging from compact C_1 and C_2 (1040 Å^2 and 1350 Å^2), intermediate I_1 and I_2 (1620 Å^2 and 1850 Å^2), and extended E_1 (2100 Å^2).

$^{DT}CCSD_{He}$ indicate that the 11+ charge state presents in one broad distribution that has been assigned to two major conformations (I_1 and I_2), with a minor extended conformation E_1. In the 10+ charge state a reduction in the intensity of I_2 can be observed along with the appearance of a more compact conformation C_2. The 9+ charge state continues this trend presenting dominantly in the C_2 conformation. All three charge states show an earlier arriving conformation labelled C_1. This early arriving conformation could be attributed to a dimer of heterodimers; however, no dimer–like peaks were in the mass spectrum.

In order to probe the stabilising effects of ligand binding, each inhibitor was added to the incubated 50 µM c–MYC:MAX solution at 100 µM, corresponding to one ligand for each molecule of protein, and incubated for one hour at 37°C. For all the ligands (with the exception of **B**), incubation induced a more extended structure in the $^{DT}CCSD_{He}$ of the unbound heterodimer (Figure 5.3). Peaks owing to bound ligand forms are present at very low intensities for all inhibitors tested for the 11+ and 10+ charge states.

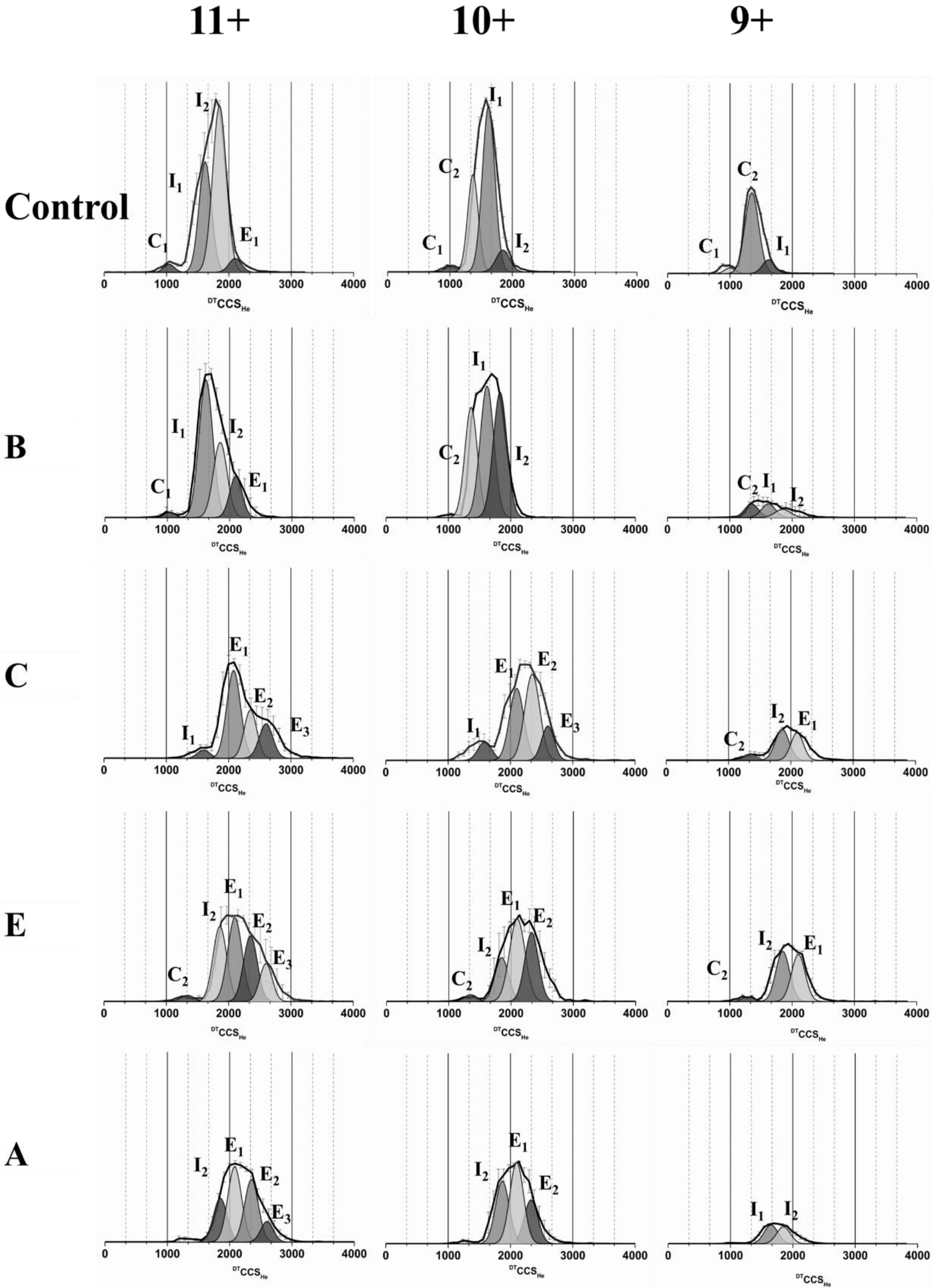


Figure 5.3: Collisional cross sectional distributions ($^{DT}CCSD_{He}$ Å^2) for the c–MYC:MAX heterodimer after incubation with each ligand. Conformational families have been generated to represent structural motifs across various interactions. Nomenclature is assigned to compact (C), intermediate (I), and extended (E) structures. Sample with no ligand added are labelled as ‘Control’. Ligands have been ordered to show similar structural trends in the order B, C, E, A. Sample conditions were: 50 µM c–MYC, 50 µM MAX and 100 µM ligand, from an aqueous solution with 100 mM ammonium acetate. Intensities are normalised within each ligand relative to intensity of *m/z*.

After incubation with 5530837 (**B**), the c–MYC:MAX heterodimer still presents in three major charge states in the mass spectra. The 9+ species presents with more extended conformations; a complete loss of the C_1 conformation is observed, reduction in the proportion of ions in the C_2 conformer, and the appearance of an intermediate I_1 conformation. This behaviour is also observed in the 10+ charge state of the complex where intensity of the most compact states is reduced while an increase in C_2 is observed. Additionally, a shift towards an intermediate state (I_2) is also shown. The 11+ charge state shows a reduction in the intermediate conformation I_2, with a corresponding increase in I_1 and E_1. Overall there is a broadening of the total distribution and a slight shift in intensity towards more extended states.

When incubated with 5793353 (**C**), a clearly visible shift in the conformational preference of the c–MYC:MAX heterodimer towards more extended states can be observed. Compared with the heterodimer alone, the 9+ charge state exhibits a greater proportion of ions in the I_2 and E_1 conformations. The 10+ charge state also exhibits a loss of both compact conformations C_1 and C_2, alongside the appearance of two extended conformations E_2 (2350 $Å^2$) and E_3 (2600 $Å^2$). These more extended states are also observed in the 11+ charge state, with complete loss of the I_2 state and a reduction in I_1.

The ligand 10058–F4 (**E**) has been extensively studied as a potential c–MYC inhibitor [17,20,41–47]. When incubated with this ligand, the $^{DT}CCSD_{He}$ of the 9+ charge state of the c–MYC:MAX heterodimer exhibits more extended conformations, with increased proportion of the ions presenting in conformers I_2 and E_1 compared with the heterodimer alone. The $^{DT}CCSD_{He}$ of the 10+ charge state also exhibits this extension, however it should be noted that this increase in extended conformers is less pronounced with compared with heterodimer incubation with ligand **C,** as it retains the I_2 conformation. The $^{DT}CCSD_{He}$ of the 11+ charge state shows the same trends, with the retention of intermediate I_2 conformation and the appearance of more extended structures including E_3.

10054–G5 (**A**), like **E**, has also been the subject of many studies [17,43,47–50]. The results presented here shows similarities to ligand **E** on its effect on the hetero–dimer conformation after incubation. The $^{DT}CCSD_{He}$ for the 9+ charge state displays loss of the compact conformations C_1 and C_2, present in the ligand free–state, with an increase in the population of intermediate conformers, although no extended states are observed. The 10+ charge state strongly resembles the distribution of **E**, however with slightly more intensity in I_2 *versus* E_2.

The absolute intensity in the mass spectrum of the heterodimer peaks can be summed across all charge states, to observe whether or not the ligands result in a reduction of heterodimer formation (Figure S5.3). Except for **B**, incubation with all the ligands resulted in a reduction in heterodimer intensity, with **A** causing a tenfold reduction. **B** appears to have little effect on the intensity of the complex and interestingly, with ion mobility, has a very different effect on the conformation of the heterodimer than other, ligands leading to only a mild extension of the complex.

5.4.2. c–MYC:MAX:dsDNA complex and its disruption by the panel of inhibitors.

The next set of experiments sought to observe if a c–MYC:MAX:dsDNA complex could be observed in the gas–phase and if the ligands **A**, **B**, **C** and **E** described above resulted in a change in conformation.

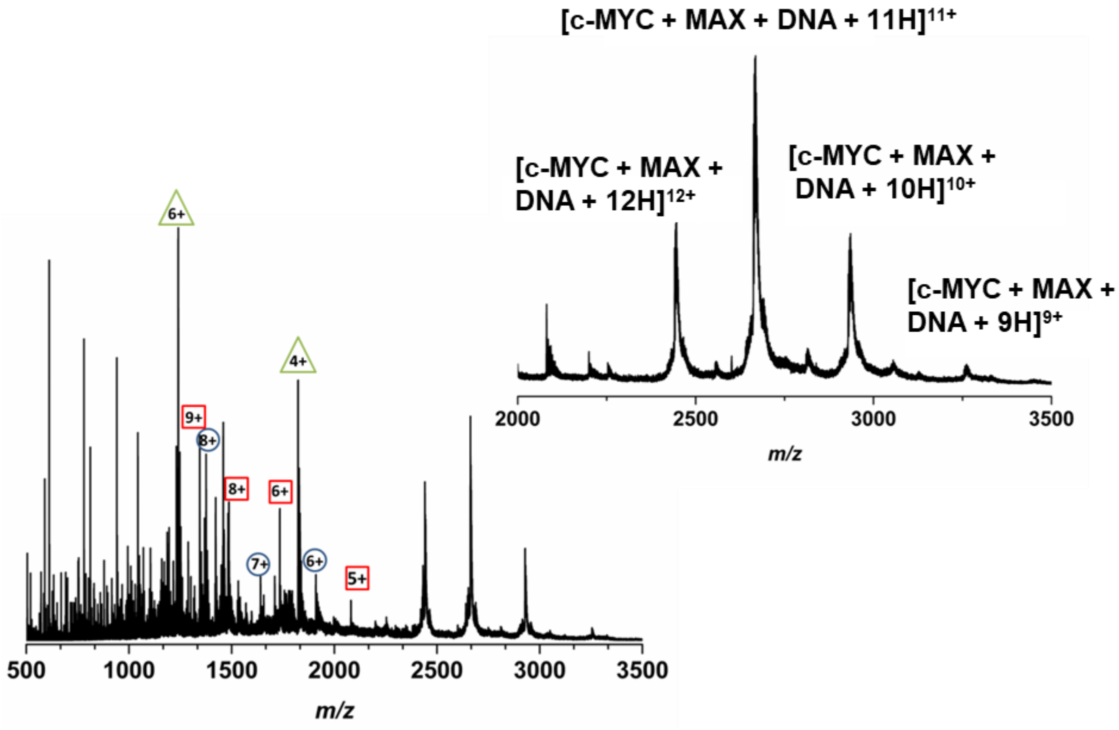


Figure 5.4: n–ESI spectra of c–MYC, MAX, and the consensus dsDNA sequence incubated for 60 minutes at 37°C. Red squares denote c–MYC charge states, blue circles MAX charge states, and Green triangles dsDNA. Inset shows the mass spectrum of the c–MYC:MAX:dsDNA complex in the 9+, 10+ ,11+, and 12 + charge. The lower end of the spectrum appears crowded due to overlapping of unbound monomer, and DNA salt adducts. Sample

conditions were 50 μM c–MYC, and 50 μM MAX, 100 μM ds DNA, from an aqueous solution with 100 mM ammonium acetate.

Figure 5.4 shows the mass spectra of c–MYC and MAX after incubation with double stranded DNA. The c–MYC:MAX heterodimer in complex with dsDNA is observed at higher intensity compared with the intensity of the complex in the absence of DNA, indicating that the presence of DNA stabilises the c–MYC:MAX heterodimer. The c–MYC:MAX:dsDNA complex (hereby referred to as MMD complex), presents across four charge states from 9+ to 12+ with the majority of the intensity in the 11+ (*m/z* 2664). Unbound protein monomers and DNA are observed indicating that either the K_d of the complex is high, or some complex is dissociated during ionisation despite the steps taken to preserve these non–covalent complexes.

The mass spectra after ligand incubations are included in the supplementary information (Figure S5.4). Figure 5.5 shows the summed $^{DT}CCSD_{He}$ for the MMD complex, across each charge state, after incubation with ligand. Figure S5.5 shows the $^{DT}CCSD_{He}$ in each individual charge state. Intensities are normalised within each ligand series *versus m/z* intensity.

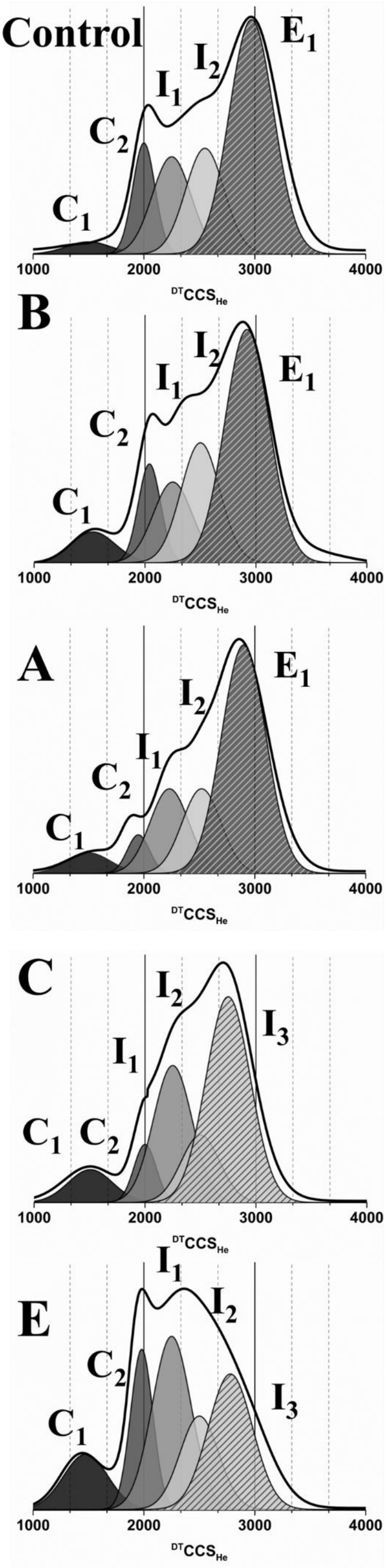


Figure 5.5: Collisional cross sectional distributions ($^{DT}CCSD_{He,}$ Å^{2}) for the c-MYC:MAX:dsDNA complex after incubation with ligands B, A, C and E. Collision cross section distributions from each charge state (9+, 10+, 11+ and 12+) were summed to give a global $^{DT}CCSD_{He}$ profile. Conformational families have been generated by fitting distributions to the $^{DT}CCSD_{He}$ to represent the structural motifs. Nomenclature is assigned to compact (C), intermediate (I), and extended (E) structures. Intensities are normalised within each ligand relative to intensity of *m/z*. Ligands have been ordered to show similar structural trends.

Before discussing this analysis, a distinction should be made between the heterodimer *versus* the MMD complex. In the first set of experiments in this chapter, it was assumed that the disordered forms of the complex are more extended than the ordered form, due to these extended forms being stabilised by the ligands. However, this may not always be the case. Due to how proteins fold, a disordered state may be a tighter conformation than one which is more ordered with internal cavities. Inversely, the disordered state may contain many tailing and flexible residues resulting in a larger collisional cross section. Therefore, the MMD complex could present as more extended or compact in its disordered state. Care should be taken when correlating between the two systems.

The $^{DT}CCSD_{He}$ of all the charge states for each ligand have been summed to provide a global collisional cross section distribution profile. Potential conformations have been fitted to these distributions relating to compact, intermediate and extended structures. The MMD complex presents primarily in the extended E_1 (2950 Å^2), with a broad intermediate section defined as I_1 and I_2 (2250 Å^2 and 2500 Å^2). There is a well–defined conformation that was labelled as C_2 which is mainly present in the lower charge states. This change from compact to extended conformations could indicate that in complex with dsDNA, the heterodimer presents in a more extended structure. There is also a small broad conformation C_1 (1500 Å^2) which could be due to a tetramer complex, however none is visible in the mass spectrum.

Incubation of the MMD complex with ligand **B** results in very minor changes in the $^{DT}CCSD_{He}$ (though the intensity of the peaks in the mass spectrum are reduced by roughly 50%, Figure S5.4). The I_2 conformation appears to become slightly more stabilised leading to a more defined intermediate between the extended and compact forms. Additionally, the compact C_1 also increases slightly in intensity.

The $^{DT}CCSD_{He}$ of the MMD complex with ligand **A** displays a much–altered distribution, with a reduction in all the compact and intermediate species. C_2 especially is of much lower intensity with also a slight loss of I_1 and I_2. E_1 remains the dominant conformation.

Incubation with ligand **C** results in a compaction from extended conformations towards more intermediate ones. E_1 has become reduced with the appearance of a more compact structure designated I_3 (2730 Å^2). Additionally, I_1 has become more intense, this likely represents that the ligand has stabilised the more compact forms.

Incubation of the MMD complex with ligand **E** results in the largest conformational changes with a strong shift to highly compact conformations. I_1 and C_2 are the most dominant peaks with a reduction in the extended I_3. There is also an increase in the highly compact $C_{1.}$ Surprisingly, the absolute intensities in the mass spectra of the MMD complex are comparable to the control.

Overall the ligands show stabilisation of more intermediate conformations, especially with loss of the extended forms. Interestingly, the $^{DT}CCSD_{He}$ range does not change strongly with the addition of ligands.

Mass spectra after incubation with ligands can be seen in Figure S5.4. Incubation with ligand **A** again appears to lead to the strongest reduction effect in complex formation, with decreased intensity in the peaks attributable to the MMD complex in the mass spectra (Figure S5.4 C). Surprisingly, incubation with ligand **E** appears to have a negligible effect on the intensity of the MMD complex (Figure S5.4 B). Incubation with ligands **B** and **C** both led to a roughly 50% reduction in observable complex intensity in the mass spectrum.

5.5. Discussion

So far, this work has detailed the observed effects of a number of small molecule inhibitors on both the c–MYC:MAX heterodimer and the c–MYC:MAX:dsDNA complex. This discussion intends to relate the observed changes in conformations, to the stabilising effects of the various SMI, linking compact and extended species to ordered and disordered conformations as gas–phase ions.

Figure 5.3 shows the $^{DT}CCSD_{He}$ of the c–MYC:MAX heterodimer across all observed charge states. Previous work has shown that this heterodimer can form either a disordered or ordered dimer [37], though this work was carried out only with the leucine zipper region. In the absence of ligands, the 9+ charge state presents in a single conformation (C_1) and the 10+ and 11+ charge states exist primarily in two conformations (C_2 and I_1, and I_1 and I_2 respectively). This shows a conformational shift to more extended forms with increasing charge state. With the addition of ligands **A**, **C** and **D,** the dominant conformer at each charge state becomes a more extended conformation and a new extended conformation can be observed. As some of these ligands have been shown to stabilise a disordered conformation, it is reasonable to postulate these new extended structures could correlate to a more disordered state of the leucine zipper domain [21, 25]. There is also evidence of a much smaller $^{DT}CCS_{He}$ conformation which could be the result of a highly compact conformation, or a gas–phase collapse. In the presence of ligand **C** this conformation does appear to present at higher intensity suggesting that ligand **C** stabilises this structure.

Considering the absolute intensity of peaks in the mass spectrum, it can be observed that incubation with ligand **A** results in the strongest inhibition of heterodimer formation. Inversely, incubation with ligand **B** has little effect on heterodimer intensity, which relates well with its structural effects which also do not follow the trend. Whilst care was taken to maintain constant instrumental parameters, comparison of absolute intensity between experiments should be made with care due to inherent differences in spray performance and ionisation efficiency, especially between instrumentation.

The addition of the DNA binding sequence to c–MYC:MAX resulted in the appearance of highly intense peaks corresponding to the c–MYC:MAX:dsDNA complex. This MMD complex is observed in the gas–phase with a much higher intensity than the corresponding c–MYC:MAX heterodimer. This strongly suggests that the presence of DNA stabilises the heterodimer formation, leading to a greater proportion bound in solution, or greater complex retention upon transfer into the gas–phase. Also, despite the addition of a negatively charged DNA sequence, the overall charge state range of the complex remains similar to the complex in the absence of DNA. In addition, there is only a slight increase in $^{DT}CCS_{He}$ size across these peaks (from 1900 Å^2 to 2000 Å^2 for the 11+) despite the addition mass (of above 25%), indicating that this interaction is specific.

Upon addition of the ligands to the MMD complex, a general trend of an increase in proportion of intermediate conformers, and the appearance of a more compact conformation in the 10+ and 11+ charge states is observed. These changes are less pronounced compared with ligand incubation with the heterodimer in the absence of dsDNA and there is no appearance of totally new conformations, nonetheless the trends are observed across all ligands. It should be noted that with the addition of DNA there are no visible peaks owing to heterodimer without DNA. This suggests that any heterodimer that does form immediately goes on to bind the DNA.

5.6. Conclusions

This work utilised ion mobility–mass spectrometry to screen the activity of small molecule inhibitors against a complex multi component disordered system. It shows how drug molecules that stabilise certain conformations can be interrogated using this technique and how differing molecules with similar mechanisms can have differing stabilising effects. Ligand **A** was found to lead to the greatest reduction in intensity of both the c–MYC:MAX and the c–MYC:MAX:dsDNA systems. Ligand **E** however, whilst having a reasonable effect in the absence of DNA, showed little reduction in it's presence and very different structural effects than the other ligands. For this particular system of c–MYC these experiments support previously performed peptide experiments (chapter 4), and provides further evidence for the efficacy of this exciting class of ligands.

5.7. Acknowledgments

Thanks, are expressed to AstraZeneca and the Jon Waltho group for gifting of the c–MYC plasmid and protein. The DTC in Cell and Proteomic Technologies EP/F500424/1, and BBSRC awards BB/L015048/1 and BB/H013636/1 are gratefully acknowledged. Thanks, are also extended to the British Mass Spectrometry Society for a grant that allowed the purchase of a nanospray tip pipette puller.

5.8. References

[1] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J. Mol. Biol.* **337** (2004) 635–645.

[2] C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions, *Annu. Rev. Biochem.* **83** (2014) 553–584.

[3] P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation, *Nat. Rev. Mol. Cell Biol.* **16** (2015) 18–29.

[4] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Lakoucheva, Z. Obradović, Intrinsic disorder and protein function, *Biochemistry.* **41** (2002) 6573–6582.

[5] P. E. Wright, H. J. Dyson, Linking folding and binding, *Curr. Opin. Struct. Biol.* **19** (2009) 31–38.

[6] H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol.* **6** (2005) 197–208.

[7] S. Pelengaris, M. Khan, G. Evan, c–MYC: more than just a matter of life and death, *Nat. Rev. Cancer.* **2** (2002) 764–776.

[8] A. L. Hsieh, Z. E. Walton, B. J. Altman, Z. E. Stine, C. V. Dang, MYC and metabolism on the path to cancer, *Semin. Cell Dev. Biol.* **43** (2015) 11–21.

[9] S. Sammak, G. Zinzalla, Targeting protein–protein interactions (PPIs) of transcription factors: challenges of intrinsically disordered proteins (IDPs) and regions (IDRs), *Prog. Biophys. Mol. Biol.* **119** (2015) 41–46.

[10] D. Horiuchi, B. Anderton, A. Goga, Taking on challenging targets: making MYC druggable, *Am. Soc. Clin. Oncol. Educ. Book.* **34** (2014) 497–502.

[11] Q. Li, C. V. Dang, c–Myc overexpression uncouples DNA replication from mitosis, *Mol. Cell. Biol.* **19** (1999) 5339–5351.

[12] L. Soucek, J. Whitfield, C. P. Martins, A. J. Finch, D. J. Murphy, N. M. Sodir, A. N. Karnezis, L. B. Swigart, S. Nasi. G. I. Evan, Modelling Myc inhibition as a cancer therapy, *Nature.* **455** (2008) 679–683.

[13] B. Eyss, M. Eilers, Addicted to Myc – but why?, *Genes Dev.* **25** (2011) 895–897.

[14] H. Wang, S. Mannava, V. Grachtchouk, D. Zhuang, M. S. Soengas, A. V. Gudkov, E. V. Prochownik, M. A. Nikiforov, c–Myc depletion inhibits proliferation of human tumor cells at various stages of the cell cycle, *Oncogene.* **27** (2008) 1905–1915.

[15] S. Walz, M. Eilers, Unlocking the mysterious mechanisms of Myc, *Nat. Med.* **19** (2013) 26–27.

[16] P. Lavigne, M. P. Crump, S. M. Gagne, R. S. Hodges, C. M. Kay, B. D. Sykes, Insights into the mechanism of heterodimerization from the ^{1}H–NMR solution structure of the c–Myc–Max heterodimeric leucine zipper, *J. Mol. Biol.* **281** (1998) 165–181.

[17] K. Y. Jung, H. Wang, P. Teriete, J. L. Yap, L. Chen, M. E. Lanning, A. Hu, L. J. Lambert, T. Holien, A. Sundan, N. D. Cosford, E. V. Prochownik, S. Fletcher, Perturbation of the c–Myc–Max protein–protein interaction *via* synthetic α–helix mimetics, *J. Med. Chem.* **58** (2015) 3002–3024.

[18] N. Sodir, G. I. Evan, Nursing some sense out of Myc, *J. Biol.* **8** (2009) 77.

[19] H. Zirath, A. Frenzel, G. Oliynyk, L. Segerström, U. K. Westermark, K. Larsson, M. Munksgaard–Persson, K. Hultenby, J. Lehtiö, C. Einvi, S. Pahlman, P. Kogner, P. J. Jakobsson, M. A. Henriksson, MYC inhibition induces metabolic changes leading to accumulation of lipid droplets in tumor cells, *Proc. Natl. Acad. Sci.* **110** (2013) 10258–10263.

[20] I. Müller, K. Larsson, A. Frenzel, G. Oliynyk, H. Zirath, E. V. Prochownik, N. J. Westwood, M. A. Henriksson, Targeting of the MYCN protein with small molecule c–MYC inhibitors, *PLoS One.* **9** (2014) e97285.

[21] L. Soucek, G. I. Evan, The ups and downs of MYC biology, *Curr. Opin. Genet. Dev.* **20** (2010) 91-95.

[22] L. Soucek, J. R. Whitfield, N. M. Sodir, D. Massó–Vallés, E. Serrano, A. N. Karnezis, L. B. Swigart, G. I. Evans*,* Inhibition of Myc family proteins eradicates KRas–driven lung cancer in mice, *Genes Dev*. **27** (2013) 504–513.

[23] D. Annibali, J. R. Whitfield, E. Favuzzi, T. Jauset, E. Serrano, I. Cuartas, S. Redondo–Campos, G. Folch, A. Gonzalez–Junca, N. M. Sodir, D. Masso–Valles, M. E. Beaulieu, L. B. Swigart, M. M. McGee, M. P. Somma, S. Nasi, J. Seoane, G. I. Evans, L. Soucek, Myc inhibition is effective against glioma and reveals a role for Myc in proficient mitosis, *Nat. Commun.* **5** (2014) 4632.

[24] J. E. Delmore, G. C. Issa, M. E. Lemieux, P. B. Rahl, J. Shi, H. M. Jacobs, E. Kastritis, T. Gilpatrick, R. M. Paranal, J. Qi, M. Chesi, A. C. Schinzel, M. R. McKeown, T. P. Heffernan, C, R. Vakoc, P. L. Bergsagel, I. M. Ghobrial, P. G. Richardson, R. A. Young, W. C. Hahn, K. C. Anderson, A. L. Kung, J. E. Bradner, C. S. Mitsiades, BET bromodomain inhibition as a therapeutic strategy target c–Myc, *Cell.* **146** (2012) 904–917.

[25] D. I. Hammoudeh, A. V. Follis, E. V. Prochownik, S. J. Metallo, Multiple independent binding sites for small–molecule inhibitors on the oncoprotein c–Myc, *J. Am. Chem. Soc.* **131** (2009) 7390–7401.

[26] V. N. Uversky, Intrinsically disordered proteins from A to Z, *Int. J. Biochem. Cell Biol.* **43** (2011) 1090–1103.

[27] N. Rangarajan, Z. Fox, A. Singh, P. Kulkarni, G. Rangarajan, Disorder, oscillatory dynamics and state switching: the role of c–Myc, *J. Theor. Biol.* **386** (2015) 105–114.

[28] S. R. Harvey, C. E. MacPhee, P. E. Barran, Ion mobility–mass spectrometry for peptide analysis, *Methods.* **54** (2011) 454–461.

[29] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, H. H. Hill, Ion mobility–mass spectrometry, *J. Mass Spectrom.* **43** (2008) 1–22.

[30] J. T. Hopper, C. V. Robinson Mass spectrometry quantifies protein interactions—from molecular chaperones to membrane porins, *Angew. Chemie. Int. Ed. Engl.* **53** (2014) 14002-14015.

[31] S. L. Bernstein, D. Liu, T. Wyttenbach, M. T. Bowers, J. C. Lee, H. B. Gray, J. R. Winkler, α–synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation, *J. Am. Soc. Mass Spectrom.* **15** (2004) 1435–1443.

[32] K. Pagel, E. Natan, Z. Hall, A. R. Fersht, C. V. Robinson, Intrinsically disordered p53 and its complexes populate compact conformations in the gas–phase, *Angew. Chemie. Int. Ed. Engl.* **52** (2013) 361–365.

[33] L. Han, B. T. Ruotolo, Ion mobility–mass spectrometry differentiates protein quaternary structures formed in solution and in electrospray droplets, *Anal. Chem*. **87** (2015) 6808–6813.

[34] R. Beveridge, Q. Chappuis, C. MacPhee, P. E. Barran, Mass spectrometry methods for intrinsically disordered proteins, *Analyst.* **138** (2012) 32–42.

[35] E. Jurneczko, F. Cruickshank, M. Porrini, P. Nikolova, I. D. G. Campuzano, M. Morris, P. E. Barran, Intrinsic disorder in proteins:a challenge for (un)structural biology met by ion mobility–mass spectrometry, *Biochem. Soc. Trans.* **40** (2012) 1021–1026.

[36] S. Niu, J. N. Rabuck, B. T. Ruotolo, Ion mobility–mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development, *Curr. Opin. Chem. Biol.* **17** (2013) 809–817.

[37] S. R. Harvey, M. Porrini, C. Stachl, D. Macmillan, G. Zinzalla, P. E. Barran, Small molecule inhibition of c–MYC:MAX leucine zipper formation is revealed by ion mobility–mass spectrometry, *J. Am. Chem. Soc.* **134** (2012) 19384–19392.

[38] R. Beveridge, S. Covill, K. J. Pacholarz, J. Kalapothakis, C. E. MacPhee, P. E. Barran, A mass–spectrometry–based framework to define the extent of disorder in proteins, *Anal. Chem.* **86** (2014) 10979–10991.

[39] A. V. Follis, D. I. Hammoudeh, A. T. Daab, S. J. Metallo, Small–molecule perturbation of competing interactions between c–Myc and Max, *Bioorg. Med. Chem. Lett*. **19** (2009) 807–810.

[40] B. J. McCullough, J. Kalapothakis, H. Eastwood, P. Kemper, D. MacMillan, K. Taylor, J. Dorrin, P. E. Barran, Development of an ion mobility quadrupole time of flight mass spectrometer, *Anal. Chem.* **80** (2008) 6336–6344.

[41] M. Zhang, H. Y. Fan, S. C. Li, Inhibition of c–Myc by 10058–F4 induces growth arrest and chemosensitivity in pancreatic ductal adenocarcinoma, *Biomed. Pharmacother*. **73** (2015) 123–128.

[42] J. Guo, R. A. Parise, E. Joseph, M. J. Egorin, J. S. Lazo, E. V. Prochownik, J. L. Eiseman, Efficacy, pharmacokinetics, tisssue distribution, and metabolism of the Myc–Max disruptor, 10058–F4 [Z, E]–5–[4–ethylbenzylidine]–2–thioxothiazolidin–4–one, in mice, *Cancer Chemother. Pharmacol.* **63** (2009) 615–625.

[43] J. Wanner, D. Romashko, D. S. Werner, E. W. May, Y. Peng, R. Schulz, K. W. Foreman, S. Russo, L. D. Arnold, M. Pinghle, D. E. Bergstrom, F. Barany, S. Thomson, Reversible linkage of two distinct small molecule inhibitors of Myc generates a dimeric inhibitor with improved potency that is active in Myc over–expressing cancer cell lines, *PLoS One*. **10** (2015) e0121793.

[44] M. R. McKeown, J. E. Bradner, Therapeutic strategies to inhibit MYC, *Cold Spring*

Harb. Perspect. Med. **4** (2014) a014266.

[45] T. Berg, Inhibition of transcription factors with small organic molecules, *Curr. Opin. Chem. Biol.* **12** (2008) 464–471.

[46] G. Mustata, A. V. Follis, D. I. Hammoudeh, S. J. Metallo, H. Wang, E. V. Prochownik, J. S. Lazo, I. Bahar, Discovery of novel MYC–MAX heterodimer disruptors with a three–dimensional pharmacophore model*, J. Med. Chem.* **52** (2009) 1247–1250.

[47] H. Wang, A. Ramakrishnan, S. Fletcher, E. V. Prochownik, A quantitative, surface plasmon resonance–based approach to evaluating DNA binding by the c–Myc oncoprotein and disruption by small molecule inhibitors, *J. Biol. Method*. **2** (2015) e18.

[48] H. Wang, J. Chauhan, A. Hu, K. Pendleton, J. L. Yap, E. Sabato, J. W. Jones, M. Perri, J. Wu, E. Cione, M. A. Kane, S. Fletcher, E. V. Prochownik, Disruption of Myc–Max heterodimerization with improved cell–penetrating analogs of the small molecule 10074–G5, *Oncotarget.* **4** (2013) 936–947.

[49] D. Clausen, J. Guo, R. Parise, J. Beumer, M. Egorin, J. S. Lazo, E. V. Prochownik, J. L. Eiseman, *In vitro* cytotoxicity and *in vivo* efficacy, pharmacokinetics, and metabolism of 10074–G5, a novel small–molecule inhibitor of c–Myc/Max dimerization, *J. Pharmacol. Exp. Ther.* **335** (2010) 715–727.

[50] T. Berg, Small–molecule modulators of c–Myc/Max and Max/Max interactions, *Curr. Top. Microbiol. Immunol.* **348** (2010) 139–149.

[51] F. Balthasart, J. Plavec, V. Gabelica, Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species?, *J. Am. Soc. Mass Spectrom.* **24** (2013) 1–8.

6.

Conclusions

Conclusions

As described at the beginning of this work, proteins function in a comparable way to cogs in machines, interacting with other proteins to perform biological processes. With increased understanding of cell pathways, the field of biochemistry is learning more and more about how many interaction partner proteins behave and the extent of how their structural features influence these interactions. Mass spectrometry is a highly advantageous technique for looking at these interacting systems as it provides information on individual complexes rather than an average across a whole system. In addition to this, mass spectrometry has a wide range of hybridised and analogous techniques which complement it, allowing greater depth and breadth of information to be determined within a single core experiment. Ion mobility–mass spectrometry has been extensively used throughout this thesis, bringing strong structural information into discussions of non–covalent interactions. Molecular systems are not static and rather exist in a constant dynamic state. Understanding structural rearrangements as two molecules interact allows a wealth of perspective on their behaviour.

The objective of the work presented in this thesis was to use mass spectrometry techniques to analyse non–covalent systems in the gas–phase, relating observed structural changes to protein:protein or molecule:ligand effects. The work spans a broad range of different systems, each representative of a different area of interest. Projects ranged from chemically synthesised molecular scaffolds, through to DNA, peptides, and up to protein complexes. The work is bound together through the analysis of non–covalent complexes which are traditionally difficult to analyse through mass spectrometry. A great deal of care and specialised handling was required to faithfully transmit these fragile structures into the gas–phase. The work shown details fine tuning of experimental conditions to obtain data on these systems.

In chapter 2 a DNA aptamer was analysed which had been designed to bind the antibiotic kanamycin. The aptamer had been designed to detect the presence of kanamycin in milk *via* an attached gold nanoparticle. The project began with the humble intention of observing the DNA:ligand complex in the gas–phase. When this was achieved, the next step was to determine if the interactions between these two molecules could be assessed in a structural manner using various mass spectrometry techniques, including multiple forms of fragmentation and ion mobility. The experiments all pointed towards the loop region being the most important region for the interaction, with the chemistry being as important as the structure. Throughout the various workflows, interesting solvent–dependent features began to develop which showed how DNA behaves differently to proteins. This work is one of a very small number of papers analysing DNA structure through ion mobility–mass spectrometry. This work was carried out in the negative ionisation mode of n–ESI which is less frequently utilised than the standard positive mode, requiring vastly different instrument tuning.

Chapter 3 results from collaboration with a group in Australia investigating a molecular anion sensor designed to bind pyrophosphate in preference to other similar molecules. A variety of structurally similar molecules with differences in bonding arm length, scaffold ring size and orientation were analysed. Mass spectrometry was used to measure the strength of these interactions through kinetic binding experiments to compare and uncover which molecules were most selective and sensitive. It was determined that an intermediate arm length was most selective for pyrophosphate and so ion mobility was employed to relate this binding information to structural changes. It was observed that the more effective molecules underwent a structural tightening in the presence of pyrophosphate. This work connected experimental mass spectrometry and ion mobility–mass spectrometry with molecular design and function, to determine why certain structural features were more favourable than others. It utilised the dynamic nature of the data acquisition to observe how features changed in the presence of ligands.

Chapter 4 moves to a peptide system including the interacting leucine zipper region of two proteins c–MYC and its partner MAX. c–MYC is an attractive target for cancer research as it is often overexpressed in human cancer; however, its disordered structure makes it difficult to target and study. This chapter investigated several small molecules that have been shown to inhibit c–MYC:MAX dimerisation. It has been postulated that this occurs through stabilisation of disordered structures of c–MYC. The ligands were analysed using mass spectrometry and ion mobility–mass spectrometry to study their effects on the dimers. Furthermore, the structural changes observed through changes in the collisional cross section distribution were related back to ligand efficacy of inhibition. It was observed that the majority of the ligands led to the reduction of the more extended states, which agrees with previous work. This thesis highlights how ion mobility–mass spectrometry can provide structural data to aid in drug design in a way no other technique could achieve on the same timescale. As IM–MS instrumentation becomes more commonly utilised, the methods used in this paper could become a staple of the pharmaceutical industry to observe the effects of many ligands on their target.

Whereas in chapter 4, peptides containing the leucine zipper domain of c–MYC and MAX had been analysed, in chapter 5 the project moved to a larger construct also containing the DNA binding domain. This opened up a much broader range of experiments to assess the extent of influence of the leucine zipper binding when other domains were introduced, along with experiments to measure the extent the complex could bind DNA. In the absence of DNA, the c–MYC:MAX dimer was observed at low intensity and became more extended in the presence of the ligands. When DNA was introduced, the observed peaks were of much higher intensity, indicating a stabilisation of the structure. The conformations appear to become more compact with the addition of the ligands, and their effects on complex intensity are larger. Protein:DNA binding is crucial to transcription and translation of the genome, and to be able to observe the effects of ligands on these interactions in a dynamic way could prove extraordinarily useful. These complexes of around 25 kDa are too large to be routinely observed in an NMR experiment, and creating a native crystal of these molecules could prove difficult. Ion mobility–mass spectrometry allowed the effects of several ligands to be analysed on a relatively short time scale, highlighting how they interacted with the complex in different ways. Linking this to work done on the peptide and computational level allows insight into how the ligands function and how best to design them in future.

Mass spectrometry is a highly flexible technique for analysing non–covalent complexes in the gas–phase, and the addition of ion mobility hybridisation generates a wealth of structural information on changing conformations. Being able to observe bound and unbound species and compare their dynamic behaviour allows measurement of interactions to be linked to molecular movement and flexibility within systems. This thesis has been conducted on several molecular species which would prove incredibly challenging to any technique that was not sufficiently gentle to preserve these fragile and dynamic interactions. Great care was taken to ensure all the data collected were as close to biological conditions as could be replicated in the gas–phase. This body of work represents a variety of molecules at the very edge of mass spectral capability with high biological relevance. The workflows presented herein could not currently be replicated on an industrial or high throughput scale, however may in the future become staples of the pharmaceutical industry as technology improves.

7.

Appendices

Table of Contents

S2.1: DNA fragmentation nomenclature outlined by McLuckey [62]. This convention was used throughout.

S2.2 A) Spectra of Ky2 aptamer (20 µM), kanamycin (100 µM), 50:50 methanol:H_2O. B) Spectra of Ky2 aptamer (20 µM) with kanamycin (100 µM), 25% (vol %) methanol:H_2O and 100 mM ammonium acetate. C) Spectra of Ky2 aptamer (20 µM) with 100 mM ammonium acetate.

S2.3: $^{DT}CCSD_{He}$ (A^2) Collisional cross section distribution for the Ky2 aptamer (20 µM) in the $[M - 4H]^{4-}$ through $[M - 6H]^{6-}$ charge states solvent conditions were 50:50 methanol:H_2O. A) Control experiment of Ky2 aptamer with no ligand added. B) Unbound Ky2 aptamer in the presence of kanamycin (100 µM kanamycin). C) Bound ligand:aptamer complex (100 µM kanamycin).

S2.4: CIU of Ky2 aptamer (20 µM) under 50:50 methanol:H_2O solution conditions for $[M - 4H]^{4-}$, $[M - 5H]^{5-}$, and $[M - 6H]^{6-}$ charge states in the presence and absence of kanamycin (100 µM).

S3.1: n–ESI spectra of ligand 1 for (A) the *apo* form, and with:(B) pyrophosphate, (C) pyrocatechol violet, (D) ATP. Conditions: 50 µM ligand 1 with 5 mM ammonium acetate and 5% propan–2–ol, pH 6.8. (B) 200 µM sodium pyrophosphate, (C) 200 µM pyrocatechol violet and (D) 100 µM adenosine 5′–triphosphate (ATP) disodium salt hydrate.

S3.2: n–ESI spectra of 4 for (A) the *apo* form, and with:(B) pyrophosphate, (C) pyrocatechol violet, (D) ATP. Conditions 50 µM 4, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 µM sodium pyrophosphate (C) 200 µM pyrocatechol violet (D) 100 µM adenosine 5′–triphosphate (ATP) disodium salt hydrate.

S3.3: n–ESI spectra of 5 for (A) the *apo* form, and with:(B) pyrophosphate, (C) pyrocatechol violet (D) ATP. Conditions 50 µM 5, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 µM sodium pyrophosphate (C) 200 µM pyrocatechol violet (D) 100 µM adenosine 5′–triphosphate (ATP) disodium salt hydrate.

S3.4: n–ESI spectra of 7 for (A) the *apo* form, and with:(B) pyrophosphate, (C) pyrocatechol violet, (D) ATP. Conditions 50 µM 7, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 µM sodium pyrophosphate (C) 200 µM pyrocatechol violet (D) 100 µM adenosine 5′–triphosphate (ATP) disodium salt hydrate.

S3.5. Experimental ($^{DT}CCS_{He}$) collisional cross sections for all macrocycles and ligands. Error bars represent standard deviation over three repeats. *Apo* form has been offset for clarity.

S3.6: Comparison between gas–phase and water molecular dynamics simulations. Top graphs represent gas–phase: a) root mean square deviation throughout the 5 ns MD run; b) collisional cross section vs. radius of gyration. Bottom graphs represent water: c) root mean square deviation throughout 2 ns MD run; d) collisional cross section vs radius of gyration. Collisional cross sections were calculated using MOBCAL's trajectory method with error bars representing the standard deviation. Experimental data is superimposed with standard deviation.

S3.7: Radius of gyration distribution through the simulation. Gas–phase molecular dynamics showed a wider range of conformational diversity than water simulations as shown by the wider spread of radius of gyration.

S4.1: n–ESI mass spectra of c–MYC (50 µM) and MAX (50 µM) in 100 mM ammonium acetate and 5% DMSO. *m/z* 1700–1900. Blue band indicates MAX homodimer, green c–MYC homodimer and pink the heterodimer.

S4.2: Collisional cross section distributions of the 5+ c–MYC:MAX heterodimer (*m/z* 1424) (5 Vcm–1). Intensity of $^{DT}CCS_{He}$ is relative to peak intensity in presence of ligand light grey **D** bound (intensity x5), grey two **D** bound (intensity x25), dark grey three **D** bound (intensity x25).

S5:1: 50 µM c–MYC collisonal cross section distributions under four solution conditions: A) 200 mM ammonium acetate, B) 100 mM ammonium acetate, C) 20 mM ammonium acetate and D) 49:49:2 methanol:H_2O:formic acid. X axis denotes $^{DT}CCS_{He}$ and Y axis denotes charge state.

S5.2: n–ESI mass spectra of 50 µM MAX in 100 mM ammonium acetate.

S5.3: n–ESI mass spectra of c–MYC (50 µM) and MAX (50 µM) in 100 mM ammonium acetate with: (A) no ligands, (B) with 100 µM **E**, (C) with 100 µM **A**, (D) with 100 µM **B** and(E) with 100 µM **C.**

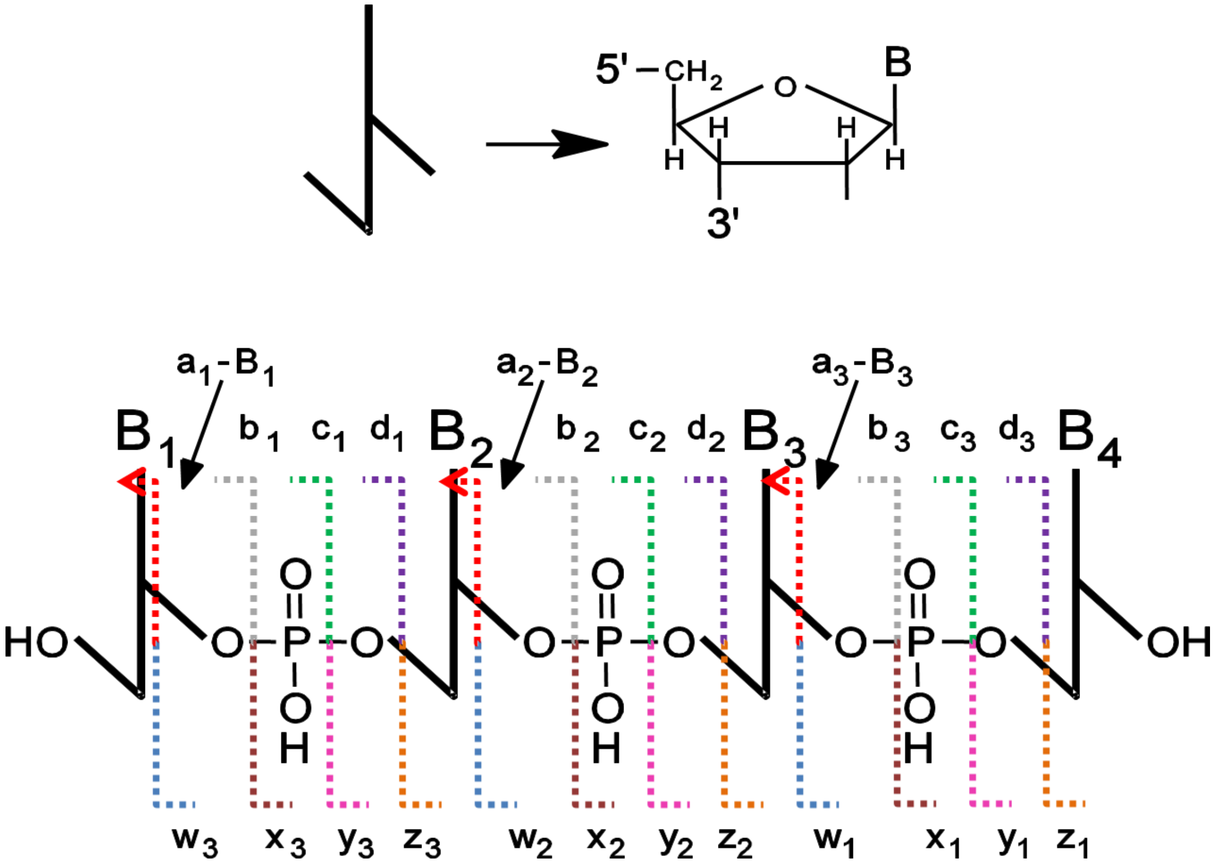
S5.4: n–ESI mass spectra of c–MYC (50 µM) and MAX (50 µM) ds DNA (100 µM) in 100 mM ammonium acetate (A) no ligands, (B) with 100 µM **E,** (C) with 100 µM **A**, (D) with 100 µM **B** and (E) with 100 µM **C.**

S5.5: Collisional cross section $^{DT}CCS_{He}$ (Å^2) for the c–MYC:MAX heterodimer $^{DT}CCS_{He}$. Sample conditions: 50 µM c–MYC, 50 µM MAX, 100 µM ligand and 100 µM DNA from an aqueous solution with 100 mM ammonium acetate. Intensities are normalised within each ligand relative to intensity of *m/z*.

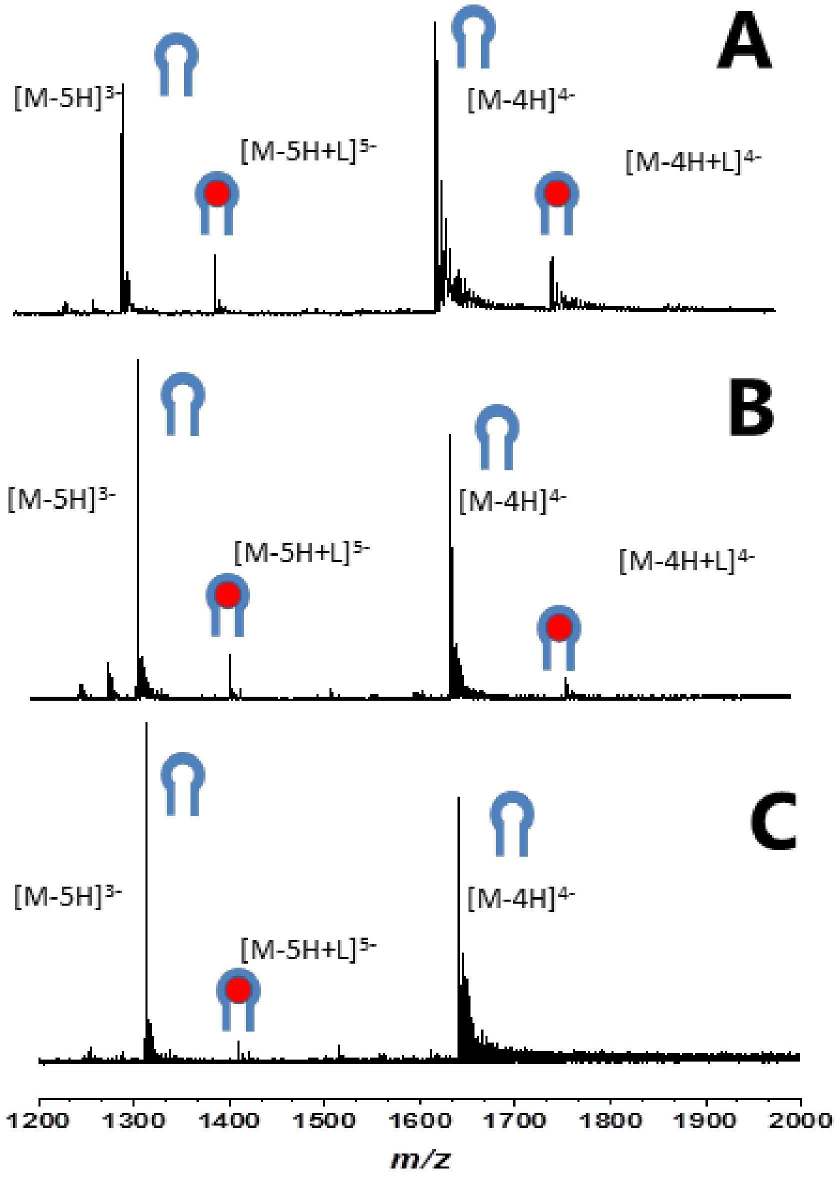
S5.6: Collisional cross section distribution $^{DT}CCSD_{He}$ (Å^2) of MMD 13+. Sample conditions: c–MYC (50 µM), MAX (50 µM), ds DNA (100 µM) in 100 mM ammonium acetate and 100 µM ligand B

Appendix 1

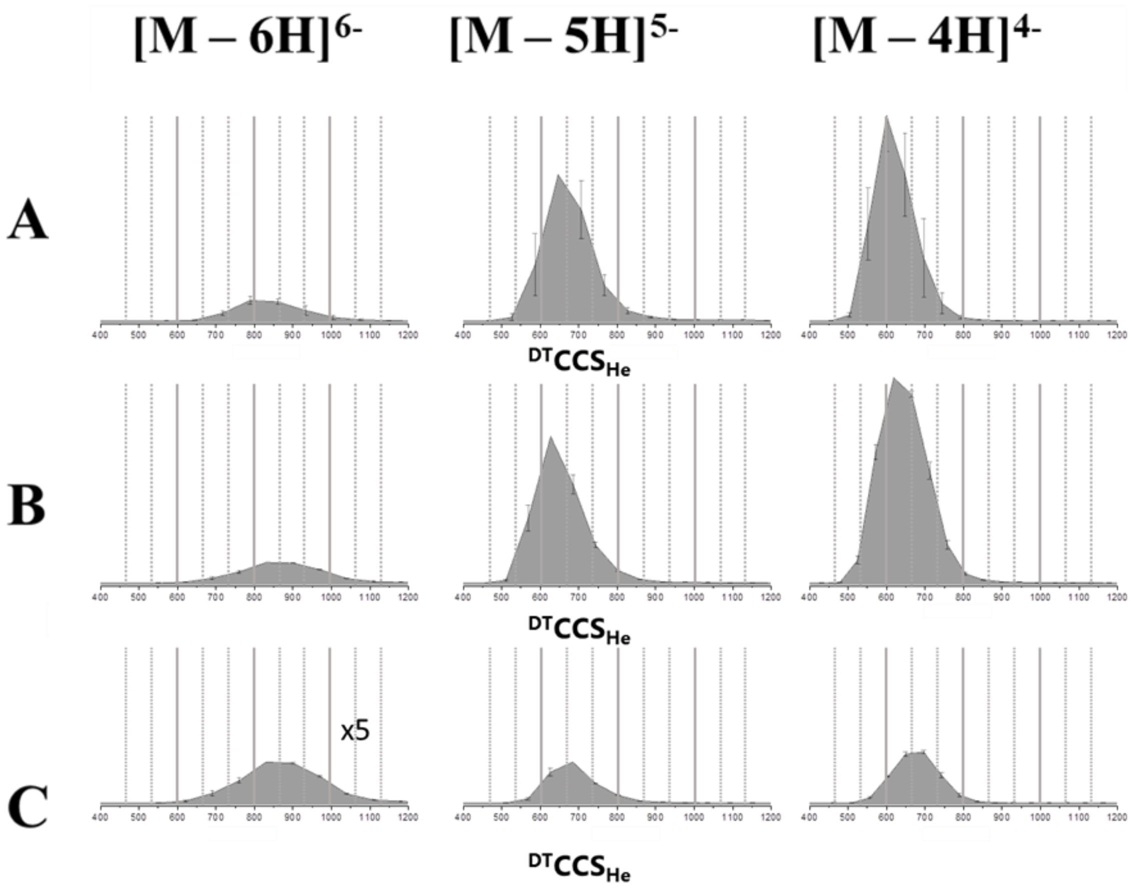
Supplementary information for chapter 2



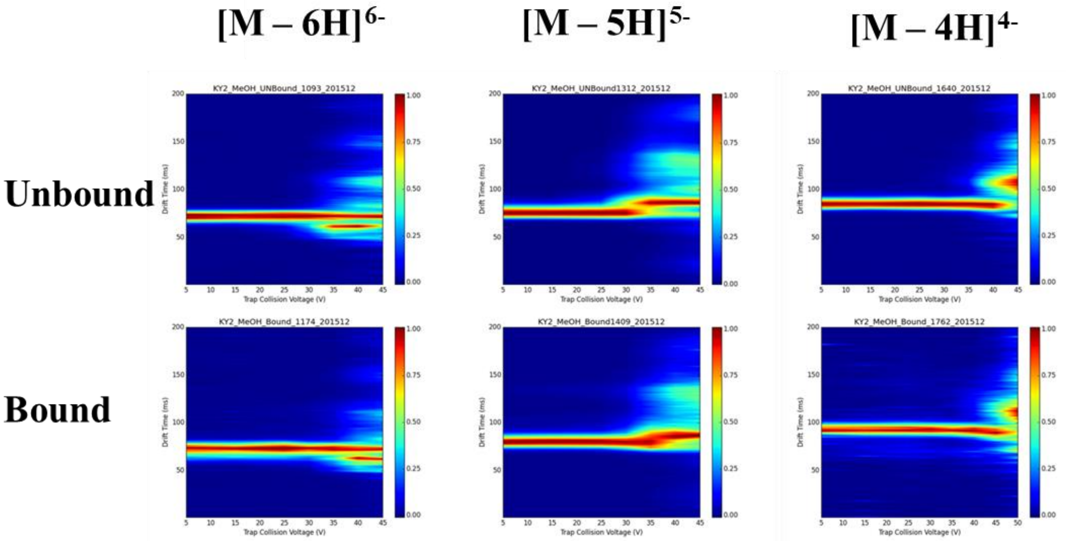
S2.1: DNA fragmentation nomenclature outlined by McLuckey [62]. This convention was used throughout.



S2.2: A) Spectra of Ky2 aptamer (20 µM), kanamycin (100 µM), 50:50 methanol:H_2O. B) Spectra of Ky2 aptamer (20 µM) with kanamycin (100 µM), 25% (vol %) methanol:H_2O and 100 mM ammonium acetate. C) Spectra of Ky2 aptamer (20 µM) with 100 mM ammonium acetate.



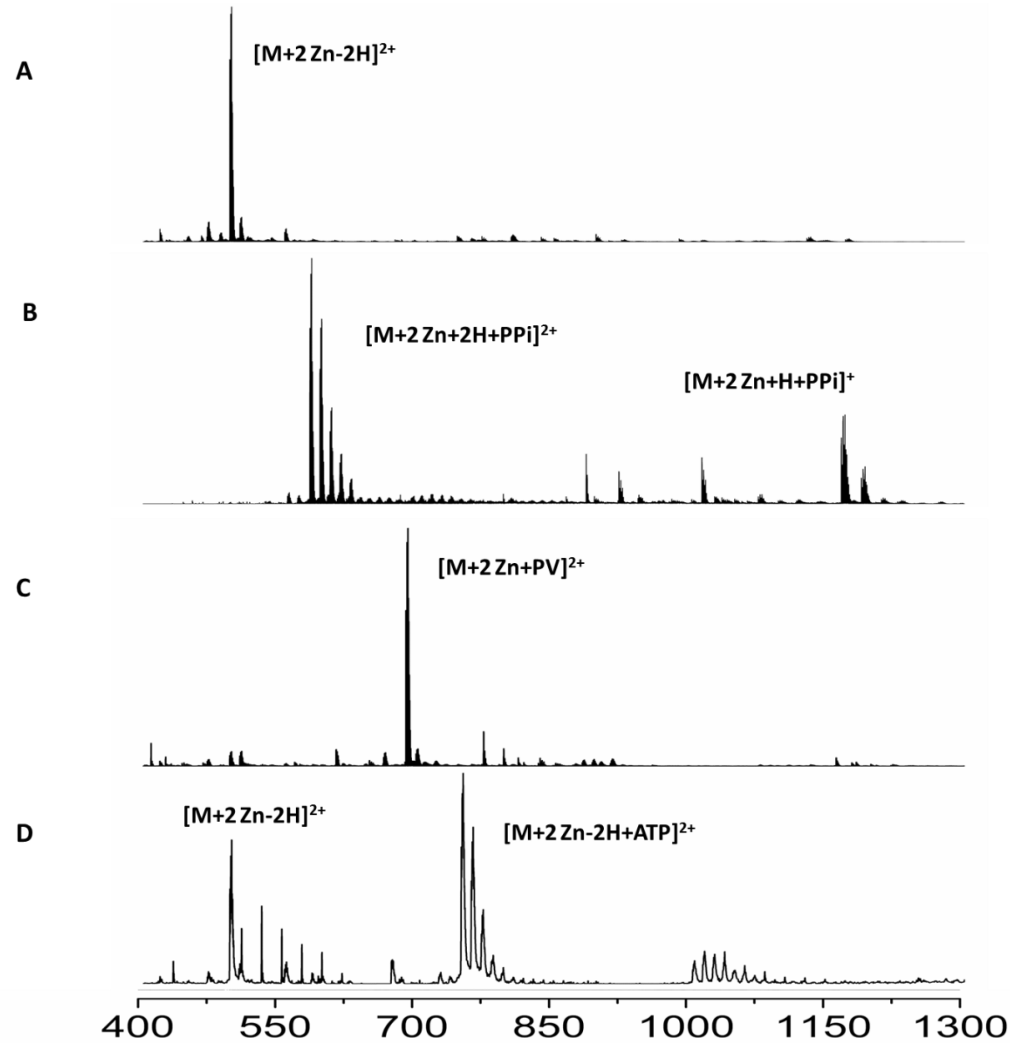
S2.3: $^{DT}CCSD_{He}$ (A^2) Collisional cross section distribution for the Ky2 aptamer (20 µM) in the $[M-4H]^{4-}$ through $[M-6H]^{6-}$ charge states solvent conditions were 50:50 methanol:H_2O. A) Control experiment of Ky2 aptamer with no ligand added. B) Unbound Ky2 aptamer in the presence of kanamycin (100 µM kanamycin). C) Bound ligand:aptamer complex (100 µM kanamycin).



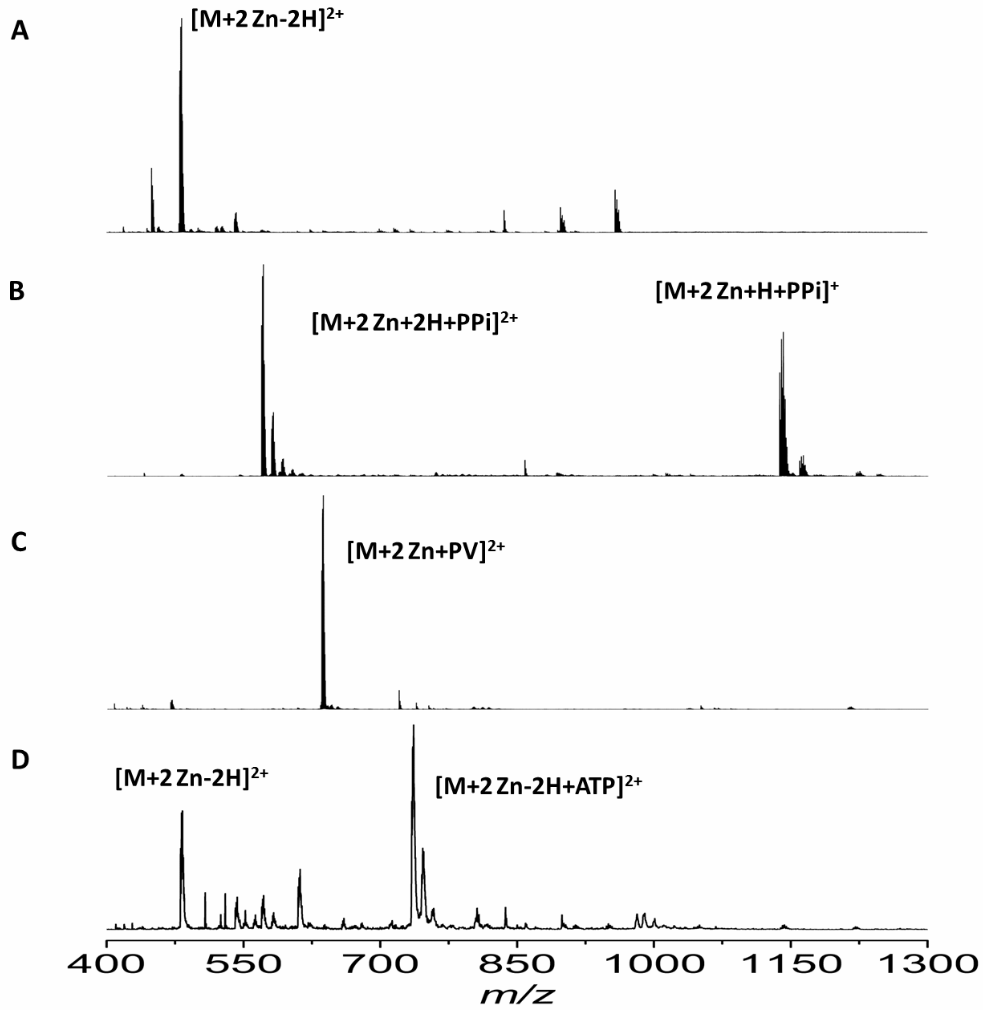
S2.4: CIU of Ky2 aptamer (20 µM) under 50:50 methanol:H_2O solution conditions for $[M – 4H]^{4–}$, $[M – 5H]^{5–}$, and $[M – 6H]^{6–}$ charge states in the presence and absence of kanamycin (100 µM).

Appendix 2

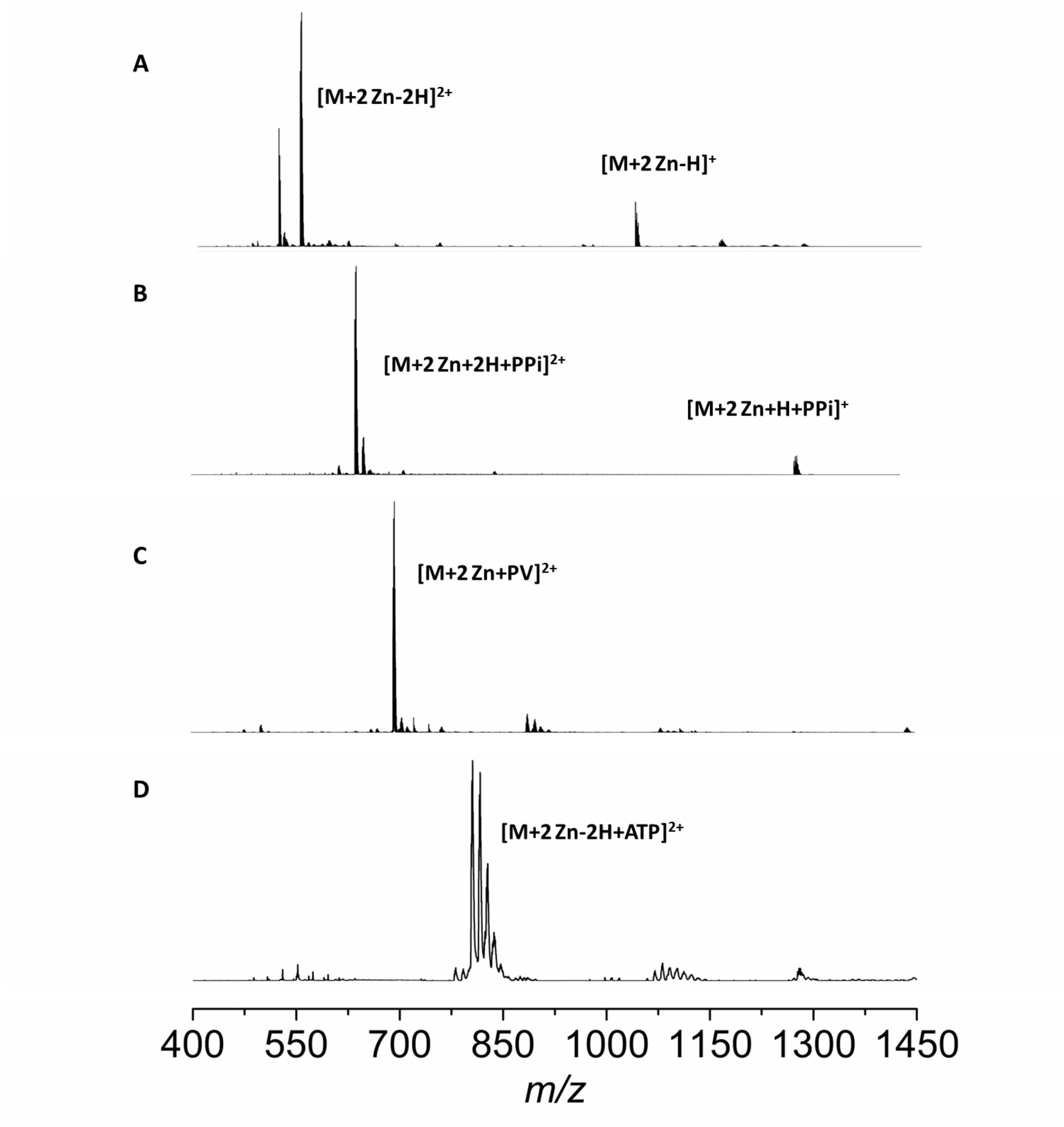
Supplementary information for chapter 3



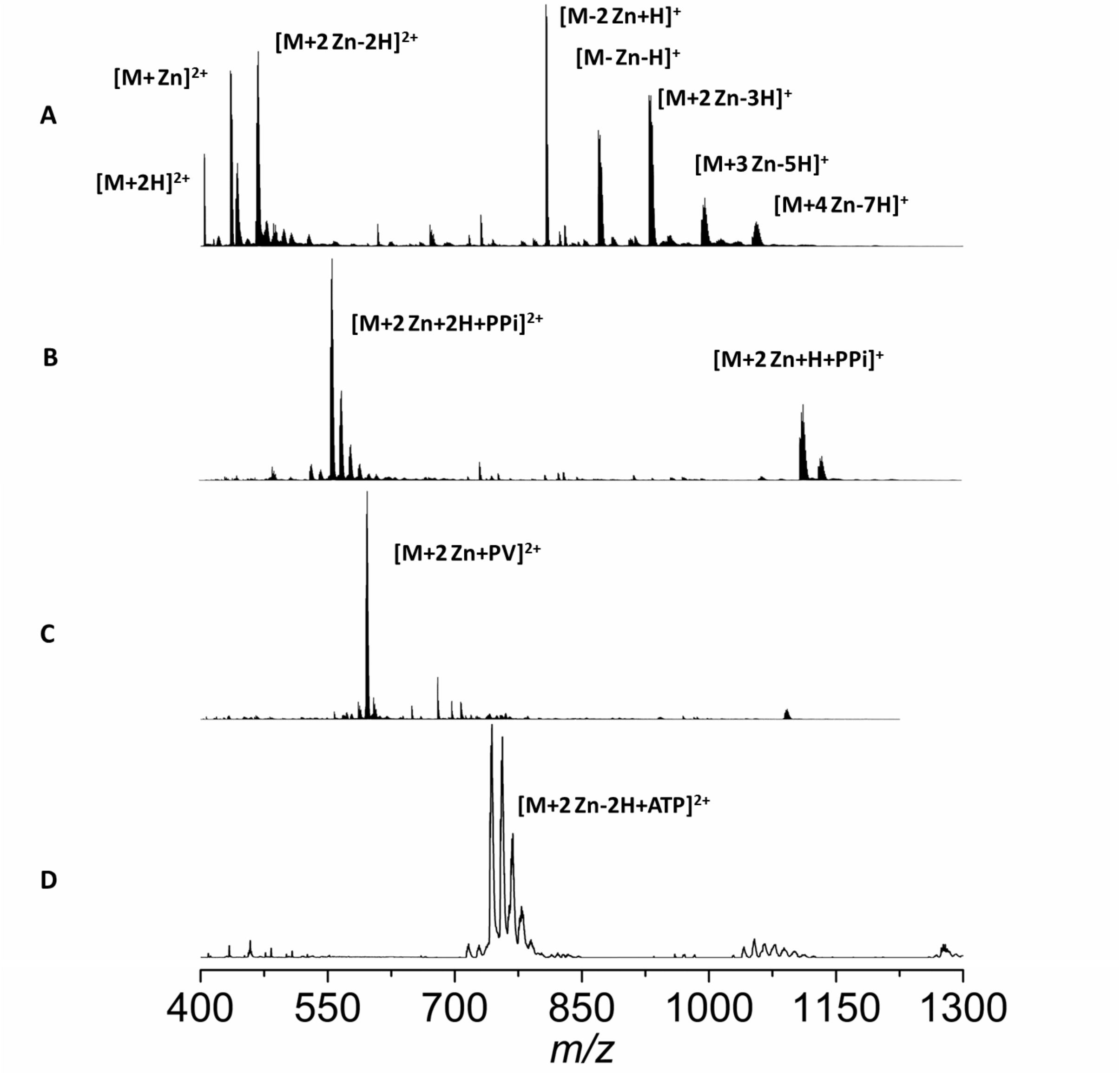
S3.1: n–ESI spectra of ligand 1 for (A) the *apo* form, and with: (B) pyrophosphate, (C) pyrocatechol violet, (D) ATP. Conditions: 50 μM ligand 1 with 5 mM ammonium acetate and 5% propan–2–ol, pH 6.8. (B) 200 μM sodium pyrophosphate, (C) 200 μM pyrocatechol violet and (D) 100 μM adenosine 5′–triphosphate (ATP) disodium salt hydrate.



S3.2: n–ESI spectra ofliagnd 4 for (A) the *apo* form, and with: (B) pyrophosphate, (C) pyrocatechol violet, (D) ATP. Conditions 50 μM ligand 4, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 μM sodium pyrophosphate (C) 200 μM pyrocatechol violet (D) 100 μM adenosine 5′–triphosphate (ATP) disodium salt hydrate.



S3.3: n–ESI spectra of ligand 5 for (A) the *apo* form, and with: (B) pyrophosphate, (C) pyrocatechol violet (D) ATP. Conditions 50 µM ligand 5, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 µM sodium pyrophosphate (C) 200 µM pyrocatechol violet (D) 100 µM adenosine 5′–triphosphate (ATP) disodium salt hydrate.



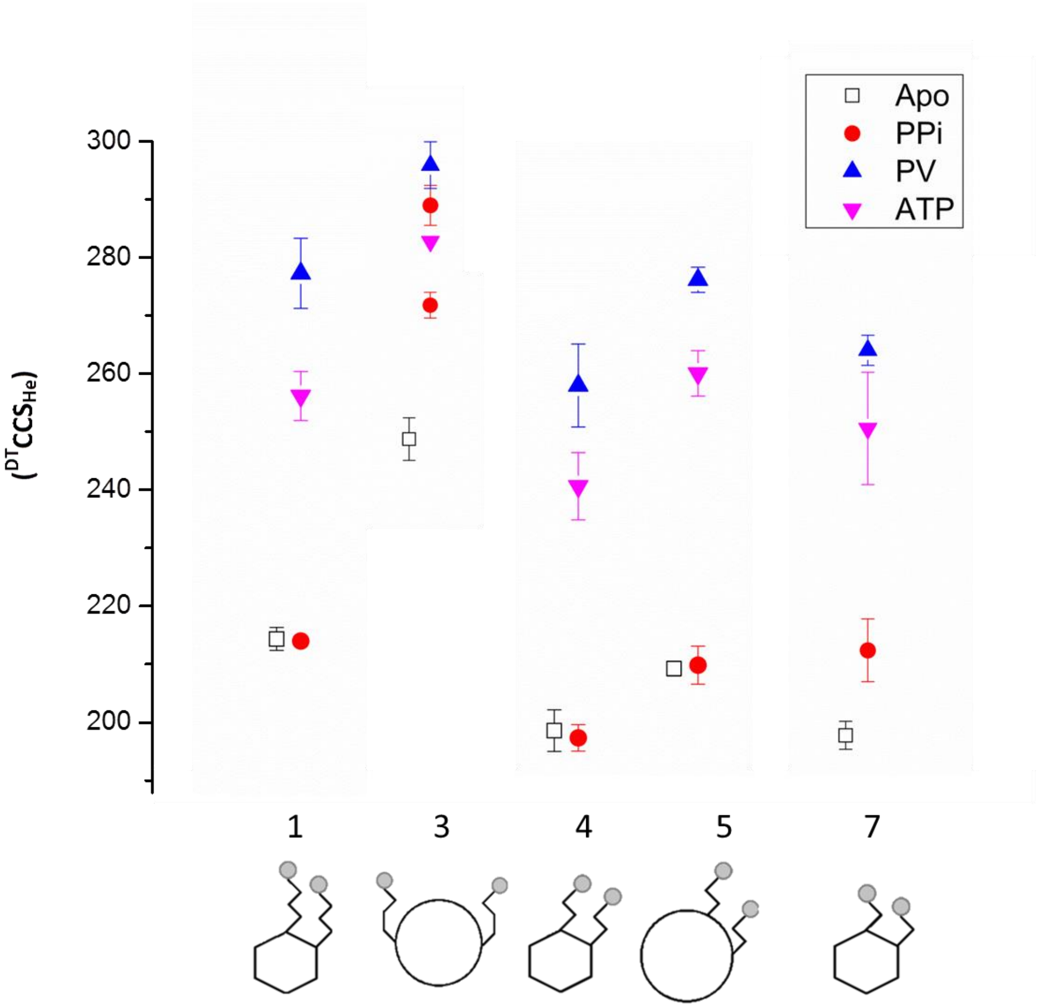
S3.4: n–ESI spectra of ligand 7 for (A) the *apo* form, and with: (B) pyrophosphate, (C) pyrocatechol violet, (D) ATP. Conditions 50 μM ligand 7, 5 mM ammonium acetate, 5% propan–2–ol, pH 6.8. (B) 200 μM sodium pyrophosphate (C) 200 μM pyrocatechol violet (D) 100 μM adenosine 5′–triphosphate (ATP) disodium salt hydrate.

ST 1: (log K_a) Values for solution UV experiments pH 7.4 5 mM HEPES

in the presence of 145 mM NaCl 8 reproduced from reference [51]

	PPi	PV	ATP
1 · Zn_2	>9	8.4	7.2
3 · Zn_2	>9	8.8	7.2
4 · Zn_2	>9	8.5	8.8
5 · Zn_2	>9	9.0	6.7
7 · Zn_2	8.4	7.4	5.4

Estimated errors in log K_a<0.2



S3.5. Experimental ($^{DT}CCS_{He}$) collisional cross sections for all macrocycles and ligands. Error bars represent standard deviation over three repeats. *Apo* form has been offset for clarity.

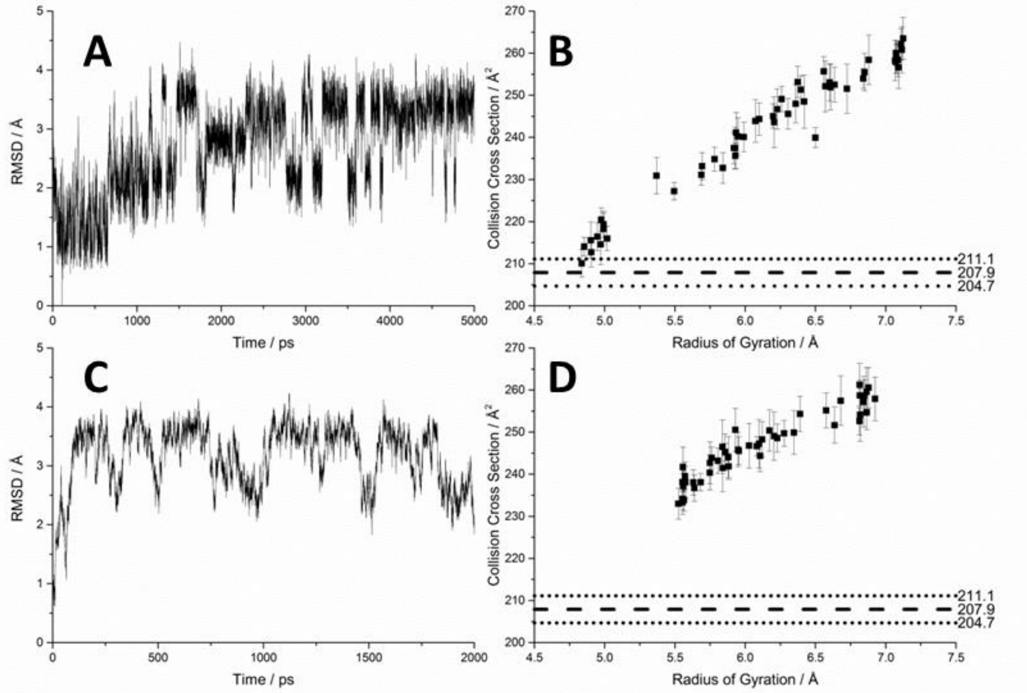
Supplementary information on calculation parameters

The molecule, **4** was built using ChemDraw [74], and initially relaxed with no restraints using AVOGADRO's MMFF94 parameters. The charges on the atoms were derived using the general AMBER force field (GAFF) and the *antechamber* AM1–BCC [76, 77].

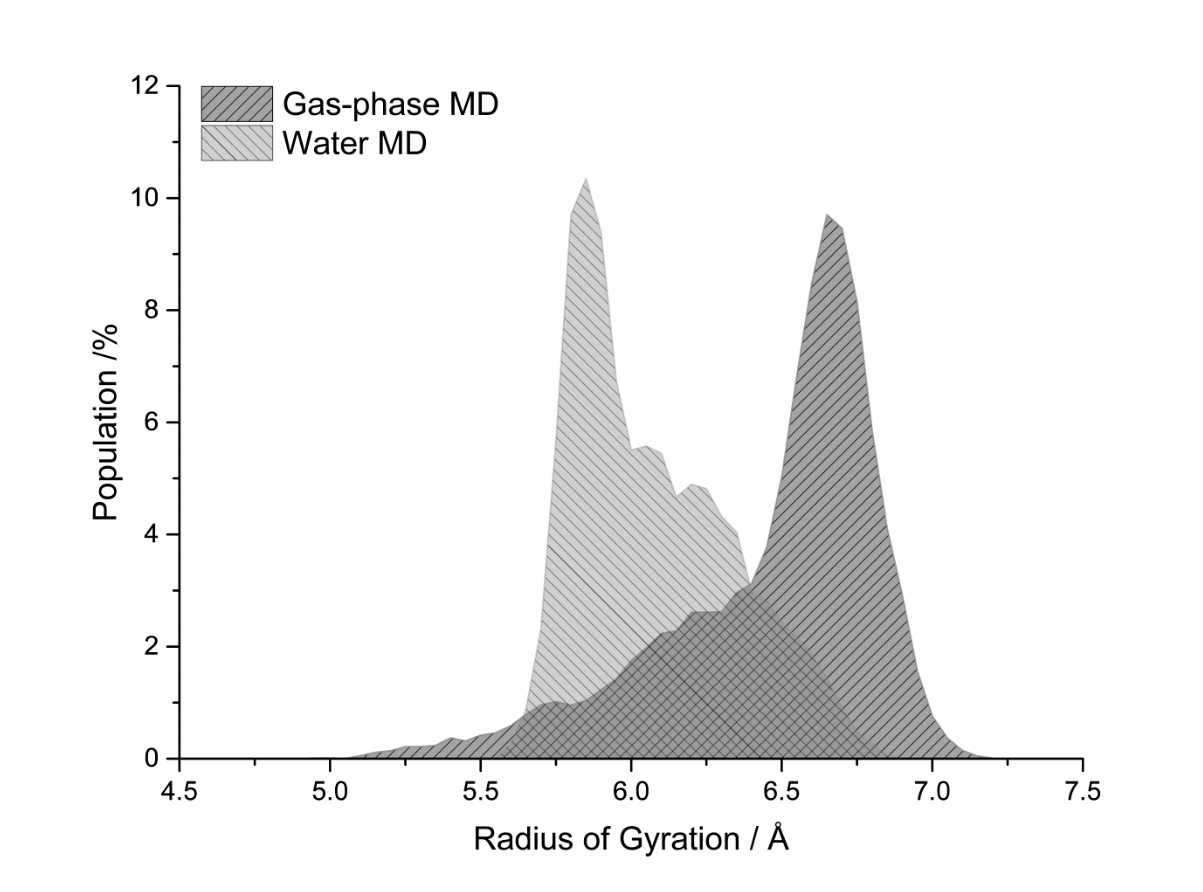
The water simulations were performed in a box and the edge of the box was ~10 Å with ~1100 TIP3P waters. The radial cut–off for non–bonded forces was 10.0 Å, and the particle mesh Ewald (PME) method was used to derive the electrostatic interactions [77]. All bonds involving hydrogen atoms were constrained at their equilibrium distance by applying the SHAKE algorithm[37], and a time step of 2.0 fs was used. The density and temperature were equilibrated prior to solvation with gradual heating up to 300 K using Langevin dynamics with constant volume and temperature and damping coefficient of 5.0 ps^{-1} and weak harmonic potentials of 10 kcal/mol/ $Å^2$ for 200 ps. Production runs were carried out for 2 ns.

The *in–vacuo* simulations were carried out with an "infinite" radial cut–off (i.e. *cut* = 999 Å) for the non–bonded interactions. All bonds involving hydrogens were constrained with SHAKE algorithm, with the time step of 2.0 fs. Production runs were carried out for 5 ns.

The trajectory was subsequently analysed with AmberTools; a total of ~50 models was extracted for each trajectory and the collisional cross sections were calculated using the trajectory method found in MOBCAL and projection superposition approximation (PSA) [81–83].



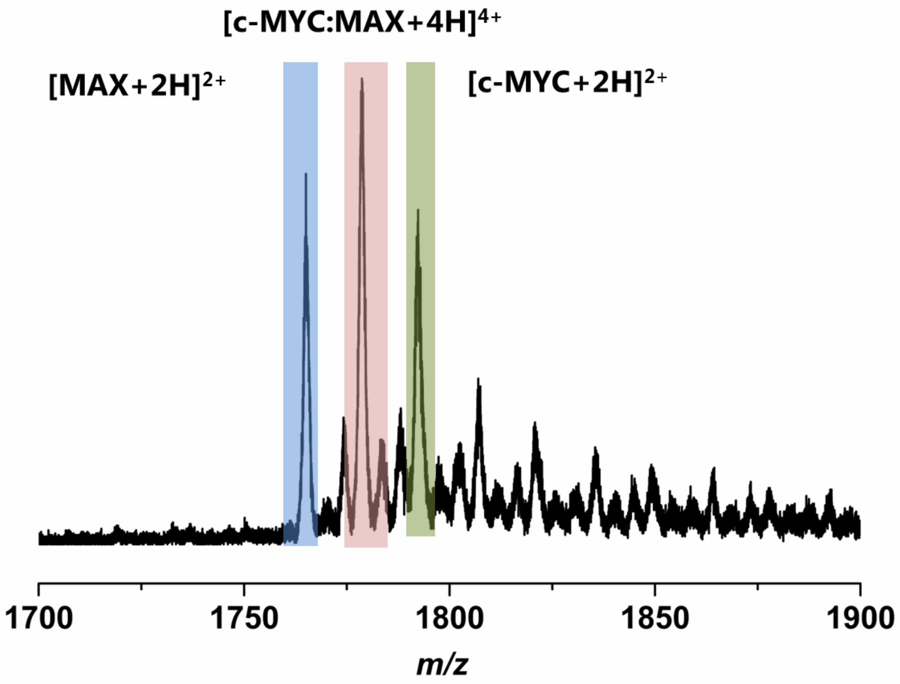
S3.6: Comparison between gas–phase and water molecular dynamics simulations. Top graphs represent gas–phase: a) root mean square deviation throughout the 5 ns MD run; b) collisional cross section vs. radius of gyration. Bottom graphs represent water: c) root mean square deviation throughout 2 ns MD run; d) collisional cross section vs radius of gyration. Collisional cross sections were calculated using MOBCAL's trajectory method with error bars representing the standard deviation. Experimental data is superimposed with standard deviation.



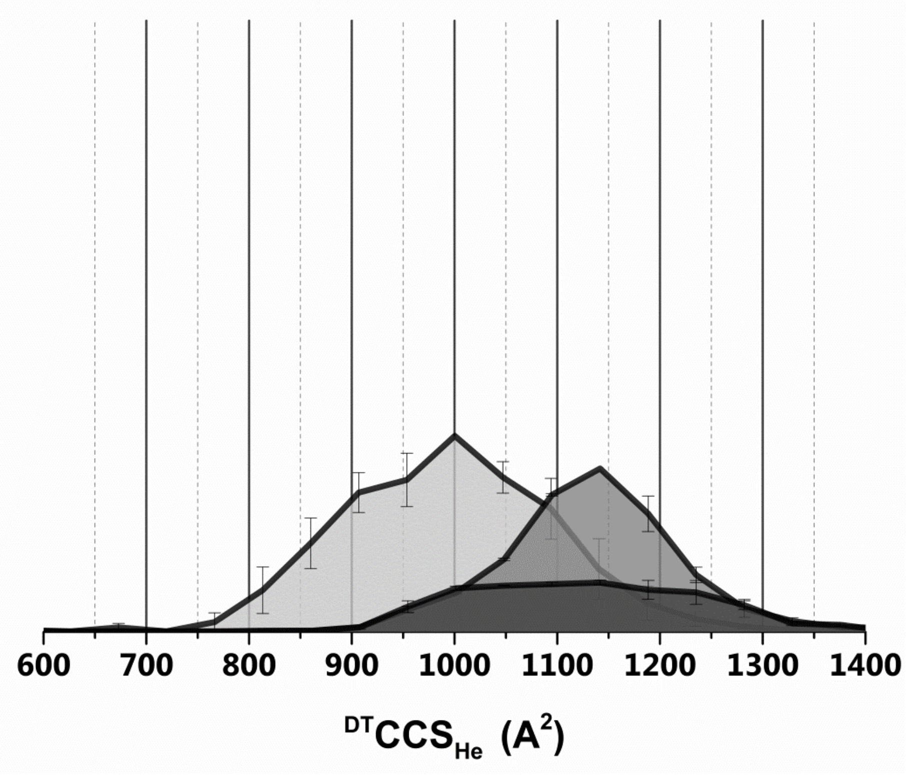
S3.7. Radius of gyration distribution (Å) through the simulation. Gas–phase molecular dynamics showed a wider range of conformational diversity than water simulations as shown by the wider spread of radius of gyration.

Appendix 3

Supplementary information for chapter 4



S4.1: n–ESI mass spectra of c–MYC (50 µM) and MAX (50 µM) in 100 mM ammonium acetate and 5% DMSO. *m/z* 1700–1900. Blue band indicates MAX homodimer, green c–MYC homodimer and pink the heterodimer.



S4.2: Collisional cross section distributions $^{DT}CCSD_{He}$ (Å^2) of the 5+ c–MYC:MAX heterodimer (*m/z* 1424) (5 Vcm–1). Intensity of $^{DT}CCS_{He}$ is relative to peak intensity in presence of ligand light grey D bound (intensity x5), grey two D bound (intensity x25), dark grey three D bound (intensity x25).

Appendix 4

Supplementary information for chapter 5

Max Sequence

SDADKRAHH NALERKRRDH IKDSFHSLRD SVPSLQGEKAS
RAQILDKATE YIQYMRRKNH THQQDIDDLK RQNALLEQQVR AL

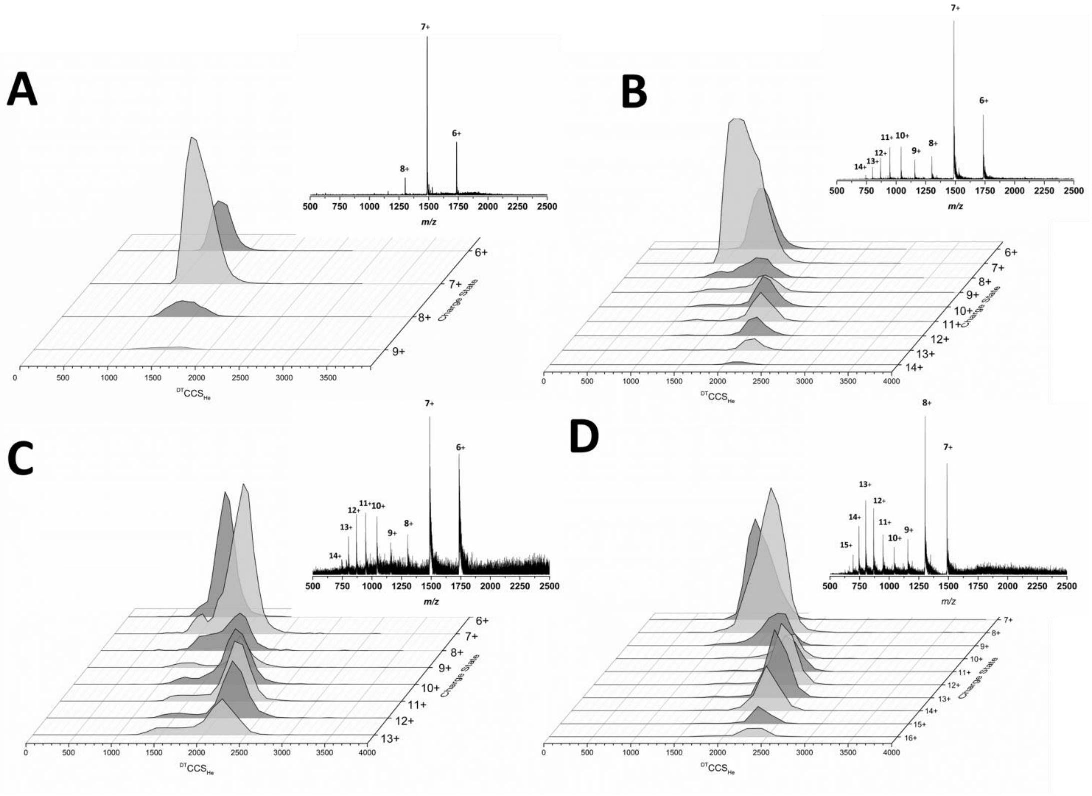
19–102

Supplementary figures

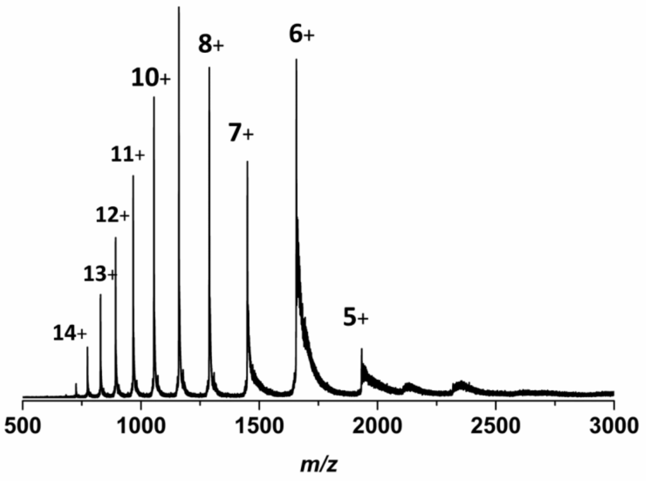
c–MYC and Max n–ESI mass spectra

c–MYC was analysed from four different solutions to ascertain the conformational flexibility. Figure S5.1A (insert) shows c–MYC in 200 mM ammonium acetate displaying a very narrow charge state distribution centred on the 7+ charge state (*m/z* 1486) with a corresponding broad but singular conformation observed through ion mobility. This same conformation is evident for the flanking higher intensity charge states 6+ and 8+. There are higher charge states visible up to 12+ however they are at very low intensity. Solution conditions were lowered to 100 mM (S1B), 20 mM (S1C), and 49:49:2 methanol:water:formic acid (S1D). Together these differing solution experiments show that c–MYC can exist across a wide range of conformations, but that under salty conditions the compact forms become more dominant.

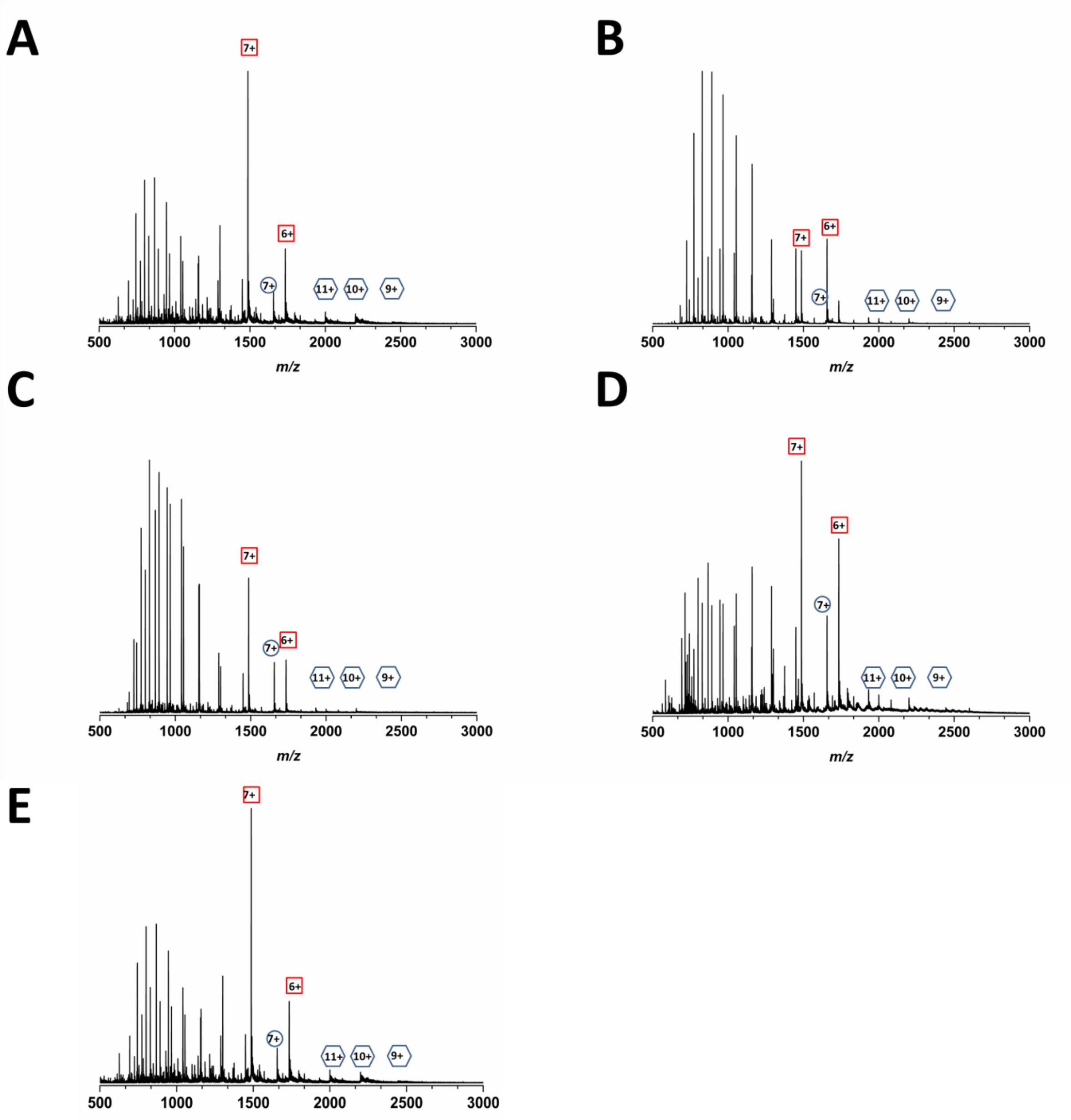
Figure S5.2 shows the spectra of MAX in 100 mM ammonium acetate. It displays a broader charge state distribution that c–MYC from 4+ through to 15+ and although there is also a bimodal distribution it is much more centred on the second envelope centred on 9+ (*m/z* 1288).



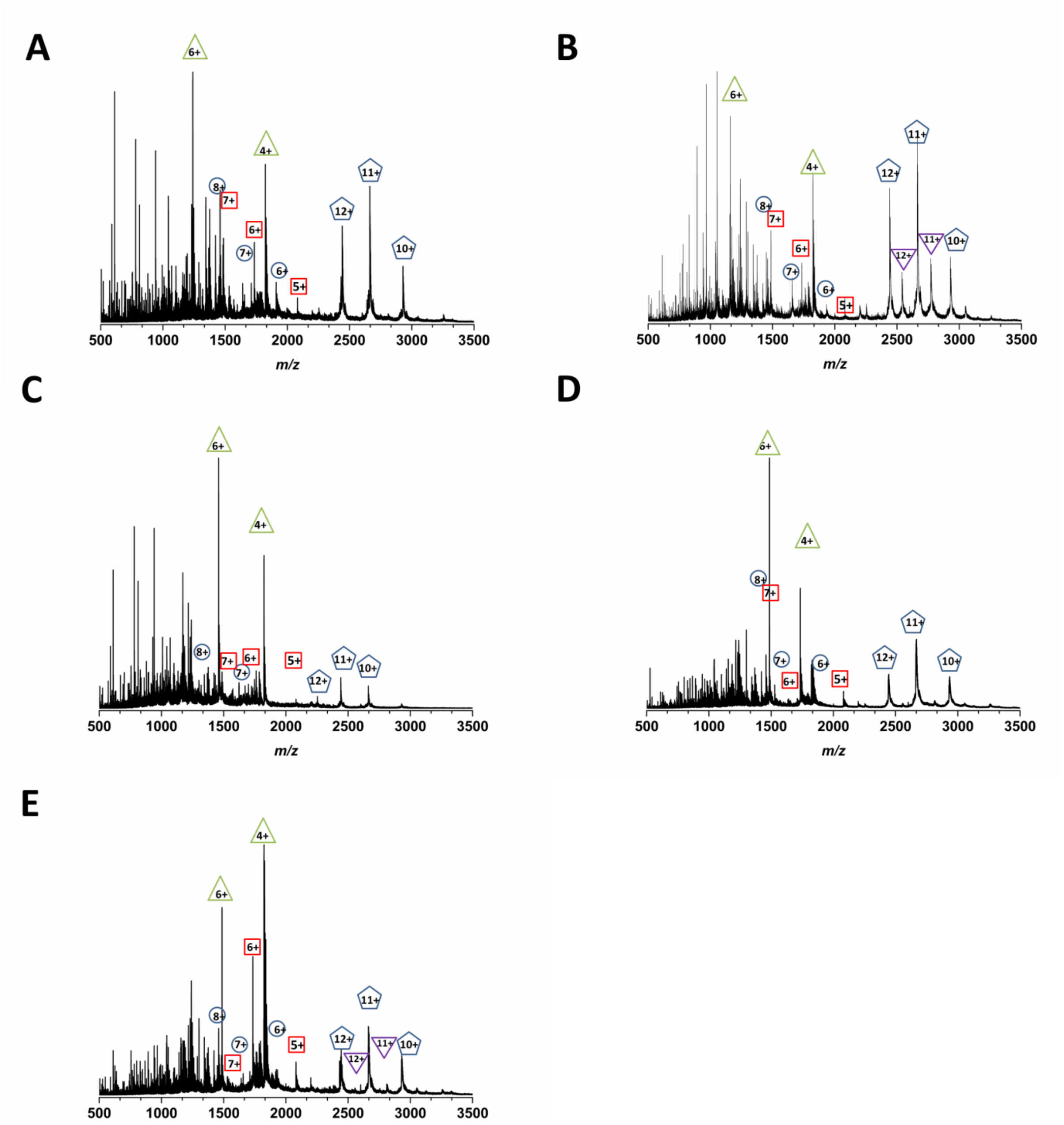
S5:1: 50 µM c–MYC collisonal cross section distributions ^{DT}CCSD$_{He}$ (Å^{2}) under four solution conditions: A) 200 mM ammonium acetate, B) 100 mM ammonium acetate, C) 20 mM ammonium acetate and D) 49:49:2 methanol:H_2O:formic acid. X axis denotes ^{DT}CCS$_{He}$ and Y axis denotes charge state.



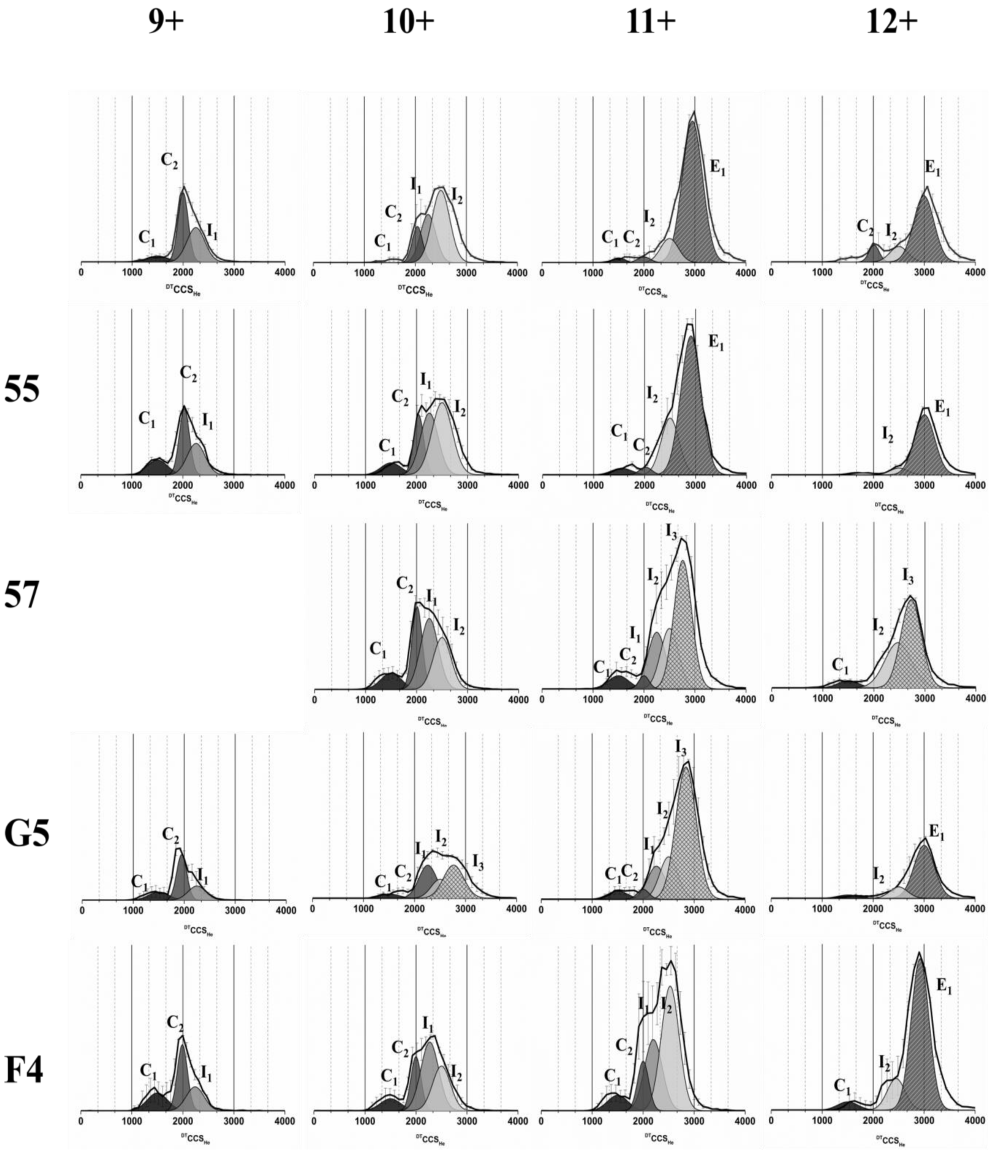
S5.2: n–ESI mass spectra of 50 µM MAX in 100 mM ammonium acetate



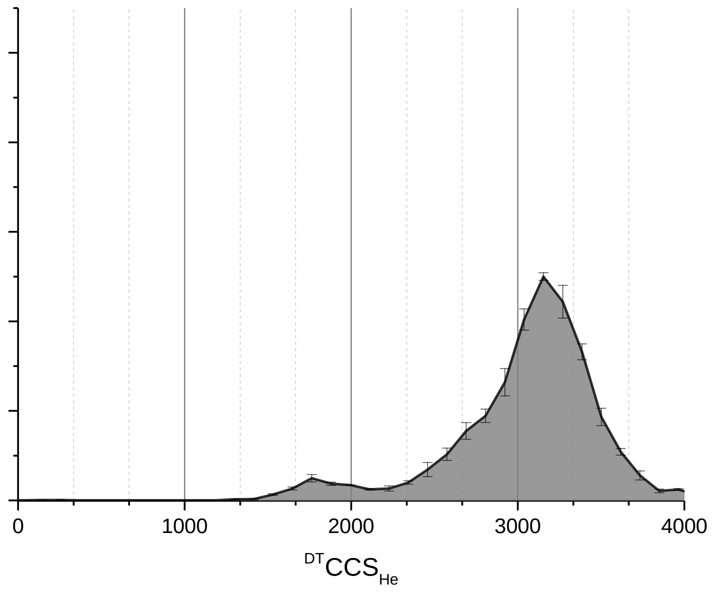
S5.3: n–ESI mass spectra of c–MYC (50 μM) and MAX (50 μM) in 100 mM ammonium acetate with: (A) no ligands, (B) with 100 μM E, (C) with 100 μM A, (D) with 100 μM B and(E) with 100 μM C. Red squares refer to c–MYC monomers, blue squares MAX monomers, and hexagons c–MYC:MAX heterodimers



S5.4: n–ESI mass spectra of c–MYC (50 μM) and MAX (50 μM) with ds DNA (100 μM) in 100 mM ammonium acetate (A) no ligands, (B) with 100 μM E, (C) with 100 μM A, (D) with 100 μM B and (E) with 100 μM C. Red squares refer to c–MYC monomers, blue squares MAX monomers, green triangles dsDNA, pentagons c–MYC:MAX:dsDNA complex, and purple inverted triangles MAX homodimers.



S5.5: Collisional cross section distributions $^{DT}CCSD_{He}$ (Å^2) for the MMD complex. Sample conditions: 50 µM c–MYC, 50 µM MAX, 100 µM ligand and 100 µM DNA from an aqueous solution with 100 mM ammonium acetate. Intensities are normalised within each ligand relative to intensity of *m/z*.



S5.6: Collisional cross section ($^{DT}CCS_{He}$, Å^2) of MMD 13+. Sample conditions: c–MYC (50 μM), MAX (50 μM), dsDNA (100 μM) in 100 mM ammonium acetate and 100 μM ligand B.