ELUCIDATING THE STRUCTURAL CONSEQUENCES OF PHOSPHORYLATION BY MASS SPECTROMETRY, ION MOBILITY AND MOLECULAR MODELLING

by

ANNA LOUISE SIMMONDS

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Physical Sciences for Health CDT
School of Chemistry
College of Engineering and Physical Sciences
University of Birmingham
September 2020

Abstract

Phosphorylation is known to affect the structure and function of proteins and peptides, affecting their intramolecular non-covalent interactions and ability to interact with other molecules. The work presented in this thesis aims to develop mass spectrometry (MS), ion mobility spectrometry (IMS) and molecular modelling (MM) methods in order to study phosphorylation in model protein and peptide systems, with a focus on the effect of phosphorylation on the non-covalent interaction pattern and structure of these model systems. An electron capture dissociation (ECD)-guided ion mobility spectrometry/molecular modelling (IMS/MM) workflow is described for the simulation of large ensembles of phosphopeptides and their fragments under mass spectrometry conditions. This workflow was facilitated by the production of forcefield parameters for several amino acid residue sets (neutral C-terminal residues, neutral N-terminal residues, neutral arginine residues, neutral phosphorylated residues and residues adjacent to the cleavage site in c^+ and $z^\bullet$ fragments). The ECD-guided IMS/MM workflow was applied to the study of the 14-3-3-ζ-derived peptide VVGARRS(pS)WRVVSSI (pS denotes phosphorylated Ser), which contains the phosphorylation motif RRSsWR. Variants were generated by successive arginine-to-leucine substitutions within the phosphorylation motif. Structures were proposed for each of the peptides investigated. In the GSK3β-derived peptide Ac-QLVRGEPNVS(pY)I(Calk)SRYYR-Am (pY denotes phosphorylated Tyr, Calk denotes cysteine-S-acetamide), substitution of specific amino acids (pTyr → pSer, Cys-alk → Ala, and Glu → Ala) allowed the ECD behaviour of the wild-type peptide to be understood, and the effect of the substituted residues to be probed. The ECD-guided IMS/MM workflow allowed insights from the ECD behaviour to be related to the structures of these peptides, and further structural features of the peptides to be probed. By studying the ECD and IMS behaviour of fragments of the GSK3β-derived peptides and the phosphoprotein β-casein, it was proposed that electron transfer dissociation (ETD) fragments might be better model systems than derived peptides to infer information about the non-covalent interactions contained in the whole peptide or protein. Applying the ECD-guided IMS/MM workflow to the z_{17}^{2+} fragment of the GSK3β-pY11 peptide showed that the fragment had a similar structure to that of the intact ion. The inclusion of travelling wave ion mobility spectrometry (TWIMS) in ETD experiments produced similar levels of fragment coverage to ECD fragmentation of β-casein, and allowed the identification of a subset of fragments in the GSK3β-pY11 spectrum that were produced by a slow dissociation process from a charge reduced species. Isolation and ECD fragmentation experiments of GSK3β-derived peptides demonstrated that the ETD- and ECD-generated z_{17}^{2+} fragment had different ECD behaviours. This observation was related to the different energetics and structures of ETD-generated and ECD-generated fragments, and set into the context of differences between the mechanisms of the two fragmentation methods.

Acknowledgments

Firstly, I'd like to acknowledge the guidance and help provided by my many supervisors: Helen Cooper, Iain Styles, David Russell, Peter Winn and John Heath, all of whom have helped to shape the work presented in this thesis.

Secondly, I'd like to thank the members of the Cooper group, Russell group and Sci-Phy CDT, and members of the wider mass spectrometry community who have provided technical and emotional support during the last 4 years. Particularly, I'd like to thank Andrea Lopez, Doyong Kim, Lisa Perez and Yuko Lam, who have assisted variously in collecting data and providing assistance with MM studies.

My profoundest gratitude goes to my network of friends and family, all of whom are wonderful, supportive people and I would not achieve a thing without you all. This includes, but is in no way limited to, The Oliver Wood Fan Club, The Simmonds Family, The Harborne Poofter Club, the Chemistry Chicas, The Office, The French and My Oldest Friends Who I Love Very Much.

A final acknowledgment must be made to Emma Sisley, with whom I have gone mad.

Table of contents

List of figures

List of tables

List of abbreviations

AI-ECD – Activated-ion electron capture dissociation

ATD – Arrival time distribution

CCS – Collision cross section

CHEF – Correlated harmonic excitation field

CI – Chemical ionisation

CID – Collision induced dissociation

CIU – Collision induced unfolding

DC – Direct current

DFT – Density functional theory

DT – Drift time

DT-IMS – Drift tube ion mobility spectrometry

FT – Fourier transform

ECD – Electron capture dissociation

EDC – Enhanced duty cycle

EHSS – Exact hard spheres scattering

ESI – Electrospray ionisation

ESP – Electrostatic potential

ETD – Electron transfer dissociation

FT-ICR – Fourier transform ion cyclotron resonance

HF – Hartree Fock

ICR – Ion cyclotron resonance

IMS – Ion mobility spectrometry

IR - Infrared

IRMPD – Infrared multiphoton dissociation

LJ – Lennard-Jones

MALDI – Matrix assisted laser desorption/ionisation

MD – Molecular dynamics

MM – Molecular mechanics

MP2 – Second-order Møller-Plesset perturbation theory

MS – Mass spectrometry

nESI – Nanoelectrospray ionisation

NMR – Nuclear magnetic resonance

PA – Projection approximation

PTM – Post-translational modification

QM – Quantum Mechanics

RF – Radio frequency

SA – Simulated annealing

SRIG – Stacked ring ion guide

SWIFT – Stored-waveform inverse Fourier transform

TIMS – Trapped ion mobility spectrometry

TM – Trajectory methods

TMLJ – A trajectory method specific to the IMoS collision cross section/mobility calculator

TOF – Time of flight

TWIMS – Travelling wave ion mobility spectrometry

Chapter 1: Introduction

The structural study of proteins allows insights into their function and interactions with other molecules.[1] Reversible phosphorylation of amino acid residues is one of the key ways that protein structure and function are modulated, so examination of the structural consequences of phosphorylation may give useful insights into biological pathways.[2] Mass spectrometry (MS), ion mobility spectrometry (IMS) and molecular modelling (MM) are tools that may be exploited to study phosphorylation.

This Chapter describes the motivation for studying protein structure and phosphorylation, and how MS, IMS and MM may be used to study the structure of proteins and peptides. A discussion of the specific MS, IMS and MM techniques employed in Chapters 3, 4 and 5 follows to indicate how these techniques may be employed in structural studies. Finally, the geometries of the MS and IMS instruments used in Chapters 3, 4 and 5 are described, as these influence experimental design.

1.1 Protein structure

Proteins are polymers of amino acid residues, which share a common backbone structure, but differ in their side chain ‘R’ group.[1] There are 20 common, naturally-occurring amino acids (Figure 1.1), each of which has a side chain with distinct properties. Broadly, amino acids may be grouped into hydrophobic, aromatic, sulfur-containing, hydroxylic, amidic, acidic and basic residues (with some overlap between these groups). The basic amino acids arginine, lysine and histidine may be positively charged under physiological conditions ($pH = 6\text{-}8$, $T = 308.15\text{-}310.15$ K), while the acidic aspartate and glutamate residues may be negatively charged. Interactions between positively- and negatively-charged amino acid residues are called salt bridges, and are one of the key interactions that drive protein folding. Interactions where only one of the acidic and basic residues is charged may be characterised as an ionic hydrogen bond. Post-translational modifications (PTM) to proteins result in the covalent binding of another chemical moiety to an amino acid residue, which changes the properties of an amino acid residue.[3] PTMs include phosphorylation, glycosylation, acetylation, nitration, and many others, with phosphorylation the most common.[4] Importantly, PTMs may be reversible, and so allow proteins to change their structure and interactions with other biomolecules. This allows protein functions to be regulated in part by their PTMs.[5]

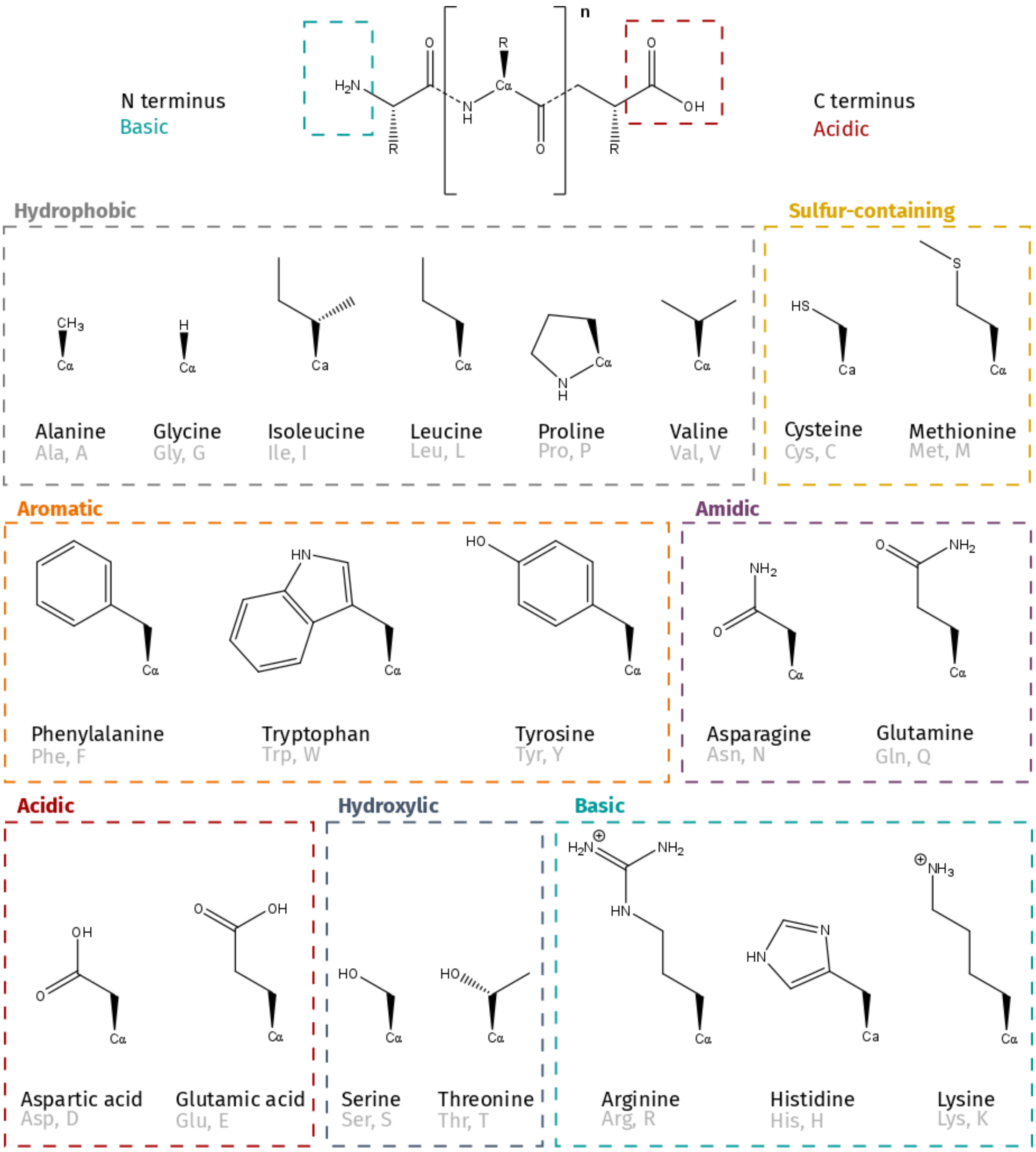


Figure 1.1 The structures of the protein/peptide backbone (top) and the twenty naturally-occurring amino acid side chains with their three- and single- letter codes, grouped by their characteristics (bottom)

Phosphorylation is the most common PTM, is found in over half of eukaryotic proteins,[6] and is instrumental in regulating protein cascades that control almost every cellular function. Typically, residues containing a hydroxyl group (Ser, Thr, Try) may be phosphorylated, with the $[PO_4H]^-$ phosphate group replacing the hydroxyl (Figure 1.2). This modification changes the nature of the side chain from hydroxylic to acidic, with phosphorylated residues having a much lower pKa than their unphosphorylated counterparts, making it much more likely that they will participate in stronger ionic non-covalent interactions.[2] Phosphorylation typically causes changes in a protein's structure, dynamics, kinetics, its interactions with other molecules, and its function.[2] The specific effects of phosphorylation are broad: in some proteins, the phosphorylation creates a motif that is recognised by other proteins that catalyse further structural or functional change (e.g. Pin1 recognises pS/pT-P motifs and catalyses cis/trans isomerisation of the proline).[7] In other proteins, the network of salt-bridges imposed by phosphorylation limits the conformational space of a protein, resulting in the conformational equilibrium being shifted towards specific conformations (e.g. Phenylalanine hydroxylase, in which phosphorylation shifts the conformational equilibrium towards active states).[8] Phosphorylation can also cause local structural effects that expose parts of the protein that bind or are recognised by other molecules (e.g. c-Src, in which phosphorylation causes a loop of the protein to be displaced, exposing the substrate binding pocket).[9]

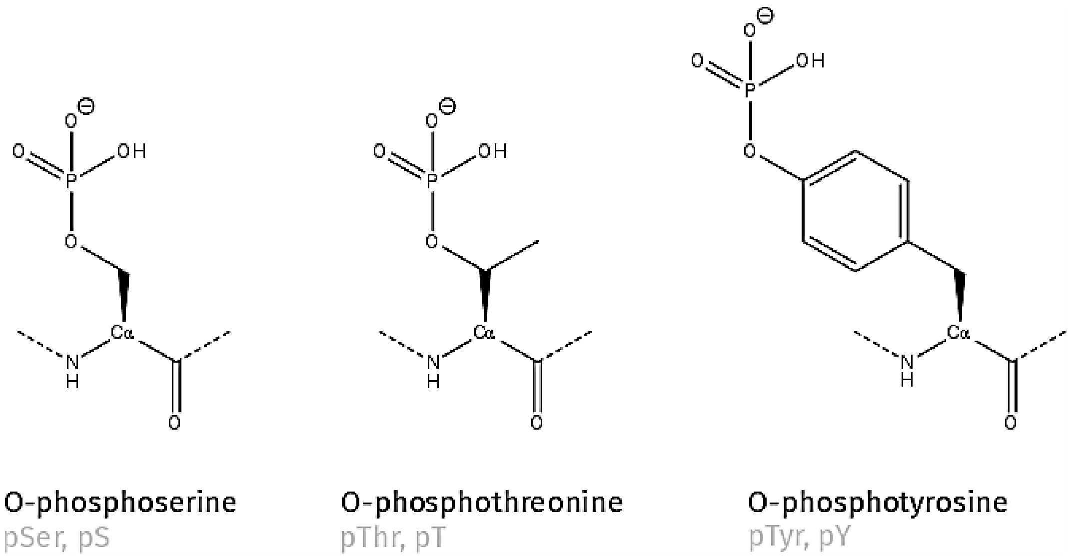


Figure 1.2 The structures and names of the three O-phosphorylated amino acid residues

X-ray crystallography, the most common method for probing solid-phase protein structure, is not well suited to the study of phosphoproteins as they are challenging to crystallise. Phosphorylation often occurs in flexible regions of proteins, the removal of which may be required before a protein can be crystallised and its structure determined by x-ray crystallography.[10, 11] Additionally, x-ray crystallography produces a single, well-resolved, model of a protein's structure, from which it is difficult to infer the conformational space that a phosphoprotein might occupy. In order to characterise protein phosphorylation, it is necessary to use other methods, a selection of which are compared in Table 1.1. Nuclear magnetic resonance (NMR) benefits from being a liquid-phase technique, meaning it can analyse proteins in conditions similar to those found in cells. 2D NMR experiments allow the through-bond and through-space structure of the protein to be analysed, and can analyse multiple conformers concurrently. However, the low sensitivity of NMR means that low-abundance conformers may not be detected, additionally, the data from large peptides, proteins, and multi-component mixtures can be very difficult to interpret.[12] Development in methods to study phosphopeptides, data processing methods, and methods with faster analysis times make NMR increasingly viable for the study of phosphopeptides and phosphoproteins despite its low sensitivity.[13, 14] Cryo-electron microscopy is a rapidly emerging biophysical technique capable of atomistic resolution of protein structure. Cryo-electron microscopy is limited by the radiation damaged that can be caused to proteins at higher electron doses over longer exposure times, and the necessity for samples to be structurally homogenous for the reconstruction of protein structure from density maps of a number of molecules to be effective. On the other hand, cryo-electron microscopy excels at determining the structure of large proteins, including phosphoproteins, that cannot be crystallised and studied by x-ray crystallography.[15, 16] Another technique, small-angle x-ray scattering, has been successfully employed to probe the structure of disordered proteins.[17] Small-angle x-ray scattering does not require proteins to be crystallised, and gives a conformational view of protein structure, but requires large amounts of sample and is usually performed using synchrotron x-ray sources, to which access is limited.[18] Mass spectrometry and ion mobility spectrometry are biophysical techniques that analyse ionised molecules in the gas phase, and are often combined in ion mobility mass spectrometers. In mass spectrometry, structural information is typically inferred from fragmentation experiments, in which the geometry of the intact ion is inferred from the identity fragments it produces.[19] Experimental workflows, such as cross-linking mass spectrometry,[20] hydrogen/deuterium exchange mass spectrometry,[21] and fast photochemical oxidation of proteins,[22] involve modification of the protein at its solvent accessible surface before its ionisation. These workflows allow solvent-accessible regions of the protein to be determined, while cross-linking mass spectrometry also probes which regions of the protein are adjacent to each other. Ion mobility spectrometry allows protein structure to be investigated by determining its ion mobility and collision cross section, which relate to a protein's mass, charge and shape. Ion mobility studies are typically coupled with molecular modelling to gain atomistic structural information, though a protein's ion mobility profile alone can be used to inform on its conformational space. Both mass spectrometry and ion mobility spectrometry may be used to study phosphoproteins, and analyse mixtures of conformers, especially as ion mobility spectrometry can separate conformers to some degree. The limitations of these methods are that the analysis of larger proteins is hindered by the complexity of the data they produce, and that collision cross section is a very coarse-grained measure of structure, requiring molecular modelling or orthogonal data to infer atomistic structural information.[19]

This thesis focusses on developing methods for studying the structural consequences of phosphorylation using mass spectrometry and ion mobility spectrometry, though it is good practice for a thorough structural study of proteins and peptides to employ a range of biophysical techniques.

Table 1.1 Comparison of several analytical techniques commonly used to study the structure of proteins[19, 23]

Method	Phase	Advantages	Disadvantages
X-ray crystallography	Solid (crystal)	Non-destructive Direct determination of 3D structure at atomic level Free electron lasers allow time-resolved x-ray diffraction studies to be performed	Crystallisation is not possible for all proteins Static structure, sample must be very structurally homogenous for structure to be resolved Not suitable for analysing mixtures
Small-angle X-ray scattering	Solid	Non-destructive Allows diffraction patterns of mixtures of conformers to be compared with those predicted by molecular modelling Does not require crystallisation	Often requires access to a synchrotron
Cryo-electron microscopy	Solid	Non-destructive (unless radiation damage occurs) Direct determination of 3D structure at atomic level Single molecule structure determination Does not require crystallisation	Static structure, sample must be very structurally homogenous for structure to be resolved High resolution cryo-EM often not possible for smaller molecules and fragile complexes
Nuclear magnetic resonance	Liquid (or solid)	Non-destructive Dynamic structure Through-space and through-bond 2D NMR methods allow connectivity and structure to be inferred Solution phase closest to the conditions found in cells	Not very sensitive Difficult to analyse mixtures Larger molecules produce complex data that can be difficult to infer structure from
Mass spectrometry	Gas	Very sensitive Tandem MS allows primary structure and complex stoichiometry to be determined, PTMs to be localised, and 3D structure to be inferred Hydrogen-deuterium exchange MS and cross-linking MS allow the solution phase structure to be inferred Dynamic structure	Destructive Larger molecules produce complex data that can be difficult to infer structure from
Ion mobility spectrometry	Gas	Very sensitive Ion mobility/collision cross section is a quantity directly related to 3D structure Coupling with molecular modelling allows atomic-level structure to be probed Dynamic structure	Destructive CCS is a very coarse-grained measure of structure and IMS relies on molecular modelling for atomic-level structure determination

Three phosphoproteins are considered in more detail in this thesis, 14-3-3-ζ, GSK3β, and β-casein, either as derived phosphopeptides or intact proteins. Each of these proteins has a number of characteristics that make them suitable for structural study by mass spectrometry, including being readily ionised, and having interesting biological features to probe. 14-3-3-ζ is studied to probe the stability of structure of the RRXpSXR motif that it contains around its pS58 residue, making it an AKT substrate.[24] This motif is usually in the dimer interface, and is only exposed when the dimer dissociates, at which point it sensitises the cell to apoptosis. The function of 14-3-3-ζ is not well defined, though it is thought to act as a signalling hub and has been identified as a possible target for cancer therapies.[25] GSK3β is a kinase that has a wide range of targets including glycogen synthase and tau protein, which leads to the formation of neurofibrillary bundles that are strongly associated with Alzheimer's disease.[26] GSK3β is also implicated in the development of insulin resistance, which leads to diabetes,[27] and a wide range of neurological and psychological conditions. Because it is implicated in so many diseases, GSK3β is often considered as a therapeutic target.[28] The phosphorylation site at the Y216 residue is located in the activation loop, and its phosphorylation leads to the activation of the protein.[29] A series of derived peptides around the Y216 residue of GSK3β are studied to observe the structural effects of this phosphorylation. β-casein is a phosphoprotein found in mammalian milk that participates in the formation of casein micelles, which are a food source of calcium phosphate.[30] Phosphorylation of β-casein promotes the formation of casein micelles and increases its ability to bind calcium cations.[31] It is often used to develop mass spectrometry techniques for phosphoproteins as it is cheap, readily available and well characterised.

1.2 Structural mass spectrometry

The analysis of a molecule by mass spectrometry requires its ionisation and transfer to the gas phase and its analysis based on its mass-to-charge ratio (m/z). The m/z alone is useful in characterising the molecule, however, additional manipulation of the molecule between ionisation and mass analysis allows more detailed information to be collected, such as its structure.[32] The native fold of proteins (i.e., the structure it adopts under physiological conditions) may be preserved to at least some extent upon transfer from the solution phase to the gas phase under certain conditions.[33] The presence of non-covalent interactions in the gas phase has been proven by studies of the peptide bradykinin using the MS technique blackbody infrared dissociation,[34] and multiple MS studies have shown the preservation of non-covalently bound protein-ligand complexes.[35-38] Phosphoproteins and phosphopeptides are routinely analysed by MS techniques,[39, 40] presenting the opportunity for the interactions between phosphorylated residues and basic amino acids to be probed. The work presented in this thesis aimed to develop methods to study the structural consequences of phosphorylation in the gas phase and apply those methods to phosphopeptides and phosphoproteins.

1.3 Ionisation

For an analyte to be manipulated and detected by mass spectrometry, it must first be ionised. A wide range of ionisation techniques exist, but the two that are commonly performed for the analysis of large biomolecules are matrix-assisted laser desorption ionisation (MALDI),[41, 42] and electrospray ionisation (ESI).[43] ESI is the preferred method for structural studies of proteins and peptides, as it is capable of preserving non-covalent interactions, allowing the native conformation of the proteins and peptides to be studied. Nevertheless, recent advances in liquid MALDI may allow its future use in probing the native conformations of peptides and proteins.[44]

1.3.1 Electrospray ionisation

In ESI, the analyte is placed in a solution containing either salts (such as ammonium acetate) or acids (such as formic acid). This solution is passed through a narrow emitter to which a potential is applied. At the end of this emitter, the solvent forms a Taylor cone, the tip of which is unstable, resulting in the formation of charged droplets (Figure 1.3a).[43] These droplets are unstable, and reduce in size via fission or solvent evaporation. Once a critical size is reached, repulsion between charged moieties, which accumulate around the droplet edge, further destabilises the droplet. At this stage the analyte may either: accumulate charge and consequently be ejected from the droplet due to the high field strengths at the edge of the droplet (ion evaporation model); accumulate charge in localised areas of the analyte, leading to unfolding of the analyte as it is ejected (chain ejection model); or else the droplet continues to undergo fission and evaporation until droplets contain a single analyte molecule, which eventually accumulates all of the charge of its droplet once the remaining solvent molecules evaporate (charged residue model) (Figure 1.3b).[45]

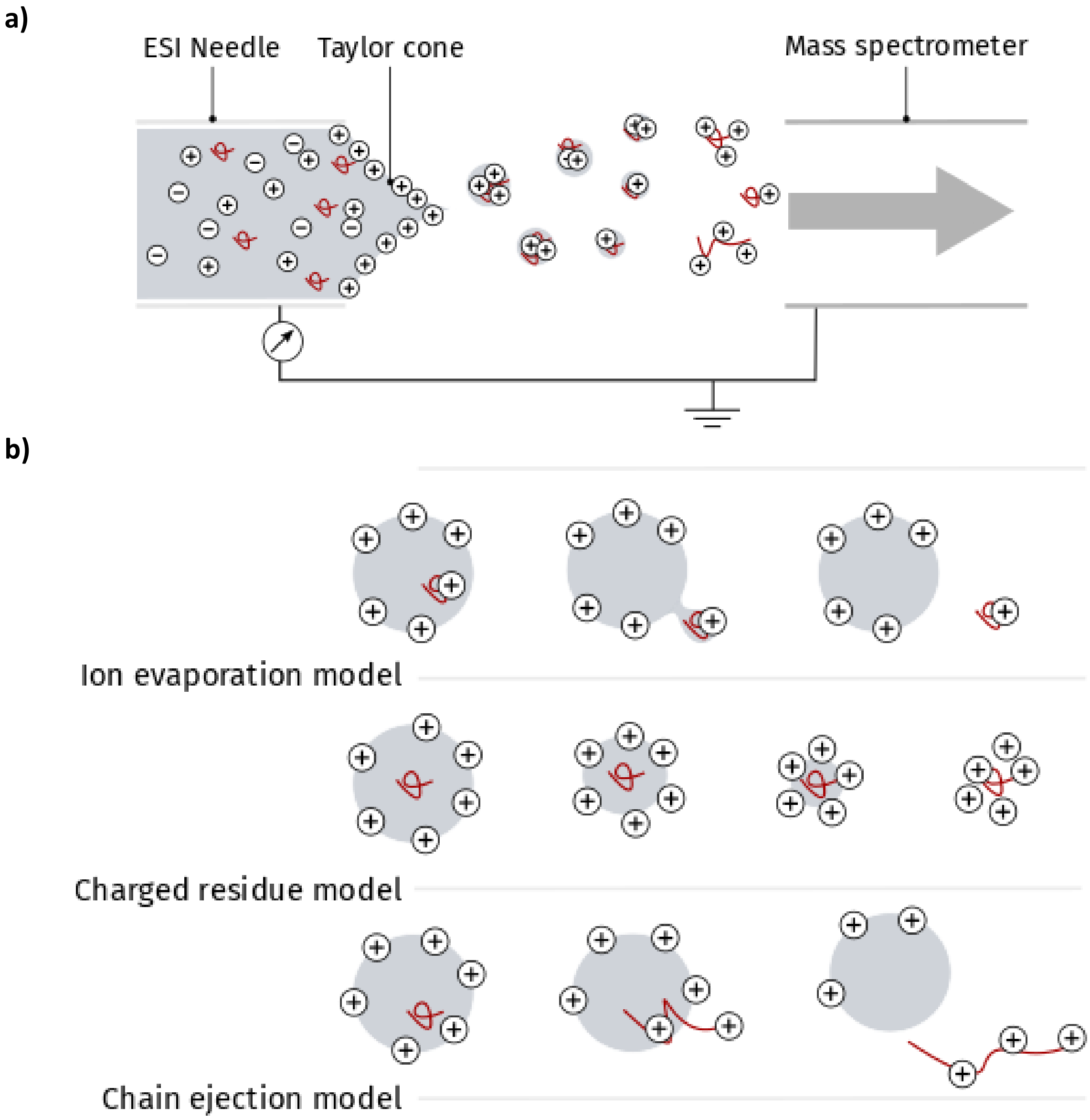


Figure 1.3 a) A schematic of the ESI apparatus operated in positive mode and b) different mechanisms for the transfer of the analyte (red lines) from the charged droplet to the gas phase

Upon entering the mass spectrometer, ions enter a region of ion optics that promote desolvation, eliminate neutral species and focus the ions into a beam.[46-49] The polarity of the potential applied to the ESI emitter determines whether the analyte ultimately forms positive or negative ions. For peptides and proteins, ESI is commonly operated in either positive or negative mode,[32, 50] although structural studies are usually performed in positive mode as it is more sensitive to changes in protein tertiary structure.[51] So, the work presented in this thesis uses positive mode ESI.

1.4 Mass analysis

With the analyte ionised and transferred to the gas phase, the analyte ions may be separated based on their mass-to-charge ratio (m/z) and detected. Mass analysers separate ions by exploiting their m/z-dependent behaviour in electric and/or magnetic fields, dispersing them in either the time or frequency domains. Once separated, a detector records the signal intensity over time, which may then be processed to produce a mass spectrum: a graph of ion abundance against m/z.[32]

1.4.1 Time of flight

In a time of flight (TOF) mass analyser, ions pass through an electric field, which causes them to accelerate to a velocity related to their m/z.[32, 52] The ions pass into a flight tube (to which no field is applied), allowing the different velocities of the ions to cause lower m/z species to arrive at the detector before higher m/z species (Figure 1.4). The m/z-dependent behaviour of an ion in the flight tube can be described by Equation 1:

$$m/z = \left(\frac{2eU}{d^2}\right)t^2 \tag{Eq 1}$$

where m is the ion's mass, z is its integer charge, U is the potential used to accelerate the ions, d is the length of the flight path of the ions, and t is time.

Therefore, with TOF analysers operated at a known acceleration potential and flight path length, m/z is proportional to the square of their flight time. Practically, the acceleration imparted on a given ion and the path it takes through the analyser may be different from an isobaric ion (an ion with the same m/z) due to a difference in their initial kinetic energies and positions in the acceleration region. This effect causes peak broadening in the mass spectrum, given that a single TOF m/z peak is typically composed of the signal from 10^3 – 10^7 ions, and each ion may be recorded with a flight time that deviates slightly from the theoretical value given by (Eq 1).[52] Various methods to homogenise the acceleration and path length of isobaric ions exist, including: delayed extraction,[53] use of a reflectron,[54] and orthogonal acceleration,[55] all of which reduce peak broadening and increase resolution between peaks.[52]

As the flight time need only be detected at a single point, detectors may be used that record the signal of each ion by its collision (and elimination) with an electron multiplier.[56] The collision causes a burst of secondary electrons, the current from which can be amplified and detected. Modern detectors are usually microchannel plates, in which each channel is an electron multiplier, allowing signal from multiple ions to be detected simultaneously.[57]

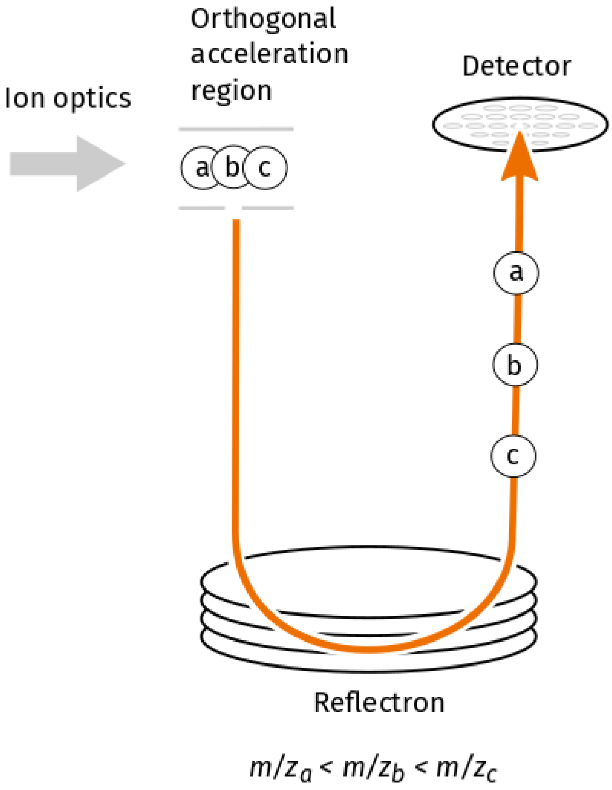


Figure 1.4 The ion path through a TOF with orthogonal acceleration and a reflectron. The lower m/z species 'a' reaches the detector before the higher m/z species 'b' and 'c'. The orthogonal acceleration and reflectron reduce peak broadening effects caused by the different initial kinetic energy of ions upon exiting the preceding ion optics

1.4.2 Fourier transform ion cyclotron resonance

In Fourier transform ion cyclotron resonance (FT-ICR) mass analysers,[58, 59] ions are gated into an ion cyclotron resonance (ICR) cell, in which a strong magnetic field induces cyclotron motion in the ions. The frequency of cyclotron motion of an ion is related to its m/z by:

$$m/z = \frac{eB}{\omega_c} \qquad \text{(Eq 2)}$$

where e is the elementary charge constant, B is the strength of the magnetic field and ω_c is the frequency of cyclotron motion.

In an ICR cell where the magnetic field is static and ions can be excited into a stable cyclotron motion, the m/z of an ion is therefore inversely proportional to the frequency of its cyclotron motion.[58-60]

To satisfy these criteria, ICR cells are operated inside of a superconducting magnet capable of maintaining a high magnetic field (between ~2 T and 21 T)[58, 61, 62] and employ excitation and detection plates (Figure 1.5).[58-60] The excitation plates deliver a radio frequency (RF) pulse that excites ions into a resonant, coherent cyclotron orbit. Following excitation, each time an ion passes the detection plate it induces an image current, which may be recorded over time. A Fourier transform can convert the time-domain trace into a frequency-domain spectrum, where the frequency of each ion can be related to its m/z[63]. More ions of the same m/z will induce a higher image current in the detector plates, so the magnitude of their signal in the time domain will be higher. The relative abundance of time-domain signals is preserved through the Fourier transform and means more abundant ions ultimately have higher signals in the mass spectrum.[58, 59]

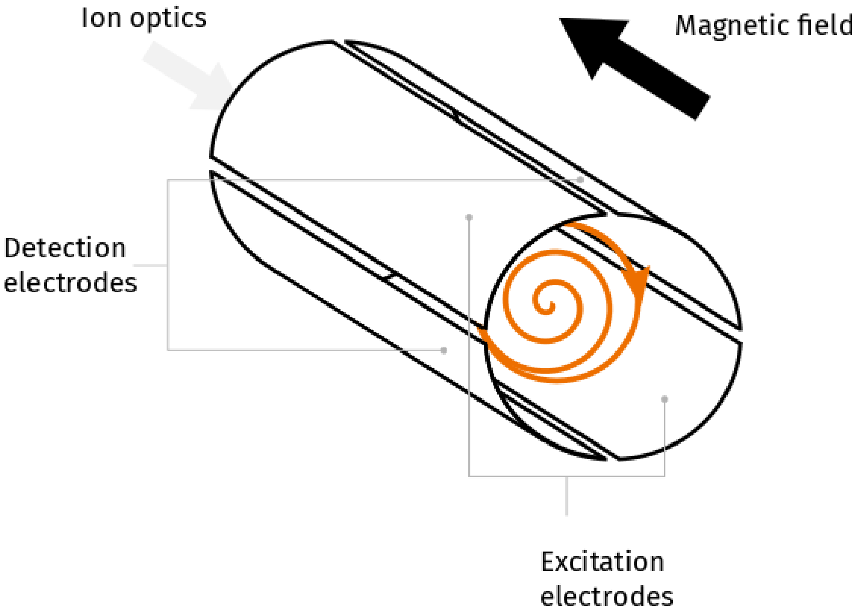


Figure 1.5 An ICR cell mass analyser. Ions are excited into a cyclotron orbit (orange) in the presence of a strong magnetic field. The frequency of cyclotron orbit depends on the m/z of an ion, which is recorded by the detection electrodes. Note: trapping plates on either end of the ICR cell are omitted to allow the ion excitation to be seen

In practice, the cyclotron motion does not confine ions along the axis of the magnetic field, so trapping plates on either end of the ICR cell are used to apply an electric field that prevents the ions spiralling out of the cell. With this trapping, the motion of the ions in the axis of the magnetic field can be described by a simple harmonic oscillator. The combination of axial and cyclotron motion results in a third type of motion: magnetron motion. The combination of cyclotron, axial and magnetron motions is shown in Figure 1.6 and described by Equations(Eq 3 3-6:

$$\omega_c' = \frac{\omega_c}{2} + \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \frac{\omega_{ax}^2}{2}} \quad \textit{(Eq 3)}$$

$$\omega_m = \frac{\omega_c}{2} - \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \frac{\omega_{ax}^2}{2}} \quad \textit{(Eq 4)}$$

where ω_c' is the reduced cyclotron frequency, ω_m is the magnetron frequency, ω_c is the cyclotron frequency (Eq 5) and ω_{ax} is the frequency of axial oscillation (Eq 6):

$$\omega_c = \frac{zeB}{m} \quad \textit{(Eq 5)}$$

$$\omega_{ax} = \sqrt{\frac{2zeV_{trap}\alpha}{ma^2}} \quad \textit{(Eq 6)}$$

in which B is the strength of the magnetic field, m is the mass of the ion, z is the integer charge of the ion, V_{trap} is the voltage applied to the trapping plates, while a and α are constants related to the size and geometry of the ICR cell.[58, 59]

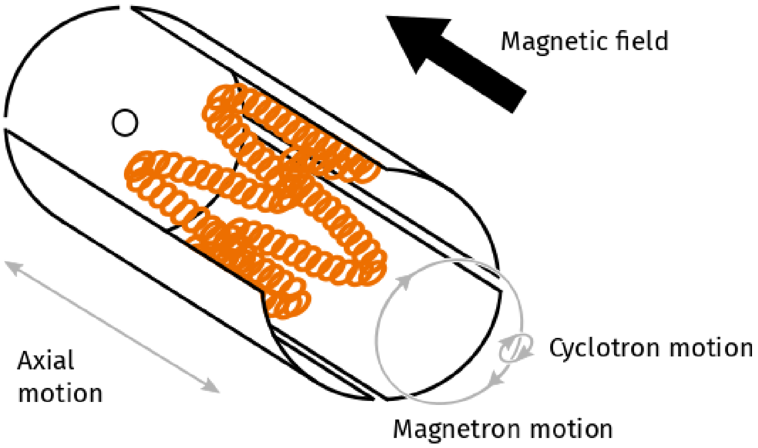


Figure 1.6 Ion motion in an FT-ICR cell, with the component cyclotron, magnetron and axial motions labelled in grey. Note: one excitation/detection plate and the front trapping plate are omitted to allow the ion motion to be seen.

The combination of reduced cyclotron, magnetron and axial motions means that the m/z of a given ion is not analytically soluble from its frequency, however, a calibration against ions with known m/z under the same experimental conditions allows the frequency-domain spectrum to be converted to an m/z-domain spectrum. [58, 59] Many calibration equations exist with different terms accounting for different phenomena that might affect the frequency of an ion, but for all, the use of more calibrant ions results in a more accurate calibration.[64]

Equations 3-6 demonstrate that the ion motion in an ICR cell is independent of the initial kinetic energy of the ions, unlike in a TOF mass analyser, where the distribution of ion kinetic energies is one of the main causes of peak broadening. This makes it possible to access higher mass spectral resolutions via FT-ICR MS than in TOF MS as peak widths are narrower. Major causes of peak broadening in FT-ICR MS include ion-neutral interactions with any background gas, which may damp cyclotron and magnetron motions; space-charge effects, in which Coulombic repulsion between ions pushes them apart and alters their trajectory; and dephasing of the ions, often caused by the instability of the axial oscillation. To overcome ion-neutral interactions, ICR cells are operated at ultra-high vacuums (~10^{-10} mbar) to eliminate any background gas, while caution must be taken not to overfill the ICR cell to limit space-charge effects.[58, 59] The instability of the axial oscillation can be mitigated by ICR cell designs that improve trapping and excitation to reduce dephasing and ejection of ions along the z axis.[65, 66]

Because ions must be accumulated, gated into the ICR cell, and detected for longer periods of time to increase resolution and signal-to-noise ratios,[58] FT-ICR MS is generally slower than TOF MS, with duty cycles (the length of time between the ionisation and detection of an ion) on the second scale, as opposed to the millisecond scale of TOF MS.[67] Additionally, the image current from an ion is lower than the signal produced by an electron multiplier, which means a greater number of ions is needed to achieve similar signal-to-noise ratios in FT-ICR MS than TOF MS. Typical FT-ICR MS peaks are produced from $10^6 – 10^{10}$ ions. These factors mean that a FT-ICR experiment would take longer than an equivalent TOF experiment, and that time-dispersive separation techniques (such as IMS) are incompatible with FT-ICR, if the time scale of the separation is shorter than the FT-ICR mass analysis time.[68]

1.4.2.1 Commercialised ICR cells

There are two ICR cell technologies that have been widely commercialised, the ‘Infinity cell’ and the ‘Paracell’ (Figure 1.7). The Infinity cell concept aims to replicate an infinitely long ICR cell by changing the potential applied by the trapping plates upon excitation. This concept prevents unwanted ejection of ions along the z-axis of the cell during excitation and increases the time that ions can be held in the ICR cell.[66] Meanwhile, the Paracell has parabolic excitation and detection electrodes, which produce an excitation field that is tailored to the stability of the ion cloud in the z-axis: the greatest excitation is applied in the centre of the ICR cell, where ions are most stable in the z-axis. This allows the axial motion to be separated from cyclotron and magnetron motions, which benefits mass spectral resolution.[65]

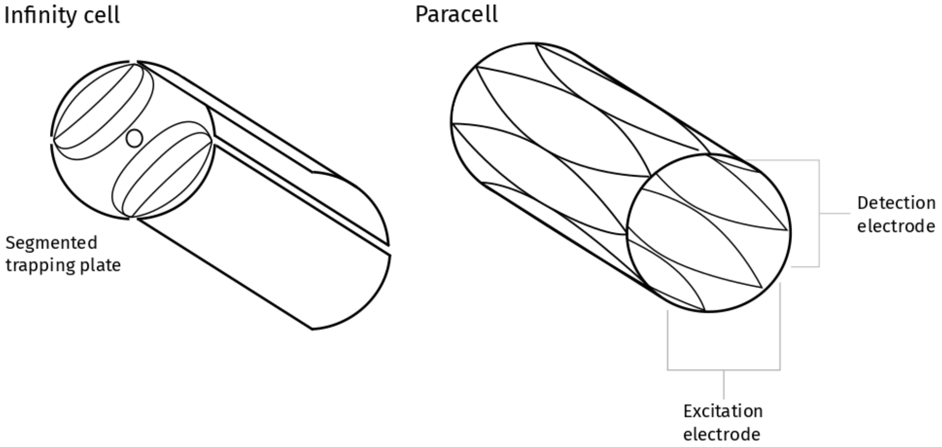


Figure 1.7 Design features of the Infinity cell and Paracell. The segmented trapping plate of the Infinity cell allows the trapping field to be varied upon the application of the excitation field. The parabolic electrodes of the Paracell allow an optimised excitation field to be applied. Note: two excitation/detection plates are omitted from the infinity cell, and the front trapping plate is omitted from both the infinity cell and Paracell to allow specific design features to be seen

1.5 Isolation

Between ionisation and detection, ions may be manipulated to obtain information about their identities or gas-phase behaviour. One of the ways ions are manipulated is by isolation of an ion of interest, achieved by eliminating all ions outside of a m/z window that contains the ion of interest. Isolation allows downstream manipulation and detection of the ion of interest without other species complicating the results.[32]

1.5.1 Quadrupole isolation

Quadrupoles are often used for isolation; they contain two sets of parallel rod electrodes arranged at 90° from each other (Figure 1.8). A potential composed of a direct current (DC) component and a time-dependent radio frequency (RF) component is applied to each electrode. The switching of the RF potential causes ions to transit the quadrupole with an oscillating motion in the planes of both sets of electrodes. The magnitude of the oscillation in each plane depends on the m/z of the ion, and those ions with large oscillations are eliminated by contact with an electrode. In this way, the potential applied to the electrodes can be tuned to permit only ions within a certain m/z window to pass through the quadrupole device, allowing their isolation. By operating the quadrupole in a scanning mode, coupled to a detector, the intensity of ions over a range of m/z values may be recorded, allowing the quadrupole to be used as a mass analyser.[32]

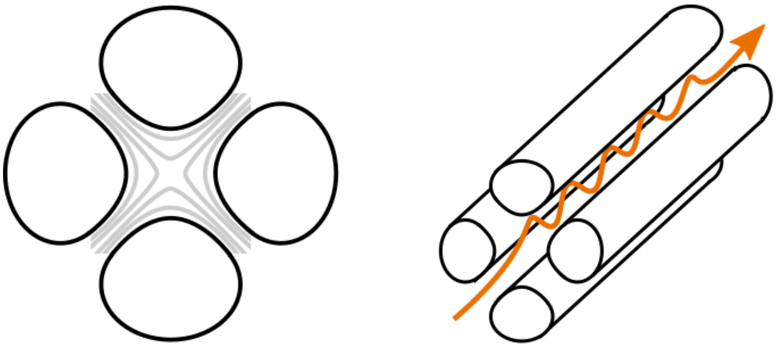


Figure 1.8 The field generated in a quadrupole with hyperbolic electrodes, and an illustration of a stable ion trajectory through the quadrupole

When operated without the DC voltage, no ion elimination occurs, so quadrupoles may be used to refocus the ion beam or trap ions. Hexapoles and octupoles are also frequently used in this way as RF-only multipoles.[48, 69]

1.5.2 In-cell isolation

Isolation may also be performed in the ICR cell of an FT-ICR mass spectrometer. This is achieved by a notched excitation, where the ion of interest is contained in the notch and all other ions are excited to an orbit that ejects them from the ICR cell; or single-pulse excitation shots that remove individual unwanted ions (Figure 1.9).[70] A notched stored-waveform inverse Fourier transform (SWIFT) excitation[70] or frequency sweep (usually implemented as the correlated harmonic excitation field (CHEF) procedure)[71] may be used for windowed excitation. In SWIFT excitation, the frequency-domain spectrum is recorded, and a desired frequency-domain excitation pattern created that preserves the ions of interest; an inverse Fourier transform of the desired frequency-domain excitation pattern produces a time-domain excitation waveform that can be applied by the excitation electrodes. In CHEF the period and duration of the excitation waveform are correlated to the cyclotron frequency of the ion of interest to reduce unwanted energy deposition in the ion of interest. When multiple notches are used for the simultaneous isolation of several ions via CHEF, this is called multi-CHEF. Ideal frequency-domain waveforms for notched and shots isolation are shown in Figure 1.9.

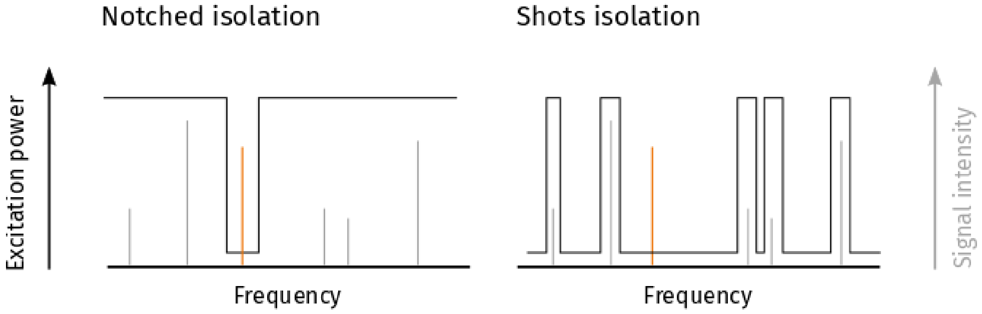


Figure 1.9 Frequency-domain excitation patterns for notch-type isolation and shots type isolation, overlaid over frequency-domain spectra where grey ions are unwanted and the orange ion is the ion of interest

1.6 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) is one of the key ways that mass spectrometry contributes to molecular biology. In a tandem mass spectrometry experiment, an ion of interest is isolated, fragmented, and the m/z of its fragments is analysed.[72] By fragmenting an analyte and measuring the m/z ratios of its fragments, the identity of the fragments informs on the identity and structure of the intact analyte. In the context of tandem MS, the intact analyte is referred to as the 'parent' or 'precursor' ion. For proteins and peptides, fragmenting into shorter amino acid chains allows the primary structure (the sequence of amino acid residues specific to the protein or peptide) and identity of the protein or peptide to be discerned. Protein and peptide fragments can be classed based on which bond between amino acid residues is cleaved, and whether the fragment contains the N- or C-terminus: a, b and c fragments contain the N-terminus, while x, y and z fragments contain the C-terminus (Figure 1.10).[73] Tandem mass spectrometry may also allow a range of PTMs on amino acid residues to be localised, provided those PTMs are preserved through the fragmentation process.[39]

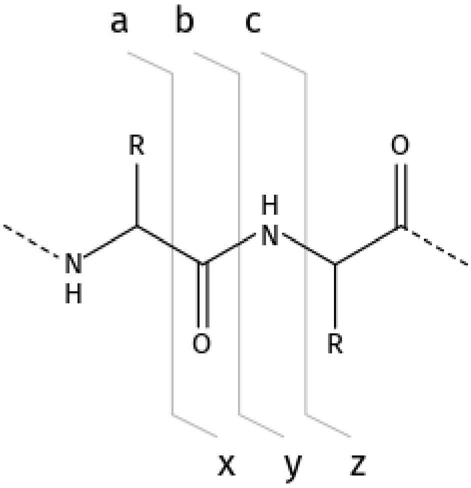


Figure 1.10 Nomenclature for fragments arising from the cleavage of peptides and proteins at the Cα-C, C-N and N-Cα bonds

1.6.1 Collision induced dissociation

The most common fragmentation method is collision induced dissociation (CID),[74] in which an inert collision gas is introduced into a cell of the mass spectrometer, and fragmentation of the analyte is achieved via collision with that gas.[75] The analyte is accelerated to a narrow kinetic energy profile, which is converted to internal energy upon inelastic collision with the collision gas. Bond cleavage occurs when this internal energy is distributed across vibrational modes in the analyte, with the weakest bonds breaking first. In proteins and peptides, this corresponds to cleavage at the amide C-N bond, forming b and y fragments. Many PTMs (including phosphorylation) are cleaved from the peptide or protein during CID, making it unsuitable for their study.[76] Tuning of the kinetic energy profile allows the energy of CID to be tuned.

1.6.2 Electron capture dissociation

In electron capture dissociation (ECD),[77] multiply-charged analyte ions are irradiated by a beam of low energy electrons (< 1 eV). Capture of these electrons by a group with high electron affinity (e.g. sulfur-containing groups or positively-charged groups) initiates a cascade of radical gas-phase chemistry that results in fragmentation of the analyte. In peptides and proteins, it is generally accepted that the electron is captured into the Rydberg orbital of a positively-charged group (such as $[Arg+H]^+$, $[Lys+H]^+$, $[His+H]^+$, or protonated backbone sites),[78] the S-S σ* orbital of disulfide bonds between cysteine amino acid residues,[79] or directly into the amide C=O π* orbital.[80] The captured electron cascades through a number of electronic states of the molecule until it populates the amide C=O π* orbital, where it causes cleavage of the N-Cα bond of the peptide backbone. This mechanism is called the Utah-Washington mechanism, and is depicted in Figure 1.11. Proton transfer occurring either before or after cleavage of the N-Cα bond results in c^+ and $z^\bullet$ fragments, which may then rearrange to $c^\bullet$ and z' fragments by abstraction of a proton from the c fragment to the z fragment.[81] Between capture and cleavage, the radical may migrate up to 15 Å through-bond, causing fragmentation up to 5 bonds away from the site of capture, or 5 Å through-space.[82, 83] Other studies have shown that some ECD products have sufficiently long lifetimes for the radical to migrate further and cause more distal fragmentation.[84, 85] Fragmentation at sites other than the peptide backbone are also frequently observed, usually causing the cleavage of amino acid side chains from the peptide backbone.[86]

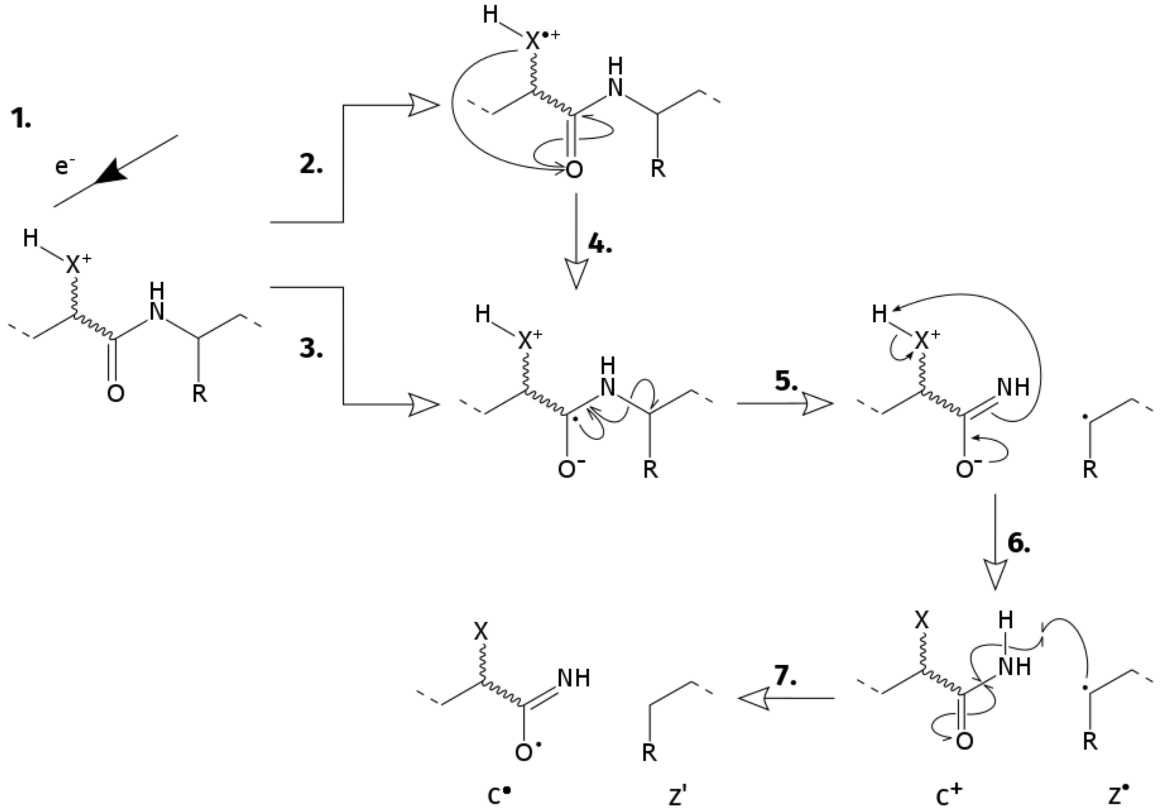


Figure 1.11 The Utah-Washington mechanism: electron capture (1) into either a Rydberg orbital (2) or an amide π orbital (3). The species formed by electron capture at a Rydberg orbital converts to the species formed by capture at the amide π* orbital (4). From the amide π* orbital, the electron cleaves the N-Cα bond forming two fragments (5). Proton transfer (which may occur either before or after cleavage of the N-Cα) forms c^+ and $z^\bullet$ fragments (6), which may then further rearrange to $c^\bullet$ and z' fragments (7) by proton abstraction. Note: 'X' denotes any protonated group and may be several residues away from the cleavage site, and be either N-terminal or C-terminal to the cleavage site*

Many PTMs, such as phosphorylation, are preserved throughout the ECD process, which enables PTMs to be localised in proteins and peptides.[87-91] Non-covalent interactions may also remain intact throughout ECD, and when backbone fragmentation occurs between two non-covalent interaction partners, the non-covalent interaction may hinder the dissociation of the two fragments and prevent them being detected separately.[92, 93] ECD may be used to map regions of proteins with more non-covalent interactions, which tend to be less flexible and have higher order structures, as these will have poorer ECD fragment coverage. ECD studies of the infrared (IR)-induced unfolding of ubiquitin ions have shown the differences between their unfolding pathways in the gas and solution phases,[94] with later work focussing on how the gas-phase structure of ubiquitin ions is affected by its desolvation.[95] ECD has also been used to show that the presence of non-covalent interactions increase the longevity of the native, solution-phase structure of the protein KIX in the gas phase.[96] The combination of ECD and IMS data has been used to study the extent and location of collision induced unfolding in human haemoglobin,[97] and also the unfolding landscape of the protein Ltn.[98] A more recent study of several denatured analytes (Substance P, melittin and ubiquitin) and native analytes (alcohol dehydrogenase and hemoglobin) used ECD and IMS to show that unfolded proteins had higher fragment coverage than folded or native proteins.[99]

Several studies have applied ECD to study the effects of non-covalent interactions in phosphopeptides. For a series of synthetic peptides based on the sequence APLSFRGSLPKSYVK with different phosphorylation sites and different numbers of basic residues, reduced ECD fragment coverage was observed, particularly in peptides containing two or more basic residues.[100] The reduced sequence coverage was attributed to salt-bridge interactions between the phosphate and the basic residues. However, a separate study showed that the presence of phosphorylated serine and basic amino acids alone (in peptides that otherwise only contained alanine and proline residues) was insufficient to inhibit ECD fragmentation, and that the conformation of the peptide (and therefore the relative positions of the interacting residues) was important in determining ECD behaviour.[101]

1.6.2.1 Activated-ion ECD

In order to overcome non-covalent interactions that may prevent the dissociation of a c-z fragment pair, ions may be activated before or after ECD is performed.[92] Activation may occur by introduction of a collision gas or irradiation with a laser (commonly an IR laser) with energies sufficient to break non-covalent interactions but below fragmentation levels. Activation is commonly used to increase fragment coverage to aid in protein identification,[91, 92] but may also be used in mechanistic studies of ECD.[102] Activated-ion ECD (AI-ECD) experiments are presented in Section 4.3, where a more detailed discussion of the rationale and implementation of AI-ECD may be found.

1.6.3 Electron transfer dissociation

Electron transfer dissociation (ETD) uses a stable radical anion reagent to transfer an electron to the analyte.[103] ETD results in the same N-Cα bond cleavage as ECD, and is generally assumed to proceed via a mechanism analogous to the Utah-Washington scheme, however, it has been noted that the lower energies imparted by ETD may result in electron attachments to lower energy orbitals and population of different states between electron attachment and bond cleavage.[104-106] Common electron transfer reagents include fluoranthene and 1,3-dicyanobenzene, both of which are shown in Figure 1.12.

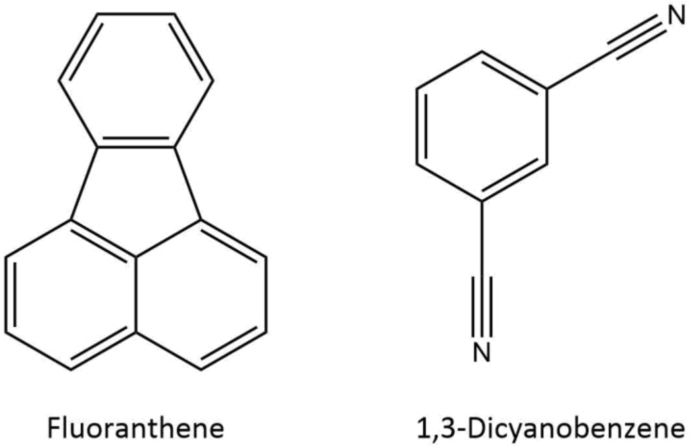


Figure 1.12 Example ETD reagents fluoranthene and 1,3-dicyanobenzene

As with ECD, PTMs and non-covalent interactions are retained throughout the ETD process. By observing regions of the peptide sequence with no fragment coverage, ETD has been used to propose intramolecular salt-bridge locations in several proteins, which were found to closely resemble the salt bridges of the native protein structures.[107, 108] It has been shown that both ECD and ETD can be used to determine the site of intermolecular interactions between acidic and basic peptides.[35] A combination of ETD and IMS enabled a study into the unfolding of several proteins.[109]

1.7 Ion mobility spectrometry

IMS is a technique in which ions are separated based on their mobility upon interaction with an inert gas and an electric field.[110] The mobility of an ion is closely linked to its size, shape, mass and charge, with larger ions of a lower charge tending to experience more collisions with the inert gas and being less affected by any electric field present. The behaviour of an ion in these conditions can be described by the Mason-Schamp equation (Eq 7):[111]

$$K = \frac{3}{16}\sqrt{\frac{2\pi}{\mu k_B T}} \times \frac{ze}{n\Omega} \tag{Eq 7}$$

where K is the ion mobility of the analyte, µ is the reduced mass of the analyte ion and the inert gas ($\mu = \frac{m_{ion}m_{gas}}{m_{ion}+m_{gas}}$), k_B is the Boltzmann constant, T is the temperature of the inert gas, z is the integer charge of the analyte ion, e is the elementary charge constant, n is the number density of the gas and Ω is the collision cross section of the analyte.

Given that ion mobility spectrometry requires the analyte to be ionised and in the gas phase, as with mass spectrometry, the two techniques are often coupled, with ion mobility cells placed between the source and mass analyser of the mass spectrometer. This allows the mobility of an ion and its m/z to be analysed in a single experiment, recording two semi-orthogonal data sets on an analyte. Additionally, if the m/z of an ion is known, the Mason-Schamp equation may be used to calculate its collision cross section (CCS), a quantity related to the size and shape of the ion that may be calculated for molecular models, providing a bridge between experimental and theoretical studies of a given analyte.[112]

There are many IMS techniques for the measurement of the mobility of an ion;[110, 113, 114] those that are commonly used in the study of peptides and proteins and allow the derivation of CCS are drift tube IMS (DT-IMS), travelling wave IMS (TWIMS), and trapped IMS (TIMS) (not discussed here).

1.7.1 Drift tube IMS

DT-IMS was the first IMS method to be implemented,[113] and remains the only primary method for calculating the mobility and CCS of an analyte.[114] All other IMS devices require calibration against a set of ions with known CCS. Drift tube IMS devices contain a cell to which a potential is applied to create a homogenous electric field. The cell is filled with an inert buffer gas, with which the ions interact (Figure 1.13).

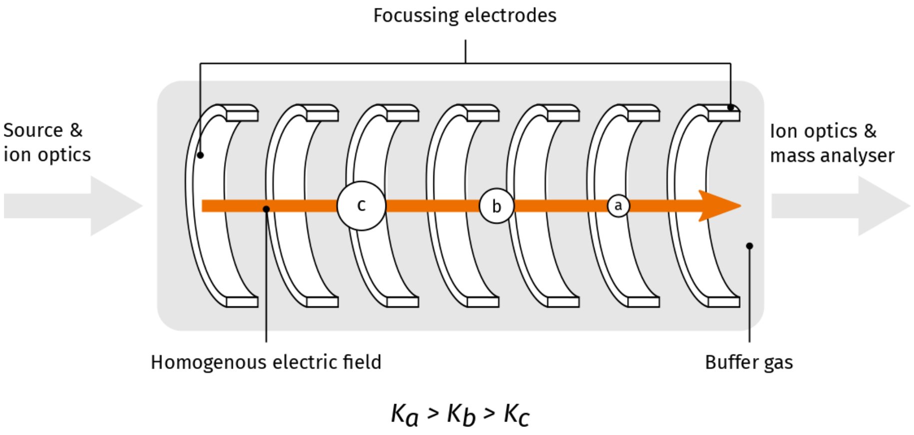


Figure 1.13 Mobility-dependent separation of ions in a drift tube IMS device. The ion 'a' has a higher mobility (K) than either ions 'b' or 'c' and will exit the drift tube first. Each ion travels at a constant velocity related to its mobility

The length of time an ion takes to traverse the drift tube under the competing influences of the electric field and the buffer gas is related to its mobility by Eq 8:

$$K = \frac{l^2}{t_D U} \qquad \text{(Eq 8)}$$

where l is the length of the drift tube, t_D is the drift time and U is the potential applied across the drift tube.

Given that drift tubes are usually followed by some ion optics and a mass analyser, drift time is not measured directly, but inferred from the arrival time of the ion at the detector. The arrival time is composed of the drift time, the transit time through any ion optics, and the mass analysis time (which may itself be dependent on the m/z of the ion if a TOF is used). To calculate the mobility of an ion, its arrival time must be measured at several drift tube potentials, with all other instrument settings remaining unchanged. This varies the drift time of the ion, such that it may be separated from the arrival time, and the mobility calculated. Otherwise, the arrival time of the analyte may be measured at a single drift tube potential and calibrated against the arrival times of a set of ions of known mobility.[110, 114]

1.7.2 Travelling wave IMS

In TWIMS, the ion mobility cell is composed of a stacked ring ion guide (SRIG) filled with an inert buffer gas.[115] The SRIG allows a travelling wave to be propagated along the length of the IMS cell, which propels the ions through the buffer gas. Interactions with the buffer gas cause ions with lower mobilities to fall behind the travelling wave, causing them to decelerate until they are reaccelerated by the next iteration of the travelling wave (Figure 1.14).

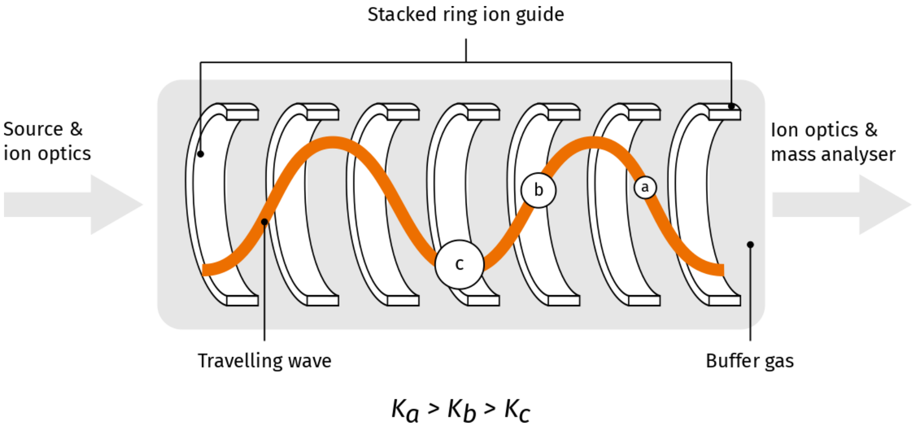


Figure 1.14 Mobility-dependent separation of ions in a TWIMS device. The stacked ring ion guide allows propagation of a travelling wave along the length of the cell. The ion 'a' has a higher mobility (K) than either ions 'b' or 'c' and will exit TWIMS cell first. This illustration shows the ions 'b' and 'c' having 'rolled over' the travelling wave: 'b' is in the process of decelerating, while 'c' is about to be re-accelerated by the next iteration of the travelling wave

The mobility of an ion affects its motion through the TWIMS cell, as described by:[116, 117]

$$\alpha \frac{d^2 z}{d\tau^2} = \frac{dz}{d\tau} - 1 + \gamma \sin z \tag{Eq 9}$$

for an ion travelling in the z axis over the time period τ, with:

$$\alpha = 2\pi \frac{\upsilon}{\lambda} \frac{Km}{ze} \tag{Eq 10}$$

$$\gamma = 2\pi \frac{V_0}{\upsilon \lambda} K \tag{Eq 11}$$

K is the mobility of the ion, υ is the wave velocity, λ is the wavelength, V_0 the wave amplitude, m the mass of the ion, z its integer charge and e the elementary charge constant.

While it is theoretically possible to calculate mobility directly from TWIMS experiments using Equations 9-11, no model has been created to date that facilitates this calculation. In practice, mobility and CCS are derived from TWIMS data by a calibration against a set of calibrant ions whose mobility has been previously determined by a drift tube experiment. Because the specific influences of mass, charge, shape, size and ion class on ion mobility in TWIMS are not well understood, it is recommended to use calibrant ions that are as closely matched to the analyte ions as possible. In the context of protein and peptide studies, this is interpreted to mean that analytes and calibrants should be ionised by the same methods, should both be natively folded or both denatured, and be of similar mass and charge.[114] Reference CCS and mobility data for many protein and peptide ions are curated in a number of online databases, including polyamino acids (e.g. polyalanine),[118, 119] tryptic peptides,[120, 121] proteins[121-125] and protein complexes.[123, 126]

Both TWIMS and DT-IMS experiments have been performed on phosphopeptides, concluding that phosphopeptides have lower CCS than would be predicted for a randomly coiled peptide of the same m/z.[127-130] The lower CCS of phosphopeptides has been suggested to be due to the phosphate group forming non-covalent interactions with positively-charged amino acid residues in the peptide sequence, resulting in the peptide having a more compact conformation.[127] An IMS/MM approach has been applied to a set of synthetic phosphopeptides to assess whether molecular models could be used to rationalise the non-covalent interactions that had been predicted from the ECD behaviour of the peptides.[131] This approach has the benefit that two experimental data sets contribute to the proposal of a molecular model as the gas-phase structure of a peptide, giving more certainty that the appropriate molecular model has been proposed than if only one experimental dataset was used.

1.8 Molecular modelling

Molecular modelling is often employed in order to relate mass spectrometry and ion mobility spectrometry data to the structure of the analyte. Molecular modelling uses either quantum mechanics or classical mechanics to simulate the structure of a molecule, from which various physical parameters can be calculated and then related to experimental data.

1.8.1 Quantum mechanics

Quantum mechanics (QM) aims to predict and explain physical phenomena by solving the Schrödinger equation (Eq 12), which relates the energy of a system to the wavefunction via the Hamiltonian operator:[132]

$$\widehat{H}\psi = E\psi \tag*{(Eq 12)}$$

where $\widehat{H}$ is the Hamiltonian operator, ψ is the wavefunction and E is the energy of the system.

The Schrödinger equation is soluble for single-electron systems but to account for electron-electron interaction, approximations must be made. Various methods for finding approximate solutions to the Schrödinger equation exist, for instance, Hartree-Fock,[133, 134] correlation methods[135] and Density Functional Theory.[136] The Hartree-Fock method assumes that electrons are independent and only experience a mean field from each other, and so specific electron-electron interactions are not considered. The method employs basis sets (combinations of atomic orbitals calculated for single-electron systems) to build molecular orbitals for a many-electron system, from which the energy of the whole system may be calculated by the Schrödinger equation. The output of this calculation can be used to build new molecular orbitals, which are input to the same process with the coefficients of the basis sets varied, forming an optimisation loop until the energy converges on a single value, giving the optimised energy for a single molecular geometry. Adjusting the molecular geometry to minimise the energy of the molecule allows the ground state structure of the molecule to be predicted.

1.8.2 Molecular mechanics

For larger systems QM becomes too computationally expensive to be a viable option for molecular modelling.[137] Additionally, QM cannot simulate the evolution of a molecule through time without great computational expense. Molecular mechanics (MM) uses classical mechanics models to define the energies of bond stretching, bond bending, bond torsions and non-covalent interactions of a molecule. The energy of a molecule is taken to be a summation of the energies of all these components, given by (Eq 13):[138]

$$V = \sum_{ij}^{n} \frac{1}{2} k_{ij}(r_{ij} - r_{ij}^{0}) + \sum_{ijk}^{n} \frac{1}{2} k_{ijk}\left(\theta_{ijk} - \theta_{ijk}^{0}\right) + \sum_{ijk}^{n} A_{ij}(1 + \cos 3\phi_{ij}) + B_{jk}(1 + \cos 3\psi_{jk}) + \sum_{ij}^{n} \frac{q_i q_j}{4\pi\varepsilon r} + \sum_{ij}^{n} \frac{C_{ij}}{{r_{ij}}^{12}} - \frac{D_{ij}}{{r_{ij}}^{6}} \tag{Eq 13}$$

Where n is the number of atoms in the molecule, k are force constants, r are bond distances, r^0 equilibrium bond distances, θ bond angles, θ^0 equilibrium bond angles, A and B are torsional energy constants, φ and ψ are dihedral angles, q are charges, ε is the permittivity of the medium surrounding the molecule, and C and D are parameters related to the equilibrium distance between a pair of atoms and the potential of their interaction at this distance that define the Lennard-Jones potential between the pair of atoms.

Different MM methods may define more terms for non-covalent interactions: Eq 13 gives the Coulombic interaction between two atoms with charges q_i and q_j, and the Lennard-Jones potential that approximates the interaction between two neutral atoms, though terms to describe hydrogen bonding and other specific interactions are often included.[139]

Some of the terms of Eq 13 are given by the geometry of the molecule (bond lengths, bond angles and dihedral angles), but the force constants (k), torsion constants (A and B), charges (q) and Lennard-Jones parameters (C and D) must be defined for unique sets of atoms. A set of these parameters is called a forcefield, and they can either be calculated by QM methods, fitted to experimental data, or a combination of the two.[139]

1.8.3 Mobility calculation

In order to link molecular modelling data to ion mobility data, the mobility or CCS of the molecular model must be calculated and compared to experimentally-derived values. CCS may be calculated by simulating gas molecules colliding with the molecular model, which is generally held still while the gas molecules are propelled towards it as this is less computationally expensive than simulating the molecular model travelling down an IMS cell filled with buffer gas.[140] Using knowledge of the trajectory of the gas molecules, the CCS and mobility may be inferred. Ions are generally considered to undergo a tumbling motion through IMS cells and may enter the cell at any rotational orientation, so experimental CCSs are rotationally averaged.[140] To approximate rotational averaging, gas molecules are fired at the molecular model from many directions, and the simulated CCS averaged over all these directions.[112]

Various methods exist for the calculation of the mobility and CCS of a given molecular model: the projection approximation (PA),[112] exact hard spheres scattering (EHSS),[141] and trajectory methods (TM).[142, 143] These methods allow different levels of interaction between the gas molecule and molecular model to describe what happens to the gas molecule after the collision: PA does not account for any scattering, EHSS uses a hard spheres model of scattering, and TM include both attractive and repulsive diffuse interactions. The differences in the interaction between gas molecules and the molecular model for these methods are shown in Figure 1.15.

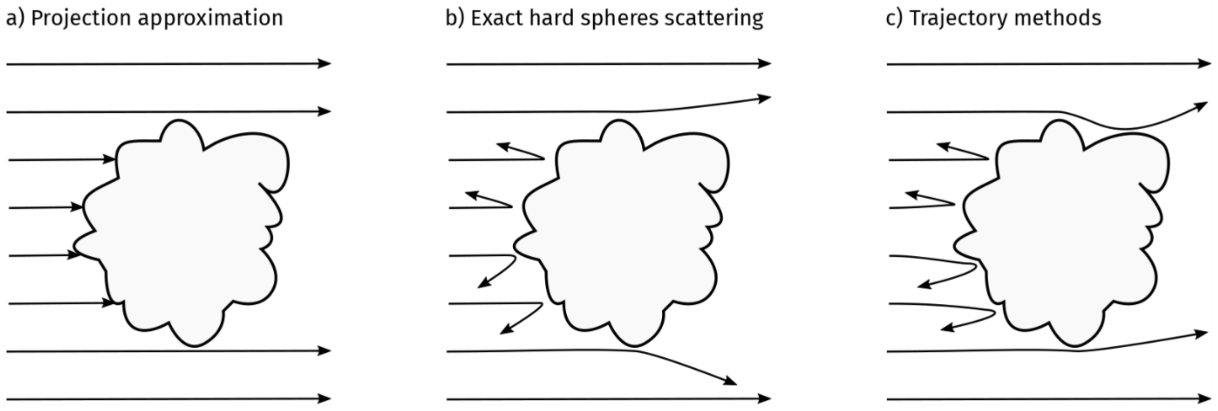


Figure 1.15 Representation of the interaction between gas molecules and a molecular model of the analyte ion as described by different CCS calculation methods: a) the projection approximation, b) the exact hard spheres scattering model and c) trajectory methods

The projection approximation does not allow any scattering after the collision and the trajectory of a gas molecule that collides with the molecular model stops at the point of collision, such that only gas molecules that miss the molecular model are recorded. Modifications to PA include the use of a 'shape factor' to improve the quality of the result.[144] EHSS allows elastic scattering of gas molecules that collide with the molecular model, describing the atoms of the gas molecule and the molecular model as hard spheres (i.e., rigid, non-overlapping spheres that only exhibit elastic interaction). Trajectory methods allow for both attractive and repulsive long-range interactions between the gas molecules and molecular model (such as those described by a Lennard-Jones (LJ) potential). Different implementations of trajectory methods employ different descriptions of these long-range interactions and the elastic or inelastic nature of the collisions that they result in.[142, 143, 145]

Many programs have been developed for the calculation of CCS/mobility, which each employing different implementations of the PA, EHSS and TM methods (Table 1.2).

Table 1.2 Details of common CCS calculators (Note: this list is not exhaustive)

Software package	Methods	Drift gases	Parallelisable?	Notes
Mobcal[141, 142]	PA, EHSS, TM	He	No	
Mobcal N2[145]	PA, EHSS, TM	N_2	No	A modification of the mobcal code to allow CCS calculation in N_2 gas. Additions include terms to describe the ion-multipole interactions with the gas and re-optimisation of LJ potentials.
IMoS[143]	PA, EHSS, TM	He, N_2	Yes	Includes diffuse versions of PA, EHSS and TM with inelastic collisions, and TMLJ, a trajectory method with explicit treatment of LJ potentials.
IMPACT[146]	PA	He	No	Optimised for rapid CCS calculation of large biomolecules.
Collidoscope[147]	TM	He, N_2	Yes	N_2 is modelled as a 'quasi-spherical' particle with an effective potential applied. Specific ion-multipole interactions with N_2 are not considered.
Sigma[148]	PA	He	No	Long-range potentials are approximated by a mass-dependent scaling factor.

TM are generally considered to give results closest to the experimental CCS because they explicitly treat more of the interactions between the analyte ion and the buffer gas that are known to affect the experimental CCS.[110] Conversely, treatment of more interactions increases the computational cost of each calculation, such that for larger systems TM become unfeasible. In these cases, scaling factors are often employed to overcome the systematic underprediction of CCS by PA and overprediction by EHSS.[149]

Most implementations of TM function by firing gas molecules from a randomly selected plane with the molecular model contained in a simulation box. The scattering angle of each gas molecule may be worked out based on the coordinates at which it entered and exited the simulation box. From initial trajectory and the scattering angle, the trajectory taken by the gas molecule through the simulation box may be inferred (Figure 1.16a). An integration of the trajectories of many gas molecules gives the

CCS at this orientation. This procedure is repeated for many more randomly selected planes (default number of planes for mobcal = 30) to give the rotationally averaged CCS. The ion mobility can then be calculated from the CCS via the Mason-Schamp equation.[141, 142, 146-148]

IMoS is unique among CCS calculators in that it calculates mobility from drag constants, not CCS from scattering angles. In this process, collisions are inelastic and allow momentum transfer to occur. The amount of momentum transfer between the molecular model and a gas molecule is calculated from the difference in the velocity of the gas molecule upon entering and exiting the simulation box (Figure 1.16b). The drag force of the collision is given by the momentum transfer per unit time. The mobility of the molecule may be worked out from the drag force of multiple collisions, and CCS may be calculated from the mobility by the Mason-Schamp equation.[143, 150]

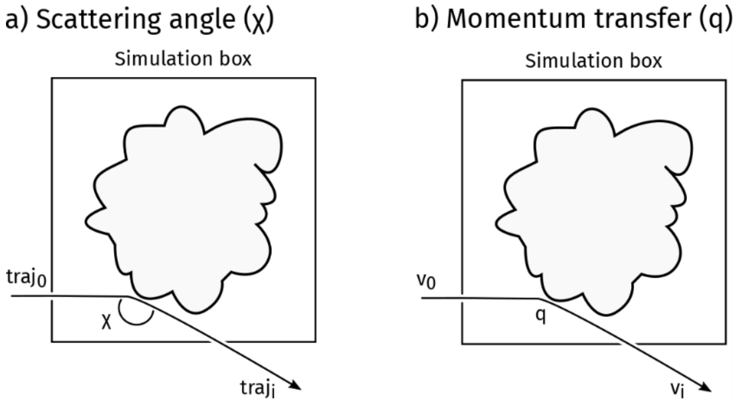


Figure 1.16 a) Scattering angles (χ) may be calculated from the trajectories of the gas molecule on entering and exiting the simulation box by the cosine rule; b) momentum transfer (q) may be calculated from the difference in the velocity of the gas molecule on entering and exiting the simulation box and the mass of the gas molecule. CCS may be calculated from a set of scattering angles, while mobility may be calculated from a set of momentum transfers

The comparison of simulated CCS to experimentally derived CCS is commonly referred to as IMS/MM. The general procedure involves producing a large ensemble of molecular models, either using MM methods alone or using structures from other experimental techniques (e.g. x-ray crystallography), calculating a simulated CCS for every structure and eliminating those that are inconsistent with the experimentally-derived CCS. IMS/MM has been employed for the structural study of many molecular classes: drugs,[145, 151, 152] nucleic acids,[153] peptides, [131, 154] proteins[155, 156] and protein complexes.[157, 158]

1.9 Instrument geometries

Commercial mass spectrometers are the most accessible way of implementing MS and IMS experiments, however the geometry of an instrument also limits the experiments that it may perform. The work presented in this thesis was performed on three different instrument geometries: all TWIMS (including ETD-TWIMS) experiments were performed on a Waters Synapt G2S instrument, while all ECD experiments were performed on a 7 T Bruker solariX XR FT-ICR mass spectrometer, with the exception of AI-ECD experiments, which were performed on a 12 T solariX instrument with a CO_2 laser installed for IR irradiation.

1.9.1 Synapt G2S

The Synapt G2S is a hybrid quadrupole TOF instrument used in the work presented in this thesis because of its ion mobility capability, provided by the inclusion of a TWIMS cell (Figure 1.17). Ions are introduced into the source region of the instrument, in which desolvation occurs. Next, ions enter the StepWave, which consists of two adjacent travelling wave ion guides: ions from the source enter the first ion guide and a DC offset potential pushes them into the adjacent second ion guide, leaving neutral moieties (i.e., unionised analyte or residual solvent) behind to be eliminated by the vacuum system.[49] Ions then pass through the quadrupole, which can be used to isolate species in a particular m/z window or bias a given m/z range, as discussed in Section 1.5.1. Following the quadrupole, ions enter the TriWave system, consisting of a trap, helium cell, TWIMS cell and transfer, all of which are travelling wave ion guides.[159, 160] The trap is used to accumulate and gate ions into the TWIMS cell; the helium cell, operated at a higher pressure than the TWIMS cell, prevents bulk movement of the nitrogen buffer gas that fills the TWIMS cell at the high N_2 pressures required for good ion mobility separation;[160] the TWIMS cell separates ions based on their mobility,[115] as described in Section 1.7.2; and finally, the transfer maintains the mobility separation of ions and delivers them into the TOF for detection (Section 1.4.1). The trap and transfer ion guides are identical, and both cells may be filled with a collision gas (the instrument used for the work presented in this thesis was equipped with helium as the collision gas) to perform CID (Section 1.6.1). When equipped with the ETD option, the Synapt G2S instrument can perform ETD fragmentation (Section 1.6.3). A make-up gas delivers the ETD reagent through a flow restrictor, which ensures an optimal flow rate, to the glow discharge electrode (operated at 25-75 µA depending on the ETD reagent used), which ionises the ETD reagent to form a beam of radical anions. The anionic ETD reagent beam coexists with the cationic analyte beam via continual switching of instrument polarity between the source and trap, where the ETD reagent and analyte are allowed to mix and the ETD reaction occurs.[161]

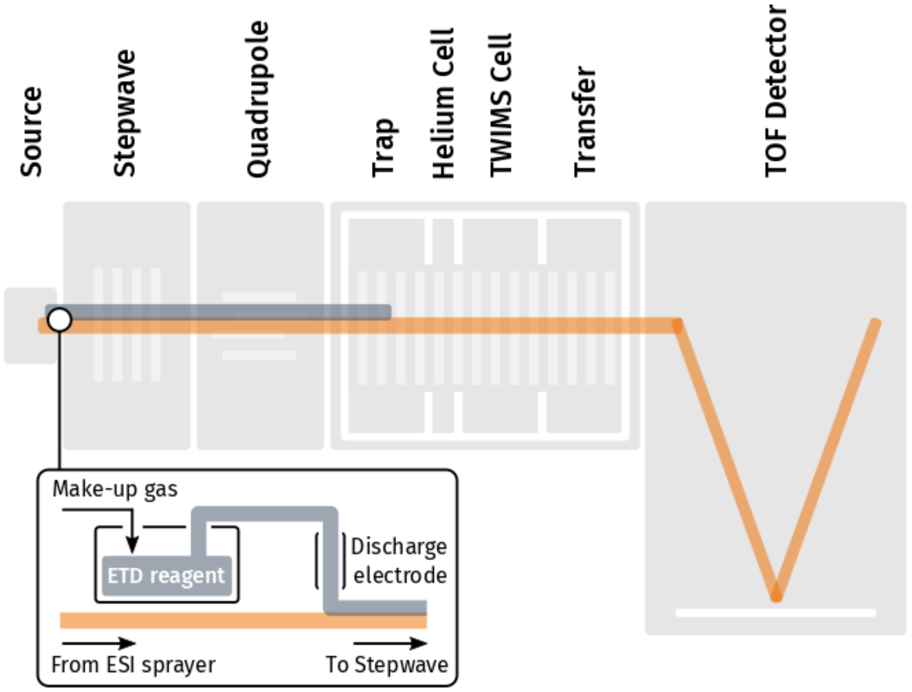


Figure 1.17 Schematic of the Synapt G2S instrument used in the work presented in this thesis. The orange trajectory represents the path that the analyte ion beam takes through the instrument, while the blue trajectory is the path the anion ETD reagent beam takes

Synapt G2S instruments are widely used in biological mass spectrometry, as the TWIMS cell and various fragmentation modalities allow CCS derivation, investigation into the conformational space of an analyte,[159] collision induced unfolding (CIU) experiments to examine unfolding dynamics,[162] separation of mixtures such as mixtures of isobaric species or mixtures of complex biological samples.[163-165] Criticisms of this instrument geometry include that the ion mobility resolution is lower than that which might be achieved by other ion mobility devices (such as DT-IMS and TIMS), and that the gas flow rates, gas pressures and field heating effects result in high internal energies of ions.[166-168]

1.9.2 7 Tesla solariX XR

The solariX XR mass spectrometer is an FT-ICR instrument employed for all ECD experiments presented in this thesis (with the exception of the AI-ECD experiments), because it is equipped with a cathode in the ICR cell that may be used to irradiate the analyte ion packet with a beam of electrons. A schematic of this instrument is shown in Figure 1.18. Ions are introduced first into a capillary heated by a bath gas that assists with desolvation. A deflector takes ions through a 90° angle, while neutral moieties continue straight ahead. The ion beam is then refocussed by a series of ion funnels and skimmers. The ions pass through a split octupole before entering a quadrupole, in which mass selection may be performed (Section 1.5.1), and a collision cell, in which the ions are collisionally cooled and may be fragmented by CID using argon gas (Section 1.6.1). ETD fragmentation (Section 1.6.3) may also be performed in the collision cell by the introduction of fluoranthene anion radicals, which are produced in the chemical ionisation (CI) source by the transfer of an electron from $CH_4^{\bullet-}$ (itself produced by applying an electrical discharge to a flow methane gas). The ETD reagent is introduced in the split octupole and coexists with the analyte ion beam until they are trapped and allowed to mix in the collision cell. Practically, some of the $CH_4^{\bullet-}$ ions survive long enough to be present in the mixture trapped in the collision cell, so some of the electron transfer occurs from this species as well as the fluoranthene. After exiting the collision cell, ions are accumulated and transferred via a hexapole to the ICR cell for detection (Section 1.4.2). The ICR cell in the solariX XR is a Paracell design (discussed in Section 1.4.2.1) equipped with a hollow cathode that produces a beam of electrons, which enters the ICR cell from the opposite side to the analyte ion packet and irradiates the ions to perform ECD fragmentation (Section 1.6.2). The hollow cathode permits short irradiation times, limiting the thermal heating of the analyte.[169] The Paracell is contained within a 7 T super-conducting magnet that creates the magnetic field that induces the ions into cyclotron motion.

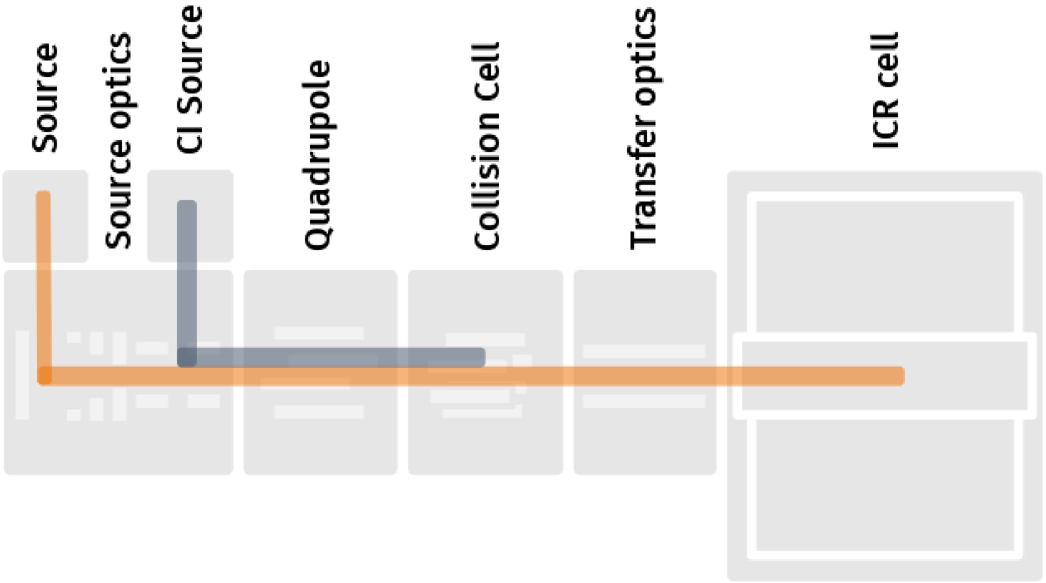


Figure 1.18 Schematic of the solariX XR instrument used in the work presented in this thesis. The orange trajectory represents the path that the analyte ion beam takes through the instrument, while the blue trajectory is the path the anion ETD reagent beam takes

1.9.3 12 Tesla solariX

The AI-ECD experiments discussed in Chapter 4 were performed on a 12 T solariX instrument. This instrument has fundamentally the same geometry as the solariX XR, but has an infinity cell in place of the Paracell (see Section 1.4.2.1), the magnet is capable of sustaining 12 T magnetic fields, and is equipped with a 10.6 µm, 25 W CO_2 laser that can irradiate ions in the ICR cell.

1.10 Aims and objectives

The work presented in this thesis aimed to develop and apply methods to study the structural consequences of phosphorylation by mass spectrometry. In Chapter 3, MM methods suited to the high-throughput modelling of phosphopeptides under mass spectrometry conditions were investigated, and combined in an IMS/MM workflow to compare the resultant molecular models to IMS data. IMS/MM is a well-established method of studying the gas-phase structure of ions. The non-covalent interactions represented in the MM structures produced from the IMS/MM workflow may then be used to rationalise the ECD behaviour of the peptide, which is known to be affected by the presence of non-covalent interactions. Chapter 4 details the application of these methods to model phosphopeptides aimed to provide insight on how phosphorylated residues interact with other residues and influence the structure of biological systems, while Chapter 5 is concerned with studying the structure of phosphopeptide and phosphoprotein fragments produced by ECD and ETD.

The specific aims that this thesis addresses are:

1. The development of MM methods for the study of peptides under mass spectrometry conditions
2. Use of the non-covalent interactions predicted from the ECD behaviour of phosphorylated peptides to refine the results of an IMS/MM workflow, and to apply this to the structural study of biologically-derived phosphopeptides
3. Assess whether these methods may be applied to the structural study of peptide or protein fragments, and whether the structure of the fragments may be used to infer information about the structure of the intact ion

Chapter 2: Materials and methods

Selected parts of this chapter have been published in: Simmonds AL, Lopez-Clajivo AF, Winn PJ, Russell DH, Styles IB, Cooper HJ, "Structural Analysis of 14-3-3-ζ-Derived Phosphopeptides Using Electron Capture Dissociation Mass Spectrometry, Travelling Wave Ion Mobility Spectrometry and Molecular Modelling", *J. Phys. Chem. B* 2020, 124, 3, 461-469, a copy of which may be found in Appendix 5.

This chapter details all methods used to generate the data discussed in subsequent chapters of this thesis. Where specific data or experiments are discussed in subsequent chapters, reference will be made to the section of this chapter that explains the methods used for that experiment.

2.1. Molecular modelling

2.1.1. Forcefield parameter optimisation

Starting geometries for 'non-standard' amino acid residues were built in Avogadro 1.2.[170] with an acetyl-glycine cap on the N-terminus and an amidated glycine cap at the C-terminus (where the residue was not itself a terminus), then optimised using Gaussian 09 (HF 6-31G (d,p)).[171] The electrostatic potential surface for the optimised geometry of each residue was calculated using the Merz-Singh-Kollman scheme[172] using the POP=MK and IOP(6/50=1) keywords. Ultrafine integrals were used throughout the quantum mechanics (QM) calculations. This scheme was used to be consistent with the ff99sb and phos10aa forcefields,[173] which were used for all MM presented in this thesis. It should be noted that this scheme is not state-of-the-art for QM, and that consistency with these pre-existing forcefields has been prioritised over using more accurate QM methods.

The residue was added to a .off library, using the antechamber, prepgen, parmchk2 and tleap utility programs contained in the AMBER distribution.[174] Backbone atoms were specifically labelled with amber-type backbone atom names and described in a 'MAINCHAIN' file to ensure that they were treated as backbone atoms. The code developed for converting the output of the QM calculation into a MM residue for use in future MM simulation is found in Appendix 1-1.

The .off libraries and corresponding .frcmod files for some of the residues produced in this work are available online (http://www.biosciences-labs.bham.ac.uk/cooper/datasets.php) and in Appendix 1-2

2.1.2 Molecular mechanics

Molecular modelling was performed in the AMBER16 suite[174] using the ff99sb and phos10aa forcefields. The phos10aa forcefield was selected as it is the only forcefield that contains parameters for phosphorylated amino acid residues, while ff99sb was selected to be consistent with phos99aa, which was optimised following the same procedure as ff99sb. It should be noted that ff99sb is not considered to be state-of-the art, and that consistency with phos10aa has been prioritised over using more accurate MM methods. Additionally, because these forcefields were optimised for solvent-phase simulations they are considered semi-quantitative for gas-phase simulations,[175] and are used due to a lack of a suitable gas-phase forcefield. Simulations were performed using a three-tiered simulated annealing (SA) approach (dt_{SA} = 1 fs, T_{max} = 1000 K, T_{min} = 300 K; tier 1, n = 1000; tier 2, n = 12,000; tier 3, n = 12,000) from an extended starting structure to produce a 25,000-strong ensemble. The code to implement three-tiered SA was provided by Dr Liza Perez (Texas A&M University). MM ensembles were processed using cpptraj.[176]

2.1.3 Ion mobility calculation

Different trajectory methods implemented by mobcal N2[145] and IMoS[143] were compared to optimise the speed and accuracy of the ion mobility calculation. For mobcal N2, all parameters except 'imp' were kept constant at default values (Appendix 2), and for IMoS, all parameters were kept constant (Appendix 1-3), but several different trajectory method calculations were used (IMoS emulating mobcal and IMoS using diffuse collisions with Lennard-Jones interactions (TMLJ)).

In all subsequent CCS calculations, IMoS was used to calculate the helium CCS using TMLJ at temperatures and pressures representative of the TWIMS cell conditions (T = 310 K, p = 300 Pa).

2.1.4 Molecular modelling ensemble analysis

A CCS-matched ensemble was created from models whose simulated CCS was within the experimental CCS ± 5%. Onward analysis of CCS-matched ensembles was performed using MATLAB R2018b (Mathworks, Natick, MA, USA), with the code used found in Appendix 1-4.

Pairwise distances between all atoms were calculated for each model of the CCS-matched ensemble, to create a $n_{atoms}^2 \times n_{structures}$ matrix, which was subjected to dimensionality reduction by random projection following the procedure described by Palmer *et al.*[177] The compression factor of the random projection was optimised on a subset of the ensemble to select the greatest compression factor from which the ensemble could still be accurately reconstructed. For this optimisation the L1 norm of the difference between the original data (M) and the reconstructed data (Q × P) was used as an error metric and calculated for several compression factors. Fitting a double exponential to these data and graphically solving for the compression factor at ⅔ the point of maximum curvature gives the optimum compression factor. The projections produced were clustered by k-means (n = 10, Euclidean) and the sum of least squares method was used to define the structures closest to the centroids of each of the 10 clusters, which would serve as a representative of each cluster. This process is represented in Figure 2.1.

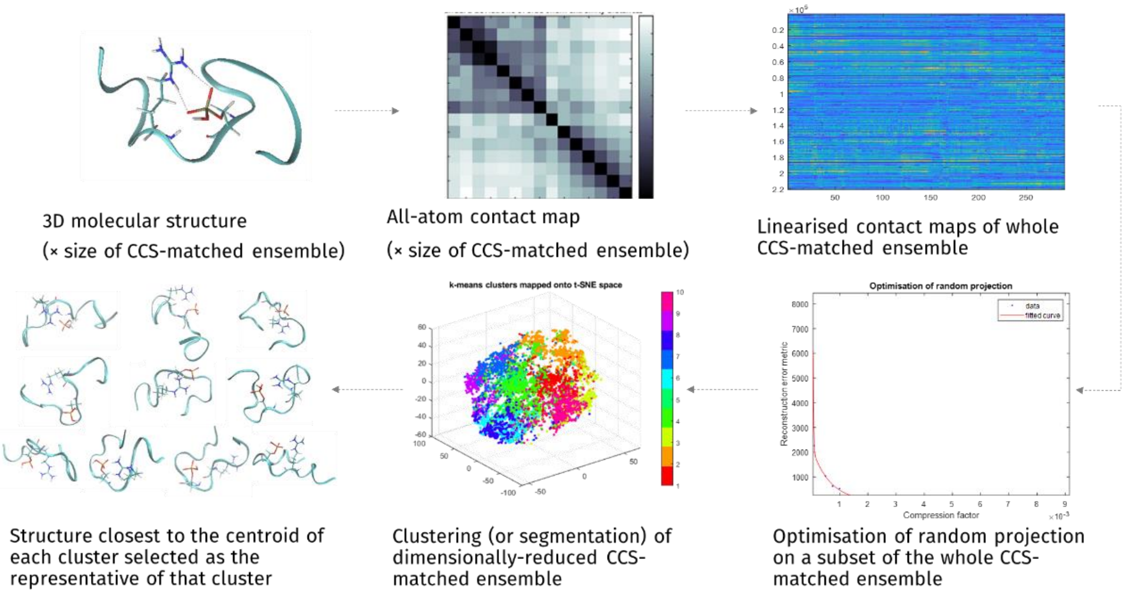


Figure 2.1 The CCS-matched ensemble analysis process. Note: The clustered CCS-matched ensemble is represented here in 3D space, where each point is a single molecular structure, though clustering occurs on the projections which have n_{atoms}^2 × *compression factor dimensions*

All QM, MM and ion mobility calculations were performed using the University of Birmingham's BlueBEAR HPC service, which provides a High Performance Computing service to the University's research community. Further details of the BlueBEAR HPC service may be found at: http://www.birmingham.ac.uk/bear.

2.2 Mass spectrometry

2.2.1 Sample preparation

LC-MS grade solvents (methanol, water, acetonitrile, formic acid, propan-2-ol) were purchased from Fisher-Scientific (Leicestershire, U.K.).

Calibrants: sodium iodide (Waters, Wilmslow, U.K.), dissolved in propan-2-ol/water (50:50), was used for mass calibration of the Synapt G2S instrument. A positive ESI tuning mix (Agilent, Santa Clara, CA, U.S.A.) was used for mass calibration of the solariX XR instrument. A 0.25 mM solution of substance P (Waters, Wilmslow, U.K.) in acetonitrile/water/formic acid (59.5:39.5:1) was used to optimise fragmentation for all ETD and ECD experiments.

14-3-3-ζ-derived peptides were synthesised by GenicBio (Shanghai, China) and used without further purification. 1 mg/mL stocks of each peptide in methanol/water/formic acid (39.9:60:0.1) were diluted in methanol/water/formic acid (49.5:49.5:1) to a final concentration of ~1.0 µM. Tryptic peptides for the calibration of TWIMS experiments were obtained from Thermo Scientific (Waltham, USA) and diluted to 100 µL in water.

GSK3β-derived peptides were synthesised by GenicBio (Shanghai, China) and used without further purification. Peptides were dissolved in acetonitrile/water/formic acid (59.5:39.5:1) to a concentration of 5 µM. Tryptic peptides for the calibration of TWIMS experiments were produced by digestion of apo-Transferrin bovine (Sigma-Aldrich, St Louis, MO, U.S.A.), chicken egg white lysozyme (Sigma-Aldrich, St Louis, MO, U.S.A.) and bovine serum albumin (Sigma-Aldrich, St Louis, MO, U.S.A.) using trypsin gold (Promega, Madison, WI, U.S.A.). ~150 µL aliquots of each protein (10 mg/mL in water) were mixed with ~150 µL of trypsin (1 mg/mL in water) and agitated overnight at 37.5 °C. The reaction was then quenched by placing the mixtures on ice, after which they were mixed and purified using a C18 ZipTip pipette tip (Millipore, Burlington, MA, U.S.A.). The purified tryptic peptide mixture was diluted to 500 µL in acetonitrile/water/formic acid (49.5:49.5:1)

β-casein (Sigma-Aldrich, St Louis, MO, U.S.A.) was dissolved in water/acetonitrile/formic acid (49.5:49.5:1) to a concentration of 5 µM. ETD-TWIMS experiments were calibrated against solutions of polyalanine (0.1 mg/mL) and ubiquitin (5 µM) in water/acetonitrile/formic acid (49.5:49.5:1).

2.2.2 ECD experiments

Samples were introduced to a 7 Tesla solariX XR (Bruker Daltonics, Bremen, Germany) mass spectrometer equipped with a Paracell via one of three ionisation methods, as indicated by Table 2.1:

Static nanoelectrospray ionisation (nESI) in positive ion mode with pulled glass capillaries (P-97 Flaming/Brown micropipette puller tip Sutter Instrument Company, Novato, CA, USA).

Triversa Nanomate nESI (Advion, Ithaca, NY, U.S.A.). in positive mode using voltages in the range 1.2-1.8 kV and gas pressures of 0.3-0.35 psi. Voltages and pressures may have been varied during a single experiment to maintain a stable spray current.

Static ESI in positive mode using the solariX XR inbuilt syringe pump at flow rates between 100 and 120 µL/hr, nebuliser gas flow rates of 0.1 – 1.0 bar.

Capillary voltages and temperatures are listed in Table 2.1. Peptide ions were isolated in the quadrupole (isolation width as indicated in Table 2.1) prior to fragmentation in the ICR cell. ECD was performed by use of thermal electrons at a current of 1.50 A, with cathode biases, lens potentials and pulse lengths as in Table 2.1. Ion transfer was optimised for each ion, further details may be found in Appendix 1-5, which contains instrument parameters for all experiments performed on the solariX XR instrument. Data analysis was performed with Data Analysis 4.2 software (Bruker Daltonics) and manually searched for *a, x,* $c^{\bullet}/c$, *b, y,* $z/z^{\bullet}$ fragment ions. For β-casein, fragments were identified by the SNAP algorithm with a signal-to-noise threshold of 2, and a maximum charge of 23+.

Table 2.1 Instrument parameters employed during ECD fragmentation on a 7 T solariX XR MS. Note: in source-CID is abbreviated to isCID, and quadrupole to Q

Sample	ESI method	Capillary		Q width / m/z	ECD parameters		
		Potential / V	Temperature / °C		Bias / eV	Lens / V	Pulse length / s
14-3-3-ζ Pep-01	Static nESI	800	120	5	0.6	10.0	0.20
14-3-3-ζ Pep-02	Static nESI	800	120	5	0.6	10.0	0.60
14-3-3-ζ Pep-03	Static nESI	800	120	5	0.6	10.0	0.60
14-3-3-ζ Pep-04	Static nESI	800	120	5	0.6	10.0	0.60
14-3-3-ζ Pep-05	Static nESI	800	120	5	0.6	10.0	0.20
14-3-3-ζ Pep-06	Static nESI	800	120	5	0.6	10.0	0.40
14-3-3-ζ Pep-07	Static nESI	800	120	5	0.6	10.0	0.60
14-3-3-ζ Pep-08	Static nESI	800	120	5	0.6	10.0	0.20
QLVRGEPNVSpYICSRYYR	Nanomate	100	120	N/A			
GSK3β-pY11	Static ESI	4500	200	3	1.2	15.0	0.05
GSK3β-pY11 + isCID (90 V)	Static ESI	4500	180	3	1.2	15.0	0.05
GSK3β-Y11	Static ESI	4000	120	5	0.8	5.0	0.80
GSK3β-pS11	Static ESI	4000	180	5	4.0	12.6	0.04
GSK3β-S11	Static ESI	4000	180	5	5.2	12.6	0.04
GSK3β-pY11-noC	Nanomate	50	120	5	4.0	12.6	0.04
GSK3β-Y11-noC	Nanomate	50	120	5	4.0	12.6	0.04
GSK3β-pY11-noEnoC	Nanomate	50	180	5	4.0	12.6	0.04
GSK3β-Y11-noEnoC	Nanomate	50	180	5	4.0	12.6	0.04
β-casein	Nanomate	50	120	3	0.8	10.0	0.10

2.2.3 AI-ECD experiments

In AI-ECD experiments, samples were introduced to a 12 T solariX (Bruker Daltonics, Bremen, Germany) mass spectrometer by static positive ion mode nESI with a 7–8 µL/hour flow rate. The capillary was maintained at 950 V and 180 °C. Ions were isolated in the quadrupole (isolation width 10 m/z) prior to fragmentation in the ICR cell. ECD was performed by use of thermal electrons at a current of 1.5 A, with the cathode bias at 1.5 eV, the ECD lens at 4 V and a pulse length of 0.1 s. Infrared ion activation was achieved using a Synrad 48–2 25 W CO_2 laser with a 10.6 µm wavelength operated at 60% power and with a pulse length of 0.08 s.

2.2.4 TWIMS experiments

TWIMS experiments were performed on a Synapt G2-S mass spectrometer (Waters Corp., Milford, MA, U.S.A.) equipped with a Triversa Nanomate (Advion, Ithaca, NY, U.S.A) nESI source. Samples were introduced using nESI voltages in the range 1.2-1.8 kV and gas pressures of 0.3-0.35 psi. The TWIMS cell was operated under ~2 mbar of nitrogen. Further instrument parameters can be found in Appendix 1-6**.** Arrival time distributions were produced using a mass range that contained all the isotopologues of each peptide and collision cross sections (CCS) calculated from the apex value. CCS calibration was performed by an adaptation of the procedure described by Ruotolo *et al.*[178] Reduced drift time (DT') of the calibrant ions were calculated by (1):

$$\boldsymbol{DT' = DT - \frac{c\sqrt{m/z}}{1000}} \qquad \textbf{\textit{(1)}}$$

where DT is drift time and c is the 'Enhanced Duty Cycle' (EDC) delay coefficient (c = 1.43 for the Synapt G2S instrument used in the work presented in this thesis). Reduced collision cross section (Ω') of the calibrant ions was calculated by (2):

$$\boldsymbol{\Omega' = \frac{{}^{DT}CCS_{He}}{z\sqrt{\frac{1}{\mu}}}} \qquad \textbf{\textit{(2)}}$$

where z is the charge of the ion, μ is the reduced mass and $^{DT}CCS_{He}$ is the reference CCS for the calibrant ion. A power regression of $\boldsymbol{\Omega'} = A \cdot \boldsymbol{DT'}^{b}$ gives values for A and b, which can be used to calculate CCS for the analyte ions by (3):

$$\boldsymbol{CCS_{analyte} = \frac{A \cdot z \cdot {}^{TW}DT'_{N2}{}^{b}}{\sqrt{\mu}}} \qquad \textbf{\textit{(3)}}$$

This process was carried out over three wave amplitudes (as in Table 2.2), from which a mean and standard deviation of $CCS_{analyte}$ was calculated. Standards used for calibration were $[M+2H]^{2+}$ tryptic peptides of BSA, cytochrome c and myoglobin, with reference drift-tube helium CCS obtained from the Clemmer database.[121] Lists of calibrant ions used can be found in Appendix 3, and the calibration curves in Appendix 4. TWIMS data and data processing are reported according to the guidelines published by Gabelica *et al.*[114] Data acquisition and processing were carried out using MassLynx 4.1 and Driftscope 2.1 software (Waters Corporation, Manchester, UK).

Table 2.2 TWIMS parameters used during TWIMS experiments on the Synapt G2S MS

Sample	TWIMS wave velocity / ms^{-1}	TWIMS wave amplitudes / V	TWIMS cell N_2 pressure
14-3-3-z-derived peptides	450	19, 20, 21	~ 2 mbar
GSK3β-derived peptides	300	18, 19, 20	~ 2 mbar

2.2.5 ETD-TWIMS experiments

ETD-TWIMS experiments were performed using the Synapt G2S instrument (Waters Corp., Milford, MA, U.S.A.) equipped with a Triversa Nanomate nESI source (Advion, Ithaca, NY, U.S.A.). Samples were introduced using nESI voltages in the range 1.2-1.8 kV and gas pressures of 0.3-0.35 psi. The ETD option was installed on the instrument, using 1,3-dicyanobenzene as the ETD reagent at a refill rate of 0.1 s and refill interval of 0.8 s. Ions were isolated in the quadrupole (LM resolution 4.9, HM resolution 15) prior to fragmentation by ETD in the trap and mobility separation of the fragments in the TWIMS cell. The TWIMS cell was maintained at ~3 mbar of nitrogen in all ETD-TWIMS experiments. ETD and TWIMS parameters are displayed in Table 2.3, while further information on instrument parameters can be found in Appendix 1-6**.** Arrival time distributions and collision cross sections (CCS) were produced using identical methods to TWIMS experiments. Lists of calibrant ions and the corresponding calibration curves may be found in Appendix 3 and Appendix 4, respectively.

Table 2.3 ETD and TWIMS parameters used during ETD-TWIMS experiments on the Synapt G2S MS.

Parent ion	ETD parameters		TWIMS parameters	
	Refill rate	Refill interval	Wave velocity / ms^{-1}	Wave amplitudes / V
GSK3β-pY11 $[M+2H]^{2+}$	0.1	0.8	600	24.5, 25, 25.5
GSK3β-pY11 $[M+3H]^{3+}$	0.1	1	800	21, 22, 23
β-casein	0.1	1	450	38, 39, 40

2.2.6 ETD-ECD experiments

ETD-ECD experiments were achieved using the 7 Tesla solariX XR mass spectrometer via Triversa Nanomate nESI in positive mode using voltages in the range 1.2-1.8 kV and gas pressures of 0.3-0.35 psi. Capillary voltages and temperatures are listed in Table 2.4. Peptide ions were isolated in the quadrupole (2 m/z isolation width, except GSK3β-pY11, where isolation width is 5 m/z) prior to ETD fragmentation in the collision cell using a fluoranthene ETD reagent (parameters as in Table 2.4). Peptide fragments were isolated in the ICR cell using a frequency sweep (2 m/z notch width). The isolated fragment was then fragmented by ECD using thermal electrons at a current of 1.50 A, with MS^3 cathode biases, lens potentials and pulse lengths as in Table 2.4. For each sample, a spectrum was collected of the isolated fragment (without MS^3 ECD fragmentation). A full list of instrument parameters used in these experiments may be found in Appendix 1-5. Data analysis was performed with Data Analysis 4.2 software (Bruker Daltonics) and manually searched for a, x, $c^{\bullet}/c$, b, y, $z/z^{\bullet}$ fragment ions.

Table 2.4 Instrument parameters used in ETD-ECD experiments. Note: accumulation is shortened to accum.

Sample	Capillary		MS^2 ETD parameters		MS^3 ECD parameters		
	Potential / V	Temperature / ˚C	Reagent accum. / s	Reaction time / s	Bias / eV	Lens / V	Pulse length / s
GSK3β-pY11	500	120	0.20	0.30	1.0	5.0	0.08
GSK3β-pY11-noC	50	180	0.25	0.10	3.5	12.0	0.04
GSK3β-Y11-noC	50	180	0.25	0.10	2.9	12.0	0.04
GSK3β-pY11-noEnoC	50	180	0.20	0.03	3.5	12.0	0.04
GSK3β-Y11-noEnoC	50	180	0.20	0.05	3.5	12.0	0.04

2.2.7 ECD-ECD experiments

ECD-ECD experiments were achieved using the 7 Tesla solariX XR mass spectrometer via Triversa Nanomate nESI in positive mode using voltages in the range 1.2-1.8 kV and gas pressures of 0.3-0.35 psi. Capillary voltages and temperatures are listed in Table 2.5. Peptide ions were isolated in the quadrupole (2 m/z isolation width, except GSK3β-pY11, where isolation width is 5 m/z) prior to ECD fragmentation in the ICR cell using thermal electrons at a current of 1.50 A, with MS^2 cathode biases, lens potentials and pulse lengths as in Table 2.5. Peptide fragments were isolated in the ICR cell using a frequency sweep (notch width 2 m/z). The isolated fragment was then fragmented by ECD using thermal electrons at a current of 1.50 A, with MS^3 cathode biases, lens potentials and pulse lengths as in Table 2.5. The time delay between CHEF isolation and MS^3 ECD fragmentation was varied as shown in Table 2.5. For each sample, a spectrum was collected of the isolated fragment (without MS^3 ECD fragmentation). A full list of instrument parameters used in these experiments may be found in Appendix 1-5. Data analysis was performed with Data Analysis 4.2 software (Bruker Daltonics) and manually searched for *a, x,* $c^\bullet$*/c, b, y, z/*$z^\bullet$ fragment ions.

Table 2.5 Instrument parameters used in ECD-ECD experiments. Note: ECD bias energy and lens potential can not be changed between MS^n steps.

Sample	Capillary		ECD parameters				
	Potential / V	Temperature / °C	Bias / eV	Lens / V	MS^2 pulse length / s	MS^3 pulse length / s	Time delay / s
GSK3β-pY11	500	120	1.2	1.0	0.5	0.2	0.8
GSK3β-pY11-noC	50	180	3.5	12.0	0.04	0.04	0.01, 0.5, 5
GSK3β-Y11-noC	50	180	2.9	12.0	0.04	0.04	0.1
GSK3β-pY11-noEnoC	50	180	3.5	12.0	0.04	0.04	0.1
GSK3β-Y11-noEnoC	50	180	3.5	12.0	0.04	0.04	0.1

Chapter 3: Development of an electron capture dissociation-guided ion mobility spectrometry/molecular modelling workflow and its application to a structural study of 14-3-3-ζ-derived phosphopeptides

Selected parts of this chapter have been published in: Simmonds AL, Lopez-Clajivo AF, Winn PJ, Russell DH, Styles IB, Cooper HJ, "Structural Analysis of 14-3-3-ζ-Derived Phosphopeptides Using Electron Capture Dissociation Mass Spectrometry, Travelling Wave Ion Mobility Spectrometry and Molecular Modelling", *J. Phys. Chem. B* 2020, 124, 3, 461-469, a copy of which may be found in Appendix 5.

3.1 Overview

As discussed in Section 1.2, structural mass spectrometry provides a means to study the structure of biomolecules, which can give insight for pharmacological studies or mechanistic biology.[179] In order to contribute to this endeavour, this thesis develops methods to study phosphopeptides and phosphoproteins. In this chapter, the development of an electron capture dissociation (ECD)-guided ion mobility spectrometry/molecular modelling (IMS/MM) workflow is presented. This workflow is subsequently applied to a structural study of a suite of phosphopeptides derived from the protein 14-3-3-ζ, a protein with diverse functions that has been identified as a possible therapeutic target in some cancers.[25]

3.2 Parameter optimisation

3.2.1 Rationale

Molecular modelling (MM) packages such as AMBER[174] and gromacs[180] were developed for and are most commonly used for solution-phase, biological pH simulations. Conversely, mass spectrometry is a gas-phase technique in which molecules must carry a net charge in order to be detected and so are often not in a biological pH range. The work presented in this thesis focusses on the analysis of proteins and peptides ionised in positive mode and detected in the gas phase. There are a number of challenges to be overcome when using molecular modelling packages to model proteins and peptides in a mass spectrometry environment:

1. Selection of forcefield
2. Limited availability of forcefield parameters for biochemically altered amino acid residues
3. Limited availability of forcefield parameters for amino acid residues in non-standard protonation states
4. Lack of forcefield parameters for amino acid residues that have undergone fragmentation

3.2.2 Selection of forcefield

There are a wide range of forcefields available for molecular dynamic simulation, all with various strengths and weaknesses, so the selection of forcefields for this work required careful consideration before any simulations could be performed. For consistency across all simulations performed, it was necessary to select a single forcefield (or set of forcefields) to use. This forcefield would also be used to determine how to calculate any additional amino acid parameters that would be required. The majority of literature studies evaluating different forcefields consider their efficacy in solution-phase simulations, though a few make recommendations for gas-phase simulations.[175] Polarisable forcefields, often considered the best solution-phase force fields in terms of their accuracy at describing electrostatic interactions,[181] remain largely untested for gas-phase simulations and are computationally expensive. But, the ‘FF’ range of AMBER forcefields (ff99, ff99SB, ff99SB-ILDN, ff14SB) are widely used in this context,[175] having been initially developed based on *ab initio* quantum mechanics (QM) gas-phase simulations of short peptides, with later forcefields from this range being refined by solvent-phase simulations.[182]

As the work in this thesis concerns phosphorylated peptides and proteins, it was also important to use a forcefield that was optimised for the simulation of phosphorylated amino acid residues. The phos10aa forcefield included in the AMBER distribution contains parameters for phosphorylated serine, threonine and tyrosine residues with doubly- and singly-deprotonated phosphate groups (i.e., $[PO_4]^{2-}$ and $[PO_4H]^-$ respectively). The phos10aa forcefield is based on parameters produced by the Sticht group,[173] created to be consistent with the ff99SB forcefield. As such, it was decided to use the ff99SB and phos10aa forcefields and to use the methods outlined by the Sticht group when developing further parameters.

3.2.3 Methods for producing parameters not available in the standard AMBER distribution

As the amino acid residues that were required did not contain any atom types that were not already included in the ff99SB and phos10aa forcefields, it was unnecessary to recalculate all force constants and non-bonding parameters, and those already incorporated in the forcefields could be used. What remained to be calculated was the starting geometry for each residue and its electrostatic potential.

The methods followed by the Sticht group were an initial Hartree-Fock optimisation with a 6-31g (d,p) basis set with fixed backbone torsions taken from the parm99 forcefield. Each residue was capped with an acetyl group on the N-terminus and a N-methylamine group on the C-terminus, which were removed prior to calculation of the electrostatic potential. Electrostatic potentials were calculated by the Merz-Singh-Kollman scheme.[172] For residues where the side chain was sufficiently flexible to interact with the backbone, the side chain geometry was optimised separately to the backbone, in order that the resultant geometry did not bias interactions between the residue's side chain and backbone when used as a starting structure in onward MM simulations.

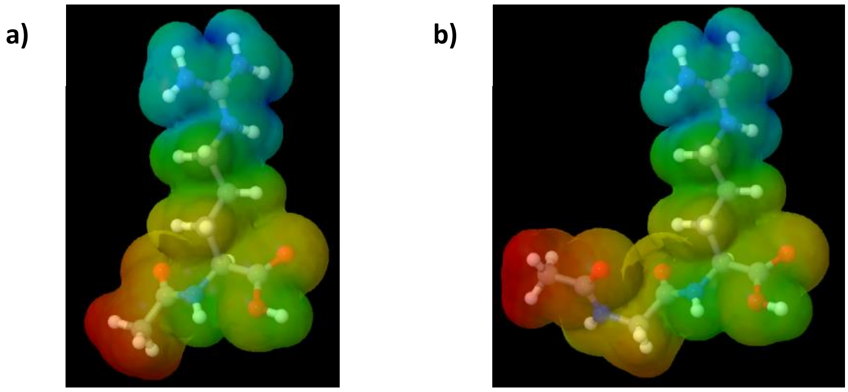


Figure 3.1 Electrostatic potential of a) [Ac-Arg+H]$^+$ and b) [Ac-Gly-Arg+H]$^+$

To investigate whether acetyl and N-methylamine caps were sufficient to replicate the environment that connecting atoms would experience in a full protein, the electrostatic potentials of [Ac-Arg+H]$^+$ and [Ac-Gly-Arg+H]$^+$ optimised at the HF (6-31g (d,p)) level were compared (Figure 3.1)**Error! Reference source not found.**. These electrostatic potentials showed minor differences at the connecting nitrogen atom. Because the Ac-Gly cap is more likely to be representative of the environment of a full protein, this N-terminal cap was used in all subsequent parameter optimisations. Given the effect of including a Gly in the N-terminal cap, it was also decided to include a Gly in the C-terminal cap, using Gly-Am in all subsequent parameter optimisations that necessitated a C-terminal cap.

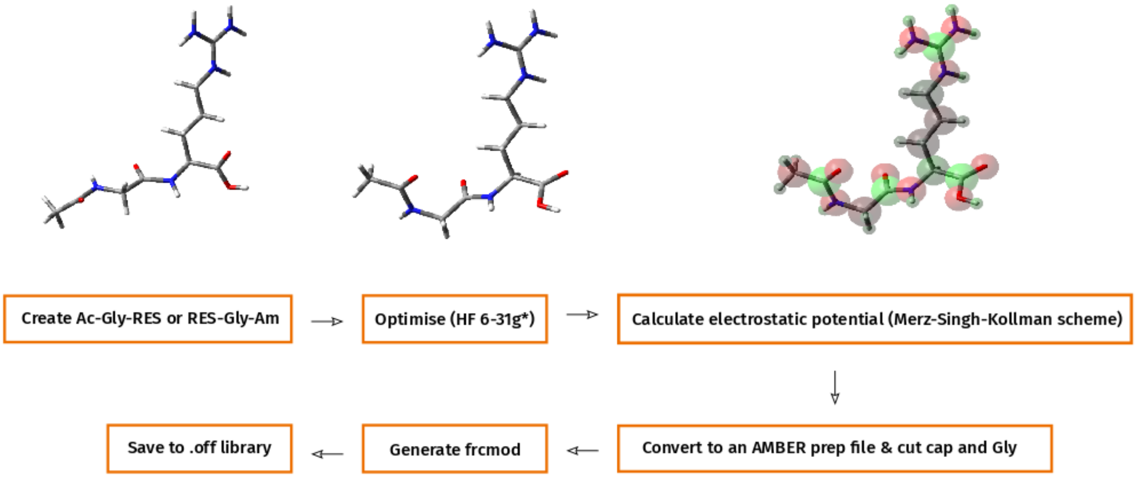
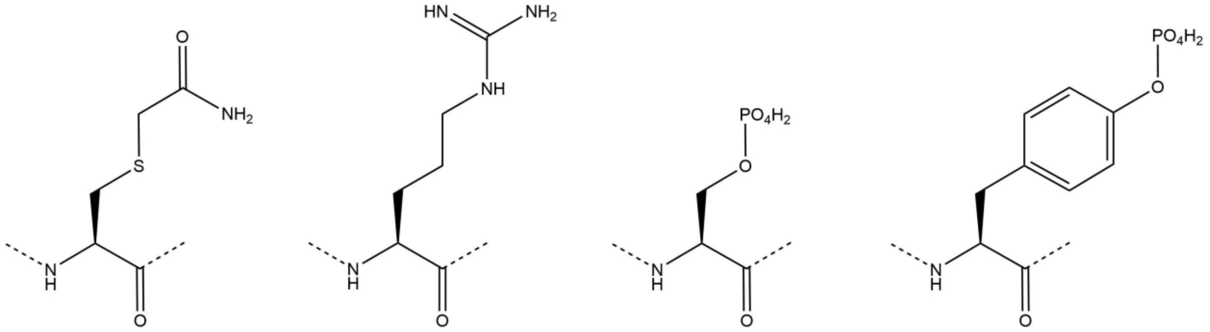
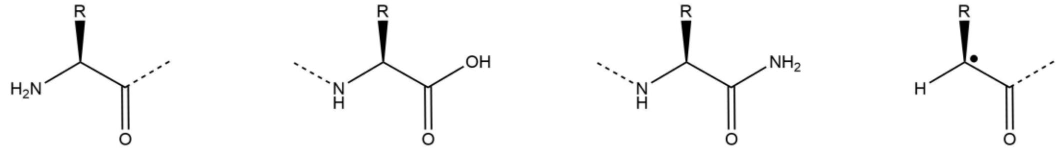


Figure 3.2 Workflow used for the creation of AMBER parameters for amino acid residues not included in the standard AMBER distribution

The optimised workflow used for creating AMBER parameters is shown in Figure 3.2. The capped residues were first optimised at the Hartree-Fock level with the 6-31g (d,p) basis set followed by calculation of the electrostatic potential (ESP) via the Merz-Singh-Kollman scheme to ensure consistency with the ff99SB forcefield. The optimised, ground state structure and ESP were then converted to an AMBER .ac file using the AMBER utility program antechamber before Ac-Gly and/or Gly-Am caps were removed. The backbone atoms were renamed in the .ac file to conform with the AMBER naming system and explicitly ensure that they use backbone parameters from the ff99SB forcefield in subsequent MM simulations. Generating a 'MAINCHAIN' file allowed the connecting atoms of the residue, its backbone atoms and its charge to be defined. The 'MAINCHAIN' file and .ac file were input to the AMBER utility program prepgen to create a .prepi file. This .prepi file was then checked by the AMBER utility program parmchk2, creating a .frcmod forcefield modification file if there were any atoms, bonds, angles or torsions not already defined by any AMBER forcefield and that needed to be created for the residue. The .prepi file for the residue was then loaded into the AMBER utility program tleap, through which its residue type was set to "protein", the connecting atoms were defined, the residue was given a unique three letter code and saved to a .off library containing similar residues (e.g. one .off library contains amino acid residues with neutral N-termini). The process of creating .off and .frcmod files for residue sets was automated via a script (Appendix 1-1) that required the input of the .gesp gaussian output files containing the ESP information for the residue(s) to be processed, the two or three letter name by which the residue(s) may be called in leap and .pdb files, the residue type ("N", "mid" or "C"), and the name of the .off file to while the residue(s) may be saved. The script was set up to process multiple residues at once by interpreting the three letter amino acid code at the start of the .gesp filename and convert it to the corresponding single letter amino acid code, such that the remaining two letters in the name used in leap and .pdb files could be used to designate the type of residue. For instance, the set of neutral N-terminal residues have three letter codes of the form "XNH", where X is the single letter amino acid code for any of the 20 naturally-occurring amino acids, and NH refers to the N-terminus being neutral. Creating comprehensive residue sets in this way enabled the .off file with any required .frcmod files to be loaded into tleap and peptides created by combining the required residues in the .off with those contained in the ff99SB and phos10aa forcefields. Given that tleap is a text-based program, the residue sets parameterised here could be used to create input structures for modelling of gas-phase peptides in a high-throughput manner through automated scripting.



Cysteine-S-acetamide Neutral arginine Neutral phosphoserine Neutral phosphotyrosine



Neutral N-terminus Neutral C-terminus c^+ fragment $z^{\bullet}$ fragment

Figure 3.3 The structures of all residues that have been parameterised and presented in this thesis. R groups indicate that residues of this type have been created for the 20 naturally-occurring amino acids

Presented in this thesis, parameters have been produced for cysteine-S-acetamide, neutral arginine, neutral phosphotyrosine and neutral phosphoserine residues. Sets of parameters have been produced for the 20 naturally-occurring amino acids residues (Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val), with neutral C-termini, neutral N-termini, and their structures when adjacent to cleavage sites for c^+ and $z^{\bullet}$ type fragments.[78, 80] The structure of all the residues that have been parameterised are shown in Figure 3.3. For the parameter sets containing all 20 naturally-occurring amino acids at least one residue from each set has been trialled by use in an MM simulation.

The results of MM simulations using these parameter sets can be found throughout Chapters 3, 4 and 5 of this thesis.

3.2.4 EXD fragment parameter validation

For the parameters produced for residues adjacent to the cleavage site in c^+ and $z^\bullet$ fragments (the formation of which is discussed in Section 1.6.2), it has been assumed that the charge/radical is localised to that residue, rather than being delocalised across the whole fragment. To determine whether this assumption is not detrimental to the production of experimentally consistent structures, it was necessary to validate these parameters against literature studies.

The 2016 study by Martens *et al*. compares infrared multiphoton dissociation (IRMPD) spectra of $z^\bullet$ fragments of the peptide $[AAHAR]^{2+}$ to simulated IR spectra.[183] In that paper, a QM density functional theory (DFT) calculation (DFT, B3LYP, 6-31++G(d,p)) is used to minimise the structure of each fragment, then the entire fragment is parameterised to perform an MM simulation by simulated annealing (SA) (T = 1000 K, dt = 1 fs, n = 500). By grouping the results of the MM simulation by structural similarity (root mean square deviation) 30-50 unique structures were selected to be minimised using QM (DFT, B3LYP, 6-31++G(d,p)) and their vibrational frequencies calculated to predict their IR spectra. Qualitative comparison of these predicted spectra to experimental IRMPD spectra of each fragment allowed the best candidate structure for each fragment to be selected. To validate the parameters for the residues adjacent to the cleavage site of $z^\bullet$ fragments, the same fragments of AAHAR were simulated using these parameters and residues from the ff99SB set by simulated annealing (T = 1000 K, dt = 1 fs, n = 1000). The results of this SA experiment were searched manually for structures similar to those proposed by the Martens *et al.* study. For each of the 4 fragments $z_4^\bullet$ ($[AHAR]^{\bullet+}$), $z_3^\bullet$ ($[HAR]^{\bullet+}$) and $z_4^\bullet$ ($[AR]^{\bullet+}$), structures similar to the IRMPD/QM-validated structures reported by Martens *et al.* were identified in the results of the SA simulation; these are shown in Figure 3.4 alongside the corresponding IRMPD/QM-validated candidate structure proposed by Martens *et al.* The $z_4^\bullet$ fragment shows the most dissimilarity to the IRMPD/QM-validated structure, in terms of the number of rotations that would need to be performed to replicate the geometry of the IRMPD/QM-validated structure, with two torsions in the arginine and histidine backbones resulting in the N-terminal alanine being rotated through 45°, and the arginine side chain having a different conformation. This observation is likely due to the high number of degrees of freedom that a molecule with 61 atoms has, making it unlikely to find an exact match to any reference structure in a sample size of 1000 candidate structures. It is notable that the differences in the structure do not affect the reproduction of the same non-covalent interactions as the IRMPD/QM-validated structure. Since these two rotations do not cause a change in the non-covalent interactions within the molecule, it is unlikely that they would cause a change in the predicted IR spectrum, so they would not be distinguishable by the workflow employed by Martens *et al*. For the $z_3^\bullet$ and $z_2^\bullet$, the most similar structures from the SA experiment using the parameters described here are near-identical to the IRMPD/QM-validated structures, with the only differences being related to subtle changes in bond lengths and angles that do not affect the gross structure of the fragments. These subtle differences likely arise because the two sets of structures are minimised using different levels of theory (DFT vs MM).

The fact that the candidate structures from the Martens *et al.* study can be reproduced implies that the assumption that the radical is localised to the residue adjacent to the cleavage site does not affect the geometries or intramolecular interactions accessed by a structure. The methods used here, a single 1000-structure SA run, are much simpler and less computationally intensive than those used by Martens *et al.*, demonstrating the potential for higher throughput by using the parameters described in Section 3.1.3.

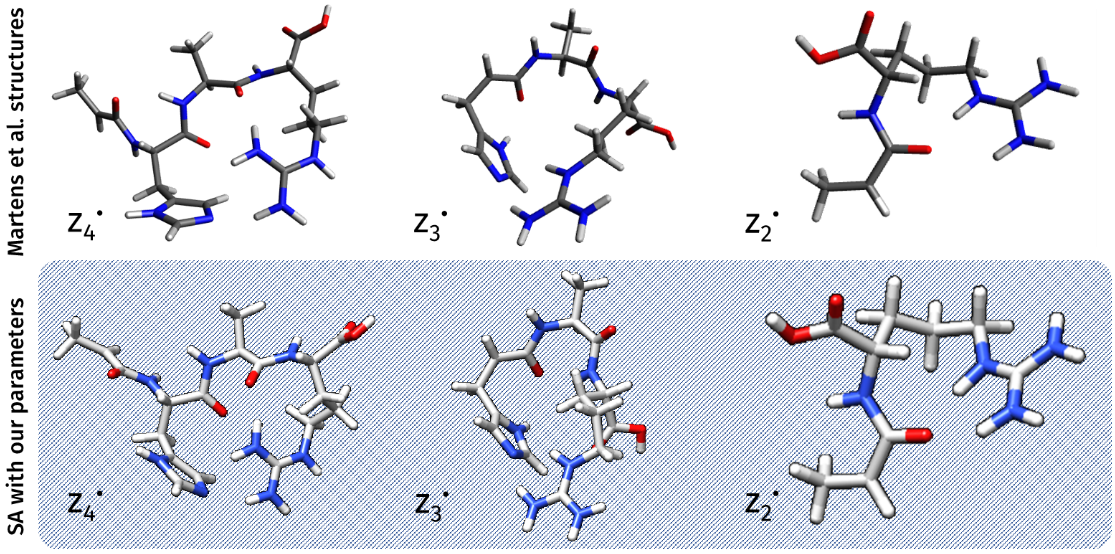


Figure 3.4 Top: AAHAR $z_4^\bullet$, $z_3^\bullet$ and $z_2^\bullet$ structures from Martens et al., 2016 IRMPD/QM study. Bottom: AAHAR $z_4^\bullet$, $z_3^\bullet$ and $z_2^\bullet$ structures produced by a single simulated annealing experiment (n = 1000, T = 1000 K) using the parameters developed in this thesis

At the time of performing this work, no similar study into the structure of c^+ fragments existed, so a similar analysis into the validity of the parameters for residues adjacent to the cleavage site in c^+ fragments could not be performed. Studies have been performed that determine that c^+ fragments have an amide group at the cleavage site,[154] as is consistent with the Utah-Washington mechanism discussed in Section 1.6.2.[78, 80] Given that amides are much more commonly simulated in AMBER than carbon radicals, most forcefields (including ff99SB) contain parameters to facilitate their simulation, with the result that the residues adjacent to the c^+ fragmentation site represent an unusual geometry that required optimisation, but can be simulated without any additional forcefield parameters. As such, it is not as pressing to validate that the c^+ fragment parameters are experimentally consistent.

3.3 Selection of molecular modelling methods

Having made parameters to facilitate the simulation of peptides in the gas phase, it was necessary to select methods for simulation. QM methods were not feasible because of the size of the peptides to be modelled (>200 atoms) and because of the need to probe their conformational landscapes. In order to decide which MM methods might be appropriate, it is always necessary to consider the experimental conditions that the MM simulations are attempting to simulate. In the mass spectra presented in this thesis, a typical, well-resolved m/z peak represents between 10^4 and 10^{10} ions, depending on the instrument used and the data acquisition time. To create representative MM ensembles that contain a similar level of conformational depth, those ensembles must contain a similar number of ions. To create large ensembles, simulated annealing or molecular dynamics (MD) simulations are commonly employed. MD simulations consist of allowing the molecule to explore its conformational landscape at a certain set of temperature and pressure values, while SA experiments involve cooling the molecule slowly from an elevated temperature. MD tends to sample a range of higher energy conformations around local minima but can struggle to access conformations with high energy barriers without the use of tools (such as accelerated MD[184] and REMD[185]) that facilitate crossing into other local minima, in contrast SA samples the minima of wells across the potential energy surface. The mass spectrometry techniques used here all have a gradient of pressures, with ionisation occurring at ambient pressure and pressures inside the mass spectrometers ranging from the source at around 1 mbar to high vacuum in the detector (between 10^{-7} and 10^{-10} mbar), which makes it quite likely that the ions are annealed during their transit through the mass spectrometer. Ions also tend to undergo collisional energy transfer during their transit through the mass spectrometer, the majority of which occurs in the TWIMS device of the Synapt G2S instrument and in the collision cell of the solariX XR instrument (instrument geometries are discussed in Section 1.9). The combination of the pressure gradient and collisional cooling likely causes ions to occupy the minima of their potential energy surface, so it is appropriate to use MM methods that sample these minima, i.e., simulated annealing.

In order that the results of the SA calculations were not overly biased by the starting structure (usually an extended structure produced in tleap), a three-tiered SA approach was adopted, as reported by Kim *et al.* in a 2015 study.[131] In this approach a simulated annealing run was performed 1000 times from the extended starting structure to give 1000 output structures. From this initial 1000 structures, 12 random structures were selected to serve as starting structures for the second tier. From these 12 random structures 12 further runs of 1000 SA simulations were performed to bring the total of the first and second tiers to 13,000 structures. Selecting a single structure from each of the 12 second tier sets of 1000 output structures gave the 12 starting structures for the third tier. As with the second tier, 12 runs of 1000 SA simulations were run from these starting structures, bringing the total to 25,000 structures. The three-tiered SA approach is represented in Figure 3.5.

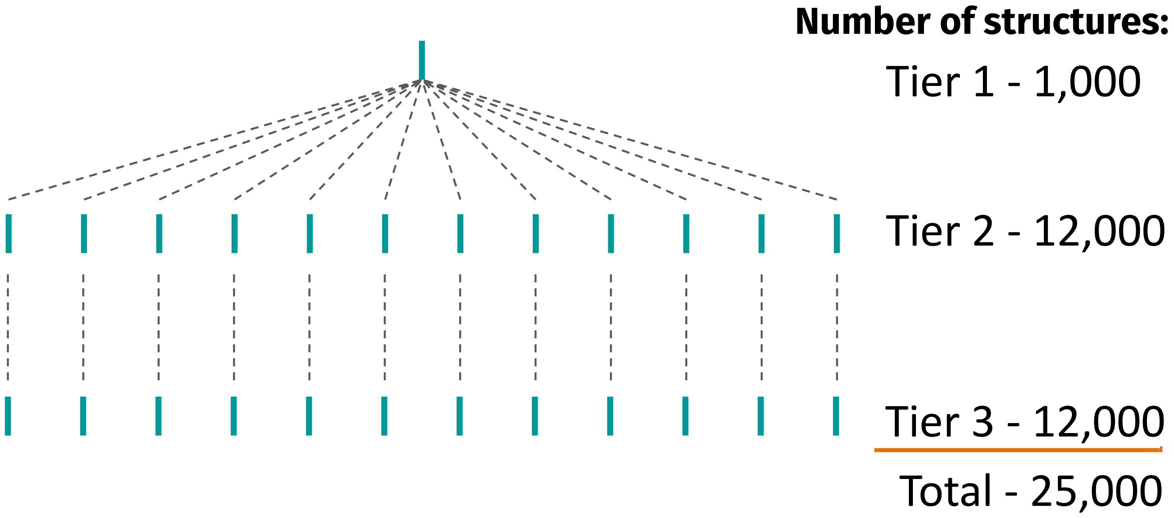


Figure 3.5 Schematic of the three-tiered simulated annealing approach. Vertical blue bars represent 1000 individual SA runs

Each single SA run was composed of two SA-minimisation cycles (SA1, minimisation, SA2, minimisation). SA1 heats from 300 K to 1000 K over 0.21168 ns and then cooling from 1000 K to 300 K for a further 0.21168 ns. SA2 heats from 300 K to 1000 K over 0.10584 ns, followed by cooling from 1000 K to 300 K over 0.10584 ns. Both SA1 and SA2 use 1 fs timesteps. The resultant structure from the full SA1-min-SA2-min run was inputted into the next SA run and the process repeated.

3.4 Ion mobility calculator comparison

Ion mobility/collision cross section calculators are vital in the ion mobility spectrometry/molecular modelling (IMS/MM) workflow as they provide a link between the experimentally derived collision cross section (CCS) and structures produced by MM. There are many different methods and software packages available for calculating the CCS of a molecular model, as discussed earlier (See section 1.8.3). The molecules discussed in this work are sufficiently small that the computational cost of performing trajectory method[142] (TM) calculations is not prohibitive. Furthermore, because the experiments are carried out in nitrogen gas, derivation of nitrogen CCS ($CCS_{N2\rightarrow N2}$) and calculation of simulated nitrogen CCS is preferable. Nevertheless, because of limitations of the calibrants available in many cases the nitrogen DT (drift time) values were calibrated against helium calibrant CCS values ($CCS_{N2\rightarrow He}$), meaning that the ability to calculate both simulated nitrogen and helium CCS is essential. Trajectory methods are the only methods that explicitly consider nitrogen dipole interactions with the analyte, and so are the most appropriate method to use in this case. Two ion mobility calculators (mobcal N2[119] and IMoS[143]), both capable of performing trajectory method calculations, were compared on the basis of computational speed, reproducibility and accuracy to determine which to use.

Mobcal N2 is the best-established mobility calculator used for calculations in nitrogen, however it is very slow, taking ~24 hours to calculate the cross section of one molecular model on the computer systems available during the course of this work (BlueBEAR HPC, see Section 2.1 for details). Given that the CCS is required for each of the 25,000 models in the three-tier SA molecular ensembles, this is too slow as it would take 600,000 core hours to calculate a set of CCS for one ensemble. The 'imp' parameter in mobcal N2 is one that is often changed to reduce the computational cost of performing mobcal simulations. This parameter determines the accuracy of the integration of the path of each simulated gas molecule and has been identified as being one of the slowest steps in the mobcal calculation. To investigate the effect of the imp parameter on the speed and accuracy of the calculation, the CCS of the phosphopeptide APLSFLGSLPK(pS)YVK was calculated with various values of the imp parameter (Figure 3.6). It was shown that computational time increases linearly with increasing imp values, whereas the standard deviation of the resulting CCS value is inversely proportional to imp, such that at high imp values, there is relatively little benefit in terms of reducing error and come at a high computational expense. Standard deviations in mobcal are calculated using the CCS resulting from a number of compete cycles of the mobility calculation subroutine MOBIL2 (in all calculation reported in this thesis, 10 MOBIL2 cycles were used); the CCS result of a single mobcal calculation is also an average of the CCS outputted by the MOBIL2 cycles.

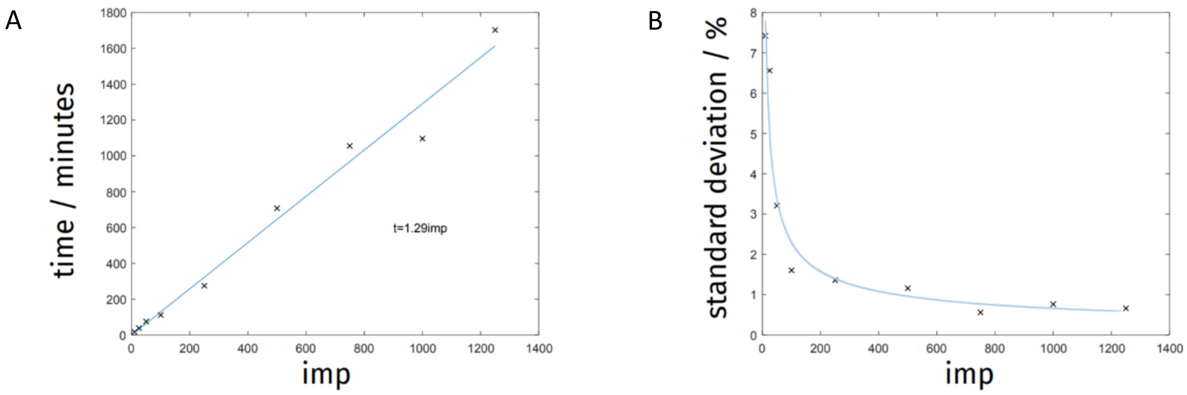


Figure 3.6 Relationship between the imp parameter of mobcal and A) the computational time taken to calculate the CCS of a single molecular model of [APLSLGSLPK(pS)YVK+2H]$^{2+}$ and B) the standard deviation of each of those calculations

An imp value of 25 was selected to reduce computational time sufficiently, such that mobcal N2 could be viably used to calculate the CCS of 25,000 molecular models. The computational time taken for each calculation also depends on the nature and number of atoms in the molecule. To ensure that an imp value of 25 also enables calculations for larger molecules within a suitable time, the CCS of two 18-mer peptides was calculated, one phosphorylated and the other not. The resultant CCS, standard deviations and computation time of these calculations are shown in Table 3.1 along with details of the imp = 25 calculation for APLSFLGSLPK(pS)YVK. All three CCS calculations completed within an hour, which is viable for calculation of 25,000 CCS for the whole MM ensemble. The standard deviations of these calculations (~5%) are unacceptably large, especially given that they must be propagated with the experimental errors (~3% for TWIMS experiments) and calibration errors (~1% for good quality ion mobility spectra), which could lead to total errors of ~6% for the whole workflow.

Table 3.1 CCS for three different peptides calculated using mobcal N2 with imp = 25, and the calculation times for each calculation.

Sequence of $[M+2H]^{2+}$	Ac-QLVRGEPNVS(pY)I(Calk)SRYYR-Am	Ac-QLVRGEPNVSYI(Calk)SRYYR-Am	APLSLGSLPK(pS)YVK
CCS (Imp-25 Mobcal N2)	609.27 ± 5.78%	666.15 ± 5.30%	486.94 ± 4.87%
Calculation time	40 min	1 hr	30 min

Given that mobcal N2 calculations either take too much computational time or give results with too much error, another mobility calculator, IMoS,[143] was investigated. As IMoS is less well established than mobcal N2, it was necessary to evaluate its performance compared to mobcal N2. The particular performance characteristics investigated here are accuracy compared to mobcal N2 and the standard deviation of the IMoS calculation using TM with diffuse collisions and Lennard-Jones interactions (TMLJ). Since this work was performed, an in-depth investigation of other metrics of performance has been published. This publication included aspects such as scalability with the size of the molecule, comparison of different methods within IMoS and the effects of parallelisation.[150]

As discussed in section 1.8.3, mobcal N2 calculates the cross section of a molecule from the deflection angles of gas molecules that collide with the analyte, while IMoS calculates mobility from the drag constant obtained by calculating the momentum transfer between the gas molecule and analyte. IMoS can perform a number of variants of the trajectory method using different collisions (specular elastic collisions, diffuse collisions or diffuse collisions with Lennard-Jones type interactions considered), as well as a method that emulates that of mobcal N2. Table 3.2 shows the results of calculations on the same molecular model of the peptide Ac-QLVRGEPNVS(pY)I(alkC)SRYYR-Am using mobcal N2 with imp = 1000, IMoS emulating mobcal (i.e., IMoS calculating CCS via deflection angles instead of its usual mode of operation in which CCS is calculated from the drag constant) and the IMoS TMLJ calculation. IMoS (TMLJ) out-performs IMoS emulating mobcal based on both the speed of the calculation and in producing a result close to that of mobcal N2, making it the more preferable option of the two. Given that IMoS emulating mobcal does not explicitly state the parameters used to control the calculation of deflection angles, such as the number of points taken in the integration of the trajectory or the number of cycles of trajectory calculation performed, it is hard to know to what extent the calculation is the same as the mobcal calculation reported here. As TMLJ involves the most rigorous treatment of collisions in IMoS, produces CCS reasonably close to those calculated by mobcal and completes calculations in a reasonable time, no other variations of the trajectory method available in IMoS were considered.

Table 3.2 CCS produced by different mobility calculation methods for the same model of [Ac-QLVRGEPNVS(pY)I(alkC)SRYYR-Am+2H]$^{2+}$and the time associated with these calculations

Method	Mobcal N2 (imp = 1000)	IMoS (emulating mobcal)	IMoS (TMLJ)
CCS	468.09	434.3593	482.382
Calculation time	22.5 hours	1 hours	40 mins
% difference in CCS to Mobcal N2 value	--	-7.2 %	3.1 %

Unlike mobcal N2, IMoS does not provide an estimate of the standard deviation of its results. Therefore, to estimate the error, the CCS calculation of a single molecular model was repeated 1000 times and the standard deviation of the resultant set of CCSs was calculated. This process was repeated on several structures of the same peptide and it was found that the standard deviations were ~1%. This standard deviation is much more acceptable than the ~5% for mobcal N2 with imp = 25, and similar to that of mobcal N2 with imp = 1000. It should be noted that this error is a measure of the repeatability of a single CCS calculation, and the accuracy of mobcal and IMoS calculations of phosphorus-containing systems cannot be assessed without comparison to some ground truth phosphorus-containing system of known structure and known CCS.

These investigations into the suitability of IMoS as a CCS calculator have shown that IMoS is as fast as mobcal N2 with imp = 25, whilst being as accurate and reproducible as mobcal N2 with imp = 1000.

IMoS also allows easier control of parameters that affect the CCS such as temperature, pressure, number of collisions, and the nature of the collision gas. For all these reasons, IMoS was used in all CCS calculations described later in this thesis.

3.5 Molecular modelling ensemble analysis

Having selected appropriate methods for molecular modelling and calculation of the CCS of the resultant 25,000 molecular models, it was necessary to decide on suitable analyses to filter out experimentally inconsistent models and establish trends in the experimentally consistent models.

To filter out models that were inconsistent with the experimental CCS (CCS_{expt}), structures were excluded whose simulated CCS (CCS_{sim}) was outside of CCS_{expt} ± 5%. The value of 5% was intended to encompass the errors from the experiment (~3%), the calibration to get CCS_{expt} (~1%) and the calculation of CCS_{sim} (~1%), which propagated to a total error of ~3.32%. A 5% margin accounts for additional error should any of these individual errors deviate from these approximate values by 1% (i.e., experimental error = 4%, CCS_{expt} calibration error = 2% and CCS_{sim} calculation error = 2%), which would lead to a total error of 4.90%. This filtering process produces a CCS-matched ensemble upon which further analysis can be performed.

The CCS-matched ensembles were typically in the range 1000-15000 structures, depending on the quality of the MM simulation and the proportion of structures that had simulated CCS values consistent with the experimental CCS, and were therefore too large for structure-by-structure analysis to be feasible. Trends in the dataset could be analysed on a whole CCS-matched ensemble scale, but in order to propose a small number (i.e. <10) of candidate structures from the ensemble, dimensionality reduction was required. Clustering was identified as an appropriate dimensionality reduction method, as the aim was to group similar structures and propose representatives of those groups as candidate structures. The k-means clustering method was chosen because of its ability to produce a consistent number of clusters and to segment data evenly if it exists in a continuum rather than as discrete clusters.[186, 187] The structures, which were represented as Cartesian coordinates, had to be converted to a common, coordinate-free format before they could be clustered. This conversion was needed because the structures were not necessarily centred at the origin and were likely oriented differently in Cartesian space, both factors that could influence the onward analysis. From the coordinates, pair-wise interatomic distances were calculated to form an all-atom contact map. This contact map was then linearised to form a single vector for each structure, which allowed the whole CCS-matched ensemble to be represented as a 2D matrix (M). Clustering based on the all-atom contact map was intended to allow more structural features to be represented in the resultant clusters than by clustering based on residue centroids, alpha-carbons or root mean squared deviation between structures.

To further optimise the clustering, it was decided to introduce a dimensionality reduction step. The linearised all-atom contact maps contained many contact distances that vary little between different structures, e.g. contact distances between atoms of the same residue. Because they were so numerous, small variations in these contact distances could have a large influence on the clustering, preventing more structurally significant features (i.e., the fold of the peptide or its inter-residue non-covalent interactions) being represented in the clusters. The introduction of a dimensionality reduction step before the clustering was intended to prevent the small changes in these contact distances from dominating the clustering. Random projection achieves this by projecting the linearised contact map for each structure onto a vector of shorter length via a random sampling of the contact distances.[188] The more the data is compressed by random projection (i.e., the fewer contact distances are sampled), the more error is associated with the reconstruction of the data. This error can be calculated by taking the L1 norm of the difference between the original data and the reconstructed data. The relationship between the reconstruction error and the compression factor takes the form of an L-shaped curve that can be described by a double exponential equation (Figure 3.7). It was necessary to select an optimum compression factor that reduces the error in the reconstruction while sufficiently limiting the effect that the contact distances over which there is little variation between structures have on the clustering. Random projection was performed multiple times on a subset of the data with different compression factors. By plotting the compression factor against the reconstruction error, the compression factor at two-thirds the value of the point of maximum curvature could be calculated and used as the compression factor for random projection of the full dataset, as was described by Palmer *et al.* in 2015.[177]

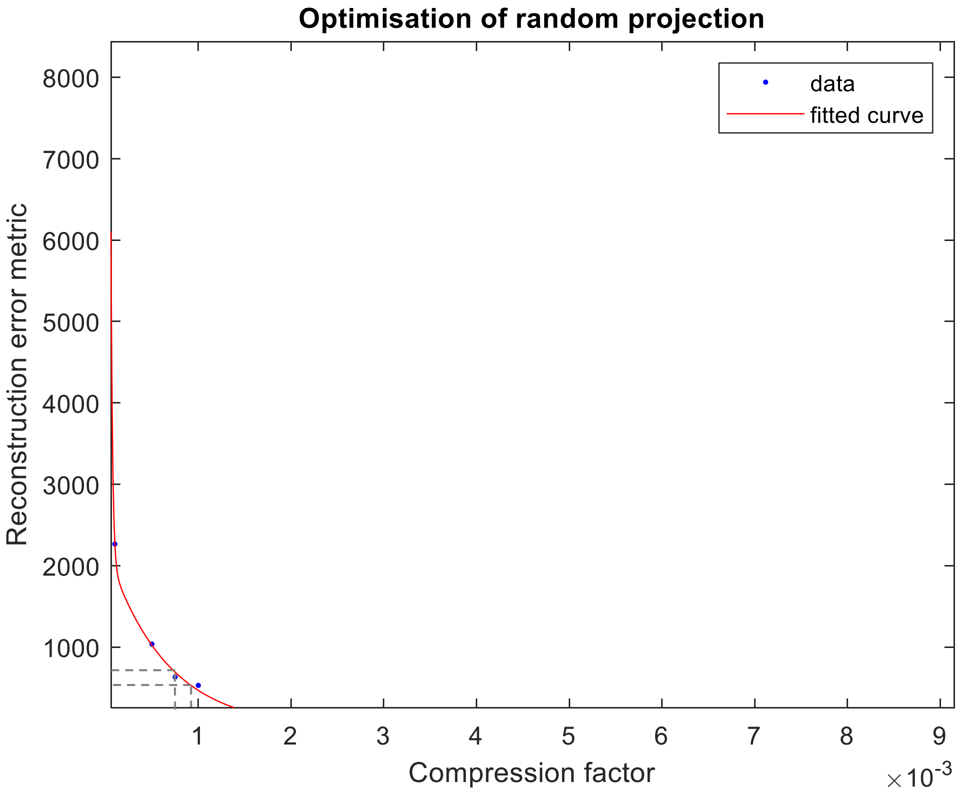


Figure 3.7 Selection of an appropriate compression factor for random projection. The optimal balance of compression vs reconstruction error is found at 2/3 the CF value of the point of maximum curvature.

Once representative structures were selected from each cluster, they could be analysed as a proxy for the cluster that they represent. The structural features of the representative structures were subsequently compared to the experimental IMS and ECD behaviour of the peptide.

Using IMS data to infer information about the structure of peptides and proteins is well established, as discussed in Section 1.7. Generally, the presence of multiple peaks in an arrival time distribution

(ATD) is interpreted as the presence of multiple conformers for a single species, with each conformer having a discrete arrival time. The relative width of the ATDs of a set of species is often used as a way to discuss the range of the conformational space of the difference species, i.e., how many unresolved conformers of different arrival times there might be in a single ATD peak, and how flexible the species might be.[114]

The ECD behaviour can also be used to infer structural information about a protein or peptide, as discussed in Section 1.6.2. In the context of comparing representative structures to the ECD behaviour of a peptide, focus will be placed on using the ECD behaviour to localise interacting partners of non-covalent interactions between acidic and basic residues. Peptides and proteins with many acidic and basic residues tend to have reduced fragment coverage compared to those that do not, as has been noted in various studies of phosphoproteins, phosphopeptides and their unphosphorylated counterparts.[100, 189] The reduction in fragment coverage is attributed to non-covalent interactions between the acidic and basic residues, which prevent fragments from dissociating and being detected separately. This results in fragmentation patterns in which fragments are detected up to an acidic or basic residue, followed by a lack of fragments up to another acidic or basic residue, after which fragments are detected again. If the lack of fragments is caused by a non-covalent interaction (or interactions), then it follows that the two acidic/basic residues that start and end the gap in fragment coverage are participating in that interaction. In this way it is possible to use the ECD behaviour of a peptide or protein to localise non-covalent interactions to specific residues in its sequence.[100, 107] The non-covalent interactions in a peptide predicted by their ECD behaviour can then be compared to the non-covalent interactions in structures produced from an IMS/MM study, as demonstrated by Kim *et al.,*[131] in an ECD-guided IMS/MM workflow.

3.6 Application of the modelling workflow to a structural study of 14-3-3-ζ-derived peptides

The full ECD-guided ion mobility spectrometry/molecular modelling workflow described above was trialled on five synthetic peptides that were derived from the sequence of 14-3-3-ζ around its S58 residue (Table 3.3), with Pep-01 being the unmodified sequence; Pep-02, Pep-03 and Pep-04 each having one of the arginine residues substituted for leucine; and Pep-05 having two arginine residues substituted for leucine. Successive arginine-to-leucine substitutions allowed the effect of multiple positive charge carriers on ECD behaviour to be examined, as well as probing the stability of the non-covalent interactions in the RRS(pS)WR phosphorylation motif. Experimental data (ECD MS/MS and TWIMS MS) were collected by Dr Andrea Lopez-Clavijo while she was working within the group, using methods described in Section 2.2.2 and Section 2.2.4, and were used to probe the modelling workflow developed here and to provide useful structural information by the combination of molecular modelling and mass spectrometry data.

Table 3.3 Sequences of the 14-3-3-ζ-derived peptides Pep-01, Pep-02, Pep-03, Pep-04 and Pep-05. Residues in blue are basic, while residues in red are acidic.

Peptide	Sequence
Pep-01	V V G A R R S pS W R V S S I
Pep-02	V V G A R L S pS W R V S S I
Pep-03	V V G A R R S pS W L V S S I
Pep-04	V V G A L R S pS W R V S S I
Pep-05	V V G A R L S pS W L V S S I

3.6.1 ECD and TWIMS mass spectrometry of 14-3-3-ζ-derived peptides

The ECD mass spectra of the $[M+2H]^{2+}$ ions of all five peptides and their corresponding fragment coverages are shown in Figure 3.8, with the corresponding fragment assignment lists available in

Appendix 6-1 to Appendix 6-8. The MS/MS data was collected and the fragment assignment lists compiled by Dr Andrea Lopez-Clavijo. There are notable gaps in the fragment coverage of Pep-01 to Pep-04, while full fragment coverage was observed for Pep-05. The gaps in the fragment coverage suggest the presence of non-covalent interactions between R5-pS8 and R10-pS8 for Pep-01, R10-pS8 for Pep-02, R6-pS8 for Pep-03, and R6-pS8 and R10-pS8 for Pep-04. The presence of low abundance fragments (i.e., an order of magnitude smaller than the next most abundant fragments) in the mass spectra of Pep-02 (c_5, c_6 and c_7), Pep-03 (c_5) and Pep-05 (c_5, c_6 and c_7) may be indicative of the presence of major conformations for these three peptides in which a R5-pS8 non-covalent interaction is present.

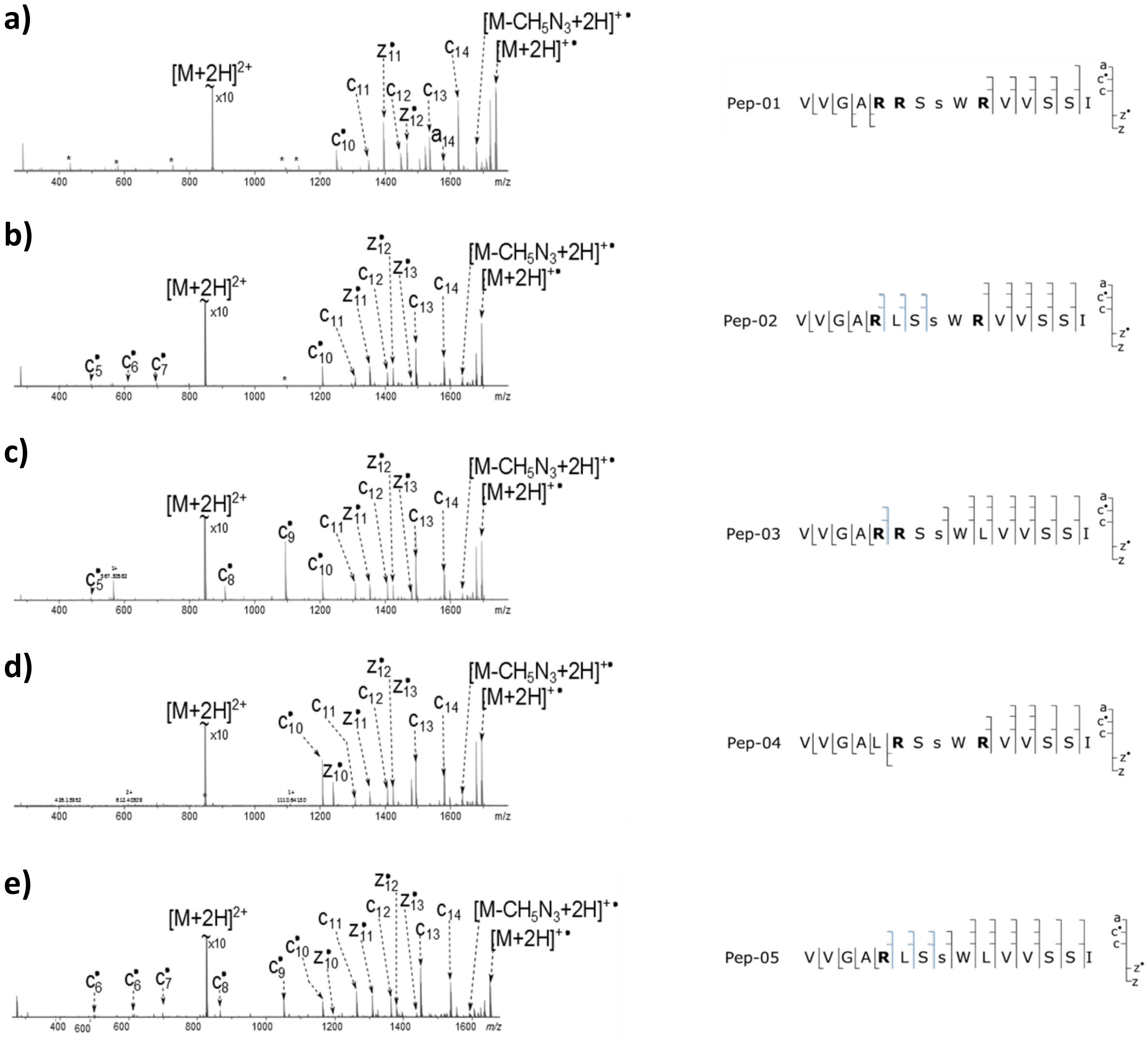


Figure 3.8 ECD MS/MS spectra and corresponding fragmentation patterns for a) Pep-01, b) Pep-02, c) Pep-03, d) Pep-04 and e) Pep-05. These spectra were collected by Dr Andrea Lopez-Clavijo

In order to investigate further the gas-phase structures of these ions, the ion mobility spectrometry-molecular modelling approach developed above was employed. TWIMS arrival time distributions of the doubly-charged ions of Pep-01 to Pep-05 are shown in Figure 3.9 (data acquired by Dr Andrea Lopez-Clavijo), from which their corresponding CCS values (Table 3.4) were calculated by means of calibration against the drift times of a reference sample of 2+ tryptic peptides of known CCS (data acquired by Dr Andrea Lopez-Clavijo, list of reference peptides in Appendix 3-1, and calibration curves in Appendix 4-1).

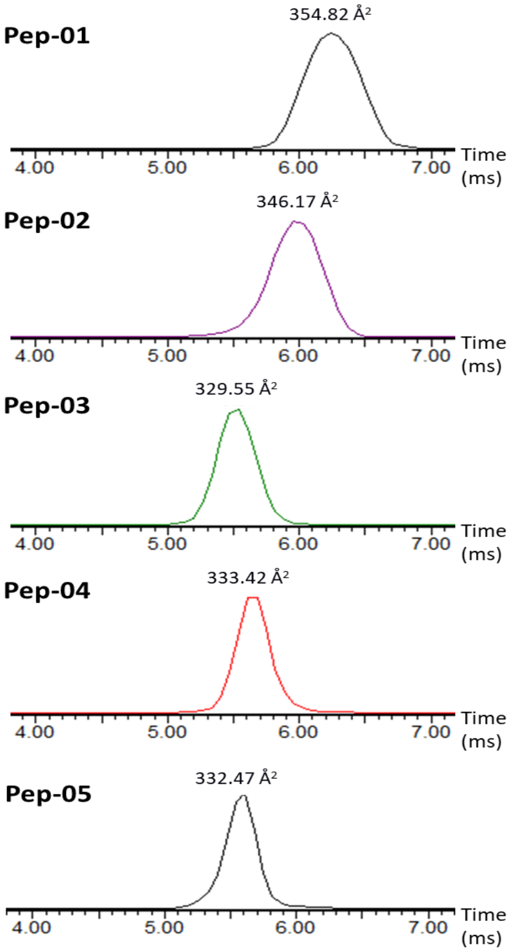


Figure 3.9 TWIMS arrival time distributions for Pep-01 to Pep-05, the CCS at the peak apex is labelled. These ion mobility spectra were collected by Dr Andrea Lopez-Clavijo

Table 3.4 Theoretical monoisotopic masses and experimentally derived CCS for Pep-01 to Pep-05

Abbreviated name	Theoretical monoisotopic mass (Da)	$^{TW}CCS_{N2\rightarrow He}$ (Å^2)
Pep-01	1737.27158	354.82 ± 0.74
Pep-02	1694.25453	346.17 ± 1.05
Pep-03	1694.25453	329.55 ± 0.65
Pep-04	1694.25453	333.42 ± 1.10
Pep-05	1609.19053	332.47 ± 0.49

3.6.2 Results of modelling workflow

It was necessary to model several protonation patterns for each peptide owing to the presence of multiple basic residues (i.e., the N-terminal valine and arginine residues) and the acidic phosphoserine and the necessity of maintaining a net charge of 2+ for all the peptides. The locations of the charges for each protonation pattern are summarised in Table 3.5, in which the basic residues and phosphoserine are labelled ^{0},$^{+}$ or $^{-}$ to denote whether they are neutral, positive or negative respectively. In all cases the C-terminal carboxylic acid is in its protonated, neutral state.

Table 3.5 Protonation patterns modelled for each peptide

Peptide	Model	Sequence
Pep-01	A	V^0 V G A R^+ R^0 S s^0 W R^+ V V S S I
	B	V^+ V G A R^+ R^0 S s^- W R^+ V V S S I
	C	V^+ V G A R^0 R^+ S s^- W R^+ V V S S I
	D	V^0 V G A R^+ R^+ S s^- W R^+ V V S S I
Pep-02	E	V^0 V G A R^+ L^0 S s^0 W R^+ V V S S I
	F	V^+ V G A R^+ L^0 S s^- W R^+ V V S S I
Pep-03	G	V^0 V G A R^+ R^+ S s^0 W L^0 V V S S I
	H	V^+ V G A R^+ R^+ S s^- W L^0 V V S S I
Pep-04	I	V^0 V G A L^0 R^+ S s^0 W R^+ V V S S I
	J	V^+ V G A L^0 R^+ S s^- W R^+ V V S S I
Pep-05	K	V^0 V G A R^+ L^0 S s^0 W L^0 V V S S I

Each of these eleven protonation patterns was modelled by simulated annealing to produce eleven molecular ensembles (n = 25,000) and simulated CCS was calculated for each of those models. By filtering the SA-ensembles by CCS_{expt} ± 5%, a CCS-matched ensemble for each protonation pattern was produced: these ranged in size from 6477 structures to 21386 structures. Dimensionality reduction was performed on each of the CCS-matched ensembles to reduce the computational cost of the onward analysis and to improve the quality of the subsequent clustering. The projections produced by the dimensionality reduction were clustered by k-means (n = 10, Euclidean), a method which should find any clusters in the data, or in the absence of any clusters (i.e., the data existing as a continuum) should segment the data into 10 equally sized portions. By finding the structure closest to the centroid of each cluster, it was possible to analyse the non-covalent interaction patterns across each CCS-matched ensemble (by assuming that the structure closest to the centroid of each cluster is representative of the whole cluster). These representative structures were grouped according to the non-covalent interactions between the phosphoserine and other pertinent residues (Val1, Arg5, Arg6, Arg10), with a non-covalent interaction being defined as occurring when a proton on one of these four residues was less than 4 Å distance from the phosphorus atom. Distances are measured between the phosphorus atom (as it is the centre of the charge on the phosphate group), and the protons on the N-terminus (Val1) and the Arg side chains. It is calculated that a P-H distance of 4 Å would bring the van der Waals radius of the proton within the van der Waals radius of at least one of the phosphate oxygen atoms, with which it may interact, while also allowing for variations in bond angles and bond lengths from their equilibrium values. A non-covalent interaction defined on these terms was considered to be a salt bridge when it involved a negatively-charged phosphorylated residue and a positively-charged residue, while a non-covalent interaction between a neutral phosphorylated residue and a positively-charged residue was considered an ionic hydrogen bond, as described by Kim *et al.* [131] Using this process, it was possible to reduce the complexity of the information in each whole molecular modelling ensemble to 10 representative structures, i.e., one structure for each of the 10 clusters. Details of the 10 clusters for each CCS-matched ensemble, including the percentage of the CCS-matched ensemble that each representative structure represents, can be found in Appendix 1-7. The representative structures themselves can be found in Appendix 1-15, where they are in ascending order of structure number, as in the corresponding file of Appendix 1-7. The information on the non-covalent interactions with pSer across the representative structures for each of the 11 sequences modelled is summarised in Appendix 1-8, in which the proton-phosphorus distances for each of the 10 representative structures for models A-K is shown. For each representative structure, proton-phosphorus distances < 4 Å are highlighted in red, while distances < 4.5 Å are outlined with a red box.

Particular importance is placed on the location and nature of non-covalent interactions between acidic and basic residues within the RRS(pS)WR motif during the analysis of the representative structures produced by the ECD-guiding IMS/MM workflow, i.e., whether they are salt-bridges between two charged residues, or ionic hydrogen bonds between one charged and one neutral residue. As described above, for each peptide, a lack of fragments in the experimental ECD mass spectrum between any arginine and the phosphoserine was interpreted as being indicative of the presence of a non-covalent interaction between these residues.

For Pep-01, models A, B, C and D had no representative structures containing non-covalent interactions consistent with those predicted from the ECD fragmentation pattern (i.e., absence of fragments between Arg5 and Arg10). Structures D2 and D6 from Model D, in which all three arginine residues were protonated and pSer was deprotonated (i.e. [Ser-PO_4H]$^-$), had non-covalent interactions between pSer and Arg5, while protons on Arg10 were just outside the range that was defined as being a non-covalent interaction (at 4.2 Å and 4.1 Å respectively). Although these two structures do not themselves have the non-covalent interactions predicted by the ECD data, they represent clusters that contain many structures in which both Arg5 and Arg10 form non-covalent interactions with pSer. Some 13% of the structures in cluster 2 contain non-covalent interactions consistent with the ECD data, while for cluster 6, 11 % of the structures have those non-covalent interactions. These representative structures cumulatively represent 18% of the model D CCS-matched ensemble (i.e., cluster 2 and cluster 6) and are shown in Figure 3.10. The interactions in these representative structures are characterised as being salt bridges, as the groups interacting are all charged. Model D is the only protonation pattern that produced any representative structures exhibiting non-covalent interactions that were close to being consistent with the ECD data, suggesting that this is the most likely protonation state for Pep-01 of those modelled, i.e., all arginine residues protonated and phosphoserine deprotonated in the gas phase. Alternatively, it is possible that a protonation state not modelled here (e.g., one with a deprotonated C-terminus) might produce representative structures with non-covalent interactions consistent with the ECD data.

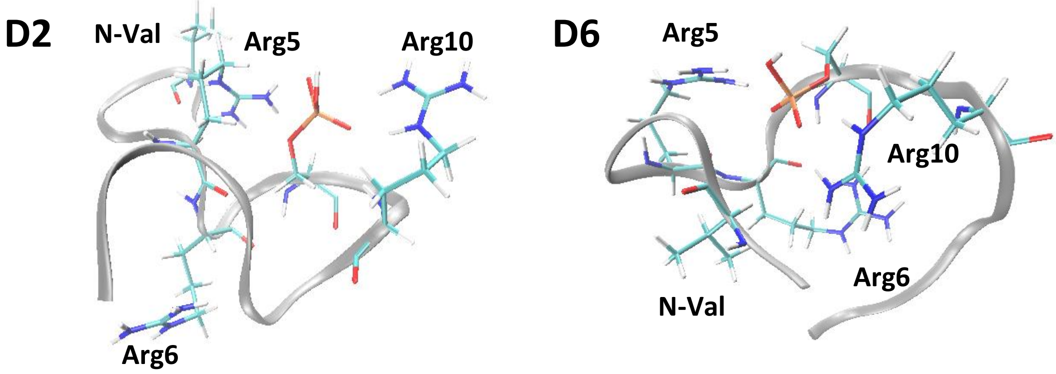


Figure 3.10 Representative structures with non-covalent interactions consistent with ECD fragmentation patterns for the clustered, CCS-matched modelling ensembles of Pep-01. Structures are labelled with their model letter and cluster number. D2 and D6 only have one of the Arg-pSer interactions predicted by the ECD data, the other being too far (> 4 Å) to be considered a non-covalent interaction

For Pep-02, Model F produced a single representative structure with the Arg10-pSer non-covalent interaction consistent with the full ECD fragmentation pattern, including the low abundance $c_5/c_5^•$, $c_6/c_6^•$ and $c_7/c_7^•$ fragments (Figure 3.11). This representative structure (F4) represented 12% of the F CCS-matched ensemble. The non-covalent interactions in model F are between two ionic residues, the negative phosphoserine and positive arginine residues, and so are characterised as salt bridges. The representative structure F4 does not have both the Arg5-pSer and Arg10-pSer non-covalent interactions that would be consistent with the ECD pattern once low-abundance fragments were excluded, however, Arg5 is in close proximity to pSer, with a proton-phosphorus distance of 4.3 Å. It is possible that this represents a weaker Arg5-pSer interaction that hinders the production of fragments in this region, but is not sufficiently strong to prevent fragmentation altogether, resulting in the low abundance fragments observed. Alternatively, it is possible that the cluster that F4 represents contains structures with both Arg5 and Arg10 interacting with pSer, and that these conformers are responsible for the low abundance of the $c_5/c_5^\bullet$, $c_6/c_6^\bullet$ and $c_7/c_7^\bullet$ fragments.

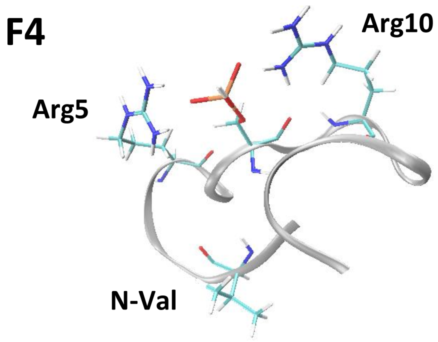


Figure 3.11 Representative structure F4 with non-covalent interactions consistent with ECD fragmentation patterns for the clustered, CCS-matched modelling ensembles of Pep-02. This structure is labelled with its model letter and cluster number

Model G produced one representative structure whose non-covalent interactions were consistent with the full fragmentation pattern of Pep-03 (Figure 3.12). This representative structure (G4) contains an interaction between Arg6 and pSer, and could be the conformer from which the c_5 fragment is generated. G4 represents 7% of the model G CCS-matched ensemble. Once the low abundance c_5 fragment is excluded, three further representative structures from model G and three representative structures from model H have the non-covalent interaction Arg5-pSer that is consistent with the ECD data. The three model G representative structures (G2, G7 and G5) all exhibit only the Arg5-pSer interaction and together represent 31% of the model G CCS-matched ensemble. The three representative structures from model H, representing 33% of the H CCS-matched ensemble, also had only the Arg5-pSer interaction, although the representative structure H10 also had protons from the N-terminal valine and Arg6 in close proximity with pSer.). The model G4 is proposed as the structure of the minor conformation from which the c_5 fragment arises, while all of the representative structures G2, G7, G5, H2, H10 and H7 are suitable candidates for the structure of the major conformation. Both models G and H produce several representative structures consistent with the proposed fragmentation pattern of the major conformer, suggesting that the two protonation patterns coexist in the gas phase. It should be noted that many of the representative structures produced from models G and H are very similar: G2, G7 and G5 are all structurally similar, as are H2 and H7. A lack of diversity in the representative structures may indicate that the G and H ensembles contain a narrower range of conformers to select from, which is consistent with the observed arrival time distribution for Pep-03, which is narrow relative to the other peptides (Figure 3.9)

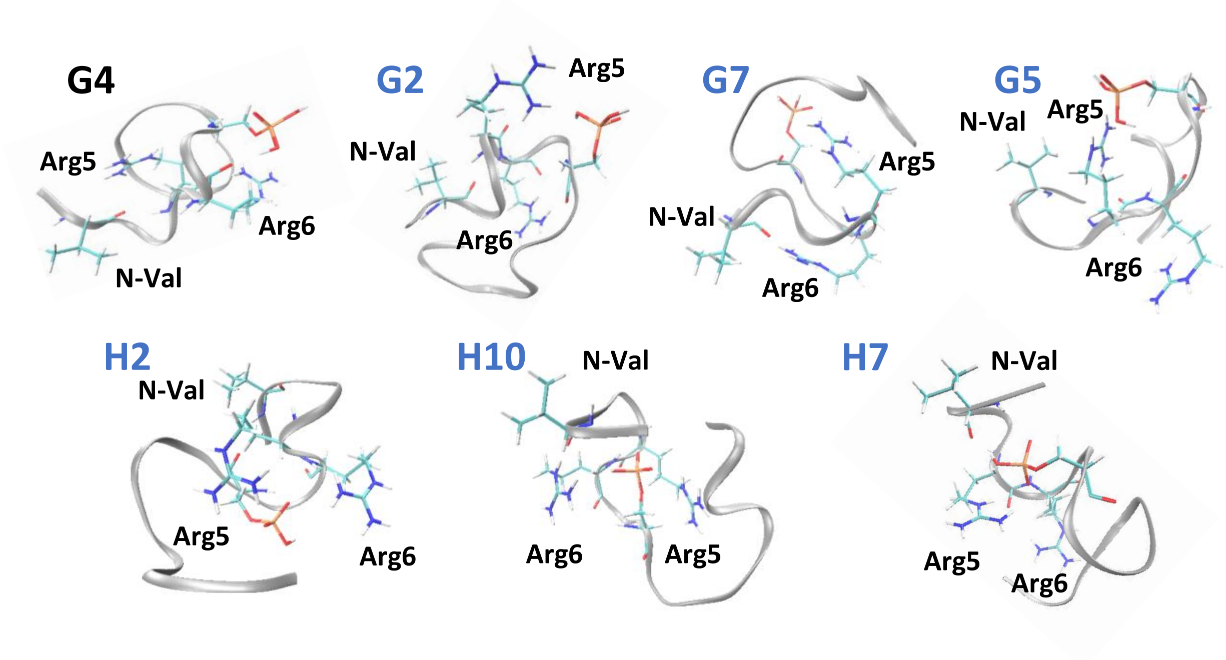


Figure 3.12 Representative structures with non-covalent interactions consistent with ECD fragmentation patterns for the clustered, CCS-matched modelling ensembles of Pep-03. Structures are labelled with their model letter and cluster number, those structures labelled in blue have non-covalent interactions consistent with the ECD data once low abundance fragments have been excluded

For Pep-04, no representative structures from either model J or I were consistent with the ECD data, in which no fragments were observed between Arg6 and Arg10, indicating Arg6-pSer and Arg10-pSer non-covalent interactions. For model J, the representative structure of cluster 2 has a salt bridge between Arg6 and pSer and at least one proton on Arg10 within 4.5 Å. While a distance of 4.5 Å is too great to be considered a non-covalent interaction, it is possible that the cluster that this structure represents contains other structures that do have the Arg6-pSer and Arg10-pSer non-covalent interactions that are consistent with the ECD. Examining the whole model J CCS-matched ensemble, 649 of the 14,442 (4%) structures have a proton on both Arg6 and Arg10 within 4 Å of the pSer phosphorus atom, while the protons on the N-terminal valine residue are further than 4 Å from the pSer phosphorus atom i.e., the salt bridges are consistent with the ECD. Cluster 2 contains 100 of these ECD-consistent structures, therefore its representative structure is the most appropriate of those produced and has been selected as the candidate structure for model J (Figure 3.13).

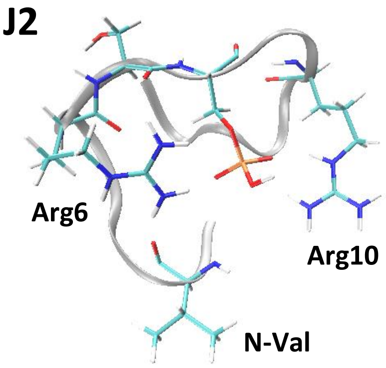


Figure 3.13 Representative structure J2 produce from the clustered, CCS-matched modelling ensembles of Pep-04. This structure is labelled with its model letter and cluster number. J2 only has only one of the Arg-pSer interactions predicted by the ECD data, the other being too far (> 4 Å) to be considered a non-covalent interaction

For Pep-05 (which contains a single arginine residue), only one sequence (model K) was modelled as there was only one likely combination of charges that produced a net charge of 2+. One representative structure (Figure 3.14) was produced that was consistent with the ECD MS/MS spectrum after the exclusion of low abundance fragments, representing 12% of the ensemble. Eight other representative structures (representing 84% of the K CCS-matched ensemble) had no non-covalent interactions between pertinent residues. That is consistent with the full ECD pattern with low-abundance fragments included, in which full fragment coverage was observed and so would not predict the presence of any non-covalent interactions.

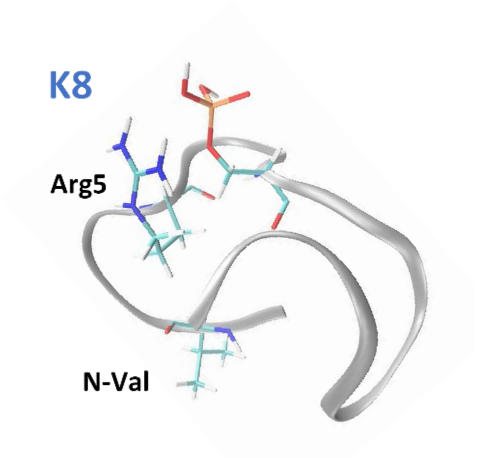


Figure 3.14 Representative structure K8 with non-covalent interactions consistent with ECD fragmentation patterns for the clustered, CCS-matched modelling ensembles of Pep-05. This structure is labelled with its model letter and cluster number, and is labelled in blue as it has non-covalent interactions consistent with the ECD data once low abundance fragments have been excluded

3.7 Conclusions

This chapter has discussed the development of an ECD-guided IMS/MM workflow. This work included the development of forcefield parameters for residues commonly found in mass spectrometry conditions: neutral C-termini, neutral N-termini, neutral arginine residues, neutral phosphorylated residues and residues adjacent to the cleavage site in c^+ and $z^\bullet$ fragments. MM methods for future simulations were established as a three-tiered SA approach, using the ff99SB and phos10aa forcefields, along with any additional forcefield parameters from the sets made to simulate peptides under mass spectrometry conditions. In order to compare the structures produced by MM to IMS data, ion mobility/CCS calculators were considered. IMoS was selected as the most appropriate ion mobility/CCS calculator following evaluation of its performance, which showed that it performs calculations sufficiently fast and with low enough errors to be suitable for calculating the CCS for each of the structures of the 25,000-strong ensembles produced from the three-tiered SA experiments. After mobility calculation, structures that had CCS_{sim} outside of CCS_{expt} ± 5% were excluded from the MM ensemble, followed by dimensionality reduction of this CCS-matched ensemble by random projection and clustering by k-means to produce 10 representative structures. These representative structures may then be compared to experimental ECD and IMS data, as demonstrated by a structural study the doubly-charged ions of five peptides derived from 14-3-3-ζ, for which regions of the peptide sequence with no ECD fragment coverage allowed the locations of non-covalent interactions to be predicted.

For Pep-02 where the ECD mass spectrum indicated the presence of conformers with two patterns of non-covalent interactions, the IMS/MM workflow produced a single candidate structure from the model F protonation pattern, this structure had non-covalent interactions consistent with the full ECD fragmentation pattern, though the Arg5 residue was sufficiently near the pSer residue to suggest that it may sometimes interact, leading to the reduced abundance of the $c_5/c_5^\bullet$, $c_6/c_6^\bullet$ and $c_7/c_7^\bullet$ fragments. One representative structure from model G was produced with non-covalent interactions consistent with the ECD data of Pep-03, however, with low abundance fragments excluded, six further representative structures had ECD-consistent non-covalent interactions. These seven representative structures could all serve as candidate structures for Pep-03. No representative structures generated for either Pep-01 or Pep-04 had non-covalent interactions that exactly matched those predicted from the ECD data, however two structures from model D and one structure from model J had one of the predicted Arg-pSer interactions and the other Arg just outside of interaction distance (<4.5 Å). Closer inspection of the associated clusters showed that they contained many structures that did have non-covalent interactions consistent with the ECD data, making them the most suitable representative structure to propose as candidates. For Pep-05, one structure was proposed as a candidate for the conformation containing the Arg5-pSer non-covalent interaction as indicated by the ECD fragmentation pattern with low abundance fragments removed. With low abundance fragments included 8 possible candidate structures were consistent with the predicted lack of non-covalent interactions between pS8 and positively-charged residues.

This structural study of 14-3-3-ζ-derived phosphopeptides successfully predicted structures for each of the peptides investigated, providing validation that the ECD-guided IMS/MM workflow developed here can plausibly be used to study the structure of phosphopeptides in the gas phase. A more thorough structural study of these peptides, including data from a range of biophysical techniques, would be necessary to lend certainty to the results of the workflow presented here, and determine how similar these structures are to solution or solid-phase structures of the same peptides.

Chapter 4: Structural study of GSK3β-derived peptides utilising an electron capture dissociation-guided ion mobility spectrometry/molecular modelling workflow to determine the effect of specific amino acid residues on gas-phase peptide structure

4.1 Overview

Having defined and validated a workflow in Chapter 3, which combined electron capture dissociation (ECD) with ion mobility spectrometry (IMS) and molecular modelling (MM) to produce candidate structures for a given peptide, the next steps were to apply this workflow to more complicated systems. This chapter describes a series of 18-mer peptides derived from the protein GSK3β, which were subjected to the workflow to further probe the effects of phosphorylation on peptide structure and their ECD behaviour. These peptides contained multiple acidic and basic residues, so had many possible combinations of non-covalent interactions that might affect their ECD behaviour. Additionally, substitution of various residues in the wild-type sequence allowed investigation into the structural and ECD-influencing effects of the identity of the phosphorylated residue (pTyr → pSer), the presence of an alkylated cysteine residue (Cys-alk → Ala) and the presence of glutamate (Glu → Ala). In each case, the unphosphorylated analogue of the peptide was also studied to gain a broader understanding of phosphorylation in the gas phase.

4.2 Selection of model system

In order to probe further the gas-phase consequences of phosphorylation, it was necessary to select a model system. The phosphoprotein GSK3β (MW = 46,744 Da, UniProt ascension number: P49841) was selected as it has five possible phosphorylation sites, two of which are represented in x-ray crystal structures. The biomedical relevance of GSK3β lies in its phosphorylation of tau protein, which leads to the formation of neurofibrillary bundles that are strongly associated with Alzheimer's disease.[26] GSK3β is also implicated in the development of insulin resistance, which leads to diabetes.[27] One of the phosphorylation sites of GSK3β is at the Y216 residue, which is located in the activation loop, and its phosphorylation leads to the activation of the protein.[29] Crystallographic data on phosphorylated and unphosphorylated GSK3β suggest that the phosphorylation at Y216 results in a 120° rotation of the Y216 side chain, and a repositioning of the R220 and R223 side chains to form non-covalent interactions between these residues and pY216.[190] The rotation of the Y216 between the phosphorylated and non-phosphorylated analogues of this protein are shown in Figure 4.1, reproduced from Dajani *et al.* 2003.

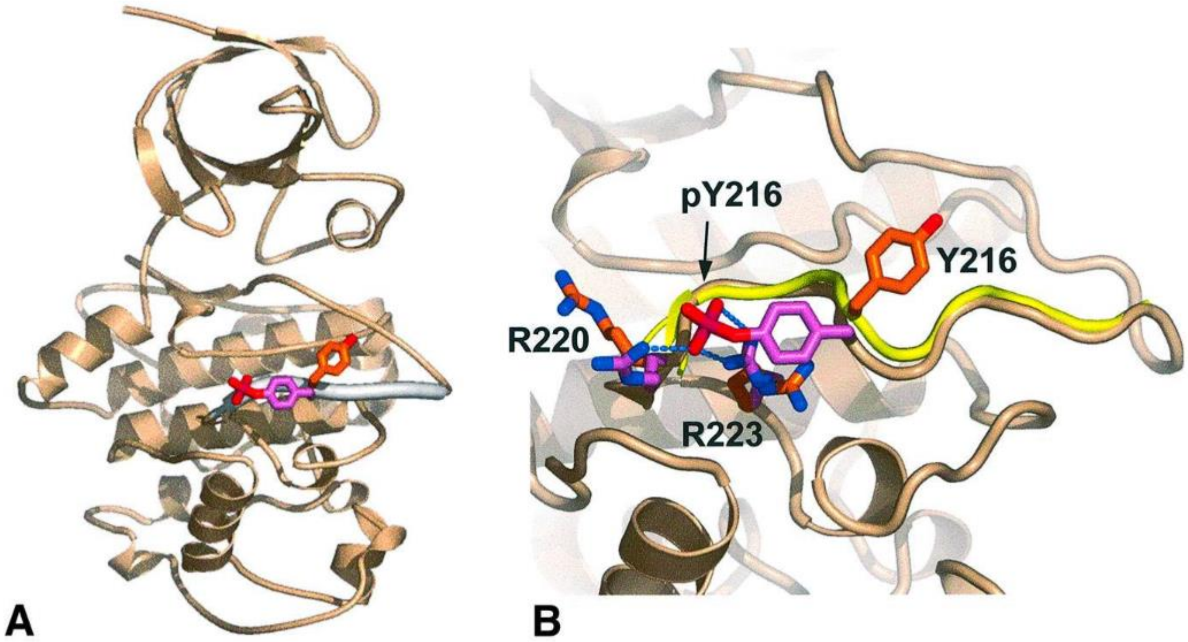


*Figure 4.1 A) Crystal structures of human GSK3β showing the difference in orientation between Y216 when phosphorylated and unphosphorylated, and B) a close up of the region around Y216, showing the non-covalent interactions that it forms with R220 and R223 when phosphorylated. Image reproduced with permission from Dajani, R.; Fraser, E.; Roe, S. M.; Yeo, M.; Good, V. M.; Thompson, V.; Dale, T. C.; Pearl, L. H., Structural Basis for Recruitment of Glycogen Synthase Kinase 3beta to the Axin-Apc Scaffold Complex. The EMBO journal **2003,** 22 (3), 494-501*

An 18-mer phosphopeptide was derived from the GSK3β sequence around the phosphorylated residue at Y216 (QLVRGEPNVSpYICSRYYR). This peptide has been observed as a tryptic peptide following the digestion of GSK3β.[191] The sequence of the derived peptide and its context in the full GSK3β sequence is shown in Figure 4.2a, and the region of the GSK3β protein that contains these 18 residues is shown in Figure 4.2b for the crystal structure with Y216 phosphorylated and Figure 4.2c for Y216 unphosphorylated. The region of the protein included in the derived phosphopeptide contains both R220 and R223 residues that interact with the phosphorylated Y216; these arginine residues are in positions 15 and 18, respectively, of the derived phosphopeptides, with the phosphotyrosine in position 11. The derived phosphopeptide also includes two further acidic and basic residues, R209 and E211, which are the fourth and sixth residues of the derived phosphopeptide. In the phosphorylated GSK3β crystal structure, E211 is involved in interactions with a water molecule, while R209 is projected away from the bulk of the protein, towards the adjacent GSK3β molecule. In some crystal structures of GSK3β, the side chain of R209 is omitted, suggesting that it has too much conformational flexibility to be resolved by x-ray crystallography.[192-196] This makes it difficult to infer whether R209 in the solution phase forms non-covalent interactions with other residues in GSK3β, whether it is solvated entirely by water, or both. Notably, in one of these crystallography studies where R209 is omitted, the side chain of E211 is also omitted,[196] suggesting that it too may adopt several conformations. The 18-mer derived phosphopeptide is limited in that it cannot probe interactions with residues contained in the rest of the GSK3β protein, and the peptide will have many more degrees of conformational freedom compared with the same residues in the context of the full protein. These factors mean that the interactions and structure of the derived phosphopeptide may not be the same as those in the same region of the peptide in the GSK3β protein, although it may be possible to make inferences between the structure of the derived phosphopeptide and the full protein.

a)

MSGRPRTTSFAESCKPVQQPSAFGSMKVSRDKDGSKVTTVVATPGQGPDRPQEVSYTDTK
VIGNGSFGVVYQAKLCDSGELVAIKKVLQDKRFKNRELQIMRKLDHCNIVRLRYFFYSSG
EKKDEVYLNLVLDYVPETVYRVARHYSRAKQTLPVIYVKLYMYQLFRSLAYIHSFGICHR
DIKPQNLLLDPDTAVLKLCDFGSAKQLVRGEPNVSYICSRYYRAPELIFGATDYTSSIDV
WSAGCVLAELLLGQPIFPGDSGVDQLVEIIKVLGTPTREQIREMNPNYTEFKFPQIKAHP
WTKVFRPRTPPEAIALCSRLLEYTPTARLTPLEACAHSFFDELRDPNVKLPNGRDTPALF
NFTTQELSSNPPLATILIPPHARIQAAASTPTNATAASDANTGDRGQTNNAASASASNST

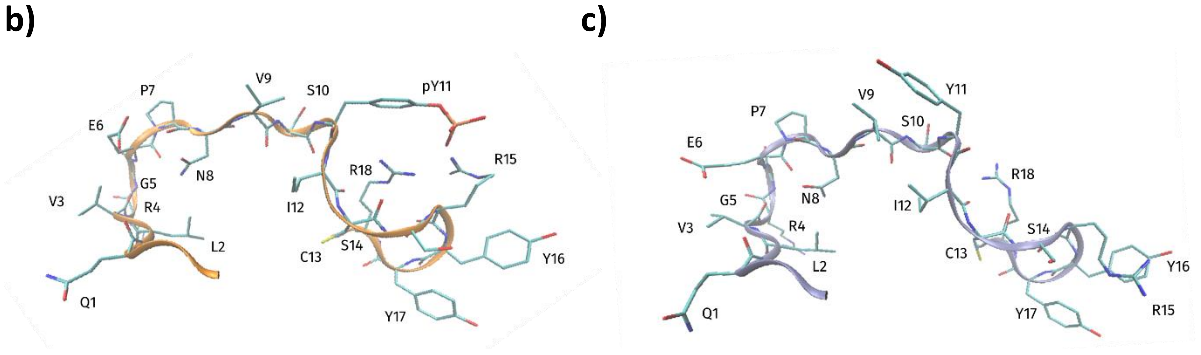


Figure 4.2 a) The sequence of the 18-mer peptide derived from GSK3β in the context of the full human protein sequence, b) the region of the crystal structure corresponding to the derived phosphopeptide with residues labelled by the numbering scheme of the peptide, and c) the unphosphorylated analogue

Initial mass spectrometry (MS) experiments (methods as in Section 2.2.2) on the peptide QLVRGEPNVSpYICSRYYR (MW = 2283.07 Da) showed that it dimerised readily. In the mass spectrum of QLVRGEPNVSpYICSRYYR the $[M+2H]^{2+}$ peak was observed to be underlaid by the $[2M+2H]^{4+}$ peak produced from dimerisation of the peptide (Figure 4.3). The monoisotopic mass of the $[2M+2H]^{4+}$ dimer (m/z = 1141.5507) is 2 Da less than that of double the $[M+2H]^{2+}$ monomer (m/z = 1142.0526), indicating that two hydrogens have been lost in the dimerisation process, which is characteristic of dimerisation by S-S bond formation between the cysteine residues. Given that this peptide is derived from the activation loop region of the full protein, it is likely to have a high affinity for binding other peptides and proteins, however, none of the crystal structures published to date show dimerisation by S-S bond formation at C218.

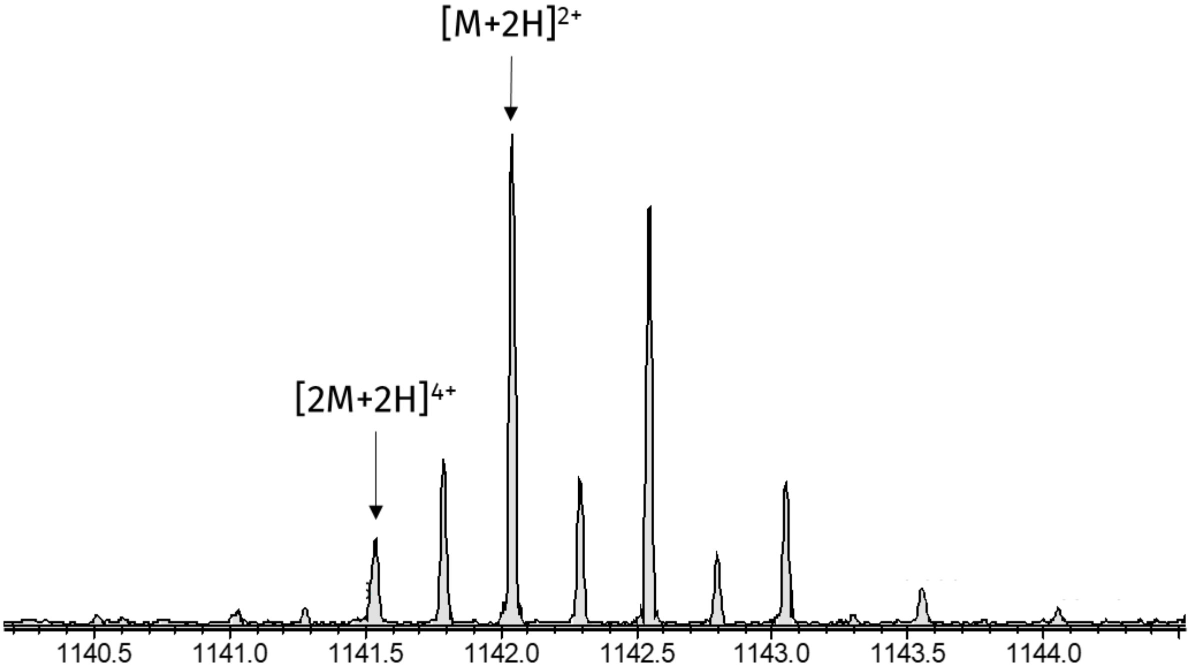


Figure 4.3 The [M+2H]$^{2+}$peak of QLVRGEPNVSpYICSRYYR, underlaid by the [2M+2H]$^{4+}$ peak of its dimer with the monoisotopic peaks of both species labelled: [2M+2H]$^{4+}$ m/z = 1141.5507, [M+2H]$^{2+}$ m/z = 1142.0526

In order to prevent dimerisation at the cysteine residue, this residue was capped with an acetamide group. This alkylated cysteine residue (alkC) does not form disulfide bridges as the acetamide group is covalently bound to the cysteine sulfur atom and sterically hinders the approach of thiol groups of other cysteine residues. Additionally, the peptide was capped at the N and C-termini with acetyl and amide groups respectively to prevent the termini existing in their ionic forms, which the Q1 and R18 residues would not be able to do in the context of the full GSK3β protein as they would be surrounded by other residues. This capped peptide (Ac-QLVRGEPNVS(pY)I(alkC)SRYYR-Am, MW = 2381.12 Da) also dimerised, as shown by the [2M+4H]$^{4+}$ peak seen alongside the [M+2H]$^{2+}$ peak in the mass spectrum (Figure 4.4a), however the monoisotopic peaks of these two species are isobaric (m/z = 1191.0605), indicating that this dimerisation process did not result in the loss of two hydrogens, and so is not caused by dimerisation by disulfide bridge formation at the cysteine residue. Application of in-source collision induced dissociation (CID) (collision energy = 90 V), resulted in the dissociation of this dimer and peaks associated with it were no longer observed (Figure 4.4b). That dissociation of this dimer can be induced in this way implies that the monomer units are associating via non-covalent interactions that are readily broken by CID. Given that the intent was to study the non-covalent interactions in the monomer of this peptide, it was not desirable to use in-source CID to remove the dimer peaks in future experiments as this could break apart the non-covalent interactions of interest in the monomer. Though, ECD products from the $[2M+4H]^{4+}$ species were likely to be higher charge than those produced from the $[M+2H]^{2+}$ species, which could only be singly-charged following the charge reduction caused by an ECD event. Therefore, singly-charged ECD products of the $[M+2H]^{2+}$/$[2M+4H]^{4+}$ mixture could be assumed to be formed from the monomer, while any doubly- or triply-charged ECD products must be formed from the dimer. Given that peaks associated with the dimer species were ~10% of the intensity of peaks associated with the monomer (based on peak heights), it was expected that fragments formed from the dimer would be similarly low abundance compared to fragments formed from the monomer, which could help to assign whether singly-charged peaks originated from the monomer or dimer.

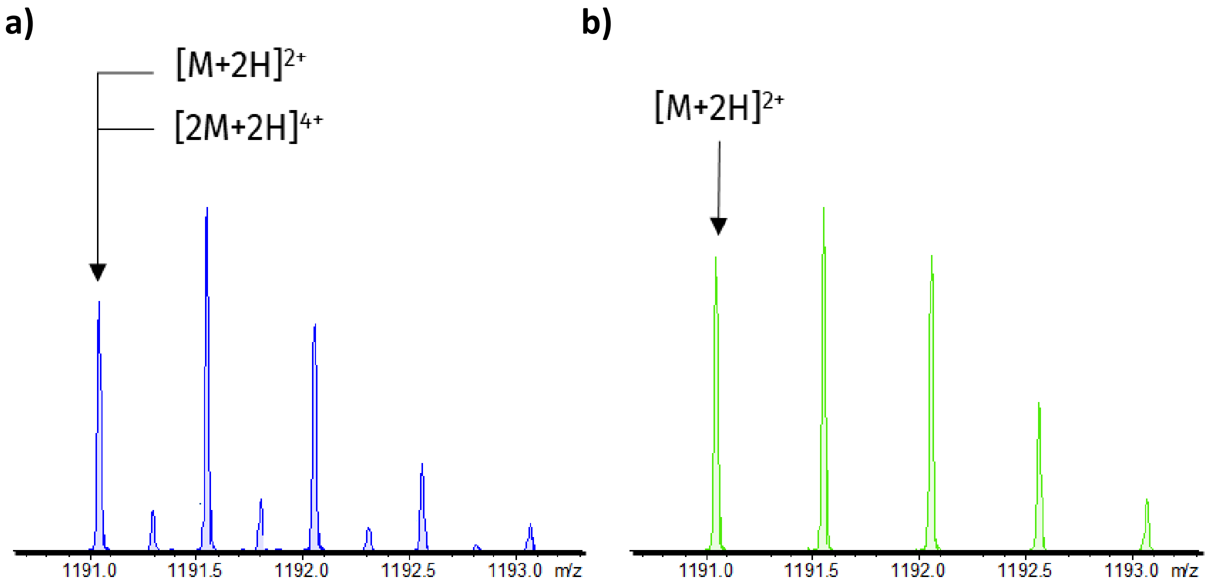


Figure 4.4 Mass spectrum of the capped peptide Ac-QLVRGEPNVS(pY)I(alkC)SRYYR-Am, showing the $[M+2H]^{2+}$peak a) with the $[2M+4H]^{4+}$underlaid (monoisotopic peak m/z = 1191.0605) and b) without the $[2M+4H]^{4+}$ species upon the application of in-source CID (90 V) with monoisotopic peak at m/z = 1191.0622. Note: the monoisotopic peaks of the dimer and monomer are assigned as being isobaric

These initial experiments suggest that the capped peptide derived from the sequence of GSK3β around its pY216 residue is an appropriate model system in which to investigate the ECD behaviour of phosphopeptides. The sequence of the first model peptide discussed in this Chapter is therefore Ac-

QLVRGEPNVS(pY)I(alkC)SRYYR-Am and is referred to as GSK3β-pY11. Several further variants of this peptide were investigated to probe the effect of substitutions of specific residues that seemed to be having a specific impact of the ECD behaviour of the peptide (pTyr→pSer, alkCys→Ala and Glu→Ala). In each case, the unphosphorylated analogue of the peptide was also studied. The sequences of all the peptides discussed in this Chapter are displayed in Table 4.1.

Table 4.1 The designated names, monoisotopic masses and sequences of the peptides derived from GSK3β, where pX indicates that residue 'X' is phosphorylated and alkC represents cysteine-S-acetamide.

Abbreviated name	Monoisotopic mass (Da)	Sequence
GSK3β-pY11	2381.12	Ac-Q L V R G E P N V S pY I alkC S R Y Y R-Am
GSK3β-Y11	2301.16	Ac-Q L V R G E P N V S Y I alkC S R Y Y R-Am
GSK3β-pS11	2305.09	Ac-Q L V R G E P N V S pS I alkC S R Y Y R-Am
GSK3β-S11	2225.12	Ac-Q L V R G E P N V S S I alkC S R Y Y R-Am
GSK3β-pY11-noC	2292.13	Ac-Q L V R G E P N V S pY I A S R Y Y R-Am
GSK3β-Y11-noC	2212.16	Ac-Q L V R G E P N V S Y I A S R Y Y R-Am
GSK3β-pY11-noEnoC	2234.12	Ac-Q L V R G A P N V S pY I A S R Y Y R-Am
GSK3β-Y11-noEnoC	2154.16	Ac-Q L V R G A P N V S Y I A S R Y Y R-Am

4.3 ECD of GSK3β-derived peptide ions

Upon performing ECD fragmentation of the GSK3β-pY11 peptide (Figure 4.5a, methods as discussed in Section 2.2.2, fragment assignment list found in Appendix 6-9) there was a notable lack of fragments observed between pY11 and Y16, however, there were also a large number of 2+ and 3+ peaks that could only be assigned to the dimer, as well as 1+ peaks that may be formed from either the monomer or the dimer. Several singly-charged species in this spectrum were much higher abundance than the rest, namely $c_{17}^{\bullet+}$, $z_{17}^{\bullet+}$, $z_{16}^{\bullet+}$ and $z_{15}^{\bullet+}$, so these were most likely to be formed from fragmentation of the monomer. None of these four singly-charged peaks were higher abundance than the peak for the $[2M+4H]^{3\bullet+}$ charge-reduced species of the dimer, so there was not a distinct difference in abundance between peaks produced from fragmentation of the dimer as opposed to the monomer, making it difficult to assign from which species the singly-charged fragments might have formed. ECD was also performed on the peptide after applying the in-source CID (90 V) that was known to dissociate the dimer (Figure 4.5b, methods described in Section 2.2.2, fragment assignment list found in Appendix 6-

9). In this spectrum, no fragments were observed between R4 and Y17, which is indicative of non-covalent interaction between R4, R15 and an acidic residue between them. The fragmentation pattern for this peptide with in-source CID contains no additional fragments compared to the fragmentation pattern of the peptide without in-source CID, which suggests that the energy of the in-source CID was sufficient to dissociate the dimer but did not disrupt non-covalent interactions within the monomer. The fragments present with the application of in-source CID are the same as the high abundance singly-charged fragments observed without in-source CID ($c_{17}^{\bullet+}$, $z_{17}^{\bullet+}$, $z_{16}^{\bullet+}$ and $z_{15}^{\bullet+}$), which suggests that these fragments are associated with the monomer, while other singly-charged fragments present in the spectrum without in-source CID are produced from the dimer. The fragmentation patterns for the dimeric and monomeric species are shown in Figure 4.5c and Figure 4.5d, respectively. For the dimeric species, all fragments are depicted on one unit of the dimer only, with one possible interface between the two units shaded in grey. Appendix 7 contains several further depictions of the dimer that would be consistent with its ECD behaviour.

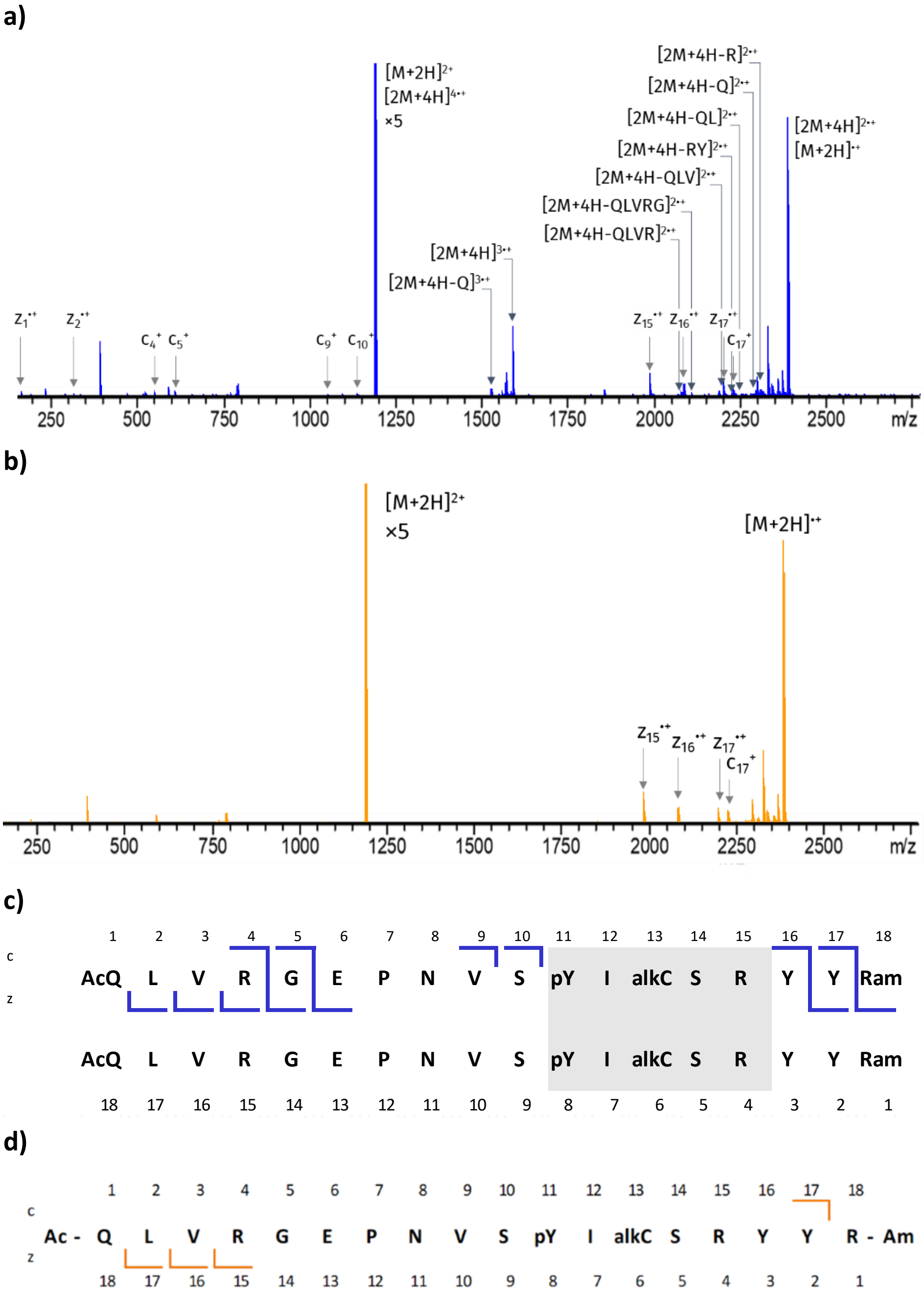


Figure 4.5 ECD MS/MS spectrum of the GSK3B-derived peptide $[M+2H]^{2+}$ ion a) without in-source CID, showing peaks formed from dissociation of the dimeric $[2M+4H]^{4+}$ and b) with 90 V in-source CID applied, showing only fragments from the monomer. The fragmentation patterns are shown for c) the dimeric species and d) the monomer. Note: the fragmentation pattern for the dimeric species assumes that all fragments in the ECD MS/MS spectrum without in-source CID are formed from the dimer except $c_{17}^{\bullet+}$, $z_{17}^{\bullet+}$, $z_{16}^{\bullet+}$ and $z_{15}^{\bullet+}$, and shows one possible interface region between the two peptide units in grey

ECD experiments were also performed on the $[M+2H]^{2+}$ ion of the non-phosphorylated analogue of this peptide (GSK3β-Y11, Ac-QLVRGEPNVSYI(Calk)SRYYR-Am), with the ECD fragmentation pattern showing a lack of fragments between R4 and Y16 (Figure 4.6). The full fragment assignment list is found in Appendix 6-10. This result suggests that the lack of fragment coverage seen for GSK3β-pY11 cannot be assigned to salt bridges with the phosphate group alone, as the GSK3β-Y11 peptide has no phosphate group, but a similar fragmentation pattern.

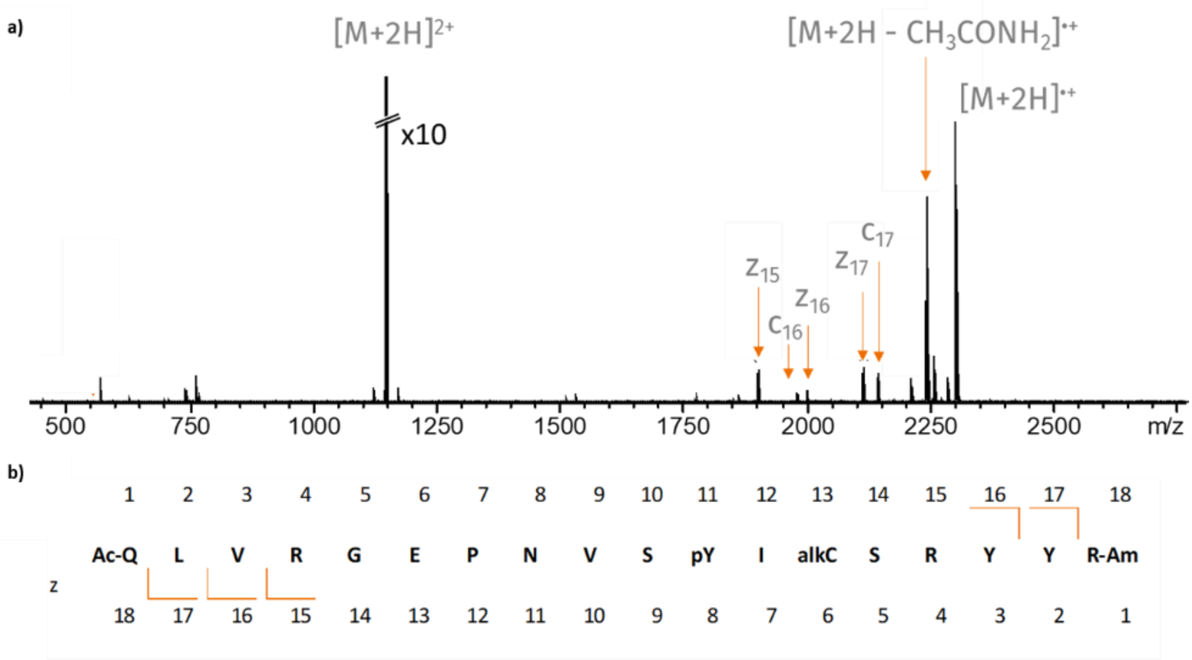


Figure 4.6 a) ECD MS/MS spectrum of GSK3β-Y11 and b) the corresponding fragmentation pattern

Despite the lack of a phosphorylated residue in GSK3β-Y11, there are several other interactions that might be present and could account for the number of missed cleavages observed in the fragmentation pattern. Other acidic residues in the sequence may be capable of forming acid-base non-covalent interactions with protonated arginine residues that may affect ECD behaviour (i.e., glutamate or deprotonated arginine residues), and the π system of the tyrosine residues may be capable of forming π-cation interactions with protonated arginine residues or π-π interactions between the tyrosine residues.[197, 198] As the gap in the fragment coverage for GSK3β-pY11 extends to Y17 and for GSK3β-Y11 extends to Y16, it was necessary to investigate whether the tyrosine residues were forming interactions that could prevent the dissociation of ECD fragments. In order to determine whether this effect was due to π-cation or π-π interactions with the Y11 residue, ECD experiments were performed on the GSK3β-pS11 and GSK3β-S11 analogues (Ac-QLVRGEPNVS(pS)I(Calk)SRYYR-Am and Ac-

QLVRGEPNVSSI(Calk)SRYYR-Am, respectively), in which the pY11/Y11 residue was substituted for pS11/S11. Details of these experiments are found in Section 2.2.2, while peak assignment lists are found in Appendix 6-11 for GSK3β-pS11 and Appendix 6-12 for GSK3β-S11. Both the GSK3β-pS11 and GSK3β-S11 peptides had identical ECD fragment coverage (Figure 4.7), with no fragments observed between R4 and R15. That is, both GSK3β-pS11 and GSK3β-S11 had increased fragment coverage relative to that observed for the GSK3β-pY11 and GSK3β-Y11 analogues, which lacked fragments between R4 and Y17, and R4 and Y16 respectively. The fragment coverage observed for GSK3β-pS11 and GSK3β-S11 suggests that tyrosine residues (Y16 and Y17) are not involved in non-covalent interactions in the GSK3β-pY11 and GSK3β-Y11 peptides and it was necessary to investigate other possibilities to explain why the non-phosphorylated GSK3β-S11 had the same fragment coverage as its phosphorylated counterpart GSK3β-pS11.

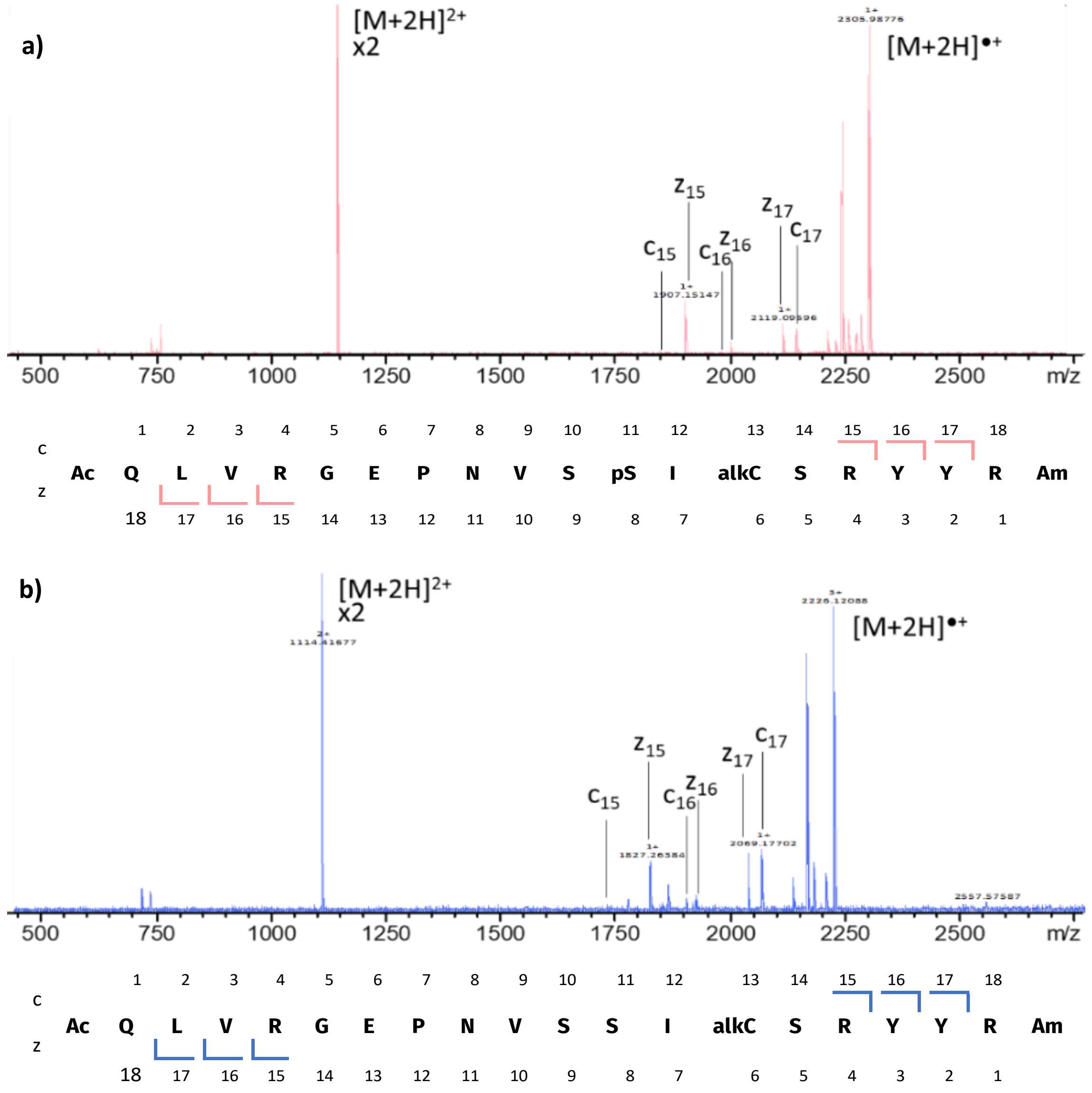


Figure 4.7 The ECD MS/MS spectra and fragmentation patterns of a) GSK3β-pS11 and b) GSK3β-S11

In order to understand why the non-phosphorylated GSK3β-Y11 and GSK3β-S11 peptides have significant gaps in their fragment coverage, it was decided to perform activated-ion ECD (AI-ECD) experiments on the GSK3β-pY11, GSK3β-Y11 and GSK3β-pS11 peptides (data for GSK3β-S11 was not collected). In these experiments, ions were activated using an infrared (IR) laser either before or after ECD. The activation should deliver sufficient thermal energy to unfold the peptides by breaking all non-covalent interactions that they contain, whether they involve a phosphorylated residue or not.[92] Any increase in fragment coverage during AI-ECD experiments would indicate that non-covalent interactions were the cause of the reduced fragment coverage observed for the peptides investigated so far. If no increase in fragment coverage was observed between ECD and AI-ECD, then it could reasonably be assumed that low fragment coverages were observed due to inefficient electron capture or cleavage being biased towards specific bonds at the expense of other possible cleavage sites.

IR activation was performed both before and after ECD. In IR-ECD, a laser pulse before the ECD is intended to unfold the analyte ion and remove non-covalent interactions that might otherwise prevent the dissociation of fragments; and ECD-IR, in which a laser pulse after the ECD should induce dissociation of any fragments held together by non-covalent interactions, allowing them to be detected separately. The difference between the two approaches is that the $[z\text{-}c]^{\bullet+}$ complex (which without activation would be detected as $[M+2H]^{\bullet+}$) is prevented from forming in IR-ECD, while in ECD-IR it can form but is subsequently broken apart. Therefore, any differences between the IR-ECD and ECD-IR spectra are likely due to processes that occur within the $[z\text{-}c]^{\bullet+}$ complex, for example the abstraction of a hydrogen from the c^{+} species to the $z^{\bullet+}$ species to form $c^{\bullet+}$ and $z^{\bullet}$.[102]

ECD (without IR activation) was also collected as the nESI and instrument configuration were different to previous experiments: nESI was achieved by using pulled glass capillaries and the instrument used was a 12 T solariX FT-ICR MS equipped with a 10.6 µm, 25 W CO_2 laser. Further details of these experiments can be found in Section 2.2.3.

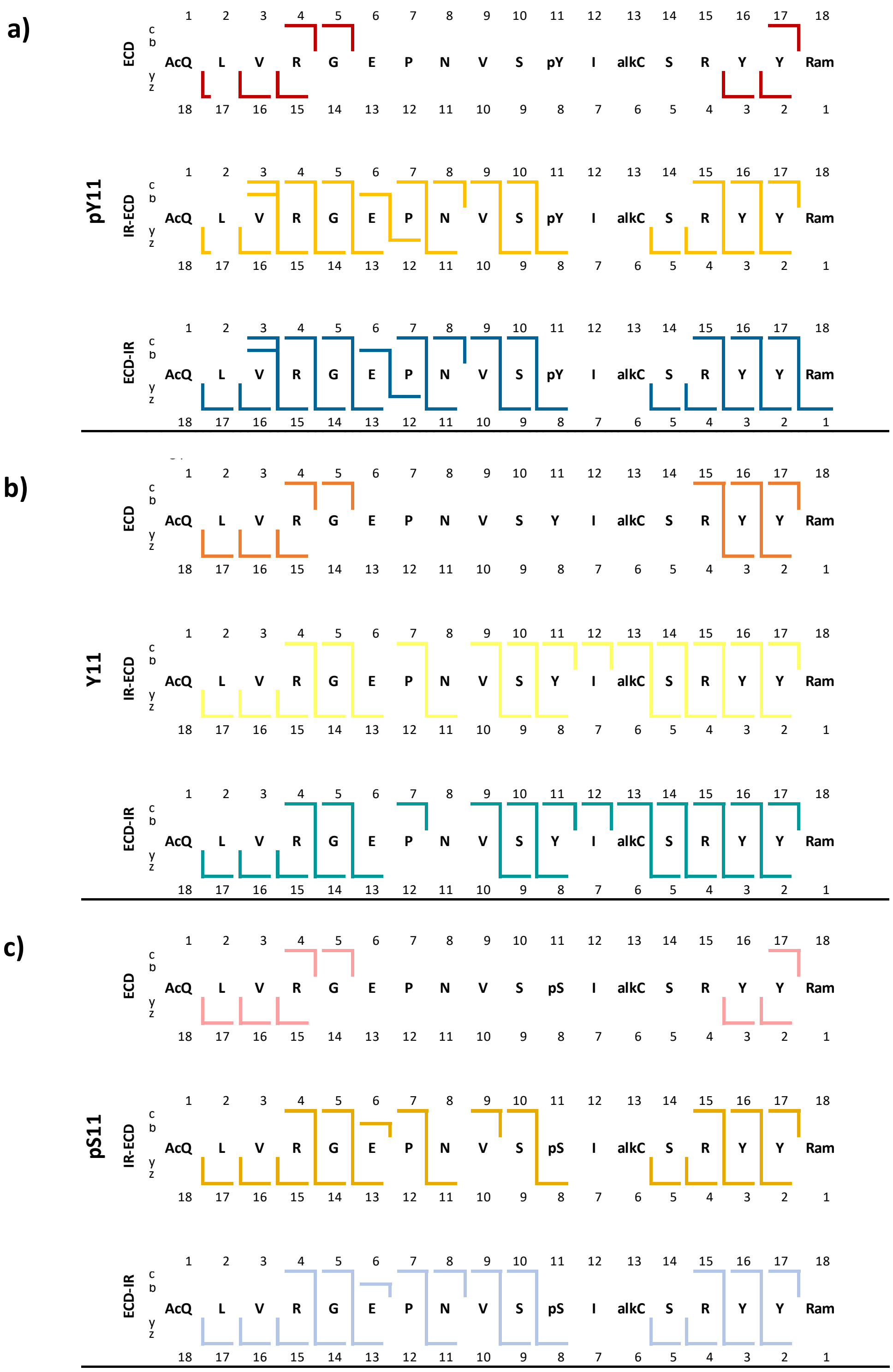


Figure 4.8 ECD, IR-ECD and ECD-IR fragment coverages for a) GSK3β-pY11 b) GSK3β-Y11 and c) GSK3β-pS11

Figure 4.8 shows the resultant fragmentation patterns from the AI-ECD experiments (annotated AI-ECD MS/MS spectra are found in Appendix 8), which demonstrate that both IR-ECD and ECD-IR achieved increased fragment coverage for the [M+2H]$^{2+}$ ions of GSK3β-pY11, GSK3β-Y11 and GSK3βpS11 compared to the ECD alone. This finding suggests that the presence of non-covalent interactions that prevent the dissociation and detection of fragments are responsible for at least some of the missed cleavages in each of these three peptides. Without IR activation, the ECD fragment coverage was identical for all 3 peptides, with no fragments observed between E6 and R15. The ECD behaviour for all 3 peptides is different to that observed in ECD studies reported earlier in this chapter, where the gaps in fragment coverage were R4-Y17 for GSK3β-pY11, R4-Y16 for GSK3β-Y11 and R4-R15 for GSK3β-pS11. The difference in ECD fragment coverage is likely due to the different instrument configuration used in these experiments. Changes in the MS instrumentation, and especially changes in ECD setup, may affect the gas-phase peptide structure, its fragmentation by ECD and the sensitivity of the mass spectrometer to detect low abundance fragments. These factors combined mean that different fragmentation patterns may be recorded on different instruments.

For the phosphopeptides GSK3β-pY11 and GSK3β-pS11 nearly full fragment coverage was observed upon activation, with only two missed cleavages between pX11 and alk-C13 (and an additional missed cleavage between N9 and S10 for the IR-ECD of GSK3β-pS11). The similarity between GSK3β-pY11 and GSK3β-pS11 further confirms that the identity of the phosphorylated residue does not have a large effect on the ECD behaviour of a peptide. For GSK3β-Y11, greater fragment coverage was observed in both the IR-ECD and ECD-IR, with only one missed cleavage between N9 and S10 in both experiments. The increased fragment coverage between Y11 and alk-C13 compared to between pY11 and alk-C13 in GSK3β-pY11 indicates that electron capture may be hindered in the phosphopeptide. The region which lacks fragment coverage in GSK3β-pY11, pY11 – I12 – alk-C13, includes two residues that could be affecting electron capture, pY11 and alk-C13. The phosphotyrosine could be affecting ECD behaviour by acting as an electron trap that forms a stable radical from which no backbone cleavage occurs, as has been noted for ECD fragmentation of peptides containing 3-nitrotyrosine.[199] This argument could not be extended to GSK3β-pS11, as phosphoserine lacks an aromatic system that would enable it to act as an electron trap. Meanwhile, the alkylated cysteine is present in both GSK3βpY11 and GSK3β-pS11 and is likely to be efficient at capturing electrons, as observed in various studies of cysteine-containing ions in which cleavage at S-X bonds is preferred.[200-202] This effect is also observed here, with many of these electrons seeming to cause cleavage of the alkyl group from the cysteine residue for all three peptides, as evidenced by the intense $[M+2H - C_2ONH_4]^{\bullet +}$ peak (acetamide, C_2ONH_4, being the alkyl group in alk-C). If most of electrons captured at alk-C13 cleave the alkyl cap, this could result in a lack of backbone cleavages nearby. The phosphotyrosine preventing bond cleavage by acting as an electron trap explains the difference in the AI-ECD experiments of GSK3β-pY11 and GSK3β-Y11, in which cleavages between Y11 and alkC13 were observed, however does not explain the lack of fragments between E6 and R15 in the ECD (no IR) of GSK3β-Y11, nor the similar fragment coverages of GSK3β-pY11 and GSK3β-pS11 by AI-ECD, in which fragment coverages were identical except for one missed cleavage at N8-V9 in the IR-ECD experiment of GSK3β-pS11. It is possible that the alk-C13 residue prevents backbone cleavages in both GSK3β-pY11 and GSK3β-Y11 as both peptides contain this residue. The regions of both peptides with fragment coverage (before E6 and after R15) could be viewed as being regions sufficiently close to other possible electron capture sites (R4, R15 and R18) that fragments here are observed, while electrons captured by alk-C13 cause cleavage of the alkyl group and prevent backbone cleavages.

In order to investigate whether the preference for cleavage of the alkyl group from the alk-C13 residue was preventing electrons captured at the cysteine residue causing backbone cleavages, analogues of GSK3β-pY11 and GSK3β-Y11 were derived that had alk-C13 substituted for an alanine residue. This alanine residue does not contain a sulfur atom nor any other moiety that might affect the ECD behaviour of the peptides. These GSK3β-pY11-noC and GSK3β-Y11-noC peptides had the sequences Ac-QLVRGEPNVS(pY)IASRYYR-Am and Ac-QLVRGEPNVSYIASRYYR-Am, respectively. The ECD mass spectra of these peptides are shown in Figure 4.9, with full peak assignment lists found in Appendix 6-13 and Appendix 6-14, with methods as discussed in Section 2.2.2. ECD of these peptides resulted in a large increase in the fragmentation efficiency compared to the GSK3β-pY11 and GSK3β-Y11 peptides that contained the alkylated cysteine residue, with Parent:Fragment peak ratios of ~1.5:1 for GSK3β-pY11-noC and GSK3β-Y11-noC compared to ~10:1 for GSK3β-pY11 and GSK3β-Y11 based on their peak heights. This increase in fragment efficiency implies that cleavage of the alkyl group from the cysteine was suppressing the formation of other fragments in those peptides where it was present. There was no difference in the fragment coverage between the GSK3β-pY11-noC and GSK3β-Y11-noC, with no fragments observed between E6 and R15 for either peptide. Given that the E6 residue is an acidic residue, it is possible that it exists in its deprotonated, negatively-charged glutamate form, and interacts via salt bridges with the R15 in both the phosphorylated and unphosphorylated peptides. A salt bridge from E6 to R15 would explain the lack of fragments between these residues.

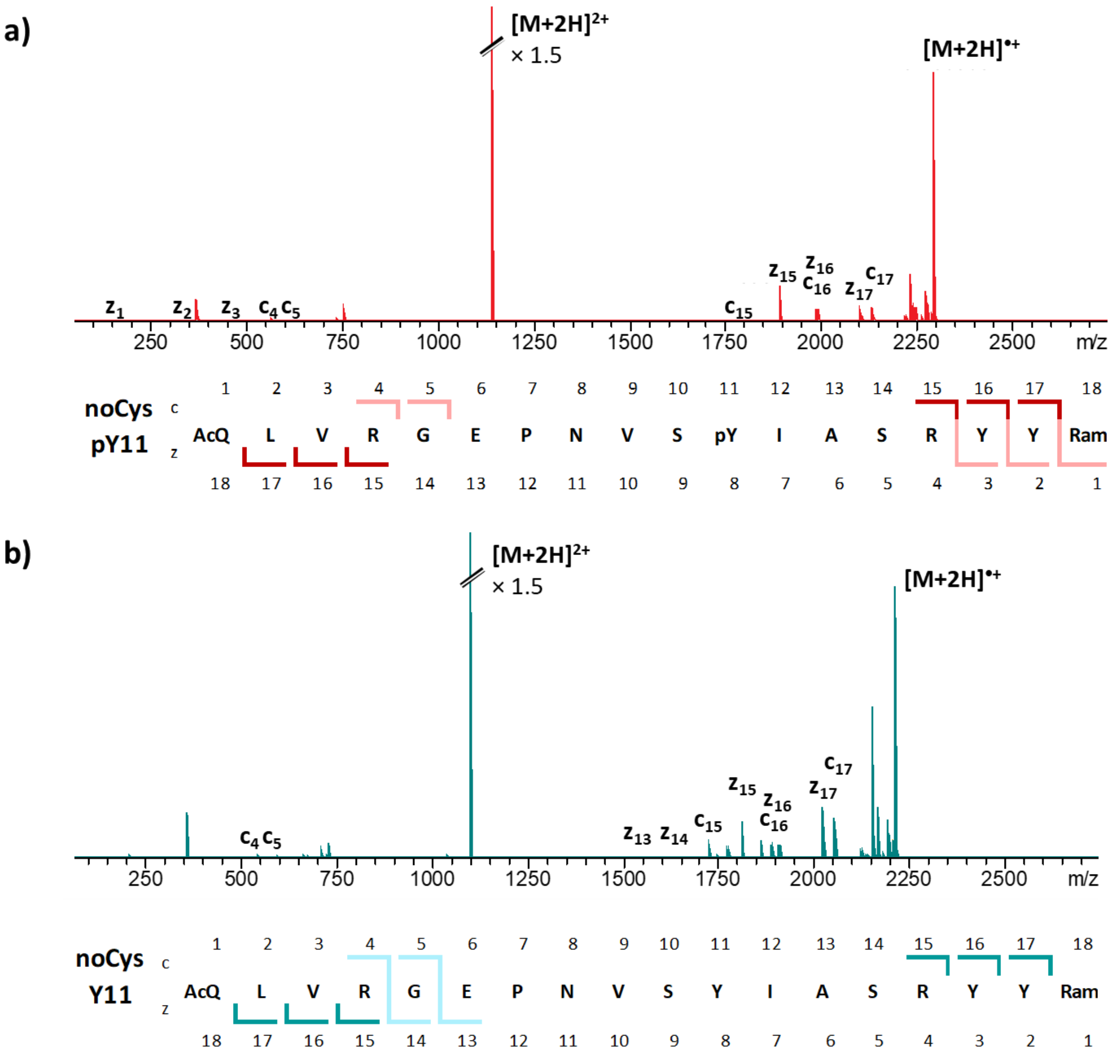


Figure 4.9 ECD MS/MS spectra and fragmentation patterns of a) GSK3β-pY11-noC and b) GSK3β-Y11-noC

Analogues were derived of the GSK3β-pY11 and GSK3β-Y11 peptides that did not contain the alk-C13 or E6 residues to study the effect of the glutamate residue. These peptides, GSK3β-pY11-noEnoC (Ac-QLVRGAPNVS(pY)IASRYYR-Am) and GSK3β-Y11-noEnoC (Ac-QLVRGAPNVSYIASRYYR-Am) were subjected to ECD fragmentation (Figure 4.10, fragment assignment lists are found in Appendix 6-15 and Appendix 6-16, methods as discussed in Section 2.2.2). For GSK3β-pY11-noEnoC, no fragments were observed between pY11 and R15, suggesting a non-covalent interaction between these residues. For GSK3β-Y11-noEnoC, two missed cleavages were observed at N8-S9 and Y11-I12, but fragments were observed between all residues likely to be involved in salt bridge or ionic hydrogen bond interactions, suggesting that no such interactions were taking place. This result suggests that in the absence of the E6 residue, the pY11 residue is capable of interacting with the R15 residue and that without either E6 or pY11 there are no other residues present that can form sufficiently strong interactions with the arginine residues to prevent the detection of ECD fragments.

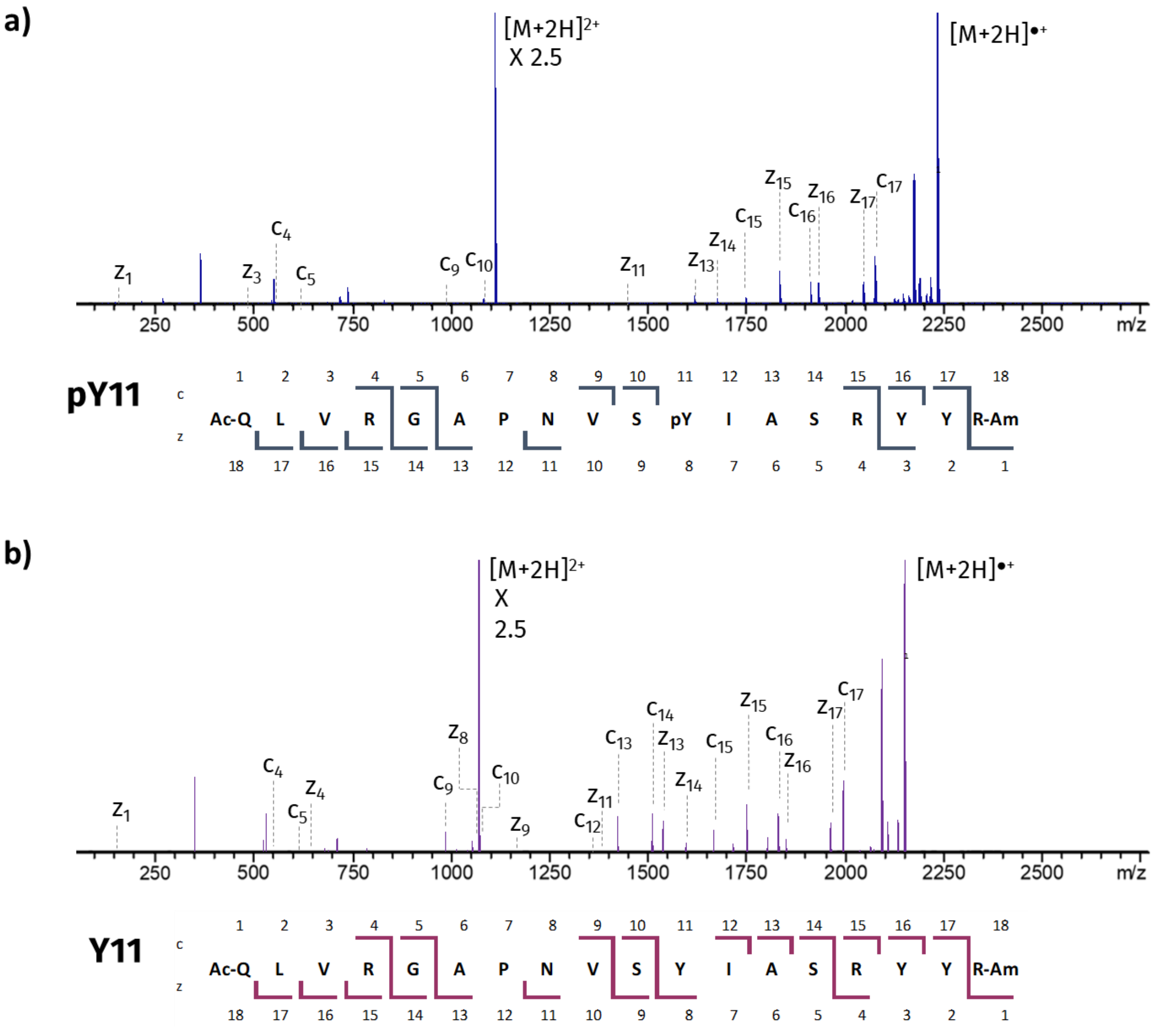


Figure 4.10 ECD MS/MS spectra and corresponding fragmentation patterns for a) GSK3β-pY11-noEnoC and b) GSK3β-Y11-noEnoC

Taking an overview of the findings from the ECD experiments of all GSK3β-derived peptides:

1. pY11 and E6 are both capable of forming interactions with R15 in the absence of the other
2. When both pY11 and E6 are present, E6 is involved in interactions with arginine residues, but pY11 may not be
3. Cleavage of the alkyl group from alk-C13 supresses all fragments to some extent

4.4 TWIMS behaviour of GSK3β-derived peptide ions

Travelling wave ion mobility spectrometry (TWIMS) experiments of all GSK3β-derived peptides (GSK3β-pY11, GSK3β-Y11, GSK3β-pS11, GSK3β-S11, GSK3β-pY11-noC, GSK3β-Y11-noC, GSK3β-pY11-noEnoC and GSK3β-Y11-noEnoC) were performed to probe their structures, using the methods outlined in Section 2.2.4. Briefly, in these TWIMS experiments, the TWIMS arrival time distribution (ATD) for each peptide is measured. By means of calibration against a mixture of tryptic peptides with known collision cross sections (CCS), the CCS of each peptide can be derived from its ATD. In each case the CCS reported is the value corresponding to the apex of its ATD peak. The CCS derived from these experiments are shown in Table 4.2 and compared graphically in Figure 4.11. The lists of calibrant ions and calibration curves used to obtain the CCS for each of these peptides are found in Appendix 3-2 and Appendices 4-2 to 4-5, respectively.

Table 4.2 Mass and CCS of GSK3β-derived peptides

Abbreviated name	Monoisotopic mass (Da)	$^{TW}CCS_{N2\rightarrow He}$ (Å^2)
GSK3β-pY11	2381.12	408.63 ± 0.94
GSK3β-Y11	2301.16	400.13 ± 0.58
GSK3β-pS11	2305.09	395.79 ± 1.95
GSK3β-S11	2225.12	389.44 ± 2.54
GSK3β-pY11-noC	2292.13	378.53 ± 0.82
GSK3β-Y11-noC	2212.16	390.57 ± 1.10
GSK3β-pY11-noEnoC	2234.12	398.53 ± 0.62
GSK3β-Y11-noEnoC	2154.16	404.18 ± 1.78

For the phosphorylated/unphosphorylated pairs GSK3β-pY11/GSK3β-Y11 and GSK3β-pS11/GSK3β-S11, the trend is that the higher mass phosphopeptides have a larger CCS than their unphosphorylated analogues, which is to be expected as CCS is proportional to mass. These four peptides (GSK3β-pY11 and GSK3β-Y11, and GSK3β-pS11 and GSK3β-S11) have a CCS lower than the trend for a random coil (represented in Figure 4.11, by extrapolation of the CCS values for polyalanine), showing that they exhibit compaction, which is common for peptides that contain non-covalent interactions. This finding is in agreement with several studies that demonstrated that phosphopeptides tend to be compacted to a greater extent than randomly-coiled peptides due to non-covalent interactions with the phosphorylated residue.[127-130] For GSK3β-pY11-noC/GSK3β-Y11-noC, and GSK3β-pY11-noEnoC/GSK3β-Y11-noEnoC, the trend is the opposite, with the lower mass unphosphorylated peptides having higher CCS than the corresponding phosphorylated peptides. This observation indicates that for these peptides the phosphorylation causes a greater extent of compaction relative to the unphosphorylated peptide than was observed for GSK3β-pY11/GSK3β-Y11 and GSK3β-pS11/GSK3β-S11. For GSK3β-Y11-noC, the lack of a phosphate group seems to result in a much smaller degree of compaction compared to GSK3β-pY11-noC, however, GSK3β-Y11-noC still exhibits greater compaction than that which would be expected for a random coil. That the GSK3β-pY11-noC/GSK3β-Y11-noC pair exhibits a different trend to GSK3β-pY11/GSK3β-Y11 and GSK3β-pS11/GSK3β-S11 indicates that the alk-C residue has a structural effect as well as the proposed effect of suppressing backbone fragmentation. GSK3β-Y11-noEnoC lies on the trendline for a randomly-coiled peptide, showing no compaction caused by non-covalent interactions within the molecule, a finding which is in agreement with its ECD data, in which fragments were observed between all acidic and basic residues.

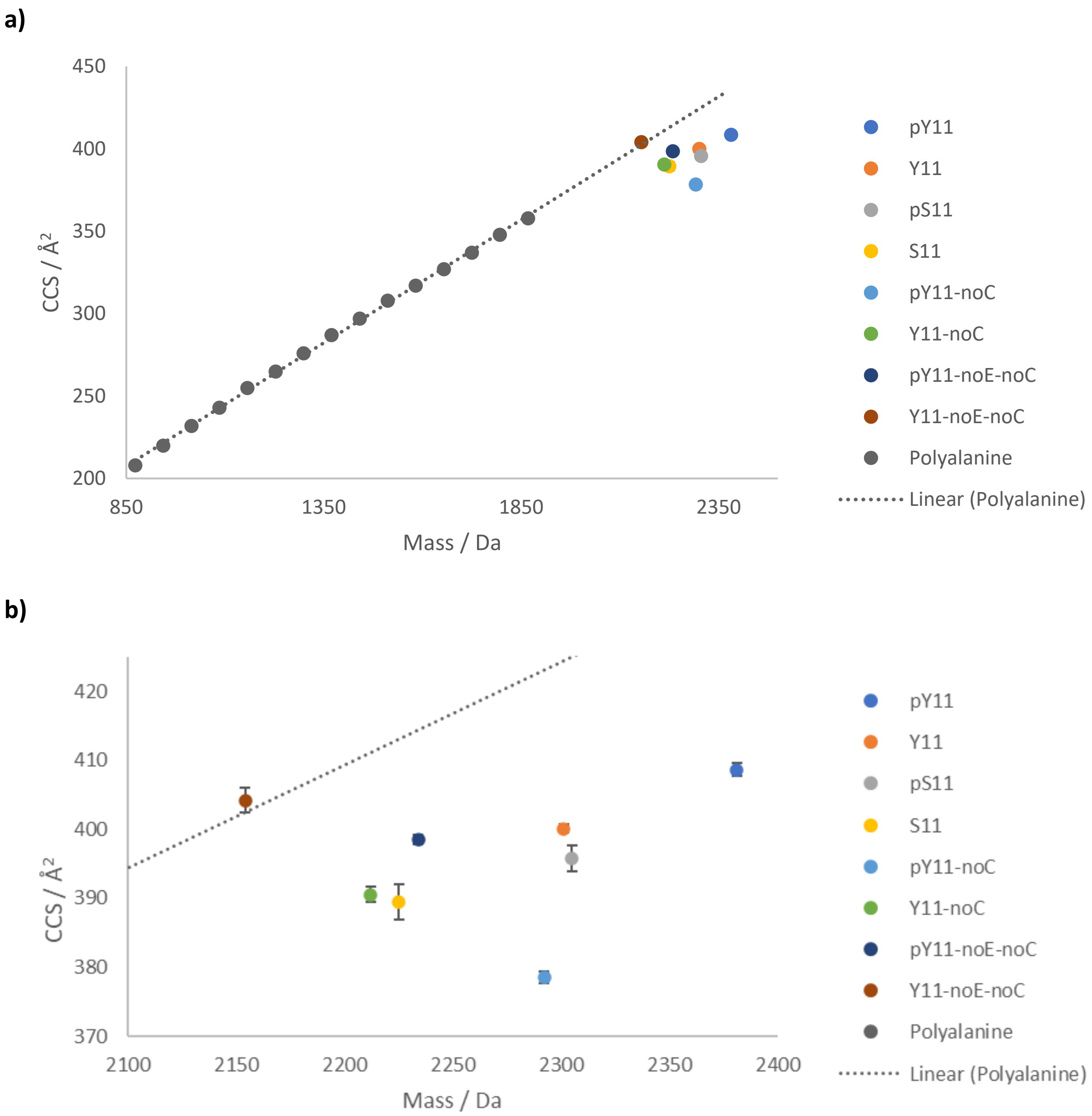


Figure 4.11 a) Comparison of GSK3β-derived peptide CCS and mass values to reference polyalanine values. The polyalanine trendline represents the relationship between CCS and mass that would be expected for a peptide adopting a random coil, and b) an expansion of the region of the graph containing the GSK3β-derived peptides

The arrival time distributions (ATD) for the GSK3β-derived peptides are shown in Figure 4.12. The phosphorylated peptides have broader ATDs than their unphosphorylated analogues, possibly suggesting that they have a broader conformational space. The ATDs for peptides which contain both a glutamate residue and a phosphorylated residue (GSK3β-pY11, GSK3β-pS11 and GSK3β-pY11noC) seem to have especially broad ATDs when compared to GSK3β-pY11noEnoC, in which the glutamate residue is not present. Many of the ATDs exhibit obvious tailing, either towards lower arrival times

(GSK3β-pY11 and GSK3β-Y11noEnoC) or higher arrival times (GSK3β-Y11, GSK3β-pS11 and GSK3β-pY11noC). Such tailing is further indication of the presence of multiple conformations for these peptides.

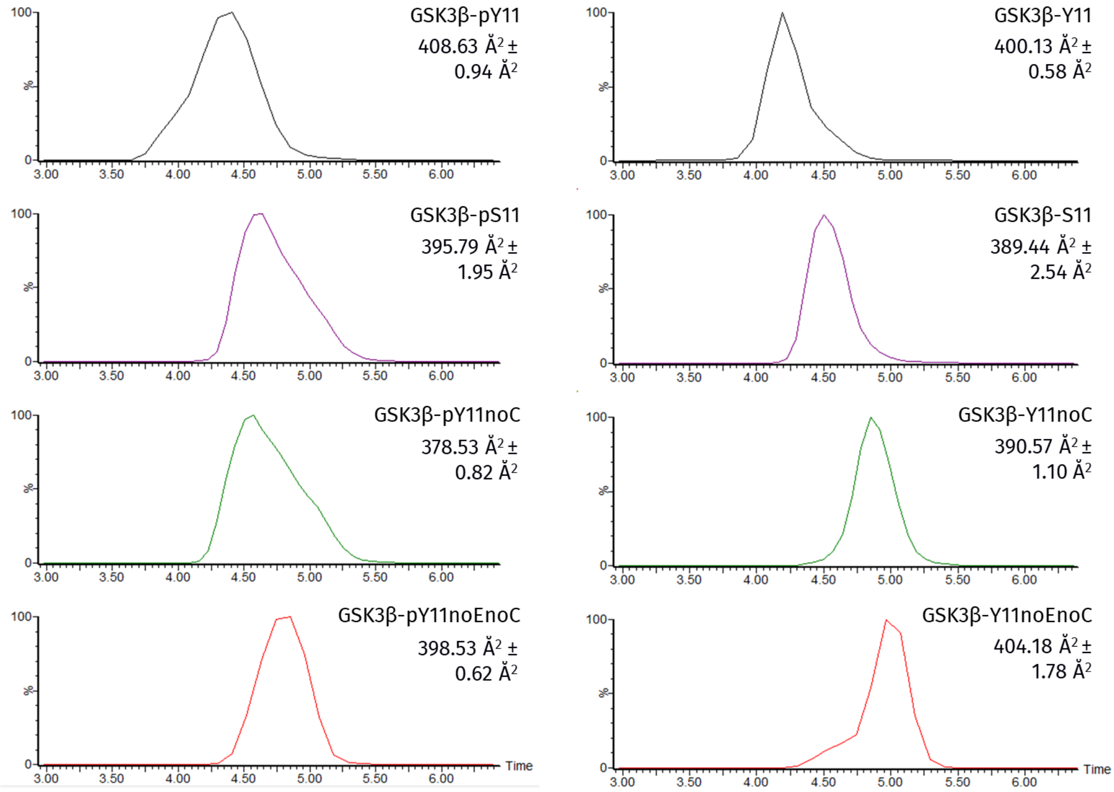


Figure 4.12 TWIMS arrival time distributions of all GSK3β-derived peptides measured at WV = 300 ms-1 and WH = 20 V, labelled with the peptide name and the CCS value corresponding to the apex of the ATD

4.5 Application of ion mobility spectrometry/molecular modelling workflow to GSK3β-derived peptides

Several protonation patterns were modelled for each peptide in order to maintain a 2+ net charge and to investigate the effect of the multiple basic residues (R4, R15 and R18) and multiple acidic residues (E6 and pX11, where X = pY or pS). The sequences of all the protonation patterns modelled are shown in Table 4.3, where acidic and basic residues are labelled with $^+$, $^-$ or ^{0} to indicate whether they are positive, negative or neutral, respectively.

Table 4.3 Protonation patterns modelled for each of the GSK3-derived peptides

Peptide	Model name	Sequence
GSK3β-pY11	pY11 Y1P	Ac- Q L V R^+ G E^0 P N V S pY^- I alkC S R^+ Y Y R^+-Am
	pY11 GLU	Ac- Q L V R^+ G E^- P N V S pY^0 I alkC S R^+ Y Y R^+-Am
GSK3β-Y11	Y11 RDP4	Ac- Q L V R^0 G E^0 P N V S Y I alkC S R^+ Y Y R^+-Am
	Y11 RDP15	Ac- Q L V R^+ G E^0 P N V S Y I alkC S R^0 Y Y R^+-Am
	Y11 RDP18	Ac- Q L V R^+ G E^0 P N V S Y I alkC S R^+ Y Y R^0-Am
	Y11 GLU	Ac- Q L V R^+ G E^- P N V S Y I alkC S R^+ Y Y R^+-Am
GSK3β-pS11	pS11 Y1P	Ac- Q L V R^+ G E^0 P N V S pS^- I alkC S R^+ Y Y R^+-Am
	pS11 GLU	Ac- Q L V R^+ G E^- P N V S pS^0 I alkC S R^+ Y Y R^+-Am
GSK3β-S11	S11 RDP4	Ac- Q L V R^0 G E^0 P N V S S I alkC S R^+ Y Y R^+-Am
	S11 RDP15	Ac- Q L V R^+ G E^0 P N V S S I alkC S R^0 Y Y R^+-Am
	S11 RDP18	Ac- Q L V R^+ G E^0 P N V S S I alkC S R^+ Y Y R^0-Am
	S11 GLU	Ac- Q L V R^+ G E^- P N V S S I alkC S R^+ Y Y R^+-Am
GSK3β-pY11-noC	pY11noC Y1P	Ac- Q L V R^+ G E^0 P N V S pY^- I A S R^+ Y Y R^+-Am
	pY11noC GLU	Ac- Q L V R^+ G E^- P N V S pY^0 I A S R^+ Y Y R^+-Am
GSK3β-Y11-noC	Y11noC RDP4	Ac- Q L V R^0 G E^0 P N V S Y I A S R^+ Y Y R^+-Am
	Y11noC RDP15	Ac- Q L V R^+ G E^0 P N V S Y I A S R^0 Y Y R^+-Am
	Y11noC RDP18	Ac- Q L V R^+ G E^0 P N V S Y I A S R^+ Y Y R^0-Am
	Y11noC GLU	Ac- Q L V R^+ G E^- P N V S Y I A S R^+ Y Y R^+-Am
GSK3β-pY11-noCnoE	pY11noEnoC Y1P	Ac- Q L V R^+ G A P N V S pY^- I A S R^+ Y Y R^+-Am
GSK3β-Y11-noCnoE	Y11noEnoC RDP4	Ac- Q L V R^0 G A P N V S Y I A S R^+ Y Y R^+-Am
	Y11noEnoC RDP15	Ac- Q L V R^+ G A P N V S Y I A S R^0 Y Y R^+-Am
	Y11noEnoC RDP18	Ac- Q L V R^+ G A P N V S Y I A S R^+ Y Y R^0-Am

Using the methods in Section 2.1, described in detail in Section 3.5, modelling each of these protonation patterns produced a 25,000-strong ensemble, which was then filtered by the experimentally derived CCS ±5% to produce a CCS-matched ensemble. These CCS-matched ensembles were then subjected to dimensionality reduction and clustering to produce 10 representative structures for each protonation pattern. The non-covalent interactions in the representative structures were compared to those predicted from the ECD experiments of each peptide (Section 4.3), with a non-covalent interaction being identified when a hydrogen atom in the guanidine group of any arginine reside is within 4 Å of the centre of an acidic residue. The centres of acidic residues considered were the phosphorous atom in pTyr and pSer, the side-chain oxygen atom in Tyr and Ser (only in residue 11 of unphosphorylated peptides in order to compare their non-covalent interactions to phosphorylated analogues) and the δ-carbon in Glu. Details of the clusters for all protonation patterns modelled are found in Appendix 1-9, while the distances between acidic and basic residues in the 10 representative structures for all protonation patterns modelled are found in Appendix 1-10. The representative structures themselves can be found in Appendix 1-15, where they are in ascending order of structure number, as in the corresponding part of Appendix 1-9.

The peptide GSK3β-pY11 lacked fragments between R4 and Y17 upon ECD fragmentation, which indicated non-covalent interactions between R4 and R15 and one of the acidic residues between them (R4-X-R15, where X is E6 or pY11). None of the representative structures produced from the model pY11-Y1P were consistent with this ECD behaviour, as all of them had R18 interacting with pY11 in addition to at least one of the other arginine residues. Five representative structures from pY11-GLU (pY11-GLU-1, pY11-GLU-10, pY11-GLU-5, pY11-GLU-8 and pY11-GLU-2) had the ECD-consistent non-covalent interactions E6-R4 and E6-R15, cumulatively representing 52.82% of the pY11-GLU CCS-matched ensemble. The ECD-consistent structures are shown in Figure 4.13. Each of these ECD-consistent structures shows the pY11 residue projected outwards from the rest of the peptide, seemingly participating in very few non-covalent interactions with the rest of the peptide, which explains the similarity of the ECD behaviour between GSK3β-pY11 and GSK3β-Y11. This orientation of the pY11 residue would also allow it to interact with other peptides were this structure preserved in the solution phase. This might be the site at which dimerisation occurs in the GSK3β-pY11 peptide, and could explain why other peptides do not exhibit dimerisation if they do not also have this feature of an acidic residue capable of forming strong non-covalent interactions projected away from the rest of the peptide. Given that only representative structures from the pY11-GLU protonation pattern had ECD-consistent non-covalent interactions, this strongly suggests that GSK3β-pY11 exists in the pY11-GLU protonation pattern in the gas phase. Although, considering that a broad ATD was observed for this peptide, this may imply a greater amount of conformational heterogeneity than is suggested by the five ECD-consistent representative structures from pY11-GLU, which all have very similar structural features. In order to explain the broad ATD, it may be necessary to consider structures with the R4-X-R18 interaction pattern, which is not inconsistent with the ECD data as these interactions extend across the region of the peptide that lacks sequence coverage, but do not explain the presence of the c_{17} fragment between Y17 and R18. There are seven representative structures with this interaction pattern from pY11-Y1P and five from the pY11-GLU protonation pattern. Proposing that GSK3β-pY11 exists as a mixture of the five structures with R4-X-R15 interactions from pY11-GLU and the 12 structures with R4-X-R18 from both pY11-Y1P and pY11-GLU would explain the broad ATD of the peptide, and also rationalises why ECD fragments are not observed between R15 and Y17, supposing that the R4-X-R18 portion of the mixture suppresses these fragments to the extent that they are not observed. In order to be more certain of the existence of such a mixture, conformational specific ECD (i.e., ion mobility separation followed by ECD) could be performed, as the R4-X-R15 and R4-X-R18 components of the mixture would have different ECD behaviour. For the ion mobility-ECD experiment to be successful, the two components of the mixture would have to have distinct CCS distributions, which may not be the case as the TWIMS ATD shows no indication of resolvable conformers. The possibility of performing conformational specific ECD is discussed as proposed future work in Section 6.2.

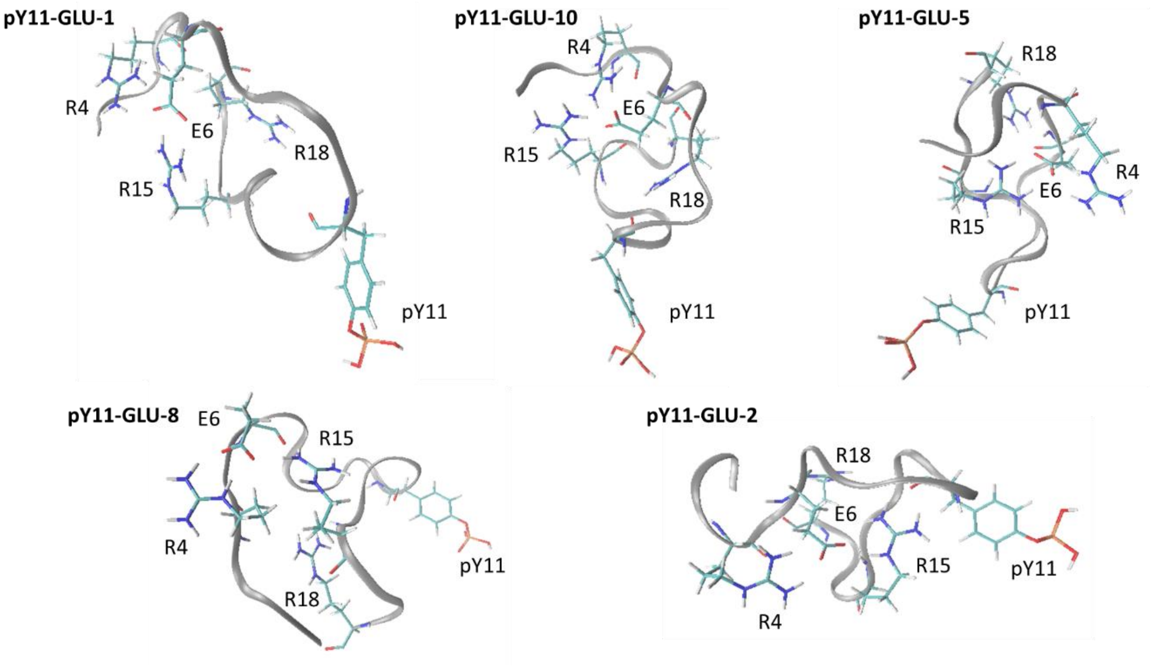


Figure 4.13 Representative structures produced from modelling GSK3β-pY11 with non-covalent interactions consistent with the ECD behaviour of the peptide

For GSK3β-Y11, no backbone cleavages were observed in the ECD data between R4 and Y16, indicating that the residues R4 and R15 are likely to be interacting with the acidic residue E6 that is contained within this region. Only the model Y11-GLU produced representative structures that had non-covalent interactions consistent with the ECD behaviour. Y11-GLU-3 and Y11-GLU-2 (Figure 4.14) both had non-covalent interactions E6-R4 and E6-R15 and represented 10.48% and 7.77% of the whole Y11-GLU CCS-matched ensemble, respectively. Similar to GSK3β-pY11, in both ECD-consistent structures the Y11 residue is projected away from the rest of the peptide, possibly due to its steric bulk. None of the protonation patterns containing a deprotonated arginine residue had more than one non-covalent interaction between E6 and any arginine residue, as such, none of these structures could be considered ECD consistent as each would allow fragments to form somewhere between R4 and Y16. Given that none of the other protonation patterns modelled produced any representative structures with ECD-consistent non-covalent interactions, it is likely that the Y11-GLU is the gas-phase protonation pattern. Five further representative structures from the Y11-GLU protonation pattern exhibit the R4-X-R18 non-covalent interaction pattern that was considered for GSK3β-pY11 to account for its broad ATD profile, however, the ATD observed for GSK3β-Y11 was narrower than the ATD for GSK3β-pY11. These five structures could be present in a mixture with the two R4-X-R15 structures and that mixture would exhibit the same ECD behaviour as that which was recorded for GSK3β-Y11. The presence of structures with a R4-X-R18 non-covalent interaction pattern could account for the missed cleavage between R15 and Y16, given that these structures could not form fragments between R4 and R18, so suppressing the signal of the fragment peaks between R15 and R18. On the other hand, the relatively narrow ATD of GSK3β-Y11 and the high intensity of the c_{16} and c_{17} ECD fragment peaks (corresponding to cleavages between Y16 and Y17, and Y17 and R18) do not support the presence of structures with the non-covalent interaction pattern R4-X-R18 in the gas-phase composition of GSK3β-Y11. Therefore, the proposed gas-phase structures for GSK3β-Y11 are the two structures with the R4-X-R15 interaction pattern (Y11-GLU-3 and Y11-GLU-2).

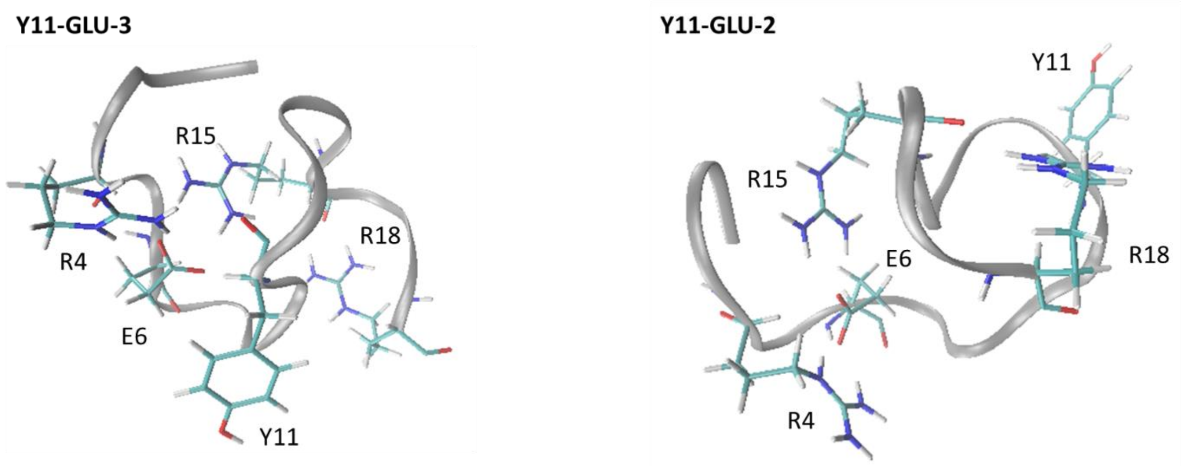
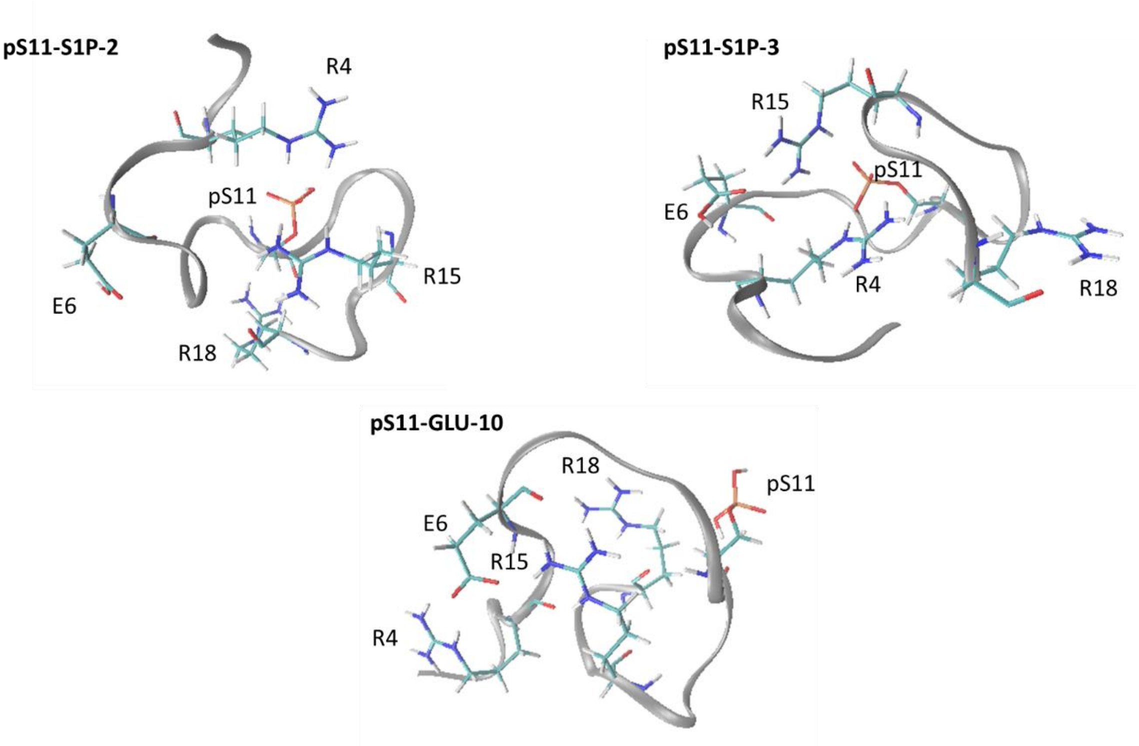


Figure 4.14 Representative structures produced from modelling GSK3β-Y11 with non-covalent interactions consistent with the ECD behaviour of the peptide

The ECD data for GSK3β-pS11 lacked fragments from R4 to R15, which leads to proposed non-covalent interactions between R4 and R15, and either E6 or pS11. Two representative structures from the pS11-S1P ensemble demonstrated these non-covalent interaction, pS11-S1P-2 and pS11-S1P-3. pS11-S1P-2 had pS11-R4 and pS11-R15 non-covalent interactions, while pS11-S1P-3 had pS11-R4, pS11-R15 and E6-R15 (Figure 4.15). The structure pS11-S1P-2 represented 9.62% of the whole pS11-S1P CCS-matched ensemble, and pS11-S1P-3 represented 8.78%. That there are ECD-consistent representative structures from pS11-S1P suggests that the phosphorylated residue in GSK3β-pS11 is more likely to be deprotonated and participate in non-covalent interactions than in GSK3β-pY11. One further structure from pS11-GLU exhibited ECD-consistent non-covalent interactions: pS11-GLU-10, which had E6-R4 and E6-R15 interactions and represented 10.80% of the pS11-GLU CCS-matched ensemble. This structure shows similarities to the pY11-GLU ECD-consistent structures, having the pS11 residue on an external loop of the peptide, but in this case pS11 can interact with the peptide backbone due to its lesser steric bulk than phosphotyrosine. Both protonation patterns modelled produced representative structures consistent with the ECD data, suggesting that GSK3β-pS11 exists as a mixture of pS11-S1P and pS11-GLU in the gas phase. GSK3β-pS11 being composed of a mixture of two protonation patterns might account for the broad TWIMS ATD observed for GSK3β-pS11, similar to GSK3β-pY11.


pS11-S1P-2

pS11-S1P-3

pS11-GLU-10

Figure 4.15 Representative structures produced from modelling GSK3β-pS11 with non-covalent interactions consistent with the ECD behaviour of the peptide.

For GSK3β-S11, the ECD data lacked fragments between R4 and R15, which leads to proposed non-covalent interactions between these residues and E6. Only the S11-GLU protonation pattern produced any representative structures with ECD consistent non-covalent interactions. These three representative structures, S11-GLU-6, S11-GLU-5 and S11-GLU-9 (Figure 4.16), all had interactions between E6-R4 and E6-R15 and cumulatively represented 26.03% of the S11-GLU CCS-matched ensemble. As with the one ECD-consistent pS11-GLU structure, the ECD-consistent S11-GLU structures have the S11 residue on an external loop, from where it interacts with the peptide backbone. Given that only the S11-GLU protonation pattern produced ECD-consistent structures, it is suggested that this is the protonation pattern in which it exists in the gas phase. This protonation pattern is analogous to the Y11-GLU protonation pattern that was proposed to be the gas-phase protonation pattern of GSK3β-Y11, which confirms the prediction from the ECD experiments of GSK3β-Y11 and GSK3β-S11 that in the absence of a phosphorylated residue, the glutamate forms non-covalent interactions with basic residues in the peptide. As with GSK3β-Y11, GSK3β-S11 had a narrow TWIMS ATD, which is consistent with the proposal that it is composed of a single protonation pattern in the gas phase.

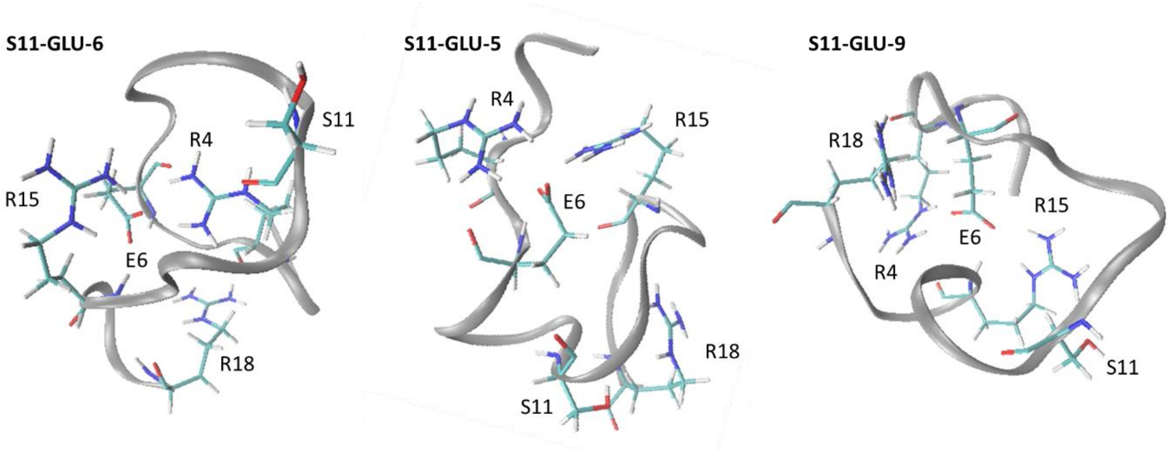


Figure 4.16 Representative structures produced from modelling GSK3β-S11 with non-covalent interactions consistent with the ECD behaviour of the peptide.

For GSK3β-pY11noC, the lack of ECD fragments between E6 and R15 suggests the non-covalent interaction E6-R15, possibly in addition to pY11-R15. No single representative structure from any of the protonation patterns modelled had this pattern of non-covalent interactions. Once the low-abundance (i.e., an order of magnitude lower in abundance than the next most intense peaks) c_4 and c_5 fragments have been excluded, the gap in fragment coverage spans from R4 to R15, indicating non-covalent interactions with these residues and either E6 or pY11. Three representative structures from pY11noC-Y1P and one from pY11noC-GLU have non-covalent interactions consistent with those predicted from the ECD data once low-abundance fragments are excluded (Figure 4.17a). The representative structure pY11noC-Y1P-1 has the non-covalent interactions pY11-R4, pY11-R15, E6-R4 and E6-R15; pY11noC-Y1P-2 has pY11-R4 and pY11-R15 interactions; pY11noC-Y1P-6 has pY11-R4, pY11-R15 and E6-R4 interactions; and pY11noC-GLU-2 has E6-R4 and E6-R15 interactions. These structures represented 32.46% of the pY11noC-Y1P CCS-matched ensemble and 12.32% of the pY11noC-GLU CCS-matched ensemble. The possible non-covalent interactions predicted once low-abundance fragments are removed are the same as those predicted for GSK3β-pY11. The GSK3β-pY11noC structures consistent with these non-covalent interactions are from both pY11noC-Y1P and pY11noC-GLU, whereas GSK3β-pY11 only had ECD-consistent structures from pY11-GLU (with a proposed contribution from 12 further structures from pY11-Y1P and pY11-GLU to account for its broad TWIMS ATD), suggesting the substitution of Cys-alk to Ala has a structural effect in addition to any effect on electron capture. These GSK3β-pY11noC ECD-consistent structures seem to have a greater deal of flexibility between pY11 and R15, with the backbone in this region adopting more diverse structures, supporting the proposal that steric hindrance from Cys-alk produced more rigid structures. In order to explain the ECD data including low-abundance fragments, it was necessary to consider combinations of multiple representative structures with at least one structure having interactions that would prevent the observation of fragments up to R15, and others which prevent the observation of fragments from E6 past R15. For example, a mixture comprising structures having an E6-R18 interaction together with those that have E6-R4 and E6-R15 interactions would produce the same ECD behaviour as a sample containing only conformers with E6-R15 interactions. Such an interaction pattern would also be consistent with the presence of low abundance fragments between R4 and E6 (c_4 and c_5), and between R15 and R18 (z_1, z_2 and z_3). For the part of the mixture with non-covalent interactions up to R15, the representative structures that explain the ECD data with low abundance fragments excluded (pY11noC-Y1P-1, pY11noC-Y1P-2, pY11noC-Y1P-6, and pY11noC-GLU-

2) are all possible candidates (Figure 4.17b). Meanwhile, the candidates for the component of the mixture that prevents detection of fragments from E6 onwards are pY11noC-GLU-4 (E6-R18), pY11noC-GLU-6 (E6-R15 and E6-R18), pY11noC-GLU-5 (E6-R18), pY11noC-GLU-1 (E6-R18) and pY11noC-GLU-10 (E6-R18). This mixture fully explains the ECD behaviour of GSK3β-pY11noC and is consistent with the broad ATD recorded for the peptide; therefore, it is proposed as the form in which the GSK3β-pY11noC peptide exists in the gas phase.

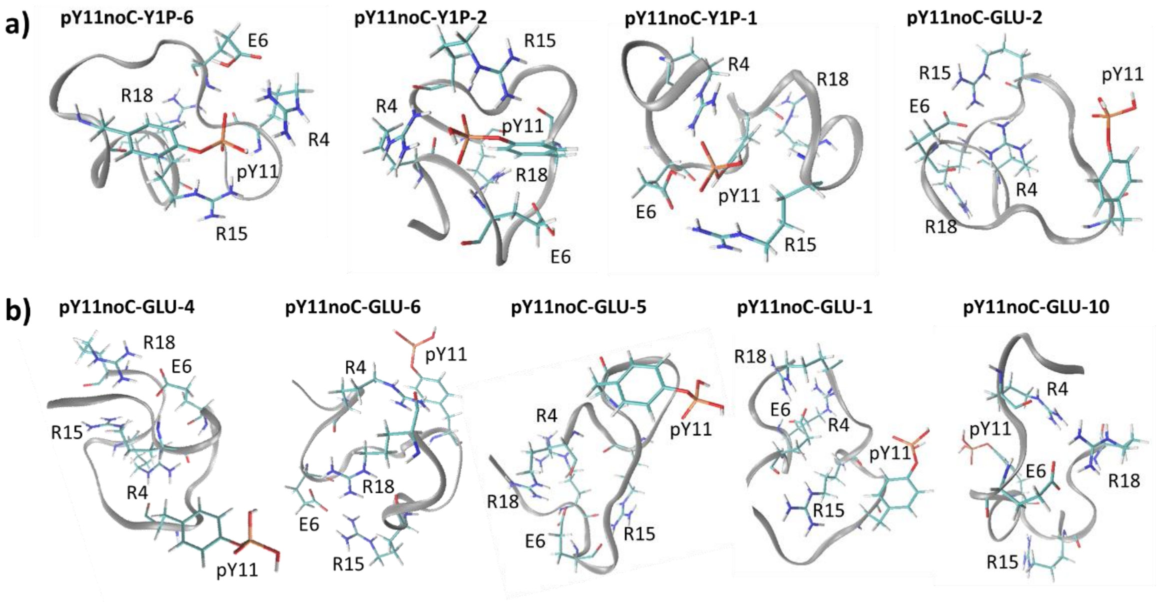


Figure 4.17 Representative structures produced from modelling GSK3β-pY11noC with non-covalent interactions consistent with the ECD behaviour of the peptide a) with all fragments considered, and b) once low-abundance fragments have been excluded

GSK3β-Y11noC exhibited the same ECD behaviour as GSK3β-pY11noC, with a lack of fragments between E6 and R15, and low abundance fragments between R4 and E6. Notably, there are no low abundance z_1, z_2 and z_3 fragments, though both the c and z fragments between R4 and E6 (c_4, c_5, z_{13} and z_{14}) are present in low abundance. One representative structure from Y11noC-RDP15 (Y11noC-RDP15-8) and one from Y11noC-RDP18 (Y11noC-RDP18-8) exhibited the E6-R15 non-covalent interaction that is consistent with the ECD fragmentation pattern with low-abundance fragments included (Figure 4.18a). Although the structure Y11noC-RDP15-8 has the E6-R15 interaction, it is notable that both interacting partners are neutral and so is characterised as a hydrogen bond. It is doubtful that this hydrogen bond interaction has the strength required to prevent dissociation of ECD fragments as hydrogen bonds between neutral, polar groups both in side chains and the peptide backbone are a ubiquitous feature of peptide structure and must be broken for any peptide to be analysed by ECD. Y11noC-RDP15-8 and Y11noC-RDP18-8 represented 9.27% and 9.85% of their whole CCS-matched ensembles respectively. With the low-abundance c_4, c_5, z_{13} and z_{14} fragments excluded, the ECD data were consistent with E6-R4 and E6-R15 non-covalent interactions. The representative structure Y11noC-GLU-7 (Figure 4.18b), representing 9.83% of the Y11noC-GLU CCS-matched ensemble, exhibited this non-covalent interaction pattern. A mixture of Y11noC-RDP15-8, Y11noC-RDP18-8 and Y11noC-GLU-7 would explain the whole ECD behaviour, with c_4, c_5, z_{13} and z_{14} fragments generated from Y11noC-RDP15-8 and Y11noC-RDP18-8, while non-covalent interactions in Y11noC-GLU-7 prevent the formation of these fragments from this conformer, resulting in their low-abundance. Unlike the non-phosphorylated peptides GSK3β-Y11 and GSK3β-S11, the ECD data for GSK3β-Y11noC cannot be rationalised by a single protonation pattern due to the low abundance fragments that were observed, despite its ATD being a similar width to those of GSK3β-Y11 and GSK3β-S11. This implies that the components of the mixture proposed for GSK3β-Y11noC in the gas phase must be relatively rigid compared to the proposed GSK3β-pY11, GSK3β-pS11 and GSK3β-pY11noC mixtures, which all had broader ATD profiles, and not deviate greatly from the apex CCS to which they were matched in this IMS/MM workflow.

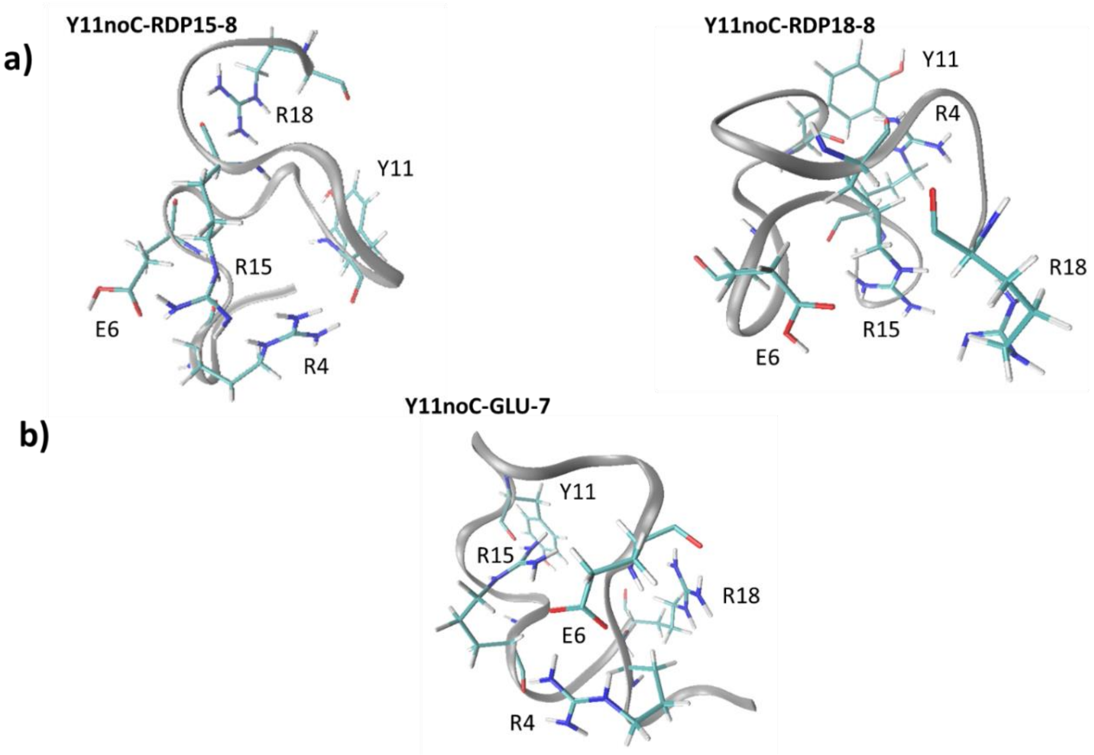


Figure 4.18 Representative structures produced from modelling GSK3β-Y11noC with non-covalent interactions consistent with the ECD behaviour of the peptide a) with all fragments considered, and b) once low-abundance fragments have been excluded

The ECD behaviour of GSK3β-pY11noEnoC indicated the presence of a pY11-R15 non-covalent interaction by a lack of backbone fragments between these residues. No single representative structure produced from the modelling workflow had this non-covalent interaction alone, with all of the representative structures having non-covalent interactions between pY11 and at least two of the arginine residues. It was possible, as with GSK3β-pY11noC, to explain the ECD-behaviour by considering a mixture of structures with two different non-covalent interaction patterns that both include the pY11-R15 interaction. In this case, the non-covalent interaction patterns were R4-pY11 in addition to pY11-R15, and pY11-R18. Structures exhibiting pY11-R4 with pY11-R15 were pY11noEnoC-Y1P-6 and pY11noEnoC-Y1P-2 (Figure 4.19a), representing 10.14% and 8.96% of the whole pY11noEnoC-Y1P CCS-matched ensemble respectively. Structures with the pY11-R18 non-covalent interaction were pY11noEnoC-Y1P-3, pY11noEnoC-Y1P-8 and pY11noEnoC-Y1P-5 (Figure 4.19b), representing 10.42%, 6.98% and 12.66% of the pY11noEnoC-Y1P CCS-matched ensemble respectively. All these structures with pY11-R18 also exhibited the pY11-R15 interaction. A mixture of these two sets of structures would explain the lack of fragments between pY11 and R15, while fragments between R4 and pY11 could be generated from structures pY11noEnoC-Y1P-3, pY11noEnoC-Y1P-8 and pY11noEnoC-Y1P-5, and those fragments observed between R15 and R18 could be generated from pY11noEnoC-Y1P-6 and pY11noEnoC-Y1P-2. On the other hand, there are no fragments in the ECD MS/MS spectrum of GSK3β-pY11noEnoC that were identified as being low abundance (defined here as fragment peaks an order of magnitude smaller than the next most abundant fragments), which would not indicate that GSK3β-pY11noEnoC exists as a mixture in the gas phase, although it is possible that a mixture with equal proportions of both components does not result in large reduction in the abundance of any fragments detected. The proposal that GSK3β-pY11noEnoC exists as a mixture in the gas phase is consistent with all the other phosphopeptides discussed (GSK3β-pY11, GSK3β-pS11 and GSK3β-pY11noC), which were also proposed to be composed of mixtures with components of different protonation patterns and non-covalent interaction patterns, and is consistent with the observation that the ATDs of the phosphorylated peptides were broader than their non-phosphorylated analogues.

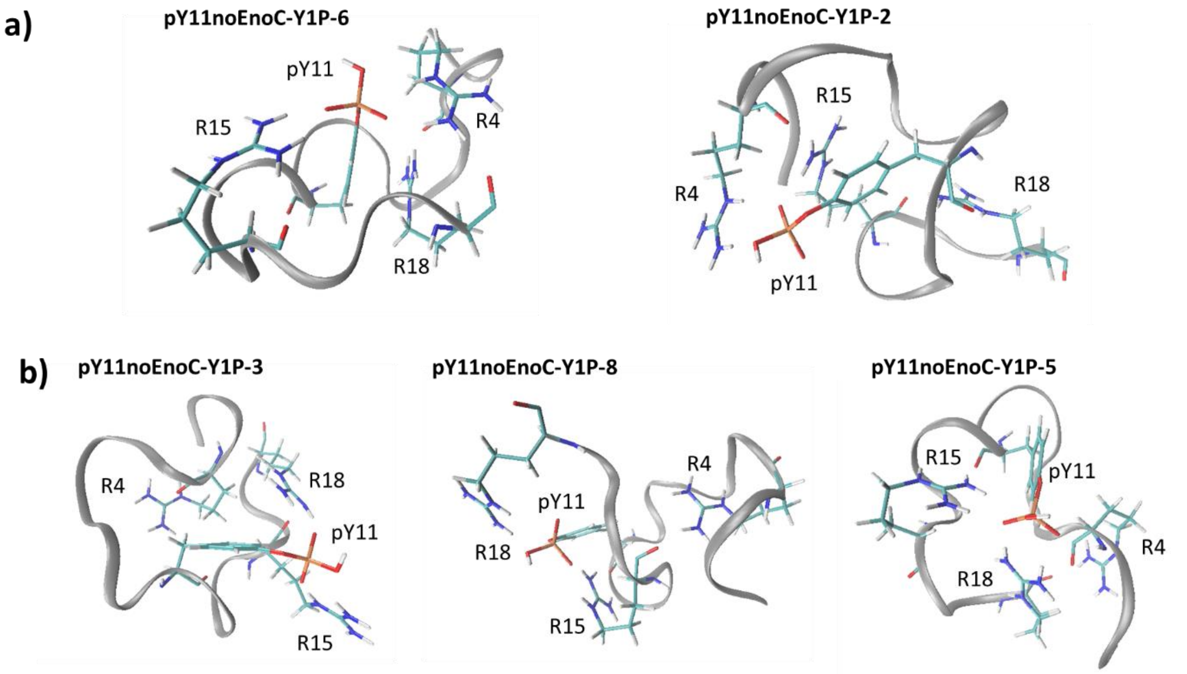


Figure 4.19 Representative structures produced from modelling GSK3β-pY11noEnoC, a mixture of which would be consistent with the ECD behaviour of the peptide: a) possible components of the mixture that prevent the detection of fragments between R4 and R15, and b) possible components of the mixture that prevent the detection of fragments between pY11 and R18

The ECD fragmentation spectrum of GSK3β-Y11noEnoC shows cleavages between all likely non-covalent interaction partners, so does not indicate the presence of any non-covalent interactions in this peptide in the gas phase. The lack of an acidic non-covalent interaction partner makes it unsurprising that there are no non-covalent interactions with the basic arginine residues that could prevent the detection of ECD fragments. The 22 representative structures produced across all of the protonation patterns modelled that had no pertinent non-covalent interactions are shown in Figure 4.20 (Y11noEnoC-RDP4-2, Y11noEnoC-RDP4-3, Y11noEnoC-RDP4-4, Y11noEnoC-RDP4-5, Y11noEnoC-RDP4-6, Y11noEnoC-RDP4-9, Y11noEnoC-RDP15-1, Y11noEnoC-RDP15-2, Y11noEnoC-RDP15-3, Y11noEnoC-RDP15-4, Y11noEnoC-RDP15-5, Y11noEnoC-RDP15-6, Y11noEnoC-RDP15-8, Y11noEnoC-RDP15-9, Y11noEnoC-RDP15-10, Y11noEnoC-RDP18-3, Y11noEnoC-RDP18-5, Y11noEnoC-RDP18-6, Y11noEnoC-RDP18-7, Y11noEnoC-RDP18-8, Y11noEnoC-RDP18-9 and Y11noEnoC-RDP18-10). These structures represented 60.22%, 90.42% and 66.66% of the Y11noEnoC-RDP4, Y11noEnoC-RDP15 and Y11noEnoC-RDP18 CCS-matched ensembles. The structures have a wide range of backbone folds, which supports the evidence from the TWIMS experiments that suggested that this peptide exists as a random coil. It is notable that most of these structures (all but Y11noEnoC-RDP4-2, Y11noEnoC-RDP4-3, Y11noEnoC-RDP4-4, Y11noEnoC-RDP4-9, and Y11noEnoC-RDP18-5) exhibit an interaction between a positively-charged arginine side chain and the deprotonated arginine residue. This interaction may give some structure to the peptide but can not be sufficiently strong to affect the peptide’s ECD behaviour. Additionally, the interaction is most prevalent between R15 and R18, which are close in the peptide sequence, indicating that the interaction is not sufficiently favoured to overcome the steric hindrance of interacting with the more distant R4 residue. Given that each protonation pattern modelled produced representative structures consistent with the ECD behaviour, the workflow employed here suggests that GSK3β-Y11noEnoC exists as a mixture of all three in the gas phase.

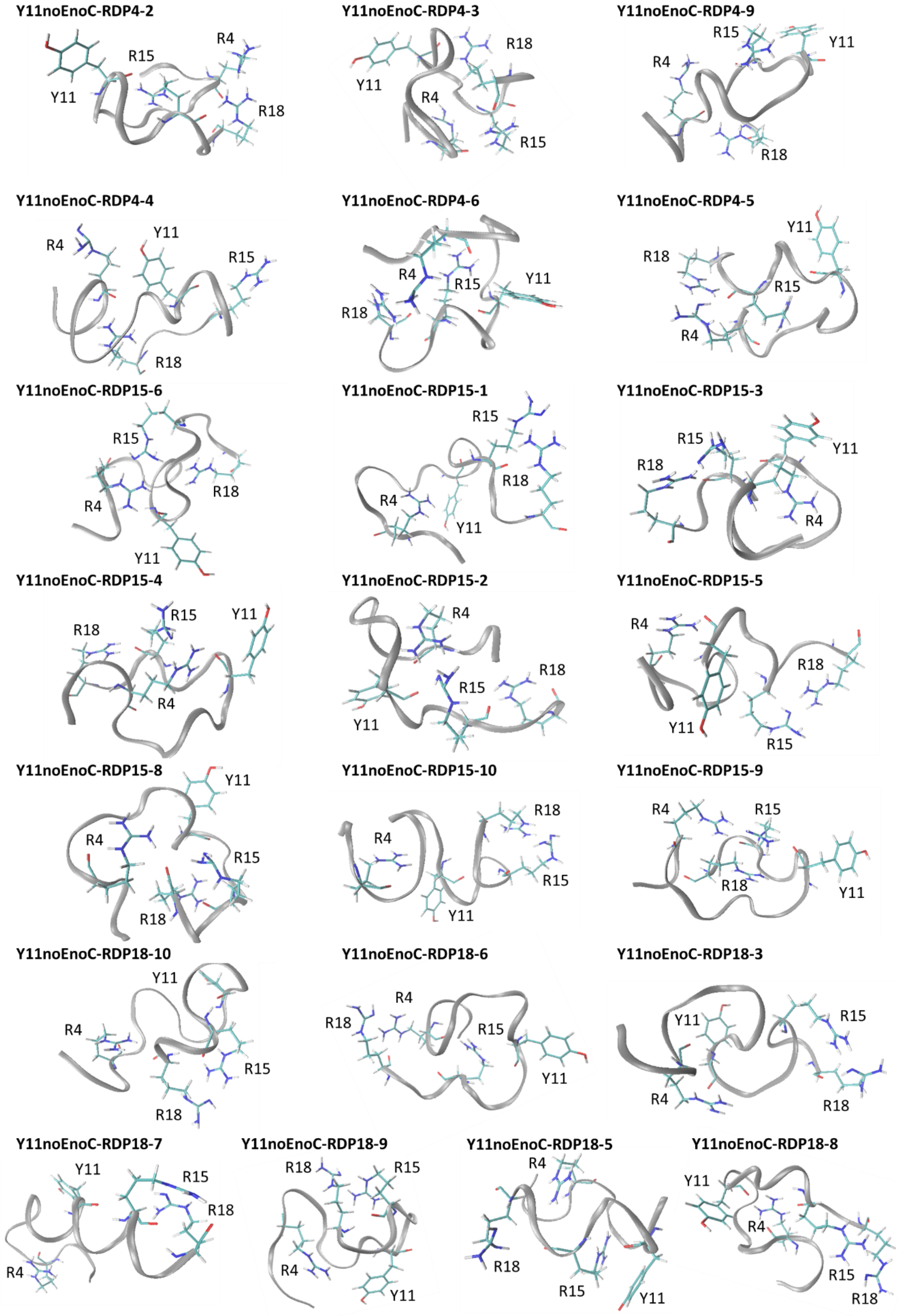


Figure 4.20 Representative structures produced from modelling GSK3β-Y11noEnoC with non-covalent interactions consistent with the ECD behaviour of the peptide

4.6 Conclusions

Application of the modelling workflow to these GSK3β-derived peptides has allowed observations made during ECD and TWIMS experiments to be related to structure. Given that these peptides had several acidic and basic residues, in many cases the non-covalent interaction pattern of a given peptide could not be conclusively assigned by experiment or simulation alone. By considering experimental data from ECD and TWIMS alongside the results of the IMS/MM workflow, a nuanced picture of the gas-phase composition for each peptide can be proposed, with some peptides best described by mixtures of structures with different non-covalent interaction patterns or from different protonation patterns. Comparison of the proposed structures for analogous peptides made it possible to suggest what impact the substitution of specific amino acid residues had on structure.

In both the GSK3β-pY11 and GSK3β-Y11 peptides, modelling showed that the (p)Y11 projected outwards, so non-covalent interactions that might affect ECD behaviour are driven by E6 in both peptides. This is consistent with the observations from the ECD experiments in which these two peptides had similar ECD behaviour. For GSK3β-pY11, the pY11 residue being in such a position that it is available to form non-covalent interactions with other molecules is consistent with the dimerisation observed for this peptide. To rationalise that GSK3β-pY11 had a broader TWIMS ATD than GSK3β-Y11, it was suggested that GSK3β-pY11 might exist as a mixture of the protonation patterns pY11-GLU and pY11-Y1P, with GSK3β-Y11 more likely to exist only as the Y11-GLU protonation pattern.

Comparison of GSK3β-pY11 and GSK3β-pS11 showed that phosphoserine seems more capable of forming non-covalent interactions with arginine residues, resulting in the population of conformations that the GSK3β-pY11 was hindered from accessing due to the steric bulk of the phosphotyrosine residue. The two had similar TWIMS and ECD behaviour, both having a large gap in fragment coverage in the middle of the sequence, which could be interpreted as them having similar structures, but the results of the ECD-guided IMS/MM workflow suggested that they have different gas-phase compositions. Meanwhile, GSK3β-Y11 and GSK3β-S11 also had similar TWIMS and ECD behaviour, and were predicted to exist in analogous protonation patterns, where all arginine residues are protonated and the glutamate residue is deprotonated (Y11-GLU and S11-GLU).

Comparison of GSK3β-pY11 and GSK3β-pY11noC candidate structures indicates that removal of the alkylated cysteine allows peptide backbone more flexibility, meaning pY11 can participate more easily in non-covalent interactions. This observation supports evidence from the AI-ECD and TWIMS that Cys-alk was having a structural effect on the peptides, where the ECD suggested that this was only an effect on electron capture. The same effect is observed for GSK3β-Y11noC, in which the candidate structures show Y11 closer to the rest of the peptide when compared to GSK3β-Y11, in which it was projected away from the rest of the peptide.

Chapter 5: Method development for the structural study of EXD fragments of phosphopeptides and phosphoproteins by EXD-ECD and ETD-TWIMS

5.1 Overview

Chapters 3 and 4 have discussed the development of an electron capture dissociation (ECD)-guided ion mobility spectrometry/molecular modelling (IMS/MM) workflow and its application to model phosphopeptides derived from the sequence of larger phosphoproteins. These studies resulted in proposed gas-phase structures for each of the peptides studied, however, comparison between these proposed gas-phase structures and x-ray crystallography structures (for those peptides where it was available), showed little similarity. It is not surprising that the gas-phase structure of a model phosphopeptide bears little resemblance to the solid-phase structure of the protein from which it was derived, given that the peptide has many more degrees of freedom than the same region in the protein and cannot form interactions with regions of the protein not included in the derived peptide. This discrepancy does mean that the model phosphopeptides cannot be relied upon as proxies for the structure of the intact protein. This chapter focusses on assessing the feasibility of using fragments produced from phosphoproteins to study the structure of the intact parent ion. Using similar methods to Chapters 3 and 4, this comprises of an ion mobility spectrometry/molecular modelling (IMS/MM) study of protein and peptide fragments, and an examination of the ECD behaviour of those fragments to probe their structure. The structures of fragments generated by the electron-mediated fragmentation methods ECD[77] and electron transfer dissociation (ETD),[103] referred to collectively as 'EXD', are investigated, as non-covalent interactions are known to survive these methods. The development and validation of parameters for the MM simulation of c^+ and $z^\bullet$ fragments is explained in Chapter 3.

This chapter describes the application of an IMS/MM workflow to ETD fragments of the GSK3β-pY11 peptide (introduced in Chapter 4) and the phosphoprotein β-casein by means of ETD-TWIMS experiments. Differences between the structure of the z_{17}^{2+} fragment of five GSK3β-derived peptides (introduced in Chapter 4) generated by ETD and ECD are probed by MS^3 (ETD-ECD and ECD-ECD) experiments.

5.2 ETD-TWIMS

IMS has been used successfully in Chapters 3 and 4 to study the structure of intact phosphopeptides. It was therefore natural to investigate whether it could also be used to study the fragments of phosphopeptides. The ETD option available on the Synapt G2S enables ETD to be performed on an instrument capable of ion mobility, with ETD reagent introduced in the source region, while the majority of the ETD reaction occurs in the trap, located before the TWIMS cell in the Tri-Wave region of the instrument. A schematic of the ETD system and how it interfaces with the rest of the instrument is shown in Figure 5.1, while a more detailed discussion of the geometry of the Synapt G2S can be found in Section 1.9.1. This configuration allows a sample to be fragmented by ETD followed by ion mobility separation of its fragments in the TWIMS device. In this way it was possible to measure the arrival times of ETD fragments, from which their collision cross sections (CCS) could be derived by means of a calibration against a sample with known CCS. The CCS of fragments could then be used to filter MM structures, as in the IMS/MM workflows previously described and implemented in Chapters 3 and 4. Using an IMS/MM workflow to study EXD fragments should allow the prediction of the structure of fragments, which could then be compared back to predictions of the intact ion structure. ETD-TWIMS was employed to perform this IMS/MM study, though the technique also allowed the observation of the ion mobility behaviour of fragments, which in itself provided information about the mechanism of ETD. At the time of performing this work, ETD was the only EXD fragmentation technique available on the Synapt G2S (used in IMS/MM studies in Chapters 3 and 4), though recent studies have implemented ECD fragmentation of proteins in a linear ECD cell both before and after the TriWave device.[99]

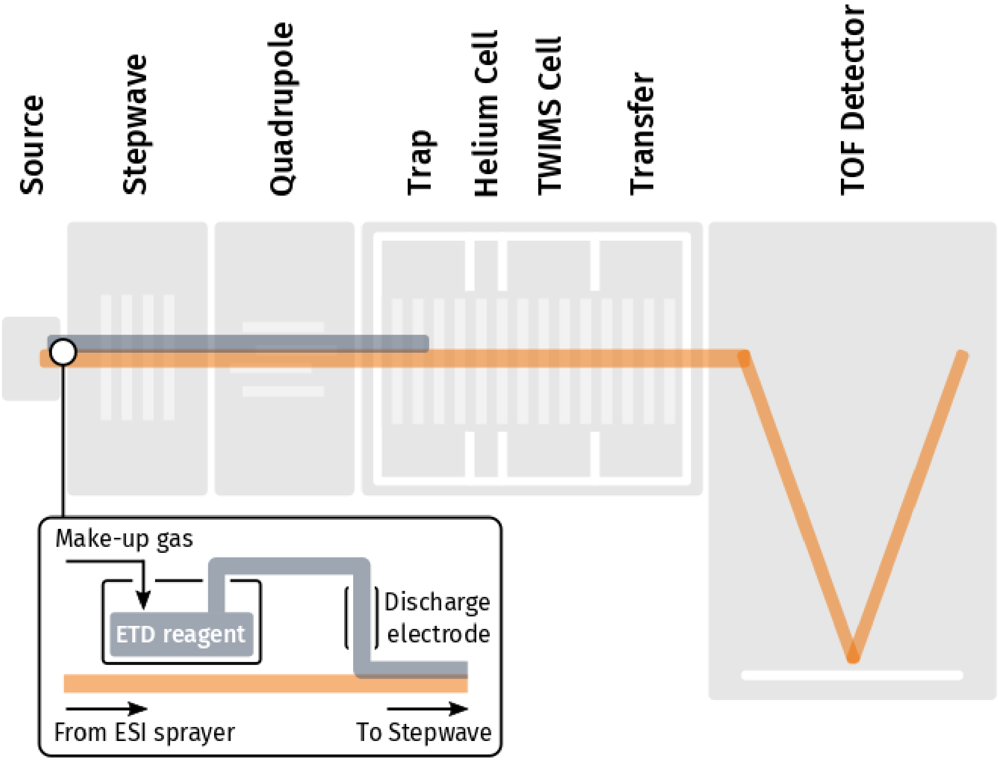


Figure 5.1 Schematic of the Synapt G2S mass spectrometer with the ETD system installed. The ETD reagent is delivered by a make-up gas to the ion beam via a discharge electrode which causes the ETD reagent molecules to become anion radicals. The ETD reagent coexists with the ion beam between the source and the trap, where most of the ETD reaction occurs

5.2.1 ETD-TWIMS behaviour of GSK3β-pY11

Initially, ETD-TWIMS experiments were performed on the GSK3β-pY11 peptide as the structure of this model system has been probed extensively during the work presented in Chapter 4. This peptide was shown to dimerise, which complicated the ECD MS/MS spectrum, as peaks could not easily be assigned to either the dimer or the monomer without performing in-source CID to dissociate the dimer into its monomer units. The ECD behaviour of the GSK3β-pY11 $[M+2H]^{2+}$ monomer ion suggested non-covalent interactions between R4, R15 and acidic residues in the peptide due to a lack of fragments observed between R4 and Y17 (with in-source CID used to dissociate the dimer). The IMS/MM study of the monomer of this peptide led to the proposal of the non-covalent interactions E6-R4 and E6-R15, where E6 is deprotonated. In order to rationalise the broad ATD profile of GSK3β-pY11, it was suggested that the peptide exists as a mixture of structures with R4-X-R15 and R4-X-R18 non-covalent interaction patterns (where X is pY11 or E6), which would have the same ECD behaviour as structures with the R4-X-R15 interaction pattern alone. The IMS/MM workflow suggested that this mixture would contain structures with E6 protonated and pY11 deprotonated as well as those with E6 deprotonated and pY11 protonated. The ECD behaviour and IMS/MM study of GSK3β-pY11 are discussed in detail in Sections 4.3 and 4.5, respectively.

The ETD MS/MS spectrum of the $[M+2H]^{2+}$ ion of GSK3β-pY11, summed over the entire drift time (DT) range recorded, is shown in Figure 5.2 (wave velocity = 600 ms^{-1}, wave height = 25 V), with data collected using the methods in Section 2.2.5. The corresponding fragment assignment list can be found in Appendix 6-17. This ETD MS/MS spectrum demonstrated the same fragment coverage as was observed for this peptide by ECD with in-source dissociation (Figure 4.5, reproduced here as Figure 5.3). The relative intensity of the charge reduced species (m/z = 2381.90) compared to the parent ion (m/z = 1191.18) was significantly lower in this ETD MS/MS spectrum than in the ECD MS/MS spectrum, as was the ratio between the charge reduced species and the fragment ions. This observation is consistent with electron transfer being a less efficient process than electron capture,[203] and also suggests that ETD products are less likely to dissociate into fragments from the charge reduced species. Despite the absence of the peak associated with the intact dimer of GSK3β-pY11, several 3+ fragments resulting from fragmentation of the dimeric species $[2M+4H]^{4+}$ were recorded between 1400-1800 m/z. The presence of these 3+ peaks produced from the $[2M+4H]^{4+}$ dimer implies that the conditions between the source and trap (where these fragments must have formed by ETD fragmentation) were insufficient to eliminate the dimer, and that the signal from the $[2M+4H]^{4+}$ ion was obscured by the high baseline of the isobaric $[M+2H]^{2+}$ peaks. There are fewer peaks resulting from the fragmentation of the dimer in this ETD spectrum than in the ECD MS/MS spectrum without in-source CID (Figure 5.3c), which may indicate that the conditions used in the ETD experiment result in a lower ratio of dimer:monomer. Because of the presence of the dimer, singly-charged fragment peaks could not easily be assigned to either the dimer or the monomer in this spectrum. Further, differences between the ECD fragment coverages of the dimer/monomer mixture or the monomer alone, and the ETD fragment coverage of the dimer/monomer mixture may be explained by the lower efficiency of ETD fragmentation rather than by structural differences between the peptide under the conditions of the this experiment and the conditions of the ECD experiments.

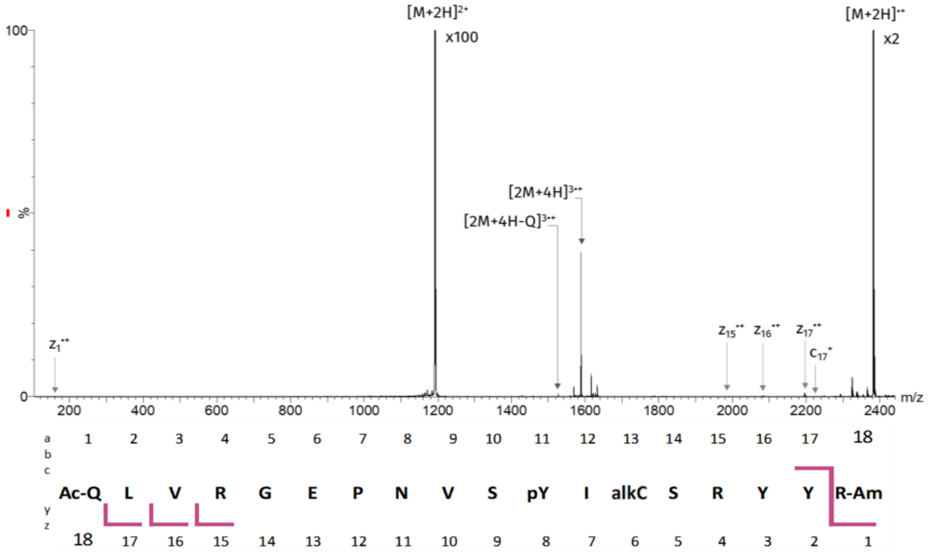


Figure 5.2 The ETD-TWIMS (wave velocity = 600 ms^{-1}, wave height = 25 V) MS/MS spectrum of GSK3β-pY11 [M+2H]$^{2+}$ summed across all drift time values and the corresponding fragmentation pattern

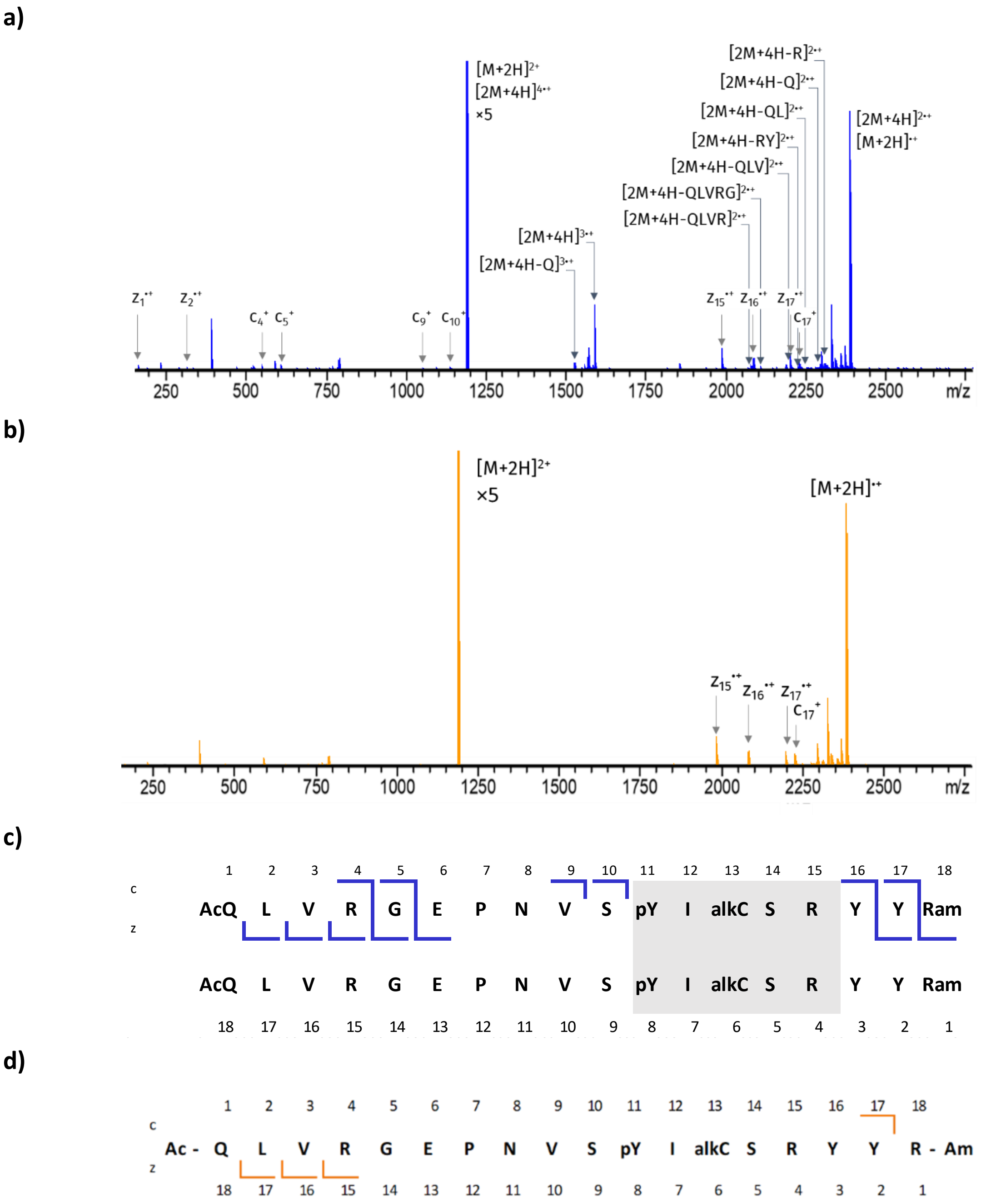
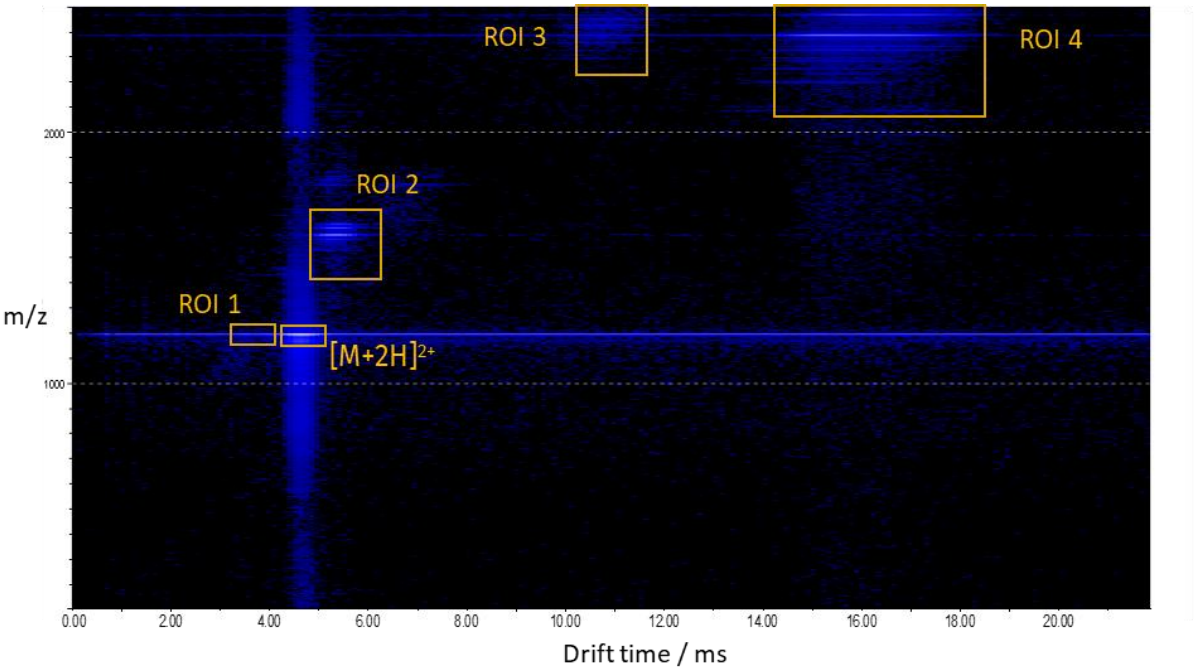


Figure 5.3 ECD MS/MS spectrum of the GSK3B-derived peptide $[M+2H]^{2+}$ ion a) without in-source CID showing peaks formed from dissociation of the dimeric $[2M+4H]^{4+}$ and b) with 90 V in-source CID applied, showing only fragments from the monomer. The fragmentation patterns are shown for c) the dimeric species and d) the monomer. Note: the fragmentation pattern for the dimeric species assumes that all fragments in the ECD MS/MS spectrum without in-source CID are formed from the dimer except $c_{17}^{\bullet+}$, $z_{17}^{\bullet+}$, $z_{16}^{\bullet+}$ and $z_{15}^{\bullet+}$, and shows one possible interface region between the two peptide units in grey

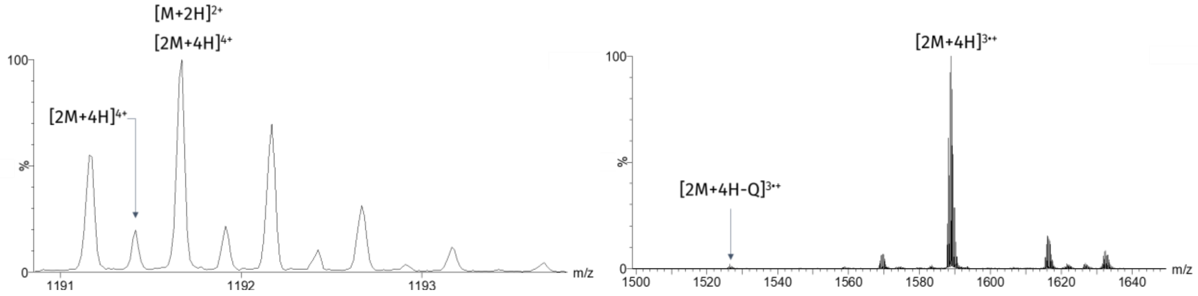
From TWIMS data, a ‘drift plot’ of m/z against drift time can be generated in which relative abundance is represented by a colour scale. The drift plot for the ETD-TWIMS data of the GSK3β-pY11 $[M+2H]^{2+}$ ion (wave velocity = 600 ms^{-1}, wave height = 25 V) is shown in Figure 5.4. The most intense peak in the drift plot was the parent $[M+2H]^{2+}$ ion, which had an apex drift time of 4.63 ms, although this peak could be observed across the full drift time range. Similarly, the $[M+2H]^{\bullet+}$ m/z peak can be observed across the full drift time range, while having distinct peaks at 15.44 ms and 16.23 ms. The parent and charge reduced species are very intense compared to any other peaks, and as such the baselines of their drift time peaks contain more signal than the apex of drift time peaks of other species. Likewise, the baseline noise generated from the $[M+2H]^{2+}$ peak was observed across the full m/z range at a drift time of 4.63 ms, although y_{18}^{2+} and $[M+2H\text{-}NH_3]^{2+}$ peaks were also present at this drift time. The baselines produced by the $[M+2H]^{2+}$ ion in the m/z and drift time dimensions can be clearly seen in Figure 5.4a as intense horizonal and vertical bands at m/z = 1192 and DT = 4.63.

The ETD-TWIMS data allows the dimeric and monomeric GSK3β-pY11 species and their fragments to be better discerned from each other. Figure 5.4b shows the mass spectrum for the drift time range at which the $[2M+4H]^{4+}$ dimer was observed (apex drift times = 3.53 ms and 4.08 ms), seen here along with the isobaric $[M+2H]^{2+}$ monomer peak. The 3+ charged fragments generated from the $[2M+4H]^{4+}$ dimer are not isobaric with any fragments generated from the $[M+2H]^{2+}$ monomer, but are also well separated in the drift time domain (Figure 5.4c), having drift times between 5.18 ms and 5.42 ms. Notably, the $z_1^{\bullet+}$ fragment has a drift time in this range (apex drift time = 5.40 ms), which is uncharacteristic for a fragment with m/z of 158.13 and suggests that this fragment has formed from the dimeric species alongside its 3+ fragments. The charge reduced species $[2M+4H]^{2\bullet+}$ generated from the $[2M+4H]^{4+}$ dimer (apex drift time = 10.70 ms) (Figure 5.4d) is isobaric with the $[M+2H]^{\bullet+}$ species generated from the $[M+2H]^{2+}$ monomer, however, mobility separation allows complete separation of the two species. The charge reduced species $[M+2H]^{\bullet+}$ and fragments ions generated from the $[M+2H]^{2+}$ monomer, shown in Figure 5.4e, are observed in a broad band between 14 and 18 ms, with drift time peaks for individual fragments not well resolved from each other.

a) Full drift plot



b) ROI 1 **c) ROI 2**



d) ROI 3 **e) ROI 4**

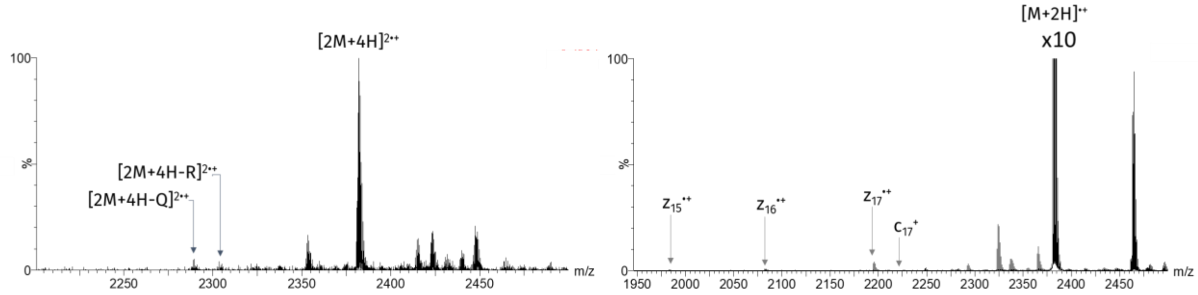


Figure 5.4 ETD-TWIMS (wave velocity = 600 ms^{-1}, wave height = 25 V) MS/MS data for GSK3β-pY11 $[M+2H]^{2+}$ a) represented as a drift plot with various regions of interest labelled: b) the mass spectrum summed over the drift time range at which the $[2M+4H]^{4+}$ dimer is observed, c) the mass spectrum of 3+ charged fragments generated from the $[2M+4H]^{4+}$ dimer, d) the charge reduced species $[2M+4H]^{3\bullet+}$ generated from the $[2M+4H]^{4+}$ dimer, and e) the charge reduced species $[M+2H]^{\bullet+}$ and fragments generated from the $[M+2H]^{2+}$ monomer

ETD-TWIMS was also performed on the $[M+3H]^{3+}$ ion of GSK3β-pY11 (wave velocity = 800 ms^{-1}, wave height = 23 V). The ETD MS/MS spectrum summed over the whole drift time range is shown in Figure 5.5a, with z and c fragments labelled. A fragment assignment list can be found in Appendix 6-18. The fragmentation pattern (Figure 5.5b) shows nearly full fragment coverage in a, b, c, y and z fragments, with only one missed cleavage between I12 and alkC13. Considering only c and z fragments, the fragment coverage remains the same, with no gaps in fragment coverage that would indicate a non-covalent interaction between acidic and basic residues (i.e., a gap extending from one acidic/basic residue to another caused by non-covalent interactions between the two that prevent the dissociation and detection of fragments). The $[M+3H]^{3+}$ ion is expected to have greater fragment coverage for two reasons: firstly, EXD efficiency increases proportionally to the square of the charge of the ion.[81] Secondly, charge repulsion between ETD fragments may be sufficient to overcome any non-covalent interactions between them, favouring their dissociation.[100]

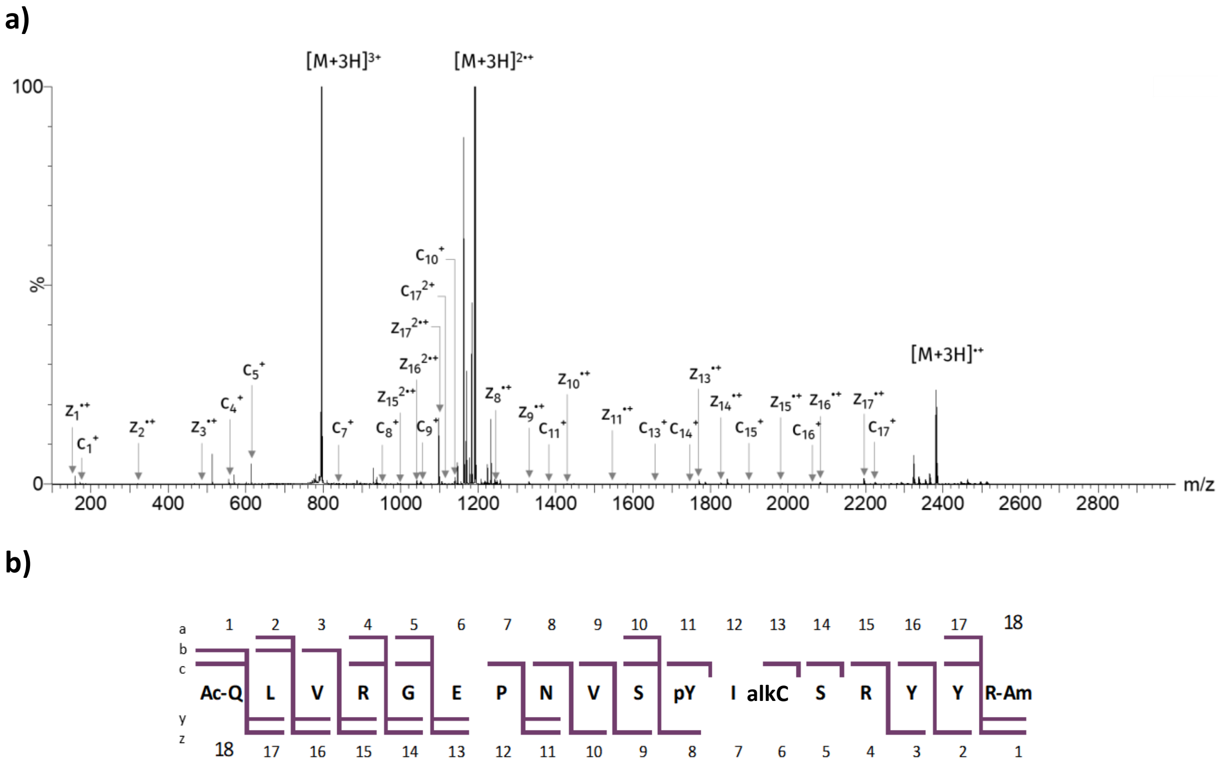


Figure 5.5 a) ETD-TWIMS (wave velocity = 800 ms^{-1}, wave height = 23 V) MS/MS spectrum of GSK3β-pY11 $[M+3H]^{3+}$ summed across all drift time values and b) the corresponding fragmentation pattern

The drift plot for GSK3β-pY11 $[M+3H]^{3+}$ (Figure 5.6a) has many more features than the $[M+2H]^{2+}$ ion, as the ETD of this charge state produces more fragments. The dimer was not observed in the $[2M+6H]^{6+}$ charge state, so the $[M+3H]^{3+}$ charge state is not underlaid by dimer peaks. As seen for the $[M+2H]^{2+}$

charge state, the baseline from m/z peaks of parent $[M+3H]^{3+}$ and charge reduced species, $[M+3H]^{2\bullet+}$ and $[M+3H]^{\bullet+}$, can be seen across the full drift time range, and the baseline from their respective drift time peaks can be seen across the full m/z range. This effect is attributed to the intensities of these peaks being orders of magnitude larger than any of the other peaks in the spectrum. However, many fragment peaks are observed at the same drift time as the parent $[M+3H]^{3+}$ and especially the charge reduced species $[M+3H]^{2\bullet+}$, which have apex drift times of 2.09 ms and 4.19 ms, respectively. The presence of these peaks at the same drift times as $[M+3H]^{3+}$ and $[M+3H]^{2\bullet+}$ is especially obvious in Figure 5.6b, where the apex drift time peak values are plotted for the monoisotopic m/z values of fragment peaks, so removing the baseline noise from $[M+3H]^{3+}$ and $[M+3H]^{2\bullet+}$. Many of the fragments observed at the same drift time as the charge reduced species $[M+3H]^{2\bullet+}$ have very different m/z values to $[M+3H]^{2\bullet+}$ and would not be expected to have the same drift time, given that drift time is proportional to mass. Most of these fragments also have a second drift time peak that is closer to that which would be expected based on their mass. For instance, the $z_2^{\bullet+}$ peak (m/z = 321.14) would be expected based on its m/z to have a drift time much lower than $[M+3H]^{2\bullet+}$ (m/z = 1191.1), but has drift time peaks at 4.3 ms and 1.32 ms. This observation suggests that the $z_2^{\bullet+}$ fragment with drift time 1.32 ms has formed before the TWIMS device (i.e., either between the source and the trap, where the analyte and the ETD reagent coexist, or in the trap immediately before TWIMS cell, in which the majority of the ETD reaction occurs) and transits the TWIMS cell with a drift time related to its mobility, whereas the $z_2^{\bullet+}$ fragment with drift time 4.3 ms is formed from the $[M+3H]^{2\bullet+}$ charge reduced species by dissociation after the TWIMS cell, and so is recorded as having the same drift time as the $[M+3H]^{2\bullet+}$ charge reduced species. By this rationale, those fragments detected at the same drift time as $[M+3H]^{2\bullet+}$ seem to be undergoing a slow dissociation process over the course of at least the transfer time between the source and the TWIMS cell as they transit through the mass spectrometer. That this dissociation process occurs during fragmentation of GSK3β-pY11 $[M+3H]^{3+}$ and not GSK3β-pY11 $[M+2H]^{2+}$ suggests that it may be due in part to charge repulsion between fragments that are bound through non-covalent interactions. Conversely, a number of the fragments observed at the same DT as the $[M+3H]^{2\bullet+}$ charged reduced species are doubly-charged species themselves, so they can only have been formed from $[M+3H]^{2\bullet+}$ by loss of a neutral moiety, which would not exhibit any charge repulsion. The population of fragments formed by dissociation after the TWIMS cell may be caused by collisional activation of these fragments in the TWIMS cell. A study by Merenbloom *et al.* showed that wave velocities of 900 ms^{-1} were sufficient to cause CID-type fragmentation of the leucine enkephalin pentapeptide,[166] so the 800 ms^{-1} wave velocity used in this ETD experiment of GSK3β-pY11 $[M+3H]^{3+}$ may cause collisional activation of $[c\text{-}z]^{2\bullet+}$ complexes, leading to their dissociation after the TWIMS device. Collisional activation in the TWIMS cell also explains the large number of a, b and y fragments observed in this experiment. The ETD experiment of GSK3β-pY11 $[M+2H]^{2+}$ was performed at a wave velocity of 600 ms^{-1}, so fragments formed in this experiment would undergo less collisional activation. Notably, in the ETD experiment of GSK3β-pY11 $[M+2H]^{2+}$, only the $z_1^{\bullet+}$ fragment seemed to be formed by dissociation of another species after the TWIMS cell.

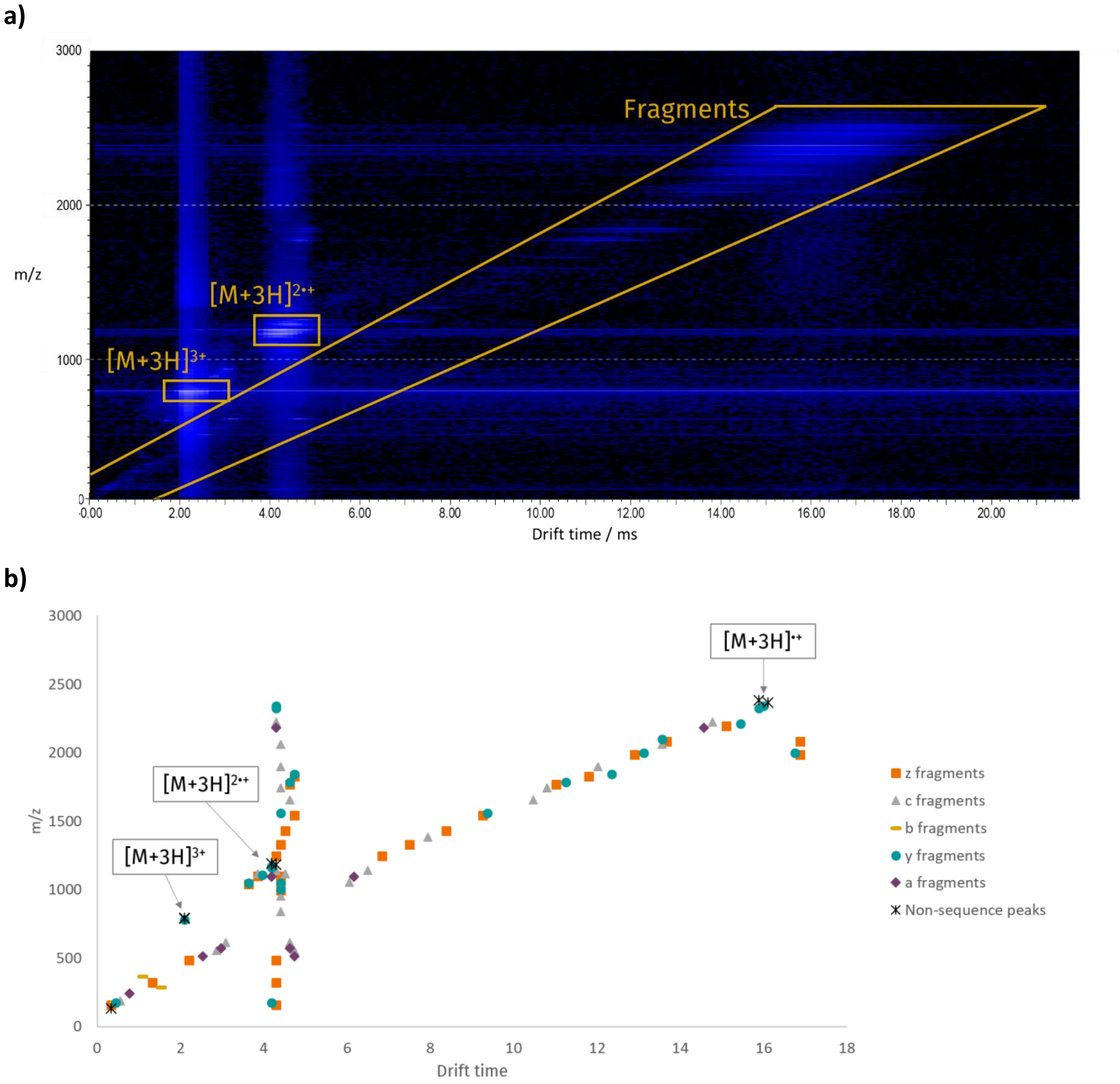


Figure 5.6 ETD-TWIMS data for GSK3β-pY11 $[M+3H]^{3+}$ (wave velocity = 800 ms^{-1}, wave height = 23 V) a) represented as a drift plot with regions corresponding to the parent $[M+3H]^{3+}$, charge reduced species $[M+3H]^{2+}$ and fragments labelled; and b) the apex drift time peak values for individual m/z peaks, stratified by fragment type

This ETD-TWIMS experiment provides a way to discern between the populations of ETD fragments formed before the TWIMS cell, and the population that dissociated from the $[c\text{-}z]^{n\bullet+}$ complex after the TWIMS cell. The latter population must be bound by non-covalent interactions that are only broken after transiting the TWIMS cell. By excluding fragments with a drift time peak the same as $[M+3H]^{2\bullet+}$, it is possible to produce a fragmentation pattern of just the population of fragments formed before the TWIMS cell which are not affected by non-covalent interactions within the parent ion (Figure 5.7).

This fragmentation pattern shows much reduced fragment coverage, with no cleavages observed between R4 and R18. This fragment coverage would suggest that the non-covalent interactions in the $[M+3H]^{3+}$ ion are between R4 and R18 and either E6 or pY11, with R15 possibly being involved.

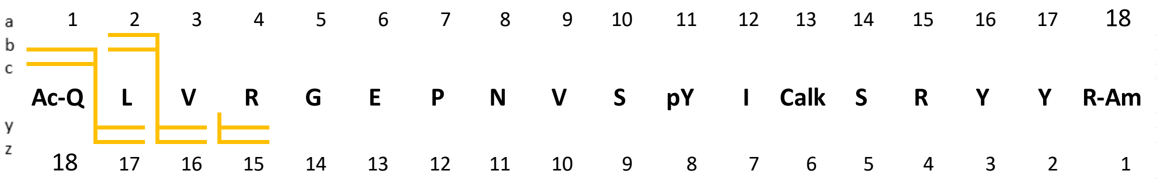


Figure 5.7 ETD-TWIMS (wave velocity = 800 ms^{-1}, wave height = 23 V) fragmentation pattern of GSK3β-pY11 $[M+3H]^{3+}$ with fragments that have a drift time peak the same as $[M+3H]^{2+}$excluded

5.2.2 Application of IMS/MM workflow to ETD fragments GSK3β-pY11 and β-casein

Having recorded drift times for the fragments of the GSK3β-pY11 $[M+2H]^{2+}$ and $[M+3H]^{3+}$ ions, it was possible to derive their collision cross sections by means of calibration to a reference sample of tryptic peptides, as described in Section 2.2.5. For the GSK3β-pY11 $[M+2H]^{2+}$ ETD-TWIMS experiment, sample and calibrant drift times were recorded at a wave velocity of 600 ms^{-1} and wave heights of 24.5 V, 25 V and 25.5 V, and calibration was performed to 1+ reference peptides; while for the GSK3β-pY11 $[M+3H]^{3+}$ experiment, these were recorded at a wave velocity of 800 ms^{-1} and wave heights of 21 V, 22 V and 23 V, and calibration was performed to 2+ reference peptides. The list of calibrant ions used for each experiment can be found in Appendix 3-3 and Appendix 3-2, while the calibrations curves are in Appendix 4-6 and Appendix 4-7. The CCS of these fragments are shown in Table 5.1. For GSK3β-pY11 $[M+3H]^{3+}$ fragments, additional drift time peaks in the same range as the charge reduced species $[M+3H]^{2\bullet+}$ are omitted from the table.

Table 5.1 Theoretical monoisotopic masses and experimentally-derived CCS for fragments of GSK3β-pY11 $[M+2H]^{2+}$ and $[M+3H]^{3+}$

Parent ion	Fragment	Monoisotopic mass (Da)	$^{TW}CCS_{N2\rightarrow He}$ (Å^2)
$[M+2H]^{2+}$	$z_1^{\bullet+}$	158.13	224.05 ± 1.24
$[M+2H]^{2+}$	$z_{15}^{\bullet+}$	1982.86	430.17 ± 4.30
$[M+2H]^{2+}$	$z_{16}^{\bullet+}$	2081.95	381.19 ± 2.26
$[M+2H]^{2+}$	$z_{16}^{\bullet+}$	2081.95	435.79 ± 4.36
$[M+2H]^{2+}$	$z_{17}^{\bullet+}$	2194.97	406.42 ± 3.98
$[M+2H]^{2+}$	c_{17}^{+}	2224.92	408.56 ± 3.46
$[M+2H]^{2+}$	$[M+2H]^{\bullet+}$	2381.90	406.48 ± 4.89
$[M+3H]^{3+}$	z_{16}^{2+}	2081.95	371.03 ± 3.29
$[M+3H]^{3+}$	z_{17}^{2+}	2194.97	381.56 ± 2.39
$[M+3H]^{3+}$	c_{17}^{2+}	2224.92	384.40 ± 0.47
$[M+3H]^{3+}$	y_{16}^{2+}	2097.21	372.45 ± 1.80
$[M+3H]^{3+}$	y_{17}^{2+}	2210.25	388.92 ± 0.79
$[M+3H]^{3+}$	$[M+3H]^{2\bullet+}$	2381.90	397.47 ± 1.41

These CCS could be used to perform an IMS/MM study of any of these fragments. In this work, the $z_{17}^{\bullet+}$ fragment of GSK3β-pY11 $[M+2H]^{2+}$ and z_{17}^{2+} fragment of GSK3β-pY11 $[M+3H]^{3+}$ were further investigated. These two fragments are the highest abundance sequence fragments in their respective ETD MS/MS spectra. The m/z peaks and corresponding arrival time distributions (ATD) for the $z_{17}^{\bullet+}$ fragment from GSK3β-pY11 $[M+2H]^{2+}$ and the z_{17}^{2+} fragment of GSK3β-pY11 $[M+3H]^{3+}$ are shown in Figure 5.8. The ATD for $z_{17}^{\bullet+}$ from $[M+2H]^{2+}$ is less smooth than the z_{17}^{2+} from $[M+3H]^{3+}$ due to the much lower signal intensity of the former. Two peaks are present in the ATD of the z_{17}^{2+} fragment from $[M+3H]^{3+}$: the more abundant peak at 3.85 ms represents the arrival time of the fragment itself, from which the CCS in Table 5.1 is calculated, while the other (apex drift time 4.41 ms) is in the same drift time range as the charge reduced species $[M+3H]^{2\bullet+}$ and so has not been used to calculate the CCS of the fragment.

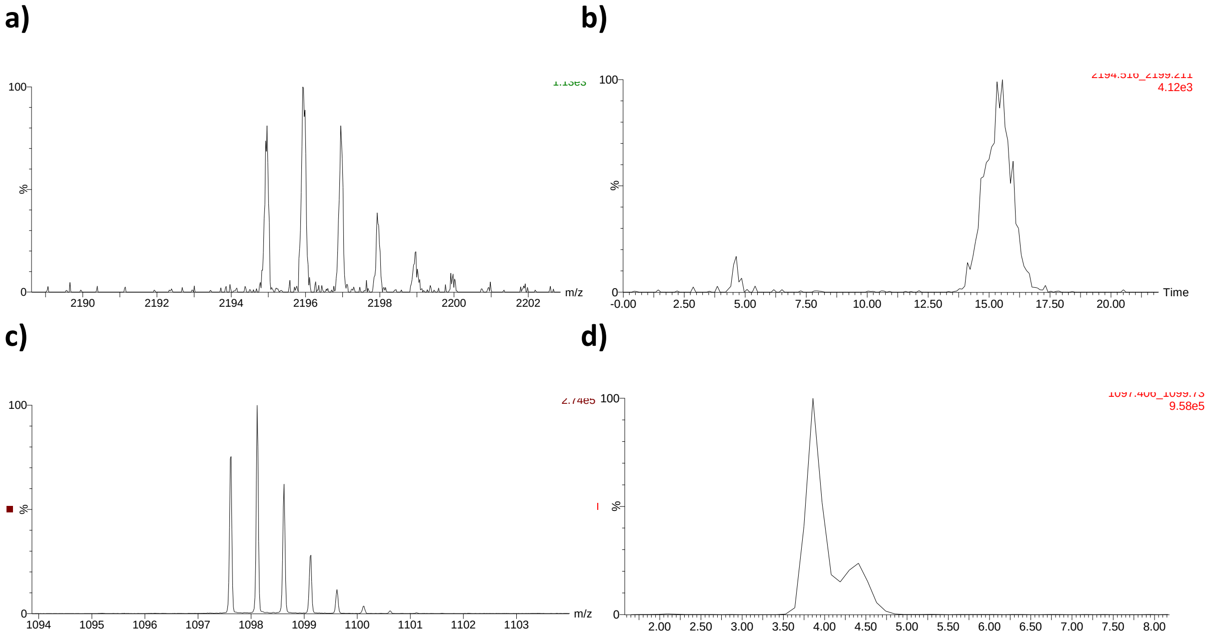


Figure 5.8 a) The m/z peak corresponding to GSK3β-pY11 $[M+2H]^{2+}$ $z_{17}^{\bullet+}$ and b) its corresponding arrival time distribution, and c) the m/z peak corresponding to GSK3β-pY11 $[M+3H]^{3+}$ z_{17}^{2+} and d) its corresponding arrival time distribution

As in Chapters 3 and 4, multiple protonation patterns for the two fragments were modelled and these are shown in Table 5.2, in which acidic and basic residues are labelled with $^+$, $^-$ or ^{0} to indicate whether they are positive, negative or neutral, respectively. As in Section 3.1.4, the fragments have been modelled with the assumption that the radical is localised to the residue adjacent to the ETD cleavage site (here leucine). The choice of protonation patterns to model was informed by the likely protonation patterns of the parent ion and the likely sites that could be deprotonated by the ETD process. It is noteworthy that this is not an exhaustive modelling study and assumptions about the likely protonation state of the parent ion and about the ETD process may influence the results of this study. The IMS/MM study of GSK3β-pY11 $[M+2H]^{2+}$, detailed in Section 4.2, gave the protonation pattern as pY11-GLU (Ac- Q L V R$^+$ G E$^-$ P N V S pY^0 I alkC S R$^+$ Y Y R$^+$-Am), in which all three arginine residues carry a positive charge, the glutamate residue is negative and the phosphotyrosine is neutral. Transfer of an electron to this peptide would result in the deprotonation of one of the arginine residues, all three of which are present in the z_{17} fragment, leading to proposed protonation patterns for $z_{17}^{\bullet+}$ with two arginine residues positively charged, one neutral arginine, a negative glutamate residue and the phosphotyrosine neutral. The GSK3β-pY11 $[M+3H]^{3+}$ parent ion has not been subjected to an IMS/MM study, however, given the gas-phase acidities of the residues involved, it most likely exists in a protonation pattern with all three arginine residues protonated and both the glutamate and phosphotyrosine neutral (Ac- Q L V R$^+$ G E^0 P N V S pY^0 I alkC S R$^+$ Y Y R$^+$-Am). As such, protonation patterns to be modelled for $z_{17}^{2+\bullet}$ contain two arginine residues positively charged, one neutral arginine, the glutamate neutral and the phosphotyrosine neutral.

Table 5.2 Protonation patterns modelled for the $z_{17}^{\bullet+}$ fragment of GSK3β-pY11 [M+2H]$^{2+}$ and z_{17}^{2+} of GSK3β-pY11 [M+3H]$^{3+}$

Parent ion	Fragment	Model name	Sequence
GSK3β-pY11 [M+2H]$^{2+}$	$z_{17}^{\bullet+}$	pY11z17_1pos_RDP4	$^{\bullet}$L V R^0 G E$^-$ P N V S pY^0 I alkC S R$^+$ Y Y R$^+$-Am
		pY11z17_1pos_RDP15	$^{\bullet}$L V R$^+$ G E$^-$ P N V S pY^0 I alkC S R^0 Y Y R$^+$-Am
		pY11z17_1pos_RDP18	$^{\bullet}$L V R$^+$ G E$^-$ P N V S pY^0 I alkC S R$^+$ Y Y R^0-Am
GSK3β-pY11 [M+3H]^3	$z_{17}^{2+\bullet}$	pY11z17_2pos_RDP4	$^{\bullet}$L V R^0 G E^0 P N V S pY^0 I alkC S R$^+$ Y Y R$^+$-Am
		pY11z17_2pos_RDP15	$^{\bullet}$L V R$^+$ G E^0 P N V S pY^0 I alkC S R^0 Y Y R$^+$-Am
		pY11z17_2pos_RDP18	$^{\bullet}$L V R$^+$ G E^0 P N V S pY^0 I alkC S R$^+$ Y Y R^0-Am

These six protonation patterns were each modelled by the workflow described in Section 3.4, i.e., first a three-tiered simulated annealing experiment to produce a 25,000-strong MM ensemble, followed by calculation of the CCS for each of these models and then filtering to include only models with a CCS within 5% of the experimentally derived CCS, producing a CCS-matched ensemble. The CCS-matched ensemble then underwent dimensionality reduction and clustering to produce 10 representative structures for each of the protonation patterns modelled. Unlike in Section 3.4, all 10 representative structures for these fragments are discussed and compared to the non-covalent interactions in their respective parent ions, rather than those in the fragments themselves. In order to predict non-covalent interactions in the fragments, they would have to be subjected to further ECD fragmentation. ECD products of $z_{17}^{\bullet+}$ would be neutral and not detectable by mass spectrometry, while the ECD behaviour of the $z_{17}^{2+\bullet}$ is discussed in Section 5.3. The definition of a non-covalent interaction here is the same as in Section 4.2: when a hydrogen atom in the guanidine group of any arginine residue is within 4 Å of the centre of an acidic residue (P in pTyr and δC in Glu). Details of the clusters produced by the IMS/MM workflow for all six protonation patterns modelled are found in Appendix 1-11, and distances between acidic and basic residues in the 10 representative structures for all six protonation patterns are found in Appendix 1-12. The representative structures themselves can be found in Appendix 1-15, where they are in ascending order of structure number, as in the corresponding part of Appendix 1-11.

The 10 representative structures for each of the models of the $z_{17}^{\bullet+}$ fragment of GSK3β-pY11 $[M+2H]^{2+}$ (pY11z17_1pos_RDP4, pY11z17_1pos_RDP15 and pY11z17_1pos_RDP18) are shown in Figure 5.9, Figure 5.10 and Figure 5.11, respectively. All these representative structures show at least one non-covalent interaction to Glu, while none show any non-covalent interactions involving pTyr. The presence of non-covalent interactions to Glu across all the representative structures implies that this interaction is maintained within fragments through the ETD process, as well as being capable of hindering the dissociation of ETD fragments from each other. Like the parent ion, many of these structures show the pTyr projected away from the rest of the peptide, implying that this feature might also be retained between the parent ion and the $z_{17}^{\bullet+}$ fragment. Only three of these representative structures have the same non-covalent interaction pattern (E6-R4 and E6-R15) as the parent ion: pY11z17_1pos_RDP4-5, pY11z17_1pos_RDP18-8 and pY11z17_1pos_RDP18-6. That no representative structures from pY11z17_1pos_RDP15 have these interactions suggests that a positively-charged R15 residue would be necessary if the fragment were to sustain these interactions.

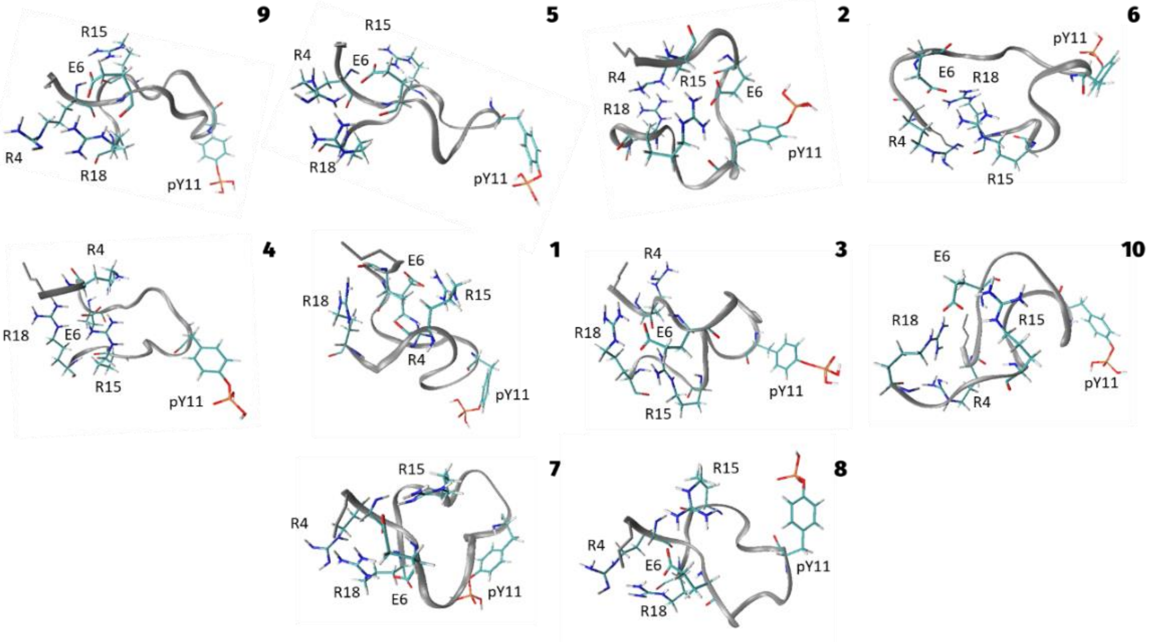


Figure 5.9 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble pY11z17_1pos_RDP4. The number in bold next to each structure is the number of the cluster that it represents

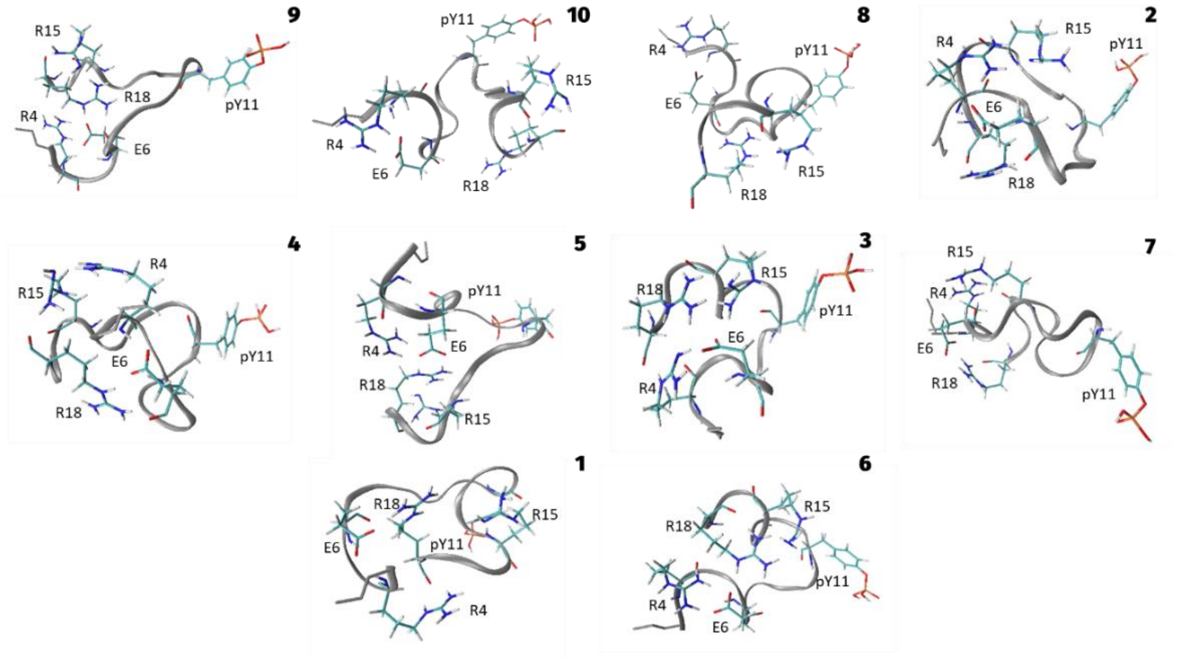


Figure 5.10 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble pY11z17_1pos_RDP15. The number in bold next to each structure is the number of the cluster that it represents

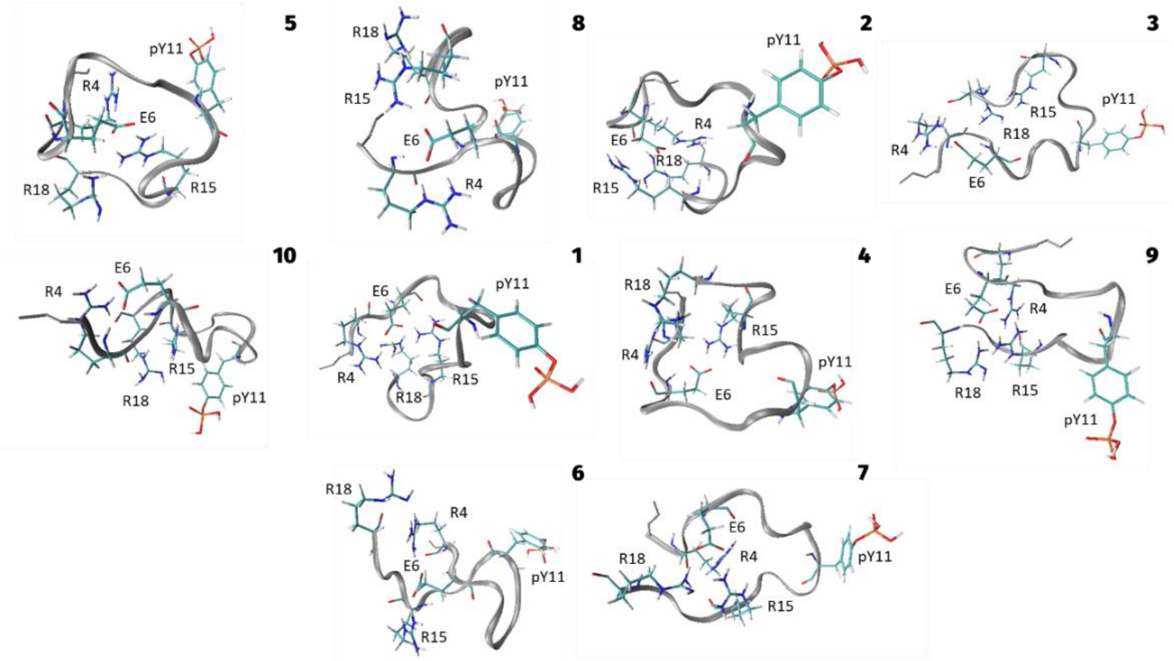


Figure 5.11 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble pY11z17_1pos_RDP18. The number in bold next to each structure is the number of the cluster that it represents

The 10 representative structures produced from modelling the three protonation patterns of the z_{17}^{2+} fragment of GSK3β-pY11 $[M+3H]^{3+}$ (pY11z17_2pos_RDP4, pY11z17_2pos_RDP15 and pY11z17_2pos_RDP18) are shown in Figure 5.12, Figure 5.13 and Figure 5.14, respectively. With Glu and pTyr both being neutral, these structures have a much lower incidence of glutamate-arginine or phosphotyrosine-arginine non-covalent interactions, with 21 out of the 30 representative structures having these non-covalent interactions. No single representative structure has non-covalent interactions consistent with those predicted from the ETD behaviour once fragments with a drift time peak the same as $[M+3H]^{2+}$ are excluded (i.e., non-covalent interactions between R4 and R18 and either E6 or pY11).

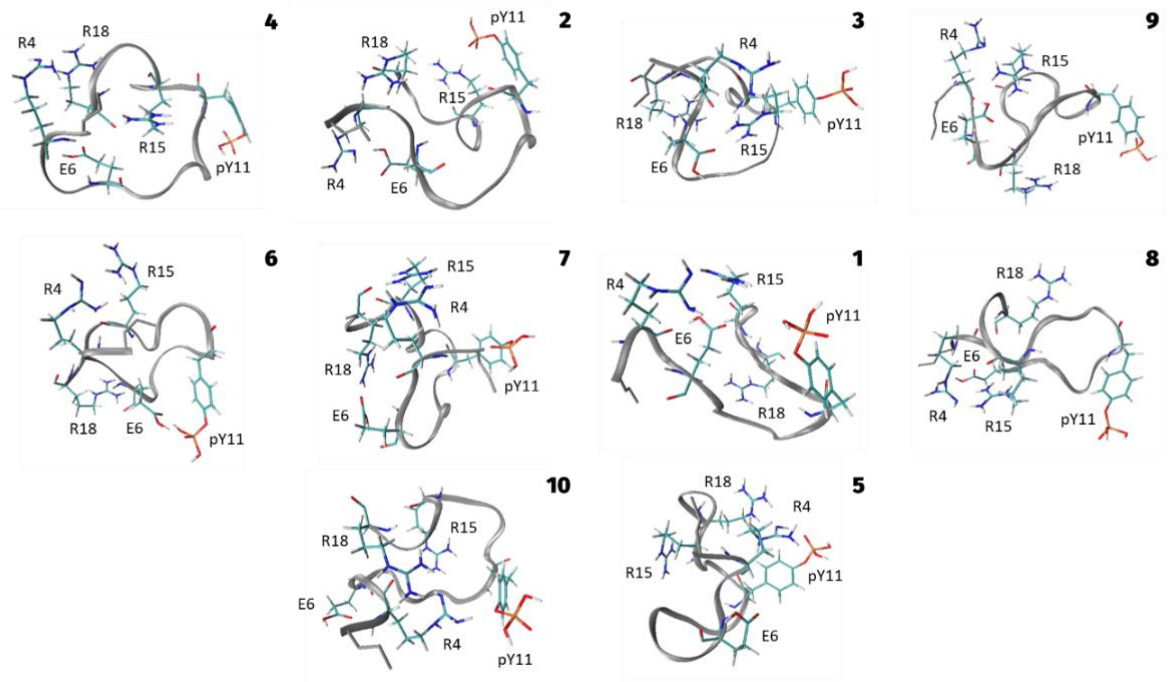


Figure 5.12 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble pY11z17_2pos_RDP4. The number in bold next to each structure is the number of the cluster that it represents

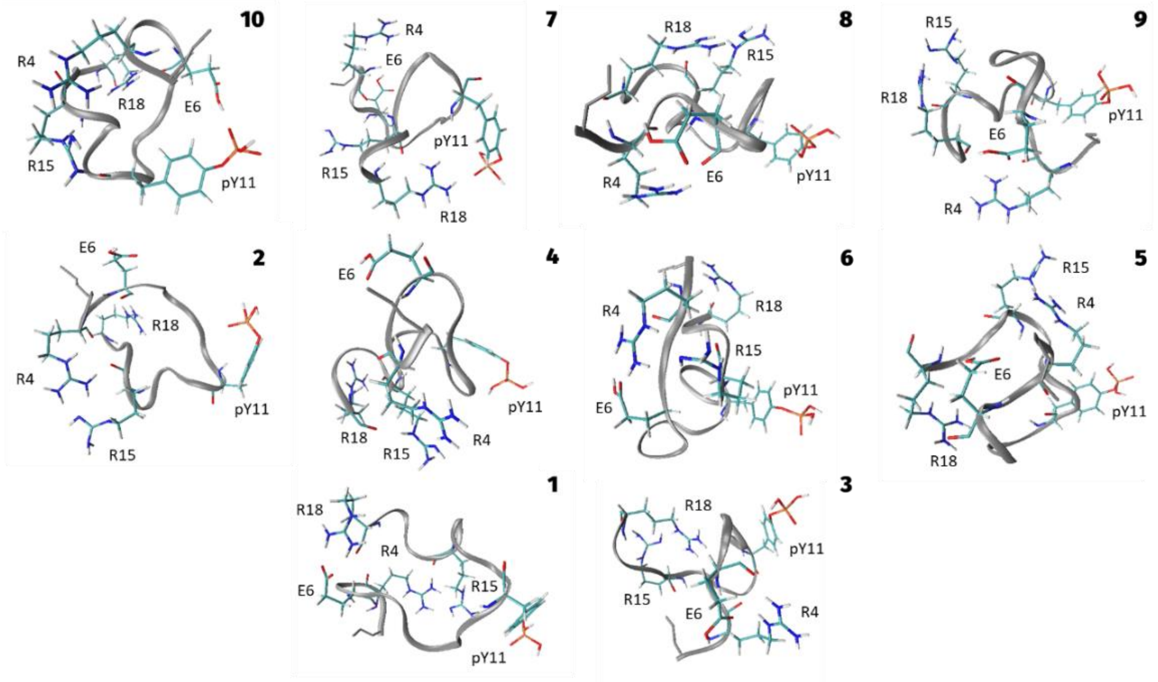


Figure 5.13 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble pY11z17_2pos_RDP15. The number in bold next to each structure is the number of the cluster that it represents

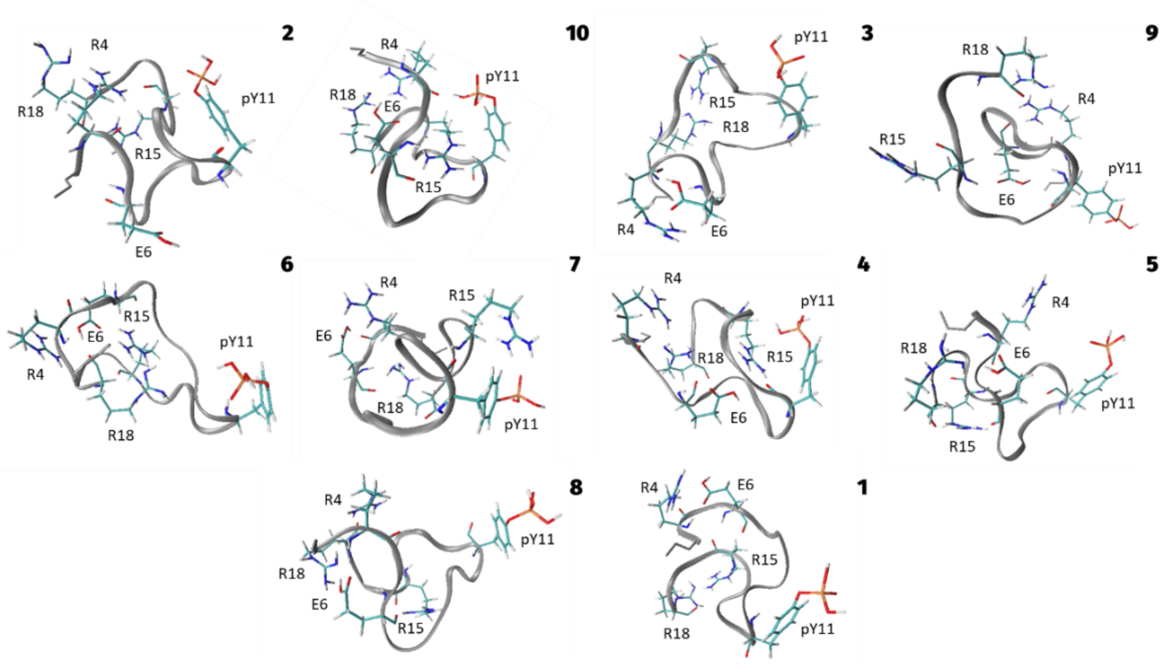


Figure 5.14 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble pY11z17_2pos_RDP18. The number in bold next to each structure is the number of the cluster that it represents

5.2.2 ETD-TWIMS behaviour of β-casein

The inefficiency of ETD is generally amplified when applied to larger proteins, with supplemental activation or ancillary fragmentation methods often used to achieve good fragment coverage.[203] It was necessary therefore to investigate whether ETD-TWIMS would be applicable to the structural study of fragments of larger proteins. The phosphoprotein β-casein was chosen for these experiments as a study performed by Chen *et al.* on the ECD behaviour of phosphorylated and unphosphorylated β-casein, showed fragments between S17 and R25 in the unphosphorylated protein, but none were observed between these residues in the phosphorylated protein,[189] which may be indicative of non-covalent interactions between these residues. β-casein (MW = 23969.16 Da, UniProt ascension number: P02666) is a phosphoprotein found in mammalian milk that participates in the formation of casein micelles, which are a food source of calcium phosphate.[30] Phosphorylation of β-casein promotes the formation of casein micelles and increases its ability to bind calcium cations.[31] The ECD MS/MS spectrum of the $[M+23H]^{23+}$ phosphoprotein β-casein is shown in Figure 5.15a, with the corresponding fragmentation pattern (Figure 5.15b) showing the same lack of fragments between pS17 and R25 that was reported by Chen *et al.*. Fragment assignment for this mass spectrum was performed using the SNAP algorithm, with the fragmentation pattern produced using Prosight Lite,[204] as discussed in Section 2.2.2.

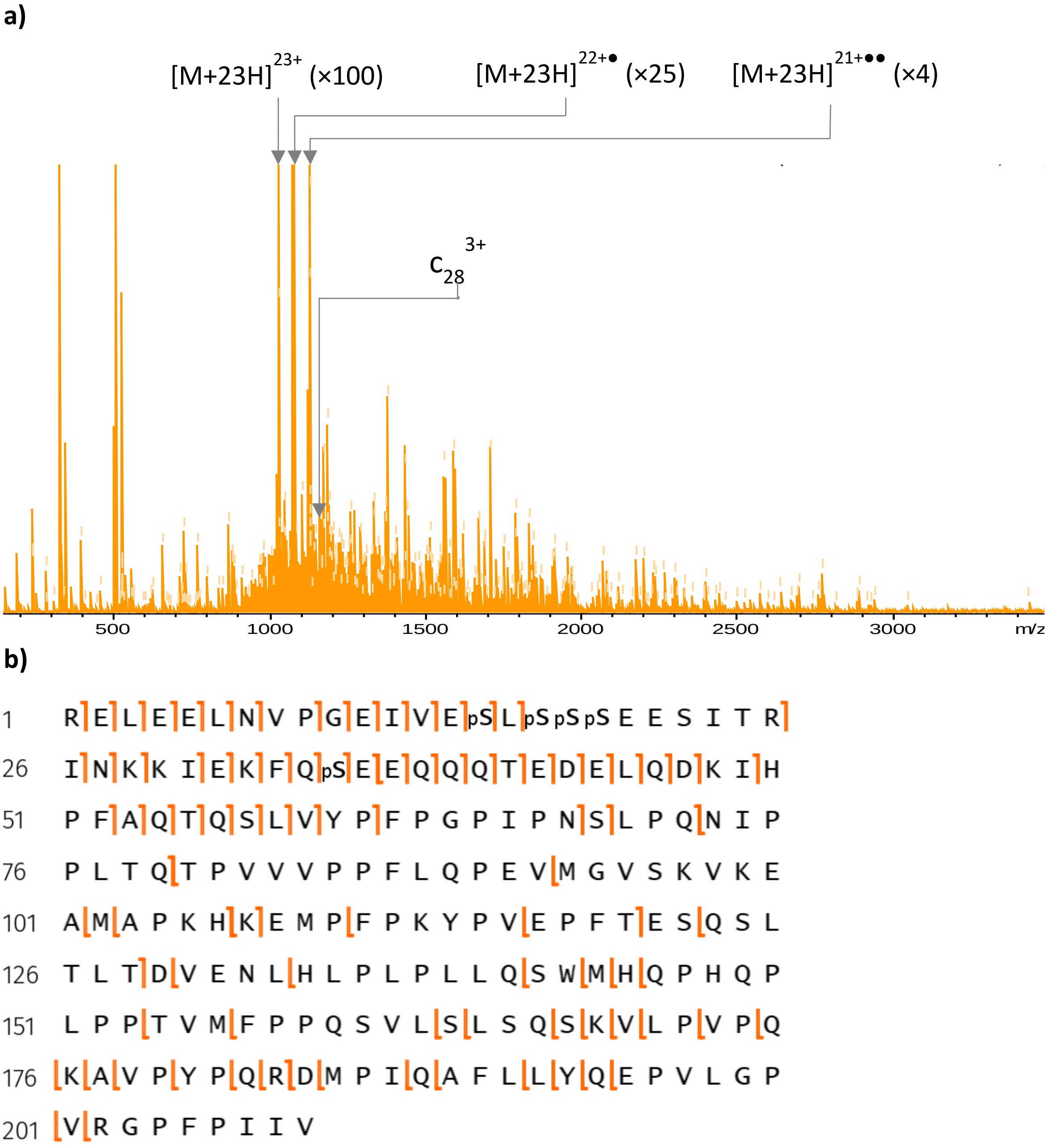


Figure 5.15 a) The ECD MS/MS spectrum of the $[M+23H]^{23+}$ ion of β-casein and b) the corresponding fragmentation pattern

The ETD-TWIMS MS/MS spectrum of the $[M+23H]^{23+}$ ion of β-casein is shown in Figure 5.16. This experiment was performed at a TWIMS wave velocity of 450 ms^{-1} and a wave height of 40 V, further experimental details can be found in Section 2.2.5. The mass spectrum produced by summing across the whole drift time range (Figure 5.16a) is typical of an ETD mass spectrum for a protein of this size, with few features other than the parent ion and several charge reduced species. The drift plot (Figure 5.16b) shows that there are fragments present, and there are drift times when those fragments are not obscured by the noise associated with the parent and charge reduced species. One such region, at a drift time of 2.4 ms, is shown in Figure 5.16c, and shows many of the fragments that were observed by ECD. In this way, ETD-TWIMS allows the detection of a similar numbers of fragments to ECD, while using ETD as the fragmentation method. Previous studies demonstrating ETD on TWIMS-enabled instruments have tended to use transfer collision energies between 25 V and 60 V to dissociate ETD products and improve fragment coverage,[109, 205, 206] or else have relied on being able to tune the ETD reaction to achieve good fragment coverage without mobility separation.[161, 207] The parameters used in this ETD-TWIMS experiment on β-casein are well below ranges that would provide supplemental activation to the ions, with trap collision energy of 4 V (NB: user control of this variable is removed when the Synapt G2S is operated in ETD mode), trap wave velocity and wave heights of 450 ms^{-1} and 0.2 V, TWIMS wave velocity and wave heights of 450 ms^{-1} and 40 V, transfer wave velocity and wave heights of 300 ms^{-1} and 4 V and transfer collision energy of 0 V. A full list of the instrument parameters used during this experiment can be found in Appendix 1-6.

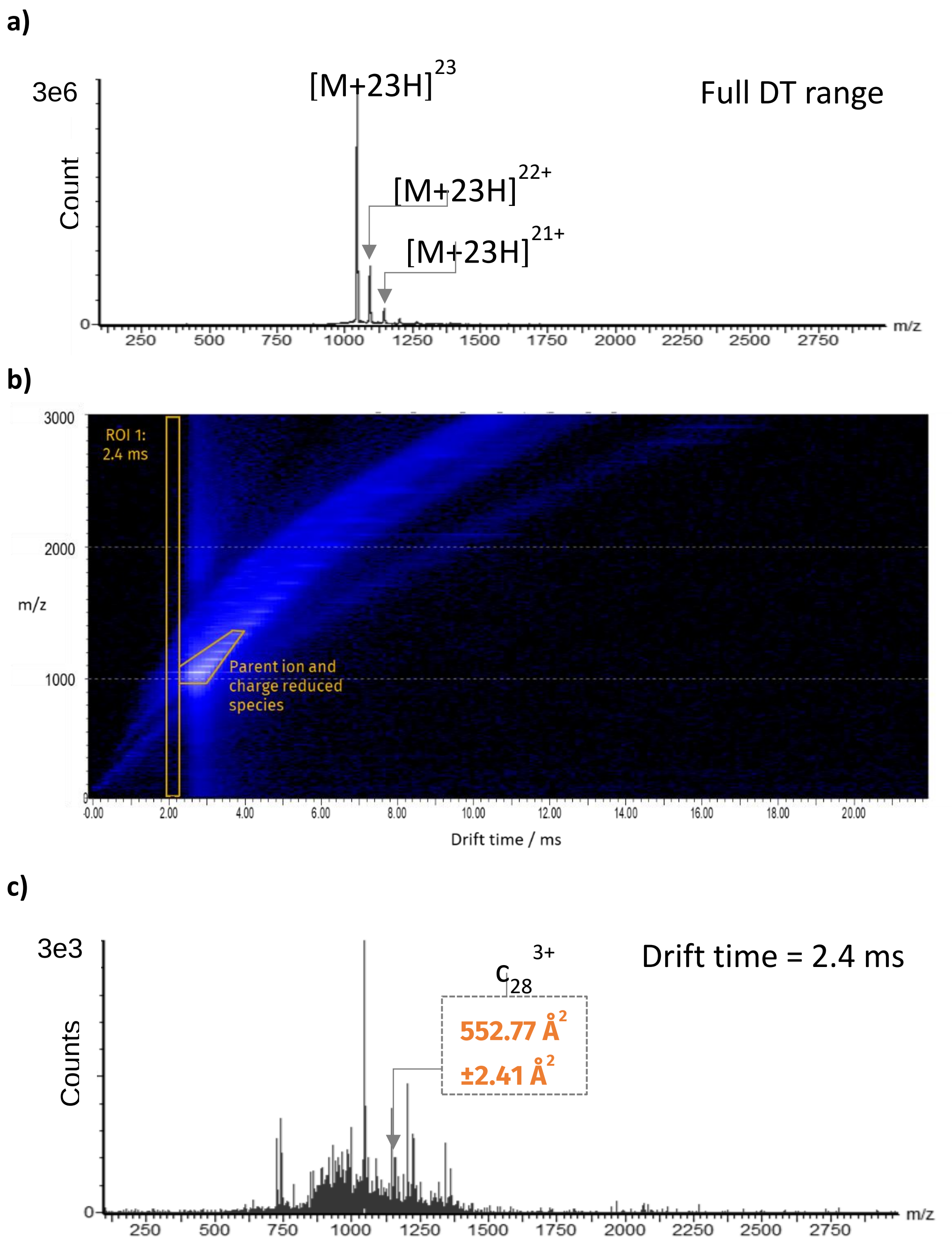


Figure 5.16 ETD-TWIMS MS/MS data for β-casein [M+23H]$^{23+}$ a) the ETD spectrum produced by summing across the full drift time range, b) the drift plot and c) the ETD spectrum at 2.4 ms containing many fragments not seen in the full drift time range ETD spectrum.

To probe the non-covalent interaction between pS17 and R25, it was necessary to select a fragment that contained this region of the protein. Given the proximity of pS17 and R25 in the amino acid sequence to the N terminus, c fragments containing these residues would be much shorter than z fragments, making them less computationally expensive to model. The shortest c fragment that contains pS17 and R25 that was measured in sufficient abundance to produce an arrival time distribution from which an apex drift time could be obtained was the c_{28}^{3+} fragment (R E L E E L N V P G E I V E pS L pS pS pS E E S I T R I N K, MW = 3476.5 Da). The c_{28}^{3+} fragment had an apex drift time of 2.4 ms, which corresponded to a CCS of 552.77 $Å^2$ ± 2.41 $Å^2$ following calibration against polyalanine and ubiquitin reference helium CCS values (calibrant ion list as in Appendix 3-4, calibration curve found in Appendix 4-8).

The three-tiered SA and trajectory method CCS calculations used previously to study phosphopeptides proved both to be very computationally expensive when applied to a fragment of 28 residues, as such only the first tier of SA (n = 1,000) was performed, and only one protonation pattern was modelled. Consequently, the IMS/MM study in this case is not exhaustive and is presented as a proof of concept that the ETD-TWIMS approach can be applied to study the fragments of intact proteins. A more in-depth IMS/MM study could be facilitated by increased computer power, or a revision of MM methods, as is discussed as possible future work in Section 6.2.

The protonation pattern modelled for the c_{28}^{3+} fragment of β-casein (Table 5.3), was informed by the necessity of maintaining a net charge of 3+, and the gap in fragment coverage between pS17 and R25 implying that the pS17 residue behaves differently to the other acidic residues in this fragment. These observations led to single positive charges being modelled on each of the N-terminus, two arginine residues (R1 and R25) and one lysine residue (K28); and one negative charge assigned to pS17. Given that there are many more acidic residues in addition to pS17 (E2, E4, E5, E11, E14, pS15, pS18, pS19, E20 and E21), it seems unlikely that this fragment exists in just one protonation pattern in the gas phase.

Table 5.3 Protonation pattern modelled for the c_{28}^{3+} fragment of β-casein $[M+23H]^{23+}$

Parent ion	Fragment	Model name	Sequence
β-casein $[M+23H]^{23+}$	c_{28}^{3+}	Bcasein-c28_3pos	R^{2+} E^0 L E^0 E^0 L N V P G E^0 I V E^0 pS^0 L pS^- pS^0 pS^0 E^0 E^0 S I T R^+ I N K^+

The CCS-matched ensemble of Bcasein-c28_3pos contained 110 structures, which produced the 10 representative structures shown in Figure 5.17, after dimensionality reduction and clustering. Details of these clusters are contained in Appendix 1-13, while distances between acidic and basic residues in the 10 representative structures are summarised in Appendix 1-14. The representative structures themselves can be found in Appendix 1-15, where they are in ascending order of structure number, as in the corresponding part of Appendix 1-13. No single representative structure, nor any combination of the representative structures, has non-covalent interactions that would produce the ECD behaviour observed for the intact protein. All the representative structures have non-covalent interactions that would produce a much larger gap in fragment coverage than the pS17-R25 gap observed by ECD. All but one of the representative structures have a non-covalent interaction pattern where the N-terminus and R1 side chain interact with acidic residues between E2 and pS18, while basic residues later in the sequence, R25 and K28, interact with residues between pS17 and E21. That is, for most of the representative structures, interactions are contained within two regions of the fragment, the region N-terminal to pS18 interacts with itself and the region from pS18 onwards interacts with itself. The only representative structure to contradict this trend belongs to cluster 9 and has non-covalent interactions extending between the two ends of the fragment, from R1 to E20 and from R25 to E4. The representative structure for cluster 9 is itself the only member of cluster 9, meaning that this non-covalent interaction pattern is not common in the CCS-matched ensemble. Given the inexhaustive nature of the MM study performed for this fragment, it is not possible to conclude that trends established in this single-tier SA experiment would extend to a 3-tiered SA approach, nor whether other protonation patterns might better represent the gas-phase ECD behaviour of the full fragment. This IMS/MM study does show that it is possible to model fragments of this size and study the non-covalent interactions within them, were a more extensive MM study performed. Following a more extensive MM study, it would be possible to ascertain whether the non-covalent interactions in protein fragments can be used to study the more complicated network of non-covalent interactions in a protein, a proposal that is discussed in more detail in Section 6.2.

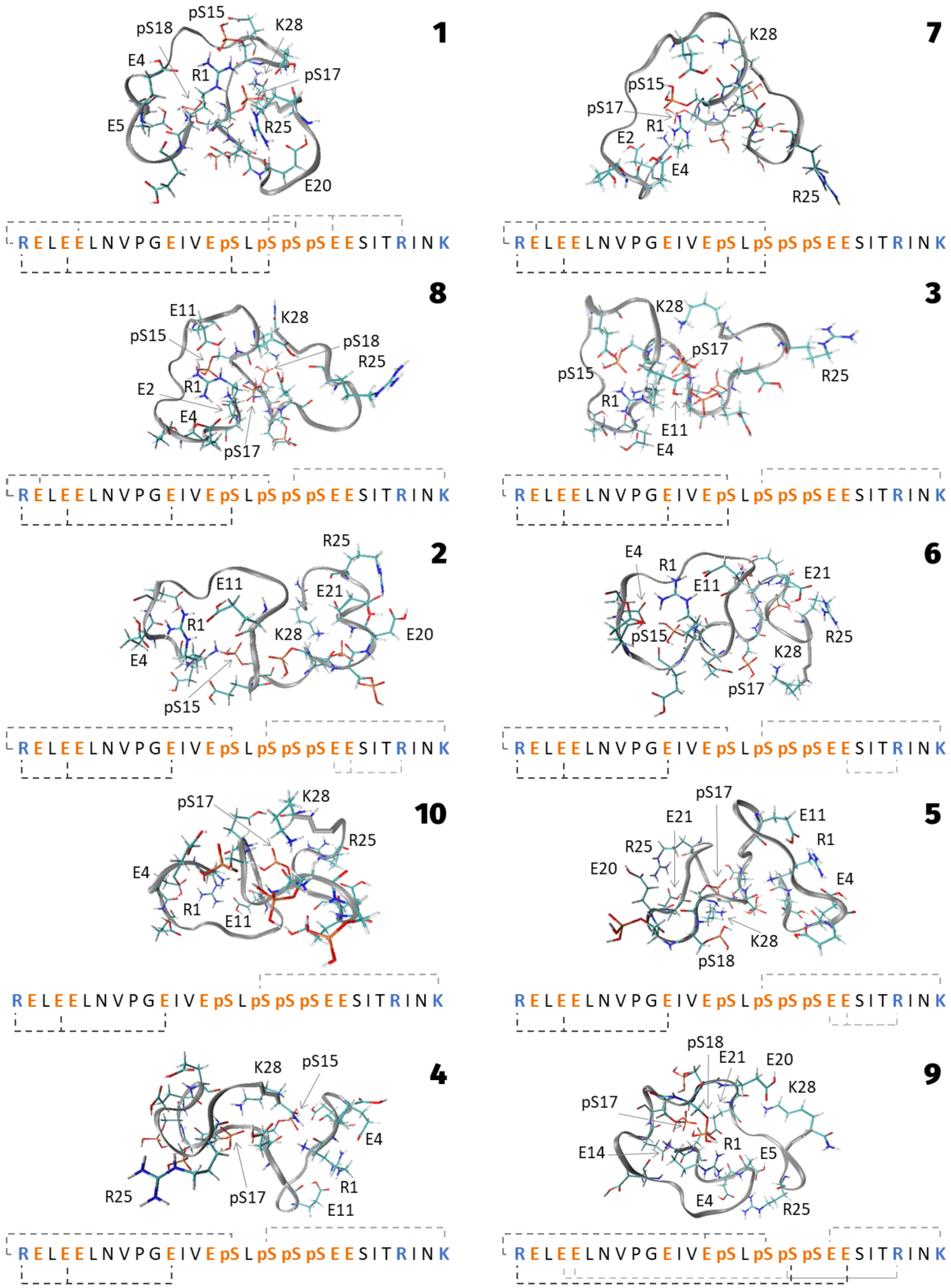


Figure 5.17 The 10 representative structures produced after dimensionality reduction and clustering the CCS-matched ensemble Bcasein-c28_3pos. The number in bold next to each structure is the number of the cluster that it represents. All acidic and basic residues are displayed, while only those involved in non-covalent interactions are labelled. Pertinent non-covalent interactions in each structure are displayed schematically below each structure, with acidic residues labelled in orange and basic residues in blue

5.3 Probing the structure of ECD and ETD fragments

In Sections 3.5.1 and 4.3 the non-covalent interactions in intact peptides have been probed by observing their ECD behaviour, while Section 5.2 presented an IMS/MM study of peptide and protein fragments. To further study the structure of peptide fragments, their ECD behaviour could be used to probe their non-covalent interactions in the same way as has been done for intact peptides. The ECD behaviour of peptide fragments can be probed by MS^3 experiments, which can be achieved using the solariX XR by ECD or ETD as the MS/MS (MS^2) level fragmentation method, followed by ECD at the MS^3 level. A schematic showing the trajectories of the ETD reagent and analyte ion beam through the instrument is shown in Figure 5.18. Previously, ETD and ECD have both been used as MS^2 level fragmentation methods to probe non-covalent interactions, with considerations about the efficiency of the two methods being the main difference discussed. ECD has been used as the preferred method due to its higher efficiency meaning that lack of detection of a given fragment is more likely to be due to a cleavage having been missed rather than its abundance being lower than the noise threshold, while limitations in instrument configuration resulted in the use of ETD in the IMS/MM study of peptide and protein fragments. Due to the difference in the mechanisms, energetics and conditions of the two methods, it is possible that the two methods produce fragments with different structures.

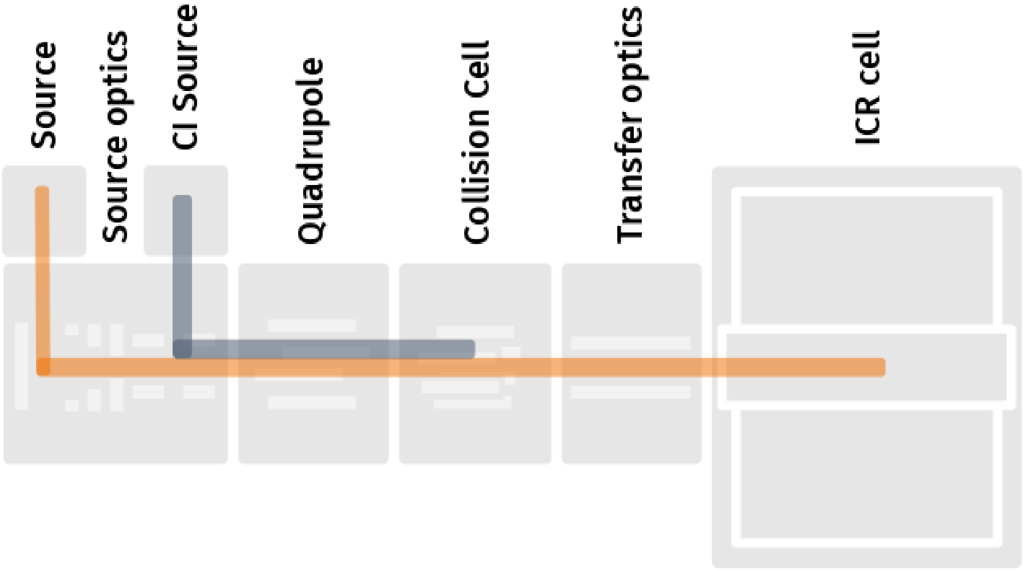


Figure 5.18 Schematic of the solariX XR mass spectrometer with the ETD system highlighted. The ETD reagent is mixed with methane gas in the CI source to form $CH_4^{\bullet-}$ *anions (blue trajectory), which travel to the collision cell, where they interact with the analyte (orange trajectory). The ETD reagent coexists with the ion beam between the split octupole in the source optics and the collision cell, where most of the ETD reaction occurs. The ECD reaction occurs in the ICR cell, in which a cathode produces a beam of thermal electrons that interact with the analyte*

The EXD-ECD experiments presented here (where EXD is either ECD or ETD) are intended to:

1. Discern whether it is possible to probe the non-covalent interactions within fragments
2. Probe any differences between the structures of fragments produced by ECD and ETD
3. Relate any differences between the ECD behaviour of ETD-generated fragments and ECD-generated fragments to the mechanisms by which they were formed

The GSK3β-derived peptides presented in Chapter 4 were chosen for analysis by EXD-ECD experiments, given that their ECD-behaviour at the MS^2 level has been well characterised and the structures of their intact $[M+2H]^{2+}$ ions probed by IMS/MM studies. Particularly, focus was placed on the z_{17}^{2+} fragment of GSK3β-pY11 as the ETD-TWIMS study presented in Section 5.2 provides orthogonal data to which EXD-ECD data could be compared and makes this the best characterised phosphopeptide fragment that contains non-covalent interactions that might affect its ECD behaviour.

The ETD and ECD MS^2 level fragmentation spectra of the GSK3β-pY11 $[M+3H]^{3+}$ ion are shown in Figure 5.19a. As discussed previously, differences in fragment coverage between the two methods may be due to differences in the efficiency of the two methods, additionally, without ion mobility separation it is not possible to know whether charge-repulsion or any other effects are promoting the dissociation of fragments that might otherwise be held together by a non-covalent interaction between the two. As such, the gap in fragment coverage between pY11 and R15 in the ETD mass spectrum cannot be assumed to be due to a non-covalent interaction; likewise, the observation of full fragment coverage by ECD cannot be interpreted as a lack of non-covalent interactions in the GSK3β-pY11 $[M+3H]^{3+}$ ion. The ETD mass spectrum produced here shows reduced fragment coverage compared to that recorded during the ETD-TWIMS experiment on the same ion (Figure 5.5), exposing differences between the Synapt G2S used for ETD-TWIMS and the solariX XR used here. The reduced fragment coverage here may be explained by a number of differences between the two instruments that may affect the formation and dissociation of fragments. The ETD in the Synapt G2S occurs in the trap region, which is a relatively high-pressure region (~5 × 10^{-2} mbar), while in the solariX XR the ETD occurs in the much lower pressure collision cell (~1 × 10^{-5} mbar). Different reagents are used in the two instruments: 1,3-dicyanobenzene in used in the Synapt G2S, while the solariX XR is equipped with fluoranthene (though some of the electron transfer occurs by reaction between the analyte and $CH_4^{\bullet-}$ anions). Further differences in the operation of the two ETD systems mean that replicating the same ETD reaction parameters is not possible, and the duty cycle between ETD reaction and detection of fragments is also different in the two instruments.

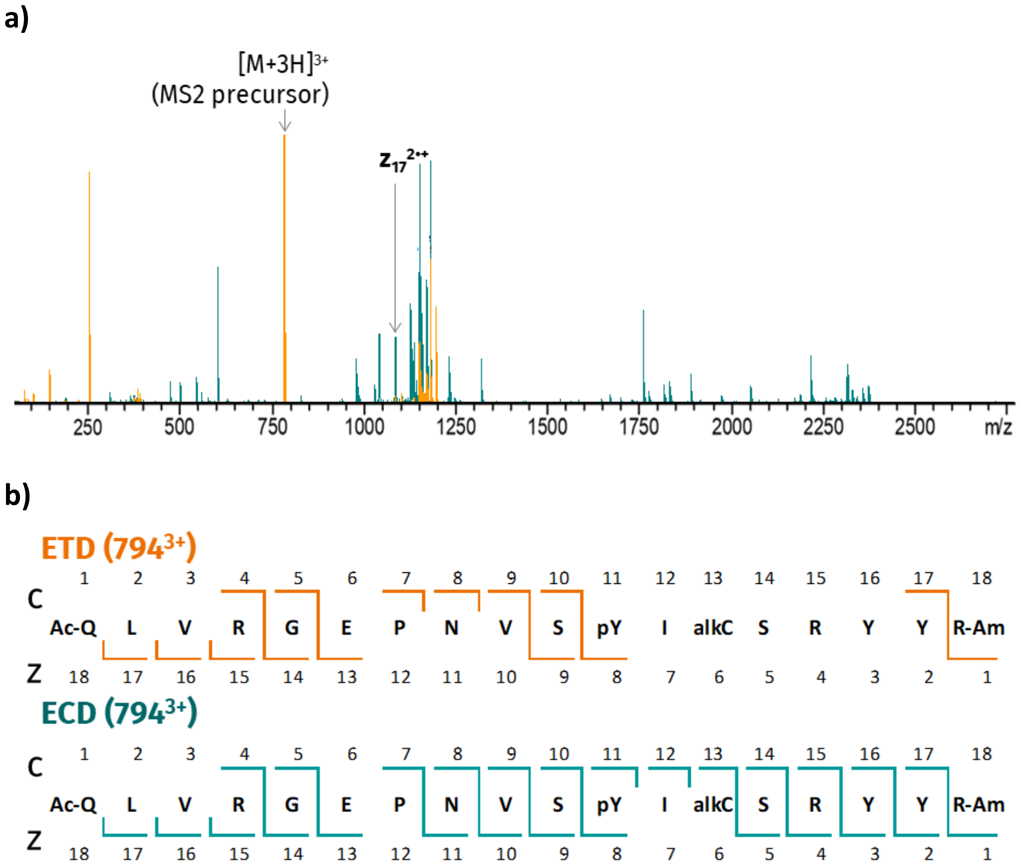


Figure 5.19 a) The ETD (orange) and ECD (teal) spectra of GSK3β-pY11 [M+3H]$^{3+}$ and b) the corresponding fragmentation patterns. The z_{17}^{2+} ion considered during the ETD-TWIMS experiments is labelled

Both ETD and ECD fragmentation of GSK3β-pY11 [M+3H]$^{3+}$ produced sufficient quantities of the z_{17}^{2+} fragment, studied previously by ETD-TWIMS, to isolate it and perform MS^3 level fragmentation. In both cases the MS^2 level parent ([M+3H]$^{3+}$) is isolated in the quadrupole, fragmented either in the collision cell (ETD) or the ICR cell (ECD), then the z_{17}^{2+} fragment is selected in the ICR cell by sweep isolation and fragmented by ECD. The mass spectra and fragmentation patterns resulting from these ETD-ECD and ECD-ECD experiments on GSK3β-pY11 [M+3H]$^{3+}$ are presented in Figure 5.20, and the methods used are described in Section 2.2.6 and Section 2.2.7. Fragment assignment lists for the ETD-ECD experiment and the ECD-ECD experiment are found in Appendix 6-19 and Appendix 6-20, respectively. Three fragments were observed during ECD of the ECD-generated z_{17}^{2+} fragment but not during ECD of the ETD-generated z_{17}^{2+} fragment: z_8, z_9 and z_{11}. Given that the MS^3 level fragmentation method is the same in both experiments, the difference in fragment coverage cannot be caused by differences in efficiency between ETD and ECD, and must instead be due to structural differences between the ETD-generated fragment and the ECD-generated fragment. The gaps in fragment coverage for both the ETD-generated z_{17}^{2+} fragment and the ECD-generated z_{17}^{2+} fragment start and end at acidic or basic residues, which could be indicative of non-covalent interactions between these residues having survived both MS^2 and MS^3 EXD steps and preventing the dissociation and detection of fragments. The ETD-generated z_{17}^{2+} fragment lacks fragments between the glutamate in position 5 (E6 in the parent GSK3β-pY11 $[M+3H]^{3+}$ ion) and the arginine at position 14 (R15 in GSK3β-pY11 $[M+3H]^{3+}$), while the ECD-generated z_{17}^{2+} fragment lacks fragments between the phosphotyrosine in position 10 (pY11 in GSK3β-pY11 $[M+3H]^{3+}$ ion) and R15. These observations suggest an E6-R15 non-covalent interaction in the ETD-generated z_{17}^{2+} fragment and a pY11-R15 interaction in the ECD-generated z_{17}^{2+} fragment, indicating that the structural differences between the two are caused by their having different non-covalent interaction patterns.

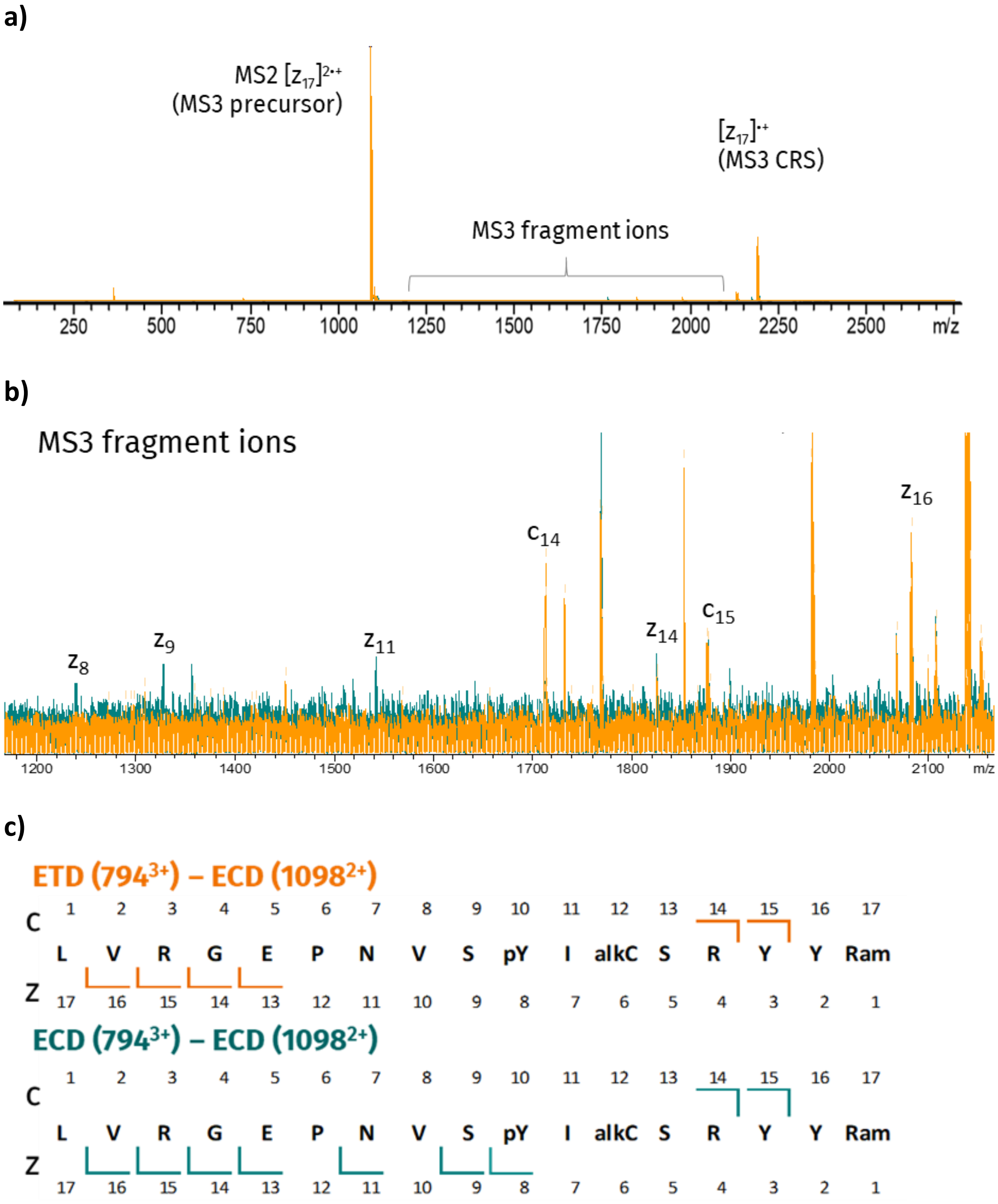


Figure 5.20 a) The ETD-ECD (orange) and ECD-ECD (teal) MS^3 spectra of the z_{17}^{2+} fragment generated from GSK3β-pY11 $[M+3H]^{3+}$, b) the region of these spectra in which MS^3 fragments are observed and c) the corresponding fragmentation patterns

There are several possible explanations for the ETD-generated and ECD-generated z_{17}^{2+} fragments having different non-covalent interactions: either non-covalent interactions in the parent ion survive the MS^2 level ETD and ECD fragmentation differently; or the fragments refold or unfold to a different extent during or after the two EXD events. If the former is true, then ETD-ECD and ECD-ECD experiments on a series of peptides with different non-covalent interaction patterns should clarify to what extent non-covalent interactions survive ETD and ECD differently. The peptides GSK3β-pY11noC, GSK3β-Y11noC, GSK3β-pY11noEnoC and GSK3β-Y11noEnoC were chosen for this work as they have successively fewer acidic charge carriers and, as shown in Chapter 4.5, a range of different non-covalent interaction patterns. The IMS/MM study of GSK3β-pY11noC suggested it exists as a mixture with a range of interactions between R4, E6, pY11, R15 and R18 across two protonation patterns and two subpopulations; for GSK3β-Y11noC, non-covalent interactions were shown to be between E6 and R15, and in a different state between R4-E6 and R15-E6; GSK3β-pY11noEnoC had interactions between R4, R15, R18 and pY11 across two sub-populations; and GSK3β-Y11noEnoC exhibited no non-covalent interactions between acidic and basic amino acid residues as it had no acidic residues with which to form them. Details of these peptides and their z_{17}^{2+} fragments can be found in Table 5.4.

Table 5.4 The m/z values for $[M+3H]^{3+}$ and z_{17}^{2+}, and the z_{17}^{2+} sequence for the GSK3β-derived peptides upon which MS^{3}-ETD-ECD and ECD-ECD experiments were performed

Peptide	$[M+3H]^{3+}$ m/z	z_{17}^{2+} m/z	z_{17}^{2+} sequence
GSK3β-pY11noC	764.6945	1053.403	$[$L V R G E P N V S pY I A S R Y Y R-Am$]^{2\bullet+}$
GSK3β-Y11noC	738.0425	1013.437	$[$L V R G E P N V S Y I A S R Y Y R-Am$]^{2\bullet+}$
GSK3β-pY11noEnoC	745.3789	1024.519	$[$L V R G A P N V S pY I A S R Y Y R-Am$]^{2\bullet+}$
GSK3β-Y11noEnoC	718.7181	984.525	$[$L V R G A P N V S Y I A S R Y Y R-Am$]^{2\bullet+}$

The ETD-ECD and ECD-ECD MS^3 spectra for the z_{17}^{2+} fragment produced from the $[M+3H]^{3+}$ ions of GSK3β-pY11noC, GSK3β-Y11noC, GSK3β-pY11noEnoC and GSK3β-Y11noEnoC are shown in Figure 5.21. Methods were used as described in Section 2.2.6 and 2.2.7, and fragment assignment lists for these spectra are found in Appendices 6-21 to 6-24. In all cases, the ECD-ECD experiment produces greater fragment coverage than the ETD-ECD, as was observed with GSK3β-pY11. For GSK3β-pY11noC and GSK3β-Y11noC the gaps in the fragment coverage begin and end at an acid or basic residue, which is indicative of the presence of non-covalent interactions. In the ETD-generated z_{17}^{2+} fragment of GSK3β-pY11noC (Figure 5.21a), the gap in fragment coverage is between E6 and R15, while for its ECD-generated z_{17}^{2+} fragment (Figure 5.21b), no fragments were observed between pY11 and R15. The

GSK3β-Y11noC ETD-generated z_{17}^{2+} fragment (Figure 5.21c) produces no fragments between E6 and R15, suggesting an interaction between these residues, but the corresponding ECD-generated z_{17}^{2+} fragment (Figure 5.21d) has full fragment coverage in c and z ions, suggesting that it contains no non-covalent interactions between acidic and basic residues. For the ETD-generated z_{17}^{2+} fragment of GSK3β-pY11noEnoC (Figure 5.21e), the gap in fragment coverage extends from A6 to R15. This gap in fragment coverage cannot easily be related to the presence of non-covalent interactions, even with the assumption that a pY11-R15 non-covalent interaction prevents the detection of fragments between these residues, because there are no sufficiently acidic residues between A6 and pY11 that could account for the missed cleavages between P7-N8, N8-V9, V9-S10 and S10-pY11. Fragments resulting from cleavage at P7-N8, V9-S10 and S10-pY11 are observed in sufficient abundance from the ECD-generated z_{17}^{2+} fragment that it seems unlikely that these missed cleavages in the ETD-generated fragment z_{17}^{2+} are due to the fragments having low signal:noise ratios and falling below the noise threshold. Nevertheless, while the fragment coverage for the GSK3β-pY11noEnoC ETD-generated z_{17}^{2+} fragment cannot be rationalised by acid-base non-covalent interactions, it does suggest structural differences between the ETD-generated z_{17}^{2+} fragment and the ECD-generated z_{17}^{2+} fragment (Figure 5.21f), for which only two missed cleavages were observed. Neither the ETD-generated z_{17}^{2+} fragment (Figure 5.21g) nor ECD-generated z_{17}^{2+} fragment (Figure 5.21h) of GSK3β-Y11noEnoC had a fragment coverage indicative of non-covalent interactions between acidic and basic residues, as was expected due to its lack of acidic residues. The fragment coverage of the GSK3β-Y11noEnoC ETD-generated z_{17}^{2+} fragment is less than that of the ECD-generated z_{17}^{2+} fragment, which implies that the structural differences between ETD-generated fragments and ECD-generated fragments are not intrinsically related to the presence of non-covalent interactions between acidic and basic residues.

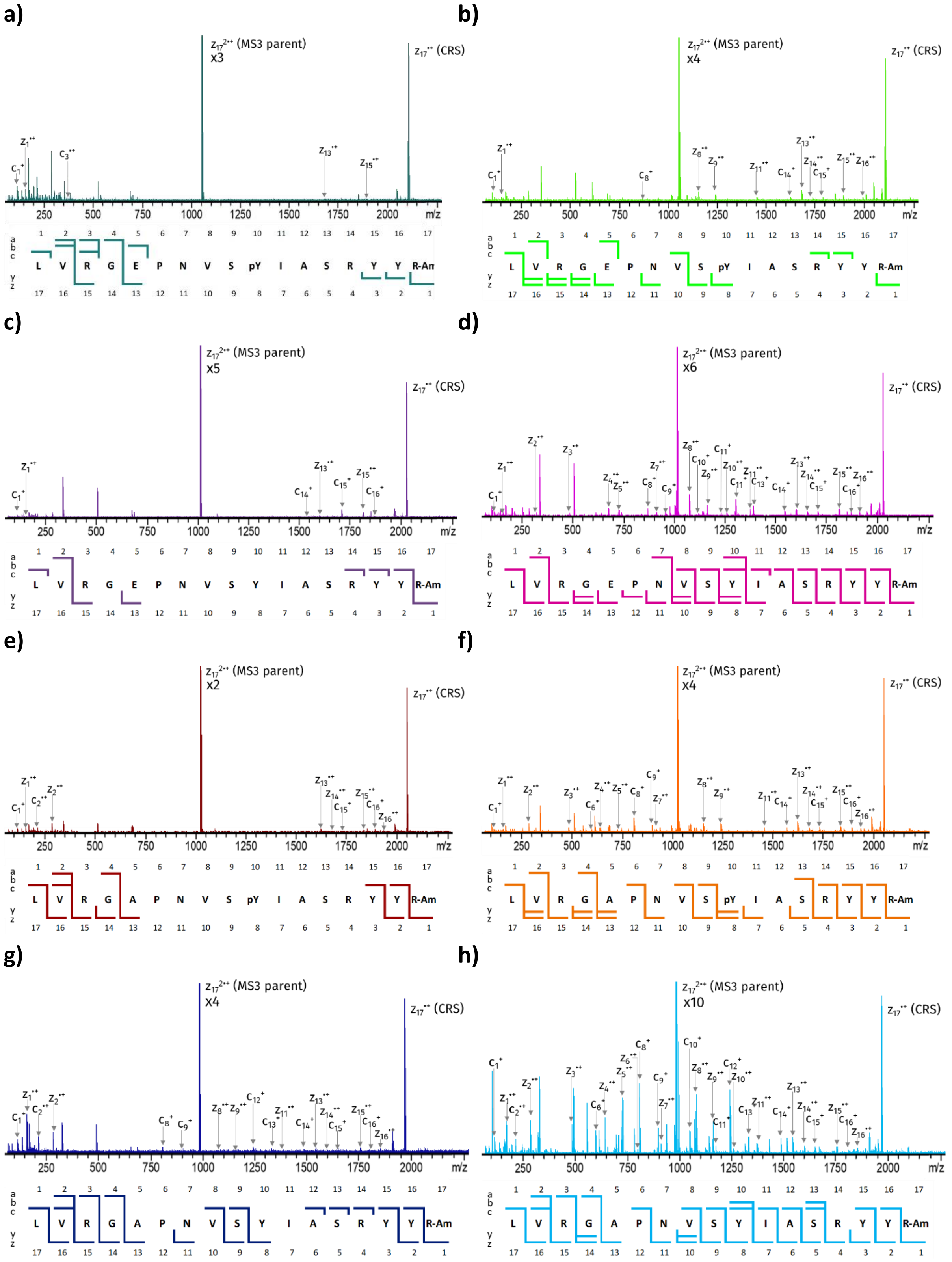


Figure 5.21 MS^3 spectra and corresponding fragmentation patterns for the z_{17}^{2+} fragment produced from the $[M+3H]^{3+}$ ions of GSK3β-derived peptides: a) GSK3β-pY11noC ETD-ECD, b) GSK3β-pY11noC ECD-ECD, c) GSK3β-Y11noC ETD-ECD, d) GSK3β-Y11noC ECD-ECD e) GSK3β-pY11noEnoC ETD-ECD, f) GSK3β-pY11noEnoC ECD-ECD, g) GSK3β-Y11noEnoC ETD-ECD and h) GSK3β-Y11noEnoC ECD-ECD. The spectra are labelled with c and z series fragments

The observation of full, or nearly full, fragment coverage for all ECD-generated z_{17}^{2+} fragments except GSK3β-pY11noC suggests that the presence of two acidic residues is necessary to preserve any non-covalent interactions between acidic and basic residues in ECD-generated fragments. Additionally, in all ECD-ECD MS^3 spectra, there are more b and y fragments than in the corresponding ETD-ECD MS^3 spectra, which suggests that the ECD-generated fragments might be more thermally activated than the ETD-generated fragments. The thermal activation of ECD-generated ions would also explain the increased abundance of their fragments relative to the charge reduced species, which shows that ECD-generated fragments more readily dissociate into MS^3 fragments upon ECD fragmentation. This thermal activation could be one source of structural differences between the two, however, these experiments do not determine whether ECD fragments have intrinsically higher thermal energy, or whether the longer delay and collisional cooling between ETD MS^2 and ECD MS^3 steps allows excess thermal energy from ETD fragments to dissipate, possibly involving a refolding step. Except for GSK3β-pY11noEnoC, the fragment coverage for the ETD-generated z_{17}^{2+} fragment is the same as that observed by ECD of the $[M+2H]^{2+}$, which could suggest structural similarities between the two. If this is the case, it would suggest refolding of the ETD-generated z_{17}^{2+} fragment given that the fragment coverages of the precursor $[M+3H]^{3+}$ ions do not resemble those of the $[M+2H]^{2+}$ ions.

Understanding the differences between these ETD-ECD and ECD-ECD experiments required an investigation into whether the ETD-generated fragments were capable of refolding, and whether ECD-generated fragments were more thermally activated and unfolded by the ECD event.

In order to determine whether the longer time delay between ETD-ECD steps (~1 s, depending on transfer and accumulation parameters) was allowing ETD fragments to refold, a series of ECD-ECD experiments with increasing time delay between MS^2 and MS^3 ECD steps was performed on the z_{17}^{2+} fragment of GSK3β-pY11noC. Time delays of 0.01 s (the time delay used in all previous ECD-ECD experiments), 0.1 s and 5 s were investigated, with the longest time of 5 s intended to mimic even the longest ETD reaction times, transfer time and ion accumulation times that were used between ETD MS^2 steps and ECD MS^3 steps, such that any refolding of the ECD-generated fragment would be observed should it occur. The ECD-ECD MS^3 spectra from this time delay experiment are shown in Figure 5.22, while details of the methods can be found in Section 2.2.7. There is no difference in the fragment coverage between the different time delays recorded, and no obvious difference between the relative abundances of any of the fragments. This implies that no refolding of this z_{17}^{2+} has occurred, regardless of the length of the time delay between MS^2 and MS^3 ECD steps. The only noticeable difference between these spectra is an increase in the $[M]^+$:$[M+H]^{\bullet+}$ ratio for all fragments at longer delays between MS^2 and MS^3 steps. The process of hydrogen loss from ECD products is well documented,[208] and seems to be the only change to the z_{17}^{2+} fragment between the MS^2 and MS^3 ECD steps.

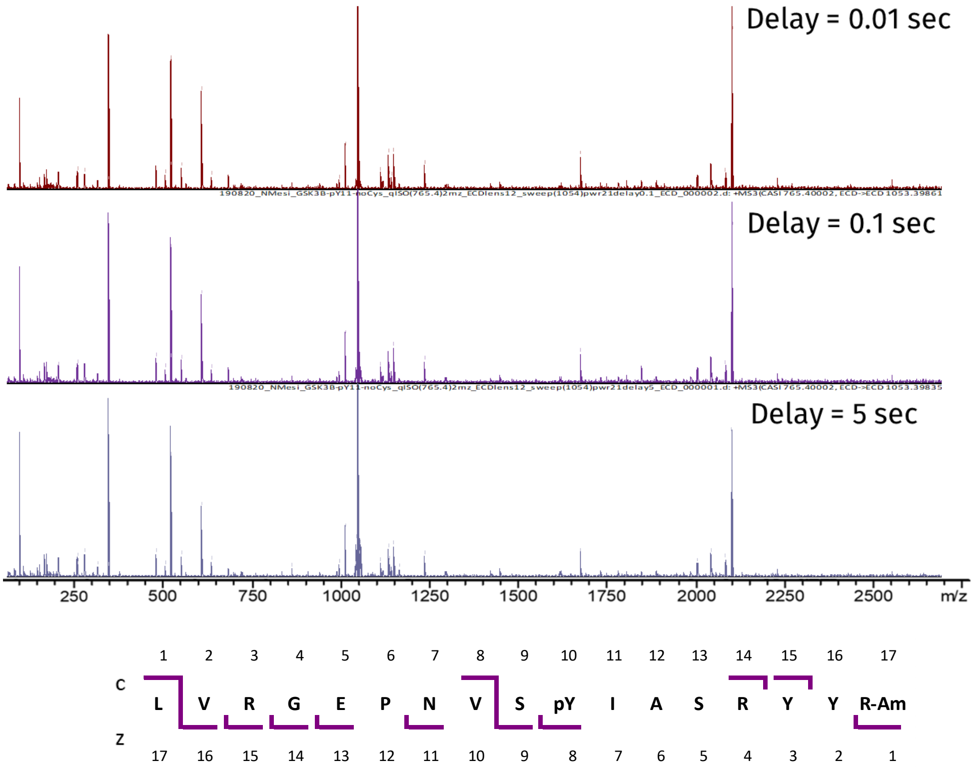


Figure 5.22 ECD-ECD MS^3 spectra of GSK3β-pY11noC with increasing time delay between MS^2 and MS^3 ECD steps. The fragmentation pattern corresponds to all three spectra as no difference was observed between them

To assess the thermal stability of the ETD-generated and ECD-generated z_{17}^{2+} fragments, an isolation mass spectrum was collected of both fragment types from each of the GSK3β-derived peptides (Figure 5.23, peak assignment lists in Appendices 6-21 to 6-24, methods as in Section 2.2.6 and Section 2.2.7). In these experiments, the MS^2 level fragmentation was performed (either by ETD or ECD) and the z_{17}^{2+} fragment isolated in the ICR cell, but no MS^3 level fragmentation was performed. Any peaks other than those present in the isolation window (in each case a 2 m/z window around the centre of the z_{17}^{2+} isotope distribution) were likely to have formed from the z_{17}^{2+} fragment. These isolation mass spectra were used to observe whether the MS^2 fragments underwent degradation or secondary fragmentation without the application of MS^3 fragmentation. It should be noted that isolation experiments alone do not conclusively prove that other species present are products of the isolated ion, and double resonance experiments would be necessary to unambiguously assign any products of the z_{17}^{2+} fragment.[209]

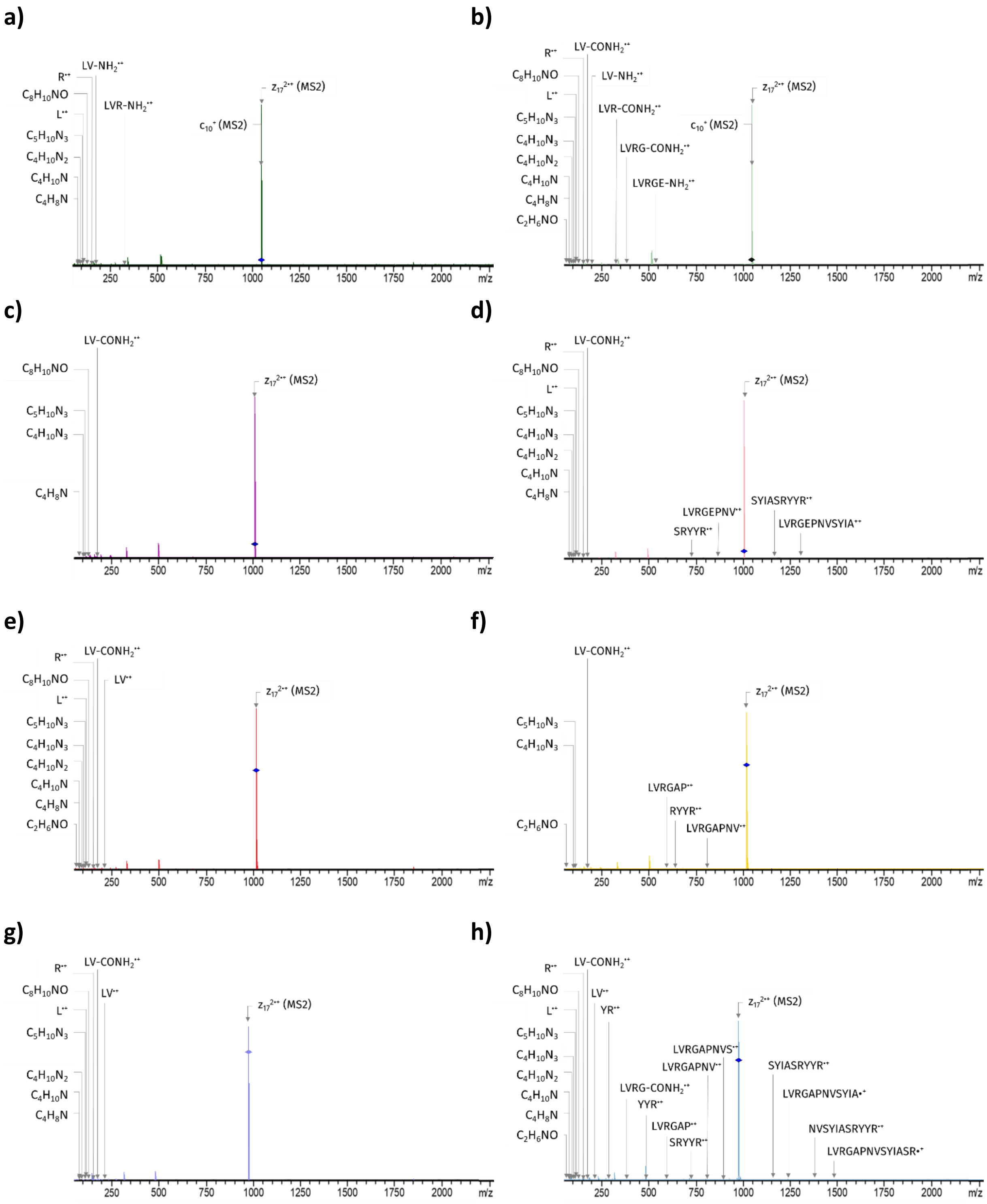


Figure 5.23 Isolation mass spectra of the z_{17}^{2+} fragment produced from the $[M+3H]^{3+}$ ions of GSK3β-derived peptides by either ETD or ECD, as indicated: a) GSK3β-pY11noC (ETD), b) GSK3β-pY11noC (ECD), c) GSK3β-Y11noC (ETD), d) GSK3β-Y11noC (ECD), e) GSK3β-pY11noEnoC (ETD), f) GSK3β-pY11noEnoC (ECD), g) GSK3β-Y11noEnoC (ETD) and h) GSK3β-Y11noEnoC (ECD). Peaks are labelled with their chemical formulae or amino acid sequence, with internal fragments taken to have c/z type termini

Each of these spectra contained a large number of low m/z ions (< 100 m/z) corresponding to side chains of serine, arginine, tyrosine and valine. Loss of side chains from the z_{17}^{2+} fragment without the application of any MS^3-level activation is indicative of these ions being excited and unstable, resulting in their decomposition. As in the ETD-ECD and ECD-ECD experiments of this set of peptides, the time delay between ETD and ECD fragmentation steps is much greater than between ECD and ECD steps (~1 s vs 0.01 s), and the ETD-generated fragment has undergone collisional cooling during its transfer to the ICR cell, so the ETD-generated fragment would be expected to have relaxed to a greater extent than the ECD-generated fragment. The secondary fragmentation observed for the ETD-generated fragment therefore suggests that these fragments are meta-stable over a long period of time and under various conditions, or that the sweep excitation deposits sufficient energy to initiate the radical gas-phase chemistry that results in the secondary fragmentation of the ETD-generated fragment.

The ECD-generated z_{17}^{2+} fragments all produce a higher number of secondary fragments resulting from the cleavage of the peptide backbone than their ETD-generated counterparts, while all but GSK3β-pY11noEnoC produce more secondary fragments resulting from loss of side chains. Most of these secondary fragments have masses corresponding to c/z-type cleavage, despite the lack of an MS^3 electron capture event to initiate this cleavage. This c/z-type secondary fragmentation as the result of a single electron capture event has been observed previously, with the suggestion that the radical fragment species undergoes an α-radical cascade that causes a second backbone cleavage.[84] It has also been shown that thermal activation of ETD fragments by CID can induce radical processes leading to side chain loss and secondary fragmentation of the peptide backbone.[210] That secondary fragmentation by ECD is relatively common, while secondary fragmentation by ETD requires activation by CID, has been interpreted as the result of ETD fragmentation of peptides depositing less energy than ECD.[211] The greater extent of secondary fragmentation observed for the ECD-generated z_{17}^{2+} fragments supports this interpretation.

Similar levels of secondary fragmentation were observed for all ETD-generated z_{17}^{2+} fragments, while the ECD-generated z_{17}^{2+} fragments with fewer acidic/basic residues tended to have more secondary fragments. This observation suggests that the formation of secondary fragments is inhibited by the presence of non-covalent interactions between acidic and basic residues, in the same way as primary fragmentation. This suggests that the non-covalent interactions that prevent the observation of MS^2 fragments survive in the z_{17}^{2+} fragments. Further, the secondary fragments observed for the ECD-generated z_{17}^{2+} fragment of GSK3β-pY11noC are exclusively a- and b-type at the secondary fragmentation site, which may suggest that the presence of two acidic residues promotes different secondary fragmentation pathways to those peptides with one or no acidic residues. One possible interpretation of this observation is that ECD-generated fragments are sufficiently unstable that they will dissipate the excess energy imparted by electron capture by other fragmentation pathways if c/z-type secondary fragmentation pathways are hindered by the presence of non-covalent interactions. While the presence of more non-covalent interactions seems to hinder secondary fragmentation of ECD-generated fragments, it is unclear whether secondary fragments could be used to localise non-covalent interactions in ECD-generated fragments or if the non-covalent interactions in these unstable ECD-generated fragments survive the MS^3 ECD event. Further, the presence of c/z-type secondary fragments makes it likely that the ECD-generated fragments undergo radical rearrangement between primary and secondary fragmentation, which could lead to structural rearrangements and changing non-covalent interactions patterns between acidic and basic residues.

The isolation experiments of ETD-generated and ECD-generated z_{17}^{2+} fragments suggest that both ETD and ECD radical fragments are meta-stable and undergo secondary fragmentation. ECD-generated fragments produced more secondary fragments as a result of peptide backbone cleavages, which is attributed to the higher energy deposited during ECD than ETD, and to the collisional cooling experience by ETD fragments, which may allow some of the energy of the ETD event to dissipate before the fragment is isolated. The ETD-generated fragments all had similar levels of secondary fragmentation, suggesting that either collisional cooling of these fragments or the longer time between the ETD event and their isolation reduces their energies to similar levels, regardless of the number of acidic/basic amino acid residues present. The presence of non-covalent interactions in the ECD-generated fragments seemed to hinder secondary fragmentation, with those peptides with more acid and basic residues exhibiting a lesser degree of secondary fragmentation.

These ETD-ECD and ECD-ECD experiments have exposed structural differences between ETD-generated fragments and ECD-generated fragments. The experiments on GSK3β-pY11noC, GSK3β-Y11noC, GSK3β-pY11noEnoC and GSK3β-Y11noEnoC expose that this structural difference seems to manifest as a difference in non-covalent interactions between acidic and basic residues, where both are present, but is also observed in peptides that do not contain acidic residues, such as GSK3β-Y11noEnoC, indicating that the structural differences are not exclusively caused by differences in non-covalent interaction patterns between acidic and basic residues. Increasing the time delay between MS^2 and MS^3 steps in ECD-ECD shows that ECD-generated fragments do not refold in the 0.01 s–5 s timeframe after fragmentation, so the differences between the ETD-ECD and ECD-ECD experiments cannot be attributed solely to the much longer time delay between MS^2 and MS^3 fragmentation steps in ETD-ECD. Isolation experiments of the ETD-generated and ECD-generated z_{17}^{2+} fragments showed much greater extents of secondary fragmentation from the ECD-generated fragments, indicating that the ECD-generated fragments have a greater energy than the corresponding ETD-generated fragments, which may contribute to the differences between the ETD-ECD and ECD-ECD experiments. If ECD-generated fragments are higher energy and less stable, then they are likely to be fragmented more easily by the MS^3 ECD event.

This difference in energy between ETD and ECD fragments may be related to intrinsic differences in the ETD process and ECD processes, with the ETD-generated fragment able to dissipate the energy imparted by electron transfer before the cleavage occurs (i.e., an ergodic process), resulting in lower energy and more stable fragments. For ECD, the opposite must be true, if the cleavage occurs before energy dissipates (i.e., a non-ergodic process), the fragment is higher energy and more unstable. This idea that explains why the ECD-generated fragment undergoes more secondary fragmentation than the ETD-generated fragment, as this pathway could allow the dissipation of excess energy. These experiments lend weight to the widely discussed idea that ECD is a non-ergodic process,[77] however less focus has been placed in the literature on the ergodicity of ETD, and it is proposed based on this evidence that ETD is an ergodic process.

Practically, this has the consequence that ETD-ECD must be used to provide information on the non-covalent interactions in ETD fragments, and ECD-ECD must be used to provide information on the structure of ECD fragments. Given that ECD-generated fragments seem to be more unstable than ETD-generated fragments, caution must be taken when inferring non-covalent interactions locations from ECD-ECD fragmentation patterns. Therefore, the results of the ETD-ECD experiment on the z_{17}^{2+} fragment of GSK3β-pY11 must be used to guide the IMS/MM workflow performed on the same fragment by ETD-TWIMS. The ETD-ECD fragmentation pattern of the z_{17}^{2+} fragment of GSK3β-pY11 lacked fragments between E6 and R15, which is interpreted as being due to the presence of a non-covalent interaction between these residues. Only three structures from the IMS/MM study of z_{17}^{2+} fragment of GSK3β-pY11 are consistent with its ECD-behaviour as observed by ETD-ECD: pY11z17_2pos_RDP4-3, pY11z17_2pos_RDP4-9 and pY11z17_2pos_RDP18-6 (Figure 5.24). These structures all exhibit the same feature as the $[M+2H]^{2+}$ ion wherein the phosphotyrosine is projected outwards from the rest of the peptide. Structural similarities between the z_{17}^{2+} fragment and $[M+2H]^{2+}$ ion are consistent with their similar ECD behaviour. Further studies on smaller fragments would elucidate to what extent ETD-generated 2+ fragments are similar in structure to the $[M+2H]^{2+}$ ion of a fragment, while IMS/MM studies on the $[M+3H]^{3+}$ ion could determine whether ECD-generated 2+ fragments are similar in structure to their parent $[M+3H]^{3+}$ ion. The ways in which ETD-TWIMS and

EXD-ECD experiments could be exploited to study the structures of both fragments and the intact ions from which they were produced are discussed as future work in Section 6.2.

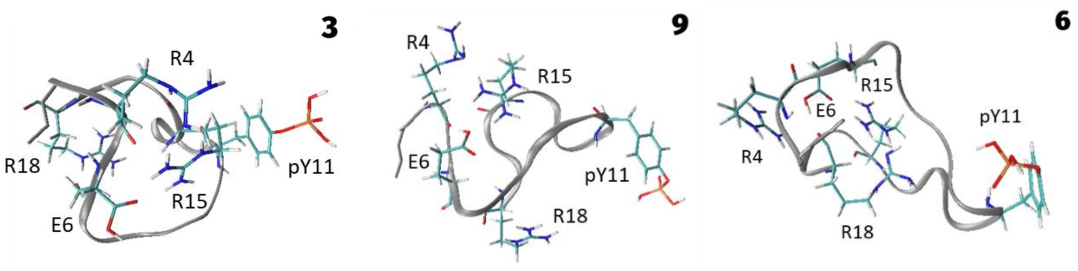


Figure 5.24 The three representative structures of the z_{17}^{2+} *fragment of GSK3β-pY11 from the IMS/MM study in Section 5.2.2 that contained non-covalent interactions consistent with its ETD-ECD fragmentation pattern: pY11z17_2pos_RDP4-3, pY11z17_2pos_RDP4-9 and pY11z17_2pos_RDP18-6*

5.4 Conclusion

In summary, various structural MS techniques have been applied to the fragments of GSK3β-derived peptides and the phosphoprotein β-casein. Of these methods, ETD-TWIMS and ETD-ECD allowed the ECD-guided IMS/MM workflow presented in previous chapters to be applied to an ETD-generated peptide fragment (z_{17}^{2+} of GSK3β-pY11), and to propose a candidate structure. It was possible to compare the structure of the fragment to the $[M+2H]^{2+}$ ion of the same peptide.

The ETD-TWIMS experiments employed here showed that the inclusion of ion mobility separation allows ETD fragment coverages to be restored to near-ECD levels for the phosphoprotein β-casein. Further, ion mobility allowed the identification of a subset of fragments in the GSK3β-pY11 spectrum that were produced by a slow dissociation process from a charge reduced species, which could be indicative of how this peptide dissociates under ETD.

MS^3 experiments allowed the differences between ETD-generated fragments and ECD-generated fragments to be probed, and these differences related back to the mechanism of the two dissociation methods. ECD imparts more energy in the analyte than ETD, and as such ECD fragments are less stable, and seem to be more easily fragmented at the MS^3 level. The energy imparted by ETD fragmentation may also be dissipated by the collisional cooling that the ETD-generated fragments experience during their transfer to the ICR cell. These processes affect the relative energetics and structures of ETD-generated fragments and ECD-generated fragments, which should be considered in the interpretation of EXD spectra and tuning of these methods.

Chapter 6: Conclusions and future work

6.1 Conclusions

The focus of this thesis has been the development of an electron capture dissociation (ECD)-guided ion mobility spectrometry/molecular modelling (IMS/MM) workflow for the study of phosphoprotein and phosphopeptide structure, and its application to biologically-derived model systems of different levels of complexity to test its ability to predict the gas-phase structure of phosphopeptides and the fragments of phosphoproteins and phosphopeptides.

To facilitate MM studies, forcefield parameters for amino acid residues commonly found under mass spectrometry conditions were developed: neutral C-terminal residues, neutral N-terminal residues, neutral arginine residues, neutral phosphorylated residues and residues adjacent to the cleavage site in c^+ and $z^\bullet$ fragments. Throughout the work presented in this thesis, at least one residue from each of these parameter sets has been used in MM simulations.

An MM workflow suitable for the production of large ensembles of peptides and their fragments was established as: a three-tiered SA approach, using the ff99SB and phos10aa forcefields, along with any additional forcefield parameters from the sets made to simulate peptides under mass spectrometry conditions. In order to filter the resultant MM ensembles by experimentally-derived collision cross section (CCS) values, ion mobility calculators were considered. IMoS was selected for ion mobility/CCS calculation following evaluation of its performance, which showed that it performs calculations sufficiently fast and with low enough errors to be suitable for calculating the CCS for each of the structures of the 25,000-strong ensembles produced from the three-tiered SA experiments. Methods to process the ensemble after mobility calculation were developed: structures that had a simulated CCS outside of the experimental CCS ± 5% were excluded from the MM ensemble, followed by dimensionality reduction of this CCS-matched ensemble by random projection and clustering by k-means to produce 10 representative structures. These representative structures may then be compared to experimental ECD and IMS data.

This IMS/MM workflow was then used to study the structures of the doubly-charged ions of five peptides derived from 14-3-3-ζ. Each of these peptides had missing regions of fragment coverage observed by ECD, which allowed the locations of non-covalent interactions to be predicted. For each peptide, the non-covalent interactions of the 10 representative structures produced by the IMS/MM methods were compared to the predicted non-covalent interactions from the ECD data, allowing ECD-consistent structures to be proposed as the gas-phase form.

The success of this structural study of 14-3-3-ζ-derived phosphopeptides in predicting structures for each of the peptides provided validation that the ECD-guided IMS/MM workflow developed here can plausibly be used to study the structure of phosphopeptides in the gas phase. This study did not make any comparison to structural data from any other techniques, so its success was not judged beyond its ability to propose viable structures.

The ECD-guided IMS/MM workflow was next applied to a series of peptides derived from the phosphoprotein GSK3β-derived peptides. In order to rationalise the ECD-behaviour of the wild-type peptide GSK3β-pY11, various amino acids were substituted and their ECD behaviour studied, along with an AI-ECD experiment of GSK3β-pY11, GSK3β-Y11 and GSK3β-pS11, which indicated that the alkylated cysteine residue was suppressing backbone cleavage following electron capture. The ECD behaviour of these variants allowed the possible non-covalent interaction partners to be identified. Even with the insight from the ECD-behaviour of all the variants, the non-covalent interaction pattern of most of these peptides could not be conclusively assigned without also performing an IMS/MM study, given that these peptides had several acidic and basic residues. The structures produced from the ECD-guided IMS/MM workflow suggested that many of the GSK3β-derived peptides existed in the gas phase as mixtures of conformers, and allowed structural features beyond the non-covalent interaction pattern to be probed. Comparison of the proposed structures for analogous peptides made it possible to suggest what impact the substitution of specific amino acid residues had on structure.

Using the ECD-guided IMS/MM workflow to study the structure of the GSK3β-derived peptides, it became clear that these derived peptides were not ideal model systems to infer structural information about the whole GSK3β protein. In the x-ray crystal structure of the whole protein, many amino acid residues contained in the derived peptides are involved in non-covalent interactions with residues not included in the derived peptide, while in the derived peptide, these residues participate in interactions with other residues in the peptide because there is no steric hinderance imposed by the rest of the protein to prevent these interactions. It was proposed that the structures of EXD fragments of the intact protein might be more representative of the structure of the intact protein than the derived peptides. As such, methods were developed to study the structures of fragments of the GSK3β-derived phosphopeptides and the phosphoprotein β-casein, the former having been characterised extensively in Chapter 4, and the latter being a commonly used model phosphoprotein whose ECD behaviour is well characterised.[189]

To apply the ECD-guided IMS/MM workflow to fragments, it was necessary to be able to measure the CCS of fragments and determine their ECD behaviour. ETD-TWIMS experiments allowed the CCS of the z_{17}^{2+} and z_{17}^{+} fragments of GSK3β-pY11 and the c_{28}^{3+} fragment of β-casein to be measured. Using forcefield parameters for residues adjacent to EXD cleavage sites, it was possible to perform an IMS/MM workflow on these peptides to probe their structures. Further, the inclusion of ion-mobility separation allowed near-ECD levels of fragment coverage to be recorded for β-casein. Ion-mobility also allowed the identification of a subset of fragments in the GSK3β-pY11 ETD MS/MS spectrum that were produced by a slow dissociation process from a charge reduced species, which could be indicative of how this peptide dissociates under ETD.

MS^3 experiments on several of the GSK3β-derived peptides showed that their ETD-generated z_{17}^{2+} fragments and ECD-generated z_{17}^{2+} fragments had different ECD behaviours, which seemed to indicate that they had different structures. The differences between ETD-generated fragments and ECD-generated fragments are likely related to differences in the mechanisms of the two dissociation methods. Isolation experiments of the ETD-generated z_{17}^{2+} fragments and ECD-generated z_{17}^{2+} fragments indicated that ECD imparts more energy in the analyte than ETD, with ECD-generated fragments undergoing more secondary fragmentation. If ECD-generated fragments are less stable, this is likely the cause of ECD-generated fragments being more easily fragmented at the MS^3 level. Additionally, any energy imparted during ETD fragmentation may be dissipated by the collisional cooling of the ETD-generated fragments between the collision cell and the ICR cell, which the ECD-generated fragments do not experience. These processes affect the energy and structure of ETD- and ECD-generated fragments, and may impact the interpretation of EXD spectra and tuning of these methods. Given that the energy imparted by ECD seems to alter the structure of ECD-generated fragments, the ETD-generated fragment is more likely to be representative of the structure of the intact ion.

The combination of ETD-TWIMS and ETD-ECD allowed the ECD-guided IMS/MM workflow developed throughout this thesis to be applied to an ETD-generated peptide fragment (z_{17}^{2+} of GSK3β-pY11), and to propose a candidate structure. It was possible to compare the structure of the fragment to the $[M+2H]^{2+}$ ion of the same peptide, with some structural features preserved between the two. This study suggests that ETD fragments may viably be used as a proxy for the intact species, which warrants further investigation in larger and more diverse systems.

6.2 Future work

To complement the work performed here, a study comparing the structures proposed by the ECD-guided IMS/MM workflow to those produced by other structural biology methods could be performed.

It is often considered that proteins and peptides in the gas phase have a ‘memory’ of their solution-phase structure.[212] Similarly, crystal and solid-phase structures can be prepared in such a way as to retain features of their solution-phase structure.[213] As such, comparison of structures proposed by the ECD-guided IMS/MM workflow to structures proposed by techniques such as NMR, x-ray crystallography and cryo-electron microscopy could be informative in appraising the ECD-guided IMS/MM workflow, and assessing the similarity of structures produced by different techniques in different phases. The phosphopeptides derived from 14-3-3-ζ and GSK3β discussed in this thesis are sufficiently small that 2D NMR studies may be possible, and it may be possible to crystallise them for x-ray crystallography studies. Such studies could be used to check the validity of the structures proposed throughout this thesis, and form the basis of a more comprehensive investigation of the structures of these phosphopeptides, allowing more concrete conclusions to be made about the biological implications of those proposed structures.

There are several challenges that must be overcome to facilitate further study of proteins or peptides by an ECD-guided IMS/MM workflow, especially in larger proteins, which have a more diverse conformational landscape and a more complicated network of non-covalent interactions.

The simulation of natively folded proteins in the gas phase is widely regarded as an area that requires further development,[110, 175, 182, 214] with the ideal solution often proposed as the development of a forcefield specifically optimised for simulation of proteins in the gas phase.[110, 175] It has been noted that until a gas-phase protein forcefield is developed, solution-phase forcefields may be used for semi-quantitative simulations of proteins in the gas phase,[175] and it is under this caveat that MM simulations of peptides and their fragments have been performed in the work presented in this thesis. The development of a gas-phase protein forcefield could be carried out by QM calculation of forcefield parameters, via re-fitting of an existing forcefield to gas-phase experimental data, or both. A modern QM approach would ideally derive forcefield parameters from a large ensemble of the largest possible model systems simulated using the highest-accuracy QM calculations that is feasible using current computing power. For example, the creation of a mass spectrometry gas-phase protein forcefield might take the form of optimising N-acetylated, C-amidated tripeptides (Ac-Xxx-**Xxx**-Xxx-Am) using the correlation method Møller-Plesset perturbation theory (MP2)[215, 216] to parameterise mid-chain amino acids (**Xxx**), using Monte Carlo methods to sample ~1000 conformations of each peptide. Considering the 20 naturally-occurring amino acids with protonated/deprotonated Arg, Asp, Glu, His and Lys, cis/trans Pro, and phosphorylated Ser, Thr and Tyr residues, there would be 21,924 permutations of Ac-Xxx-Xxx-Xxx-Am tripeptides. If **Xxx**-Xxx-Am and Ac-Xxx-**Xxx** dipeptides were used to characterise terminal **Xxx** residues (protonated/deprotonated N- and C- termini, and residues adjacent to a/x, b/y and $c^+/z^\bullet$ and $c^\bullet/z'$ cleavage sites), simulation of a further 9,744 dipeptides would be required. From simulation of these ~32,000 peptides, bond lengths, bond angles, backbone torsion angles and electrostatic potentials could be easily defined. Following the methods described in the development of the new DES-Amber forcefield, specifically optimised for the simulation of solution-phase proteins and protein complexes, non-bonding parameters (i.e., Lennard-Jones parameters) could be calculated via MP2 simulation of ~10,000 Ac-Gly-Am and N-methylacetamide dimer conformations.[217] The same study also simulated 20 Ac-[AAXAA]$_3$-Am helical peptides (i.e., one for each of the 20 'X' amino acids considered) to ensure that helical conformations are represented in the parameterisation of backbone torsion angles. Assuming that each of these calculations took ~24 hours, this *ab initio* parameterisation of a mass spectrometry gas-phase protein forcefield would take ~1,008,000 core hours, which would represent an ambitious endeavour, but is not beyond the capabilities of high-performance computing systems available today. Refining this *ab initio* forcefield by experimental gas-phase data presents the problem that there are few methods to directly study molecular geometry in the gas phase. A number of MS methods may be used to quantify the strength of inter- and intra-molecular interactions, for instance threshold CID and BIRD,[218-220] which could be used to refine non-bonding parameters. To refine geometric parameters such as bond lengths, bond angles and torsion angles, either IRMPD spectroscopy or soft-landing could be used. In the former workflow, the gas-phase IR spectrum of a peptide is measured, and compared to the calculated IR spectrum of peptides simulated by the proposed mass spectrometry gas-phase protein forcefield. Mismatches between the experimental and calculated IR spectra could indicate bond types that needed to be refined. In a soft-landing approach, ions are deposited on a target following mass analysis. This target may then be analysed by single-molecule microscopy techniques such as atomic force microscopy, cryo-electron microscopy or low-energy electron microscopy.[221-223] Proteins and protein complexes analysed following soft landing have been observed to retain at least some of the structural features acquired in the gas phase, and also those from the solution phase from which they were sampled, so analysis of soft-landed structures could give insights into gas-phase structure.[222] These techniques allow the gas-phase geometries of peptides and proteins to be recorded, which could be used to refine geometric parameters for the proposed mass spectrometry gas-phase protein forcefield.

Simulating proteins in the gas phase also introduces the problem of determining the likely protonation pattern, as seen in the simulation of the c_{28}^{3+} fragment of β-casein, where the presence of four basic sites and 11 acidic sites resulted in many possible protonation patterns. This problem may be resolved by using the gas-phase acidities/basicities of each amino acid and its chemical environment to assign the most likely protonation pattern,[224-226] or by simulation using mobile protons to find the protonation pattern with the lowest energy.[227] Because of their size, it has been possible to use extended structures as the starting geometries for the peptides and fragments simulated throughout this thesis, however, for larger proteins, this strategy would require long computational times to fold the protein into a realistic conformation.[228] X-ray crystal structures are often used as the starting structure for the simulation of larger proteins, however, this inevitably biases the results of the simulation.[229] A better approach would be to base the starting structure on gas-phase structural data. Hydrogen-deuterium exchange MS and crosslinking MS can be used to probe the solvent-accessible regions of the protein, and regions of the protein in close contact with each other, respectively.[20, 21] By arranging the protein chain in a starting geometry in agreement with gas-phase MS data, the gas-phase conformations of the protein are more likely to be sampled.

Once the gas-phase structure of a protein has been simulated using the approach outline above, the CCS of the MM ensemble may be calculated and filtered against experimental CCS values, followed by comparison of the non-covalent interactions in the CCS-matched MM ensemble and the ECD behaviour of the protein. The preliminary ECD studies of β-casein suggest that this is unlikely to be very informative, given the number of possible interaction partners and the resulting complexity of the non-covalent interaction network in a protein. But, given that results presented on the z_{17}^{2+} fragment of GSK3β-pY11 suggest that the non-covalent interactions in the fragment were indicative of the non-covalent interaction in the intact peptide, it may be possible to map the non-covalent interactions in an intact protein by studying the non-covalent interactions in its ETD fragments. Starting with an ECD-guided IMS/MM study of smaller fragments with fewer non-covalent interaction partners, the presence or absence of specific, local non-covalent interactions could be probed, with the possibility of determining whether conformers are present with/without a specific non-covalent interaction. Considering the structure of larger fragments, longer distance interactions could be probed, with insight from smaller fragments used to infer what non-covalent interactions might be present where there is uncertainty. By performing structural analysis of enough fragments from all regions of a protein, a map of the non-covalent interactions it contains could be produced.

Better characterisation of peptides, proteins and their fragments that have conformers with different non-covalent interactions could be achieved by separation of these conformers and performing ECD fragmentation on each conformer individually. Such an experiment would remove the need to infer the presence of these conformers from the presence of low abundance ECD fragments that correspond to a different non-covalent interaction pattern, as in Chapters 3 and 4. If the conformers had different CCS, they could be separated via IMS. Several instruments can perform IMS followed by ECD, so the

ECD behaviour of conformers could be probed. These instruments are: a modified Synapt G2Si MS with an ECD cell between the IMS cell and the TOF analyser, reported by Williams *et al.*;[99] and an FT-ICR MS instrument with TIMS installed,[230, 231] which has been used to characterise isobaric glycan geometric isomers by ECD.[232] Theoretically, the linear magnetic ECD cell developed by Voinov *et al.* could be incorporated into any IMS-MS instrument to perform ECD of IMS separated conformers.[233] The ultraviolet photon dissociation behaviour of IMS-separated peptide geometric isomers has been achieved using a Synapt G2S modified to allow irradiation of ions after IMS separation using a UV laser. This study demonstrated that IMS separation of the peptides RPPGSPF and PPGSPFR allowed them to be fragmented separately.[234] To date, IMS-separated fragmentation has been used to yield sequence information about sets of isobaric geometric isomers, but with sufficient IMS resolution between conformers with differing non-covalent interaction patterns, ECD could be used to characterise the difference between them. The ratio of the IMS peak of each conformer could be used to quantify the ratio of the two conformers, and the frequency that their non-covalent interaction patterns occur. The frequency of each proposed non-covalent interaction could be included in the non-covalent interaction map of the protein. Proposed non-covalent interactions with a high frequency should be observed as a gap in the ECD fragment coverage of the intact protein, while less frequent interactions may have to be inferred from the ECD-guided IMS/MM workflow applied to fragments of the protein that contain those interactions. Including the interaction frequency in the non-covalent interaction map of the protein would produce a truly conformational picture of the gas-phase structure of the protein, allowing the dynamic structure of the protein to be studied, with the results of the IMS/MM workflow allowing protein dynamics to be probed at an atomistic level.

List of references

1. Liljas, A., *Textbook of Structural Biology*. World Scientific: 2009.

2. Nishi, H.; Shaytan, A.; Panchenko, A. R., Physicochemical Mechanisms of Protein Regulation by Phosphorylation. *Frontiers in Genetics* **2014,** *5* (270).

3. Gajadhar, A. S.; White, F. M., System Level Dynamics of Post-Translational Modifications. *Curr Opin Biotechnol* **2014,** *28*, 83-7.

4. Khoury, G. A.; Baliban, R. C.; Floudas, C. A., Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Scientific reports* **2011,** *1*, 90.

5. Cooper, G. M., *The Cell: A Molecular Approach*. ASM Press: 2000.

6. Olsen, J. V.; Vermeulen, M.; Santamaria, A.; Kumar, C.; Miller, M. L.; Jensen, L. J.; Gnad, F.; Cox, J.; Jensen, T. S.; Nigg, E. A., et al., Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. *Science Signaling* **2010,** *3* (104), ra3.

7. Liou, Y.-C.; Zhou, X. Z.; Lu, K. P., Prolyl Isomerase Pin1 as a Molecular Switch to Determine the Fate of Phosphoproteins. *Trends in biochemical sciences* **2011,** *36* (10), 501-514.

8. Li, J.; Fitzpatrick, P. F., Regulation of Phenylalanine Hydroxylase: Conformational Changes Upon Phosphorylation Detected by H/D Exchange and Mass Spectrometry. *Archives of biochemistry and biophysics* **2013,** *535* (2), 115-119.

9. Bjorge, J. D.; Jakymiw, A.; Fujita, D. J., Selected Glimpses into the Activation and Function of Src Kinase. *Oncogene* **2000,** *19* (49), 5620-5635.

10. Ota, M.; Koike, R.; Amemiya, T.; Tenno, T.; Romero, P. R.; Hiroaki, H.; Dunker, A. K.; Fukuchi, S., An Assignment of Intrinsically Disordered Regions of Proteins Based on Nmr Structures. *Journal of structural biology* **2013,** *181* (1), 29-36.

11. Oldfield, C. J.; Xue, B.; Van, Y.-Y.; Ulrich, E. L.; Markley, J. L.; Dunker, A. K.; Uversky, V. N., Utilization of Protein Intrinsic Disorder Knowledge in Structural Proteomics. *Biochimica et biophysica acta* **2013,** *1834* (2), 487-498.

12. Marion, D., An Introduction to Biological Nmr Spectroscopy. *Mol Cell Proteomics* **2013,** *12* (11), 3006-25.

13. Raeck, C.; Berger, S., A 2d Nmr Method to Study Peptide Phosphorylation. *Analytical and Bioanalytical Chemistry* **2007,** *389* (7), 2161-2165.

14. Abdul-Hamid Emwas, M. A., Samah Al-Harthi, Benjamin Gabriel Poulson, Kacper Szczepski, Kousik Chandra and Mariusz Jaremko, New Advances in Fast Methods of 2d Nmr Experiments. In *Nuclear Magnetic Resonance*, Khaneja, N., Ed. IntechOpen, 2019.

15. Frank, J., Single-Particle Imaging of Macromolecules by Cryo-Electron Microscopy. *Annual Review of Biophysics and Biomolecular Structure* **2002,** *31* (1), 303-319.

16. Murata, K.; Wolf, M., Cryo-Electron Microscopy for Structural Analysis of Dynamic Biological Macromolecules. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2018,** *1862* (2), 324-334.

17. Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I., Structural Characterization of Flexible Proteins Using Small-Angle X-Ray Scattering. *Journal of the American Chemical Society* **2007,** *129* (17), 5656-5664.

18. Kikhney, A. G.; Svergun, D. I., A Practical Guide to Small Angle X-Ray Scattering (Saxs) of Flexible and Intrinsically Disordered Proteins. *FEBS Letters* **2015,** *589* (19, Part A), 2570-2577.

19. Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E., The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nat Chem* **2014,** *6* (4), 281-94.

20. O'Reilly, F. J.; Rappsilber, J., Cross-Linking Mass Spectrometry: Methods and Applications in Structural, Molecular and Systems Biology. *Nature Structural & Molecular Biology* **2018,** *25* (11), 1000-1008.

21. Konermann, L.; Pan, J.; Liu, Y. H., Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem Soc Rev* **2011,** *40* (3), 1224-34.

22. Johnson, D. T.; Di Stefano, L. H.; Jones, L. M., Fast Photochemical Oxidation of Proteins (Fpop): A Powerful Mass Spectrometry-Based Structural Proteomics Tool. *The Journal of biological chemistry* **2019,** *294* (32), 11969-11979.

23. Dobson, C. M., Biophysical Techniques in Structural Biology. *Annual Review of Biochemistry* **2019,** *88* (1), 25-33.

24. Powell, D. W.; Rane, M. J.; Chen, Q.; Singh, S.; McLeish, K. R., Identification of 14-3-3zeta as a Protein Kinase B/Akt Substrate. *J Biol Chem* **2002,** *277* (24), 21639-42.

25. Pennington, K. L.; Chan, T. Y.; Torres, M. P.; Andersen, J. L., The Dynamic and Stress-Adaptive Signaling Hub of 14-3-3: Emerging Mechanisms of Regulation and Context-Dependent Protein–Protein Interactions. *Oncogene* **2018,** *37* (42), 5587-5604.

26. Michel, G.; Mercken, M.; Murayama, M.; Noguchi, K.; Ishiguro, K.; Imahori, K.; Takashima, A., Characterization of Tau Phosphorylation in Glycogen Synthase Kinase-3β and Cyclin Dependent Kinase-5 Activator (P23) Transfected Cells. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1998,** *1380* (2), 177-182.

27. Lee, J.; Kim, M.-S., The Role of Gsk3 in Glucose Homeostasis and the Development of Insulin Resistance. *Diabetes Research and Clinical Practice* **2007,** *77* (3), S49-S57.

28. Beurel, E.; Grieco, S. F.; Jope, R. S., Glycogen Synthase Kinase-3 (Gsk3): Regulation, Actions, and Diseases. *Pharmacology & therapeutics* **2015,** *148*, 114-131.

29. Ali, A.; Hoeflich, K. P.; Woodgett, J. R., Glycogen Synthase Kinase-3: Properties, Functions, and Regulation. *Chemical Reviews* **2001,** *101* (8), 2527-2540.

30. De Kruif, C.; Holt, C., Casein Micelle Structure, Functions and Interactions. In *Advanced Dairy Chemistry—1 Proteins*, Springer: 2003; pp 233-276.

31. Broyard, C.; Gaucheron, F., Modifications of Structures and Functions of Caseins: A Scientific and Technological Challenge. *Dairy Science & Technology* **2015,** *95* (6), 831-862.

32. Gross, J. H.; Roepstorff, P., *Mass Spectrometry: A Textbook*. Springer Berlin Heidelberg: 2011.

33. Leney, A. C.; Heck, A. J. R., Native Mass Spectrometry: What Is in the Name? *Journal of the American Society for Mass Spectrometry* **2017,** *28* (1), 5-13.

34. Schnier, P. D.; Price, W. D.; Jockusch, R. A.; Williams, E. R., Blackbody Infrared Radiative Dissociation of Bradykinin and Its Analogues: Energetics, Dynamics, and Evidence for Salt-Bridge Structures in the Gas Phase. *J Am Chem Soc* **1996,** *118* (30), 7178-89.

35. Jackson, S. N.; Dutta, S.; Woods, A. S., The Use of Ecd/Etd to Identify the Site of Electrostatic Interaction in Noncovalent Complexes. *J Am Soc Mass Spectrom* **2009,** *20* (2), 176-9.

36. Robinson, C. V.; Chung, E. W.; Kragelund, B. B.; Knudsen, J.; Aplin, R. T.; Poulsen, F. M.; Dobson, C. M., Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein−Coa Ligand Binding and Assembly. *Journal of the American Chemical Society* **1996,** *118* (36), 8646-8653.

37. Kitova, E. N.; Seo, M.; Roy, P. N.; Klassen, J. S., Elucidating the Intermolecular Interactions within a Desolvated Protein-Ligand Complex. An Experimental and Computational Study. *J Am Chem Soc* **2008,** *130* (4), 1214-26.

38. Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A., Top-Down Esi-Ecd-Ft-Icr Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites. *J Am Chem Soc* **2006,** *128* (45), 14432-3.

39. Cui, W.; Rohrs, H. W.; Gross, M. L., Top-Down Mass Spectrometry: Recent Developments, Applications and Perspectives. *The Analyst* **2011,** *136* (19), 3854-3864.

40. Potel, C. M.; Lemeer, S.; Heck, A. J. R., Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Analytical chemistry* **2019,** *91* (1), 126-141.

41. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T.; Matsuo, T., Protein and Polymer Analyses up to M/Z 100 000 by Laser Ionization Time-of-Flight Mass Spectrometry. *Rapid Communications in Mass Spectrometry* **1988,** *2* (8), 151-153.

42. Karas, M.; Hillenkamp, F., Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Analytical Chemistry* **1988,** *60* (20), 2299-2301.

43. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *246* (4926), 64-71.
44. Beaufour, M.; Ginguené, D.; Le Meur, R.; Castaing, B.; Cadene, M., Liquid Native Maldi Mass Spectrometry for the Detection of Protein-Protein Complexes. *J Am Soc Mass Spectrom* **2018,** *29* (10), 1981-1994.
45. Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S., Unraveling the Mechanism of Electrospray Ionization. *Analytical Chemistry* **2013,** *85* (1), 2-9.
46. Kelly, R. T.; Tolmachev, A. V.; Page, J. S.; Tang, K.; Smith, R. D., The Ion Funnel: Theory, Implementations, and Applications. *Mass spectrometry reviews* **2010,** *29* (2), 294-312.
47. Guan, S.; Marshall, A. G., Stacked-Ring Electrostatic Ion Guide. *Journal of the American Society for Mass Spectrometry* **1996,** *7* (1), 101-106.
48. Gerlich, D., Inhomogeneous Rf Fields: A Versatile Tool for the Study of Processes with Slow Ions. In *Advances in Chemical Physics*, C.-Y. Ng, M. B., I. Prigogine and S.A. Rice, Ed. 1992; pp 1-176.
49. Guttman, M.; Wales, T. E.; Whittington, D.; Engen, J. R.; Brown, J. M.; Lee, K. K., Tuning a High Transmission Ion Guide to Prevent Gas-Phase Proton Exchange During H/D Exchange Ms Analysis. *Journal of the American Society for Mass Spectrometry* **2016,** *27* (4), 662-668.
50. Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A., Think Negative: Finding the Best Electrospray Ionization/Ms Mode for Your Analyte. *Analytical Chemistry* **2017,** *89* (11), 5665-5668.
51. Konermann, L.; Douglas, D. J., Unfolding of Proteins Monitored by Electrospray Ionization Mass Spectrometry: A Comparison of Positive and Negative Ion Modes. *Journal of the American Society for Mass Spectrometry* **1998,** *9* (12), 1248-1254.
52. Boesl, U., Time-of-Flight Mass Spectrometry: Introduction to the Basics. *Mass Spectrometry Reviews* **2017,** *36* (1), 86-109.
53. Wiley, W. C.; McLaren, I. H., Time-of-Flight Mass Spectrometer with Improved Resolution. *Review of Scientific Instruments* **1955,** *26* (12), 1150-1157.
54. Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A., The Mass-Reflectron, a New Nonmagnetic Time-of-Flight Mass Spectrometer with High Resolution. *Soviet Journal of Experimental and Theoretical Physics* **1973,** *37*, 45.
55. Guilhaus, M.; Selby, D.; Mlynski, V., Orthogonal Acceleration Time-of-Flight Mass Spectrometry. *Mass Spectrom Rev* **2000,** *19* (2), 65-107.
56. Dietz, L. A., Basic Properties of Electron Multiplier Ion Detection and Pulse Counting Methods in Mass Spectrometry. *Review of Scientific Instruments* **1965,** *36* (12), 1763-1770.
57. Ladislas Wiza, J., Microchannel Plate Detectors. *Nuclear Instruments and Methods* **1979,** *162* (1), 587-601.
58. Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S., Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrometry Reviews* **1998,** *17* (1), 1-35.
59. Amster, I. J., Fourier Transform Mass Spectrometry. *Journal of Mass Spectrometry* **1996,** *31* (12), 1325-1337.
60. Beauchamp, J. L., Ion Cyclotron Resonance Spectroscopy. *Annual Review of Physical Chemistry* **1971,** *22* (1), 527-561.
61. Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C., 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *Journal of The American Society for Mass Spectrometry* **2015,** *26* (9), 1626-1632.
62. Karabacak, N. M.; Easterling, M. L.; Agar, N. Y. R.; Agar, J. N., Transformative Effects of Higher Magnetic Field in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2010,** *21* (7), 1218-1222.
63. Comisarow, M. B.; Marshall, A. G., Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. *Chemical Physics Letters* **1974,** *25* (2), 282-283.

64. Muddiman, D. C.; Oberg, A. L., Statistical Evaluation of Internal and External Mass Calibration Laws Utilized in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **2005,** *77* (8), 2406-2414.

65. Boldin, I. A.; Nikolaev, E. N., Fourier Transform Ion Cyclotron Resonance Cell with Dynamic Harmonization of the Electric Field in the Whole Volume by Shaping of the Excitation and Detection Electrode Assembly. *Rapid Communications in Mass Spectrometry* **2011,** *25* (1), 122-126.

66. Caravatti, P.; Allemann, M., The ‘Infinity Cell’: A New Trapped-Ion Cell with Radiofrequency Covered Trapping Electrodes for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Organic Mass Spectrometry* **1991,** *26* (5), 514-518.

67. Richardson, K.; Hoyes, J., A Novel Multipass Oa-Tof Mass Spectrometer. *International Journal of Mass Spectrometry* **2015,** *377*, 309-315.

68. Ghaste, M.; Mistrik, R.; Shulaev, V., Applications of Fourier Transform Ion Cyclotron Resonance (Ft-Icr) and Orbitrap Based High Resolution Mass Spectrometry in Metabolomics and Lipidomics. *International journal of molecular sciences* **2016,** *17* (6), 816.

69. Thomson, B. A., 1997 Mcbryde Medal Award Lecture Radio Frequency Quadrupole Ion Guides in Modern Mass Spectrometry. *Canadian Journal of Chemistry* **1998,** *76* (5), 499-505.

70. Marshall, A. G.; Wang, T. C. L.; Ricca, T. L., Tailored Excitation for Fourier Transform Ion Cyclotron Mass Spectrometry. *Journal of the American Chemical Society* **1985,** *107* (26), 7893-7897.

71. de Koning, L. J.; Nibbering, N. M. M.; van Orden, S. L.; Laukien, F. H., Mass Selection of Ions in a Fourier Transform Ion Cyclotron Resonance Trap Using Correlated Harmonic Excitation Fields (Chef). *International Journal of Mass Spectrometry and Ion Processes* **1997,** *165-166*, 209-219.

72. McLafferty, F. W., Tandem Mass Spectrometry. *Science* **1981,** *214* (4518), 280.

73. Roepstorff, P.; Fohlman, J., Letter to the Editors. *Biomedical Mass Spectrometry* **1984,** *11* (11), 601-601.

74. Mitchell Wells, J.; McLuckey, S. A., Collision-Induced Dissociation (Cid) of Peptides and Proteins. In *Methods in Enzymology*, Academic Press: 2005; Vol. 402, pp 148-185.

75. McLafferty, F. W.; Bente, P. F.; Kornfeld, R.; Tsai, S.-C.; Howe, I., Metastable Ion Characteristics. Xxii. Collisional Activation Spectra of Organic Ions. *Journal of the American Chemical Society* **1973,** *95* (7), 2120-2129.

76. Kruger, N. A.; Zubarev, R. A.; Carpenter, B. K.; Kelleher, N. L.; Horn, D. M.; McLafferty, F. W., Electron Capture Versus Energetic Dissociation of Protein Ions. *International Journal of Mass Spectrometry* **1999,** *182-183*, 1-5.

77. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society* **1998,** *120* (13), 3265-3266.

78. Sobczyk, M.; Anusiewicz, I.; Berdys-Kochanska, J.; Sawicka, A.; Skurski, P.; Simons, J., Coulomb-Assisted Dissociative Electron Attachment: Application to a Model Peptide. *The Journal of Physical Chemistry A* **2005,** *109* (1), 250-258.

79. Sawicka, A.; Berdys-Kochańska, J.; Skurski, P.; Simons, J., Low-Energy (0.1 Ev) Electron Attachment S-S Bond Cleavage Assisted by Coulomb Stabilization. *International Journal of Quantum Chemistry* **2005,** *102* (5), 838-846.

80. Syrstad, E. A.; Turecek, F., Toward a General Mechanism of Electron Capture Dissociation. *J Am Soc Mass Spectrom* **2005,** *16* (2), 208-24.

81. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W., Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Analytical Chemistry* **2000,** *72* (3), 563-573.

82. Sobczyk, M.; Simons, J., Distance Dependence of through-Bond Electron Transfer Rates in Electron-Capture and Electron-Transfer Dissociation. *International Journal of Mass Spectrometry* **2006,** *253* (3), 274-280.

83. Sobczyk, M.; Neff, D.; Simons, J., Theoretical Study of through-Space and through-Bond Electron Transfer within Positively Charged Peptides in the Gas Phase. *International Journal of Mass Spectrometry* **2008,** *269* (3), 149-164.

84. Leymarie, N.; Costello, C. E.; O'Connor, P. B., Electron Capture Dissociation Initiates a Free Radical Reaction Cascade. *Journal of the American Chemical Society* **2003,** *125* (29), 8949-8958.

85. O’Connor, P. B.; Lin, C.; Cournoyer, J. J.; Pittman, J. L.; Belyayev, M.; Budnik, B. A., Long-Lived Electron Capture Dissociation Product Ions Experience Radical Migration Via Hydrogen Abstraction. *Journal of the American Society for Mass Spectrometry* **2006,** *17* (4), 576-585.

86. Cooper, H. J.; Hudgins, R. R.; Håkansson, K.; Marshall, A. G., Characterization of Amino Acid Side Chain Losses in Electron Capture Dissociation. *Journal of the American Society for Mass Spectrometry* **2002,** *13* (3), 241-249.

87. Shi, S. D.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W., Phosphopeptide/Phosphoprotein Mapping by Electron Capture Dissociation Mass Spectrometry. *Anal Chem* **2001,** *73* (1), 19-22.

88. Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A., Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun Mass Spectrom* **2000,** *14* (19), 1793-800.

89. Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L., Electron Capture Dissociation and Infrared Multiphoton Dissociation Ms/Ms of an N-Glycosylated Tryptic Peptic to Yield Complementary Sequence Information. *Anal Chem* **2001,** *73* (18), 4530-6.

90. Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T., Localization of Labile Posttranslational Modifications by Electron Capture Dissociation: The Case of Gamma-Carboxyglutamic Acid. *Anal Chem* **1999,** *71* (19), 4250-3.

91. Chalmers, M. J.; Hakansson, K.; Johnson, R.; Smith, R.; Shen, J.; Emmett, M. R.; Marshall, A. G., Protein Kinase a Phosphorylation Characterized by Tandem Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Proteomics* **2004,** *4* (4), 970-81.

92. Horn, D. M.; Ge, Y.; McLafferty, F. W., Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 Kda) Proteins. *Anal Chem* **2000,** *72* (20), 4778-84.

93. Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W., Plasma Electron Capture Dissociation for the Characterization of Large Proteins by Top Down Mass Spectrometry. *Anal Chem* **2003,** *75* (7), 1599-603.

94. Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W., Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation. *J Am Chem Soc* **2002,** *124* (22), 6407-20.

95. Skinner, O. S.; McLafferty, F. W.; Breuker, K., How Ubiquitin Unfolds after Transfer into the Gas Phase. *J Am Soc Mass Spectrom* **2012,** *23* (6), 1011-4.

96. Breuker, K.; Bruschweiler, S.; Tollinger, M., Electrostatic Stabilization of a Native Protein Structure in the Gas Phase. *Angew Chem Int Ed Engl* **2011,** *50* (4), 873-7.

97. Cui, W.; Zhang, H.; Blankenship, R. E.; Gross, M. L., Electron-Capture Dissociation and Ion Mobility Mass Spectrometry for Characterization of the Hemoglobin Protein Assembly. *Protein Sci* **2015,** *24* (8), 1325-32.

98. Harvey, S. R.; Porrini, M.; Tyler, R. C.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E., Electron Capture Dissociation and Drift Tube Ion Mobility-Mass Spectrometry Coupled with Site Directed Mutations Provide Insights into the Conformational Diversity of a Metamorphic Protein. *Phys Chem Chem Phys* **2015,** *17* (16), 10538-50.

99. Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F., Top-Down Characterization of Denatured Proteins and Native Protein Complexes Using Electron Capture Dissociation Implemented within a Modified Ion Mobility-Mass Spectrometer. *Analytical Chemistry* **2020,** *92* (5), 3674-3681.

100. Creese, A. J.; Cooper, H. J., The Effect of Phosphorylation on the Electron Capture Dissociation of Peptide Ions. *J Am Soc Mass Spectrom* **2008,** *19* (9), 1263-74.

101. Lopez-Clavijo, A. F.; Duque-Daza, C. A.; Creese, A. J.; Cooper, H. J., Electron Capture Dissociation Mass Spectrometry of Phosphopeptides: Arginine and Phosphoserine. *International Journal of Mass Spectrometry* **2015,** *390*, 63-70.

102. Lin, C.; Cournoyer, J. J.; O'Connor, P. B., Probing the Gas-Phase Folding Kinetics of Peptide Ions by Ir Activated Dr-Ecd. *Journal of the American Society for Mass Spectrometry* **2008,** *19* (6), 780-789.

103. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F., Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (26), 9528-9533.

104. Sohn, C. H.; Chung, C. K.; Yin, S.; Ramachandran, P.; Loo, J. A.; Beauchamp, J. L., Probing the Mechanism of Electron Capture and Electron Transfer Dissociation Using Tags with Variable Electron Affinity. *Journal of the American Chemical Society* **2009,** *131* (15), 5444-5459.

105. Mentinova, M.; Crizer, D. M.; Baba, T.; McGee, W. M.; Glish, G. L.; McLuckey, S. A., Cation Recombination Energy/Coulomb Repulsion Effects in Etd/Ecd as Revealed by Variation of Charge Per Residue at Fixed Total Charge. *Journal of the American Society for Mass Spectrometry* **2013,** *24* (11), 1676-1689.

106. Asakawa, D.; De Pauw, E., Difference of Electron Capture and Transfer Dissociation Mass Spectrometry on Ni2+-, Cu2+-, and Zn2+-Polyhistidine Complexes in the Absence of Remote Protons. *Journal of The American Society for Mass Spectrometry* **2016,** *27* (7), 1165-1175.

107. Zhang, Z.; Browne, S. J.; Vachet, R. W., Exploring Salt Bridge Structures of Gas-Phase Protein Ions Using Multiple Stages of Electron Transfer and Collision Induced Dissociation. *J Am Soc Mass Spectrom* **2014,** *25* (4), 604-13.

108. Zhang, Z.; Vachet, R. W., Gas-Phase Protein Salt Bridge Stabilities from Collisional Activation and Electron Transfer Dissociation. *Int J Mass Spectrom* **2017,** *420*, 51-56.

109. Lermyte, F.; Sobott, F., Electron Transfer Dissociation Provides Higher-Order Structural Information of Native and Partially Unfolded Protein Complexes. *Proteomics* **2015,** *15* (16), 2813-22.

110. Gabelica, V.; Marklund, E., Fundamentals of Ion Mobility Spectrometry. *Current Opinion in Chemical Biology* **2018,** *42*, 51-59.

111. Revercomb, H. E.; Mason, E. A., Theory of Plasma Chromatography/Gaseous Electrophoresis. Review. *Analytical Chemistry* **1975,** *47* (7), 970-983.

112. Mack, E., Average Cross-Sectional Areas of Molecules by Gaseous Diffusion Methods. *Journal of the American Chemical Society* **1925,** *47* (10), 2468-2482.

113. May, J. C.; McLean, J. A., Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry* **2015,** *87* (3), 1422-1436.

114. Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L., et al., Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrom Rev* **2019,** *38* (3), 291-320.

115. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., Applications of a Travelling Wave-Based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Commun Mass Spectrom* **2004,** *18* (20), 2401-14.

116. Keith Richardson, D. L., Kevin Giles, Sugyan Dixit, Jakub Ujma, Brandon Ruotolo, An Improved Calibration Approach for Travelling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. In *American Society for Mass Spectrometry Annual Conference*, Waters Corportation: Atlanta, 2019; Vol. PSTR135022197.

117. Richardson, K.; Langridge, D.; Giles, K., Fundamentals of Travelling Wave Ion Mobility Revisited: I. Smoothly Moving Waves. *International Journal of Mass Spectrometry* **2018,** *428*, 71-80.

118. Henderson, S. C.; Li, J.; Counterman, A. E.; Clemmer, D. E., Intrinsic Size Parameters for Val, Ile, Leu, Gln, Thr, Phe, and Trp Residues from Ion Mobility Measurements of Polyamino Acid Ions. *The Journal of Physical Chemistry B* **1999,** *103* (41), 8780-8785.

119. Bush, M. F.; Campuzano, I. D.; Robinson, C. V., Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal Chem* **2012,** *84* (16), 7124-30.

120. Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Reilly, J. P.; Clemmer, D. E., Gas-Phase Separations of Protease Digests. *J Am Soc Mass Spectrom* **1998,** *9* (11), 1213-6.

121. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E., Three-Dimensional Ion Mobility/Tofms Analysis of Electrosprayed Biomolecules. *Anal Chem* **1998,** *70* (11), 2236-42.

122. Valentine, S. J.; Anderson, J. G.; Ellington, A. D.; Clemmer, D. E., Disulfide-Intact and -Reduced Lysozyme in the Gas Phase: Conformations and Pathways of Folding and Unfolding. *The Journal of Physical Chemistry B* **1997,** *101* (19), 3891-3900.

123. Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T., Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Analytical Chemistry* **2010,** *82* (22), 9557-9565.

124. Valentine, S. J.; Counterman, A. E.; Clemmer, D. E., Conformer-Dependent Proton-Transfer Reactions of Ubiquitin Ions. *Journal of the American Society for Mass Spectrometry* **1997,** *8* (9), 954-961.

125. Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F., Protein Structure in Vacuo: Gas-Phase Conformations of Bpti and Cytochrome C. *Journal of the American Chemical Society* **1997,** *119* (9), 2240-2248.

126. Salbo, R.; Bush, M. F.; Naver, H.; Campuzano, I.; Robinson, C. V.; Pettersson, I.; Jørgensen, T. J. D.; Haselmann, K. F., Traveling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers. *Rapid Communications in Mass Spectrometry* **2012,** *26* (10), 1181-1193.

127. Ruotolo, B. T.; Gillig, K. J.; Woods, A. S.; Egan, T. F.; Ugarov, M. V.; Schultz, J. A.; Russell, D. H., Analysis of Phosphorylated Peptides by Ion Mobility-Mass Spectrometry. *Anal Chem* **2004,** *76* (22), 6727-33.

128. Ruotolo, B. T.; Verbeck; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H., Distinguishing between Phosphorylated and Nonphosphorylated Peptides with Ion Mobility−Mass Spectrometry. *Journal of Proteome Research* **2002,** *1* (4), 303-306.

129. Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H., Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal Chem* **2009,** *81* (1), 248-54.

130. Glover, M. S.; Dilger, J. M.; Acton, M. D.; Arnold, R. J.; Radivojac, P.; Clemmer, D. E., Examining the Influence of Phosphorylation on Peptide Ion Structure by Ion Mobility Spectrometry-Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2016,** *27* (5), 786-794.

131. Kim, D.; Pai, P. J.; Creese, A. J.; Jones, A. W.; Russell, D. H.; Cooper, H. J., Probing the Electron Capture Dissociation Mass Spectrometry of Phosphopeptides with Traveling Wave Ion Mobility Spectrometry and Molecular Dynamics Simulations. *J Am Soc Mass Spectrom* **2015,** *26* (6), 1004-13.

132. Atkins, P. W.; De Paula, J., *Atkins' Physical Chemistry*. Oxford University Press: 2002.

133. Slater, J. C., A Simplification of the Hartree-Fock Method. *Physical Review* **1951,** *81* (3), 385-390.

134. Roothaan, C. C. J., New Developments in Molecular Orbital Theory. *Reviews of Modern Physics* **1951,** *23* (2), 69-89.

135. Urban, M.; Černušák, I.; Kellö, V.; Noga, J., Electron Correlation in Molecules. In *Methods in Computational Chemistry: Volume 1 Electron Correlation in Atoms and Molecules*, Wilson, S., Ed. Springer US: Boston, MA, 1987; pp 117-250.

136. Parr, R. G. In *Density Functional Theory of Atoms and Molecules*, Horizons of Quantum Chemistry, Dordrecht, 1980//; Fukui, K.; Pullman, B., Eds. Springer Netherlands: Dordrecht, 1980; pp 5-15.

137. Tozzini, V., Multiscale Modeling of Proteins. *Accounts of Chemical Research* **2010,** *43* (2), 220-230.

138. Weiner, P. K.; Kollman, P. A., Amber: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and Their Interactions. *Journal of Computational Chemistry* **1981,** *2* (3), 287-303.

139. Poltev, V., Molecular Mechanics: Principles, History, and Current Status. 2017; pp 21-67.

140. Lai, R.; Dodds, E. D.; Li, H., Molecular Dynamics Simulation of Ion Mobility in Gases. *The Journal of chemical physics* **2018,** *148* (6), 064109-064109.

141. Shvartsburg, A. A.; Jarrold, M. F., An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chemical Physics Letters* **1996,** *261* (1), 86-91.

142. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F., Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *The Journal of Physical Chemistry* **1996,** *100* (40), 16082-16086.

143. Larriba, C.; Hogan, C. J., Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *Journal of Computational Physics* **2013,** *251*, 344-363.

144. Bleiholder, C.; Wyttenbach, T.; Bowers, M. T., A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (I). Method. *International Journal of Mass Spectrometry* **2011,** *308* (1), 1-10.

145. Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I., Structural Characterization of Drug-Like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections. *Anal Chem* **2012,** *84* (2), 1026-33.

146. Marklund, Erik G.; Degiacomi, Matteo T.; Robinson, Carol V.; Baldwin, Andrew J.; Benesch, Justin L. P., Collision Cross Sections for Structural Proteomics. *Structure* **2015,** *23* (4), 791-799.

147. Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S., Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *Journal of the American Society for Mass Spectrometry* **2017,** *28* (4), 587-596.

148. von Helden, G.; Hsu, M. T.; Gotts, N.; Bowers, M. T., Carbon Cluster Cations with up to 84 Atoms: Structures, Formation Mechanism, and Reactivity. *The Journal of Physical Chemistry* **1993,** *97* (31), 8182-8192.

149. Jurneczko, E.; Barran, P. E., How Useful Is Ion Mobility Mass Spectrometry for Structural Biology? The Relationship between Protein Crystal Structures and Their Collision Cross Sections in the Gas Phase. *Analyst* **2011,** *136* (1), 20-28.

150. Shrivastav, V.; Nahin, M.; Hogan, C. J.; Larriba-Andaluz, C., Benchmark Comparison for a Multi-Processing Ion Mobility Calculator in the Free Molecular Regime. *J Am Soc Mass Spectrom* **2017,** *28* (8), 1540-1551.

151. Lapthorn, C.; Pullen, F. S.; Chowdhry, B. Z.; Wright, P.; Perkins, G. L.; Heredia, Y., How Useful Is Molecular Modelling in Combination with Ion Mobility Mass Spectrometry for ‘Small Molecule’ Ion Mobility Collision Cross-Sections? *Analyst* **2015,** *140* (20), 6814-6823.

152. Ouyang, H.; Bo, T.; Zhang, Z.; Guo, X.; He, M.; Li, J.; Yang, S.; Ma, X.; Feng, Y., Ion Mobility Mass Spectrometry with Molecular Modelling to Reveal Bioactive Isomer Conformations and Underlying Relationship with Isomerization. *Rapid Communications in Mass Spectrometry* **2018,** *32* (22), 1931-1940.

153. D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V., Linking Molecular Models with Ion Mobility Experiments. Illustration with a Rigid Nucleic Acid Structure. *J Mass Spectrom* **2015,** *50* (5), 711-26.

154. Moss, C. L.; Chamot-Rooke, J.; Nicol, E.; Brown, J.; Campuzano, I.; Richardson, K.; Williams, J. P.; Bush, M. F.; Bythell, B.; Paizs, B., et al., Assigning Structures to Gas-Phase Peptide Cations and Cation-Radicals. An Infrared Multiphoton Dissociation, Ion Mobility, Electron Transfer, and Computational Study of a Histidine Peptide Ion. *J Phys Chem B* **2012,** *116* (10), 3445-56.

155. Eschweiler, J. D.; Farrugia, M. A.; Dixit, S. M.; Hausinger, R. P.; Ruotolo, B. T., A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. *Structure* **2018,** *26* (4), 599-606.e3.

156. Allison, T. M.; Bechara, C., Structural Mass Spectrometry Comes of Age: New Insight into Protein Structure, Function and Interactions. *Biochemical Society Transactions* **2019,** *47* (1), 317-327.

157. Bereszczak, J. Z.; Barbu, I. M.; Tan, M.; Xia, M.; Jiang, X.; van Duijn, E.; Heck, A. J., Structure, Stability and Dynamics of Norovirus P Domain Derived Protein Complexes Studied by Native Mass Spectrometry. *J Struct Biol* **2012,** *177* (2), 273-82.

158. Politis, A.; Park, A. Y.; Hyung, S.-J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V., Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLOS ONE* **2010,** *5* (8), e12080.

159. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An Investigation of the Mobility Separation of Some Peptide and Protein Ions Using a New Hybrid Quadrupole/Travelling Wave Ims/Oa-Tof Instrument. *International Journal of Mass Spectrometry* **2007,** *261* (1), 1-12.

160. Giles, K.; Williams, J. P.; Campuzano, I., Enhancements in Travelling Wave Ion Mobility Resolution. *Rapid Communications in Mass Spectrometry* **2011,** *25* (11), 1559-1566.

161. Williams, J. P.; Brown, J. M.; Campuzano, I.; Sadler, P. J., Identifying Drug Metallation Sites on Peptides Using Electron Transfer Dissociation (Etd), Collision Induced Dissociation (Cid) and Ion Mobility-Mass Spectrometry (Im-Ms). *Chemical Communications* **2010,** *46* (30), 5458-5460.

162. Zhao, Y.; Yang, J. Y.; Thieker, D. F.; Xu, Y.; Zong, C.; Boons, G.-J.; Liu, J.; Woods, R. J.; Moremen, K. W.; Amster, I. J., A Traveling Wave Ion Mobility Spectrometry (Twims) Study of the Robo1-Heparan Sulfate Interaction. *Journal of the American Society for Mass Spectrometry* **2018,** *29* (6), 1153-1165.

163. Paglia, G.; Astarita, G., Metabolomics and Lipidomics Using Traveling-Wave Ion Mobility Mass Spectrometry. *Nature Protocols* **2017,** *12* (4), 797-813.

164. Hale, O. J.; Sisley, E. K.; Griffiths, R. L.; Styles, I. B.; Cooper, H. J., Native Lesa Twims-Msi: Spatial, Conformational, and Mass Analysis of Proteins and Protein Complexes. *Journal of the American Society for Mass Spectrometry* **2020,** *31* (4), 873-879.

165. Griffiths, R. L.; Sisley, E. K.; Lopez-Clavijo, A. F.; Simmonds, A. L.; Styles, I. B.; Cooper, H. J., Native Mass Spectrometry Imaging of Intact Proteins and Protein Complexes in Thin Tissue Sections. *International Journal of Mass Spectrometry* **2019,** *437*, 23-29.

166. Merenbloom, S. I.; Flick, T. G.; Williams, E. R., How Hot Are Your Ions in Twave Ion Mobility Spectrometry? *J Am Soc Mass Spectrom* **2012,** *23* (3), 553-62.

167. Morsa, D.; Gabelica, V.; De Pauw, E., Effective Temperature of Ions in Traveling Wave Ion Mobility Spectrometry. *Analytical Chemistry* **2011,** *83* (14), 5775-5782.

168. Morsa, D.; Gabelica, V.; De Pauw, E., Fragmentation and Isomerization Due to Field Heating in Traveling Wave Ion Mobility Spectrometry. *Journal of the American Society for Mass Spectrometry* **2014,** *25* (8), 1384-1393.

169. Tsybin, Y. O.; Baykut, G. Method and Device for Irradiating Ions in an Ion Cyclotron Resonance Trap with Photons and Electrons. Office, U. S. P. a. T., US6803569B2, 2003.

170. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *Journal of Cheminformatics* **2012,** *4* (1), 17.

171. M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R.

Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian 09, Revision D.01. *Gaussian, Inc., Wallingford CT* **2016**.

172. Singh, U. C.; Kollman, P. A., An Approach to Computing Electrostatic Charges for Molecules. *Journal of Computational Chemistry* **1984,** *5* (2), 129-145.

173. Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H., Amber Force-Field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *Journal of Molecular Modeling* **2006,** *12* (3), 281-289.

174. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, Amber 2016. *University of California, San Francisco* **2016**.

175. Lee, J. H.; Pollert, K.; Konermann, L., Testing the Robustness of Solution Force Fields for Md Simulations on Gaseous Protein Ions. *The Journal of Physical Chemistry B* **2019,** *123* (31), 6705-6715.

176. Roe, D. R.; Cheatham, T. E., Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013,** *9* (7), 3084-3095.

177. Palmer, A. D.; Bunch, J.; Styles, I. B., The Use of Random Projections for the Analysis of Mass Spectrometry Imaging Data. *J Am Soc Mass Spectrom* **2015,** *26* (2), 315-22.

178. Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V., Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nat Protoc* **2008,** *3* (7), 1139-52.

179. Lössl, P.; van de Waterbeemd, M.; Heck, A., Jr., The Diverse and Expanding Role of Mass Spectrometry in Structural and Molecular Biology. *The EMBO journal* **2016,** *35* (24), 2634-2657.

180. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., Gromacs: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995,** *91* (1), 43-56.

181. Lopes, P. E. M.; Guvench, O.; MacKerell, A. D., Jr., Current Status of Protein Force Fields for Molecular Dynamics Simulations. *Methods in molecular biology (Clifton, N.J.)* **2015,** *1215*, 47-71.

182. Konermann, L.; Metwally, H.; McAllister, R. G.; Popa, V., How to Run Molecular Dynamics Simulations on Electrospray Droplets and Gas Phase Proteins: Basic Guidelines and Selected Applications. *Methods* **2018,** *144*, 104-112.

183. Martens, J.; Grzetic, J.; Berden, G.; Oomens, J., Structural Identification of Electron Transfer Dissociation Products in Mass Spectrometry Using Infrared Ion Spectroscopy. *Nat Commun* **2016,** *7*, 11754.

184. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *The Journal of Chemical Physics* **2004,** *120* (24), 11919-11929.

185. Hukushima, K.; Nemoto, K., Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal of the Physical Society of Japan* **1996,** *65* (6), 1604-1608.

186. Forgy, E., Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classification. *Biometrics* **1965,** *21* (3), 768-769.

187. Lloyd, S., Least Squares Quantization in Pcm. *IEEE Transactions on Information Theory* **1982,** *28* (2), 129-137.

188. Bingham, E.; Mannila, H., Random Projection in Dimensionality Reduction: Applications to Image and Text Data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, Association for Computing Machinery: San Francisco, California, 2001; pp 245–250.

189. Chen, B.; Guo, X.; Tucholski, T.; Lin, Z.; McIlwain, S.; Ge, Y., The Impact of Phosphorylation on Electron Capture Dissociation of Proteins: A Top-Down Perspective. *J Am Soc Mass Spectrom* **2017,** *28* (9), 1805-1814.

190. Dajani, R.; Fraser, E.; Roe, S. M.; Yeo, M.; Good, V. M.; Thompson, V.; Dale, T. C.; Pearl, L. H., Structural Basis for Recruitment of Glycogen Synthase Kinase 3beta to the Axin-Apc Scaffold Complex. *The EMBO journal* **2003,** *22* (3), 494-501.

191. Shinde, M. Y.; Sidoli, S.; Kulej, K.; Mallory, M. J.; Radens, C. M.; Reicherter, A. L.; Myers, R. L.; Barash, Y.; Lynch, K. W.; Garcia, B. A., et al., Phosphoproteomics Reveals That Glycogen Synthase Kinase-3 Phosphorylates Multiple Splicing Factors and Is Associated with Alternative Splicing. *The Journal of biological chemistry* **2017,** *292* (44), 18240-18255.

192. Wagner, F. F.; Bishop, J. A.; Gale, J. P.; Shi, X.; Walk, M.; Ketterman, J.; Patnaik, D.; Barker, D.; Walpita, D.; Campbell, A. J., et al., Inhibitors of Glycogen Synthase Kinase 3 with Exquisite Kinome-Wide Selectivity and Their Functional Effects. *ACS Chemical Biology* **2016,** *11* (7), 1952-1963.

193. Wagner, F. F.; Benajiba, L.; Campbell, A. J.; Weïwer, M.; Sacher, J. R.; Gale, J. P.; Ross, L.; Puissant, A.; Alexe, G.; Conway, A., et al., Exploiting an Asp-Glu “Switch” in Glycogen Synthase Kinase 3 to Design Paralog-Selective Inhibitors for Use in Acute Myeloid Leukemia. *Science Translational Medicine* **2018,** *10* (431), eaam8460.

194. Ramurthy, S.; Pfister, K. B.; Boyce, R. S.; Brown, S. P.; Costales, A. Q.; Desai, M. C.; Fang, E.; Levine, B. H.; Ng, S. C.; Nuss, J. M., et al., Discovery and Optimization of Novel Pyridines as Highly Potent and Selective Glycogen Synthase Kinase 3 Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2020,** *30* (4), 126930.

195. Henley, Z. A.; Bax, B. D.; Inglesby, L. M.; Champigny, A.; Gaines, S.; Faulder, P.; Le, J.; Thomas, D. A.; Washio, Y.; Baldwin, I. R., From Pim1 to Pi3kδ Via Gsk3β: Target Hopping through the Kinome. *ACS Medicinal Chemistry Letters* **2017,** *8* (10), 1093-1098.

196. Gobbo, D.; Piretti, V.; Di Martino, R. M. C.; Tripathi, S. K.; Giabbai, B.; Storici, P.; Demitri, N.; Girotto, S.; Decherchi, S.; Cavalli, A., Investigating Drug–Target Residence Time in Kinases through Enhanced Sampling Simulations. *Journal of Chemical Theory and Computation* **2019,** *15* (8), 4646-4659.

197. Singh, J.; Thornton, J. M., Sirius: An Automated Method for the Analysis of the Preferred Packing Arrangements between Protein Groups. *Journal of Molecular Biology* **1990,** *211* (3), 595-615.

198. Gallivan, J. P.; Dougherty, D. A., Cation-Pi Interactions in Structural Biology. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (17), 9459-9464.

199. Jones, A. W.; Mikhailov, V. A.; Iniesta, J.; Cooper, H. J., Electron Capture Dissociation Mass Spectrometry of Tyrosine Nitrated Peptides. *Journal of the American Society for Mass Spectrometry* **2010,** *21* (2), 268-277.

200. Chrisman, P. A.; Pitteri, S. J.; Hogan, J. M.; McLuckey, S. A., So2−· Electron Transfer Ion/Ion Reactions with Disulfide Linked Polypeptide Ions. *Journal of the American Society for Mass Spectrometry* **2005,** *16* (7), 1020-1030.

201. Chalkley, R. J.; Brinkworth, C. S.; Burlingame, A. L., Side-Chain Fragmentation of Alkylated Cysteine Residues in Electron Capture Dissociation Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **2006,** *17* (9), 1271-1274.

202. Simons, J., Mechanisms for S–S and N–Cα Bond Cleavage in Peptide Ecd and Etd Mass Spectrometry. *Chemical Physics Letters* **2010,** *484* (4-6), 81-95.

203. Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O., Principles of Electron Capture and Transfer Dissociation Mass Spectrometry Applied to Peptide and Protein Structure Analysis. *Chemical Society Reviews* **2013,** *42* (12), 5014-5030.

204. Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M., Prosight Lite: Graphical Software to Analyze Top-Down Mass Spectrometry Data. *PROTEOMICS* **2015,** *15* (7), 1235-1238.

205. Kolli, V.; Schumacher, K. N.; Dodds, E. D., Ion Mobility-Resolved Collision-Induced Dissociation and Electron Transfer Dissociation of N-Glycopeptides: Gathering Orthogonal Connectivity Information from a Single Mass-Selected Precursor Ion Population. *Analyst* **2017,** *142* (24), 4691-4702.
206. Lermyte, F.; Konijnenberg, A.; Williams, J. P.; Brown, J. M.; Valkenborg, D.; Sobott, F., Etd Allows for Native Surface Mapping of a 150 Kda Noncovalent Complex on a Commercial Q-Twims-Tof Instrument. *Journal of the American Society for Mass Spectrometry* **2014,** *25* (3), 343-350.
207. Lermyte, F.; Verschueren, T.; Brown, J. M.; Williams, J. P.; Valkenborg, D.; Sobott, F., Characterization of Top-Down Etd in a Travelling-Wave Ion Guide. *Methods* **2015,** *89*, 22-29.
208. Breuker, K.; Oh, H.; Cerda, B. A.; Horn, D. M.; McLafferty, F. W., Hydrogen Atom Loss in Electron-Capture Dissociation: A Fourier Transform-Ion Cyclotron Resonance Study with Single Isotopomeric Ubiquitin Ions. *European Journal of Mass Spectrometry* **2002,** *8* (2), 177-180.
209. Comisarow, M. B.; Grassi, V.; Parisod, G., Fourier Transform Ion Cyclotron Double Resonance. *Chemical Physics Letters* **1978,** *57* (3), 413-416.
210. Han, H.; Xia, Y.; McLuckey, S. A., Ion Trap Collisional Activation of C and Z* Ions Formed Via Gas-Phase Ion/Ion Electron-Transfer Dissociation. *Journal of proteome research* **2007,** *6* (8), 3062-3069.
211. Li, X.; Lin, C.; Han, L.; Costello, C. E.; O'Connor, P. B., Charge Remote Fragmentation in Electron Capture and Electron Transfer Dissociation. *J Am Soc Mass Spectrom* **2010,** *21* (4), 646-56.
212. Breuker, K.; McLafferty, F. W., Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10≪Sup≫−12≪/Sup≫ to 10≪Sup≫2≪/Sup≫ S. *Proceedings of the National Academy of Sciences* **2008,** *105* (47), 18145.
213. McPherson, A.; Gavira, J. A., Introduction to Protein Crystallization. *Acta crystallographica. Section F, Structural biology communications* **2014,** *70* (Pt 1), 2-20.
214. Allison, T. M.; Barran, P.; Cianférani, S.; Degiacomi, M. T.; Gabelica, V.; Grandori, R.; Marklund, E. G.; Menneteau, T.; Migas, L. G.; Politis, A., et al., Computational Strategies and Challenges for Using Native Ion Mobility Mass Spectrometry in Biophysics and Structural Biology. *Analytical Chemistry* **2020**.
215. Møller, C.; Plesset, M. S., Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934,** *46* (7), 618-622.
216. Schütz, M.; Hetzer, G.; Werner, H.-J., Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local Mp2. *The Journal of Chemical Physics* **1999,** *111* (13), 5691-5705.
217. Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E., Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *Journal of Chemical Theory and Computation* **2020,** *16* (4), 2494-2507.
218. Rodgers, M. T.; Ervin, K. M.; Armentrout, P. B., Statistical Modeling of Collision-Induced Dissociation Thresholds. *The Journal of Chemical Physics* **1997,** *106* (11), 4499-4508.
219. Gross, D. S.; Zhao, Y.; Williams, E. R., Dissociation of Heme-Globin Complexes by Blackbody Infrared Radiative Dissociation: Molecular Specificity in the Gas Phase? *Journal of the American Society for Mass Spectrometry* **1997,** *8* (5), 519-524.
220. Schnier, P. D.; Price, W. D.; Strittmatter, E. F.; Williams, E. R., Dissociation Energetics and Mechanisms of Leucine Enkephalin (M + H)+ and (2m + X)+ Ions (X = H, Li, Na, K, and Rb) Measured by Blackbody Infrared Radiative Dissociation. *Journal of the American Society for Mass Spectrometry* **1997,** *8* (8), 771-780.
221. Longchamp, J.-N.; Rauschenbach, S.; Abb, S.; Escher, C.; Latychevskaia, T.; Kern, K.; Fink, H.-W., Imaging Proteins at the Single-Molecule Level. *Proceedings of the National Academy of Sciences* **2017,** *114* (7), 1474.
222. Mikhailov, V. A.; Mize, T. H.; Benesch, J. L.; Robinson, C. V., Mass-Selective Soft-Landing of Protein Assemblies with Controlled Landing Energies. *Anal Chem* **2014,** *86* (16), 8321-8.
223. Verbeck, G.; Hoffmann, W.; Walton, B., Soft-Landing Preparative Mass Spectrometry. *Analyst* **2012,** *137* (19), 4393-4407.

224. Zenaidee, M. A.; Leeming, M. G.; Zhang, F.; Funston, T. T.; Donald, W. A., Highly Charged Protein Ions: The Strongest Organic Acids to Date. *Angewandte Chemie International Edition* **2017,** *56* (29), 8522-8526.

225. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: A Server for Estimating Pkas and Adding Missing Hydrogens to Macromolecules. *Nucleic acids research* **2005,** *33* (Web Server issue), W368-W371.

226. Tan, K. P.; Nguyen, T. B.; Patel, S.; Varadarajan, R.; Madhusudhan, M. S., Depth: A Web Server to Compute Depth, Cavity Sizes, Detect Potential Small-Molecule Ligand-Binding Cavities and Predict the Pka of Ionizable Residues in Proteins. *Nucleic acids research* **2013,** *41* (Web Server issue), W314-W321.

227. Konermann, L., Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation. *The Journal of Physical Chemistry B* **2017,** *121* (34), 8102-8112.

228. Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K., Challenges in Protein-Folding Simulations. *Nature Physics* **2010,** *6* (10), 751-758.

229. Garcia-Viloca, M.; Poulsen, T. D.; Truhlar, D. G.; Gao, J., Sensitivity of Molecular Dynamics Simulations to the Choice of the X-Ray Structure Used to Model an Enzymatic Reaction. *Protein science : a publication of the Protein Society* **2004,** *13* (9), 2341-2354.

230. Ridgeway, M. E.; Wolff, J. J.; Silveira, J. A.; Lin, C.; Costello, C. E.; Park, M. A., Gated Trapped Ion Mobility Spectrometry Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Int J Ion Mobil Spectrom* **2016,** *19* (2), 77-85.

231. Benigni, P.; Thompson, C. J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F., Targeted High-Resolution Ion Mobility Separation Coupled to Ultrahigh-Resolution Mass Spectrometry of Endocrine Disruptors in Complex Mixtures. *Anal Chem* **2015,** *87* (8), 4321-5.

232. Pu, Y.; Ridgeway, M. E.; Glaskin, R. S.; Park, M. A.; Costello, C. E.; Lin, C., Separation and Identification of Isomeric Glycans by Selected Accumulation-Trapped Ion Mobility Spectrometry-Electron Activated Dissociation Tandem Mass Spectrometry. *Analytical chemistry* **2016,** *88* (7), 3440-3443.

233. Voinov, V. G.; Deinzer, M. L.; Barofsky, D. F., Electron Capture Dissociation in a Linear Radiofrequency-Free Magnetic Cell. *Rapid Commun Mass Spectrom* **2008,** *22* (19), 3087-8.

234. Stiving, A. H., S.; Jones, B. J.; Bellina, B.; Brown, J.; Barran, P. E.; Wysocki, V., Coupling 193 Nm Ultraviolet Photodissociation and Ion Mobility for Sequence Characterization of Conformationally-Selected Peptides. *chemrxiv* **2020,** *https://doi.org/10.26434/chemrxiv.12636668.v1*.

Appendix 1: Electronic appendices

A .zip file available for download from:
https://beardatashare.bham.ac.uk/getlink/fiJa5uRVsdodADAKym9B2G8p/Electronic_Appendices%20Corrections.zip

Contents:

- 1-1_gesptoooff_code.txt
- 1-2_MS_ff_parameters
- 1-3_IMoS_input.cla
- 1-4_kmeans_opt.m
- 1-5_solariX_XR_instrument_parameters.xlsx
- 1-6_Synapt_G2S_instrument_paramters.xlsx
- 1-7_14-3-3-z_clustering_information.xlsx
- 1-8_14-3-3-z_non-covalent_interactions.xlsx
- 1-9_GSK3B_clustering_information.xlsx
- 1-10_GSK3B_non-covalent_interactions.xlsx
- 1-11_GSK3B-z17_clustering_information.xlsx
- 1-12_GSK3B-z17_non-covalent_interactions.xlsx
- 1-13_Bcasein-c28_clustering_information.xlsx
- 1-14_Bcasein-c28_non-covalent_interactions.xlsx
- 1-15_Representative_structures

Appendix 2: Mobcal variables

```
MOBIL2
      itn=10
      inp=40
      imp=1000
      cmin=0.0005
      sw1=0.00005d0
      sw2=0.005d0
      dtsf1=0.5d0
      dtsf2=0.1d0
      inwr=1
      ifail=100
      ifailc=0
MOBIL4
      inum=250000
      inor=30
```

Appendix 3: TWIMS calibrant ion lists

Table A3-1 List of calibrant ions used for $^{TW}CCS_{N2\rightarrow He}$ calibration of all 14-3-3-ζ-derived peptide ions

Database CCS (Å²)	**Sequence**	***m/z***	***z***
231.34	DIPVPKPK	447.2764	2
252.18	ANELLINVK	507.3031	2
253.92	EKDIVGAVLK	536.3241	2
282.95	FKDLGEEHFK	625.3142	2
259.73	GVIFYESHGK	568.7904	2
267.73	KQTALVELLK	571.8608	2
266.65	LVNELTEFAK	582.319	2
290.36	HLVDEPQNLIK	653.3617	2
267.56	TGPNLHGLFGR	584.8147	2
308.04	DTDFKLNELRGK	718.3806	2
312.71	EETLMEYLENPK	748.3529	2
313.44	RHPEYAVSVLLR	720.4095	2
282.95	SIGGEVFIDFTK	656.8428	2
278.59	SISIVGSYVGNR	626.3382	2
319.03	CASIQKFGERALK	725.8954	2
313.94	LGEYGFQNALIVR	740.4014	2
283.88	TCVADESHAGCEK	675.2767	2
310.43	VVGLSTLPEIYEK	724.4058	2
297.61	TGQAPGFTYTDANK	735.8466	2
302.28	ANGTTVLVGMPAGAK	693.8741	2
354.9	VLGIDGGEGKEELFR	809.9254	2
334.59	SISIVGSYVGNRADTR	847.9446	2
389.55	AAANFFSASCVPCADQSSFPK	1074.48	2
412.48	WSGFSGGAIECETAENTEECIAK	1216.523	2
415.18	ATDGGAHGVINVSVSEAAIEASTR	1156.577	2

Table A3-2 List of calibrant ions used for $^{TW}CCS_{N2\rightarrow He}$ calibration of all GSK3β-derived peptide $[M+2H]^{2+}$ ions and z_n^{2+} fragments of the GSK3β-pY11 $[M+3H]^{3+}$ ion

Database CCS (Å²)	**Sequence**	**m/z**	**z**
210.68	LVTDLTK	395.24	2
267.73	KQTALVELLK	571.86	2
266.65	LVNELTEFAK	582.32	2
282.95	FKDLGEEHFK	625.31	2
290.36	HLVDEPQNLIK	653.36	2
283.88	TCVADESHAGCEK	675.28	2
313.44	RHPEYAVSVLLR	720.41	2
319.03	CASIQKFGERALK	725.9	2
313.94	LGEYGFQNALIVR	740.4	2
389.55	AAANFFSASCVPCADQSSFPK	1073.977	2
412.48	WSGFSGGAIECETAENTEECIAK	1216.019	2
432	TVGGKEDVIWELLNHAQEHFGK	1254.138	2
330.83	NTDGSTDYGILQINSR	877.4218	2
412.61	NLCNIPCSALLSSDITASVNCAKK	1233.114	2

Table A3-3 List of calibrant ions used for $^{TW}CCS_{N2\rightarrow He}$ *calibration of* z_n^+ *fragments of the GSK3β-pY11* $[M+3H]^{3+}$ *and* $[M+2H]^{2+}$ *ions*

Database CCS (Å²)	*Sequence*	*m/z*	*z*
97.89	EK	276.1559	1
110.03	LR	289.2114	1
101.76	NR	290.1703	1
104.07	QR	304.1859	1
129.74	QIK	388.256	1
123.14	FPK	391.2345	1
132.39	SLGK	404.2509	1
128.01	GTDK	420.2094	1
136.71	VGTR	433.2649	1
131.96	YTR	440.2383	1
139.78	IPSK	444.2822	1
140.18	TVGGK	461.2724	1
138.92	ADEK	462.22	1
149.68	LSQK	475.288	1
142.74	DTHK	500.2469	1
149.94	HKPK	510.3278	1
159.31	ADLAK	517.2986	1
160.46	FWGK	537.2825	1
163.7	VASLR	546.3489	1
168.36	AFDEK	609.2884	1
183.11	DLLFK	635.3768	1
182.27	LNELR	645.381	1
181.29	IETMR	650.3421	1
175.98	TPVSEK	660.3568	1
185.85	KFWGK	666.3853	1
185.37	AWSVAR	690.3813	1
200.45	GDVAFVK	735.4041	1
205.76	LVTDLTK	789.4722	1
206.4	ATEEQLK	818.426	1
209.3	DSADGFLK	852.4103	1
223.42	AEFVEVTK	922.4886	1
228.03	YLYEIAR	928.5018	1
225.29	DLGEEHFK	974.4583	1
239.27	LVVSTQTALA	1001.576	1
242.91	YYGYTGAFR	1098.513	1
267.51	LVNELTEFAK	1163.631	1
282.63	HLVDEPQNLIK	1305.717	1
323.84	TAGWNIPMGLLYSK	1550.804	1

Table A3-4 List of calibrant ions used for $^{TW}CCS_{N2\rightarrow He}$ calibration of c and z fragments of the β-casein $[M+23H]^{23+}$ ion

Database CCS (Å²)	*Calibrant*	*m/z*	*z*
89	$(Ala)_3$	233.1413	1
100	$(Ala)_4$	304.1784	1
114	$(Ala)_5$	375.2156	1
128	$(Ala)_6$	446.2527	1
141	$(Ala)_7$	517.2898	1
157	$(Ala)_8$	588.3269	1
170	$(Ala)_9$	659.364	1
181	$(Ala)_{10}$	730.4011	1
194	$(Ala)_{11}$	801.4382	1
206	$(Ala)_{12}$	872.4753	1
217	$(Ala)_{13}$	943.5124	1
228	$(Ala)_{14}$	1014.55	1
208	$(Ala)_{12}$	436.2377	2
220	$(Ala)_{13}$	471.7562	2
232	$(Ala)_{14}$	507.2748	2
243	$(Ala)_{15}$	542.7933	2
255	$(Ala)_{16}$	578.3119	2
265	$(Ala)_{17}$	613.8304	2
276	$(Ala)_{18}$	649.349	2
287	$(Ala)_{19}$	684.8675	2
297	$(Ala)_{20}$	720.3861	2
308	$(Ala)_{21}$	755.9047	2
317	$(Ala)_{22}$	791.4232	2
327	$(Ala)_{23}$	826.9418	2
337	$(Ala)_{24}$	862.4603	2
348	$(Ala)_{25}$	897.9789	2
358	$(Ala)_{26}$	933.4974	2
386	$(Ala)_{23}$	551.2945	3
399	$(Ala)_{24}$	574.9735	3
412	$(Ala)_{25}$	598.6526	3
425	$(Ala)_{26}$	622.3316	3
438	$(Ala)_{27}$	646.0107	3
452	$(Ala)_{28}$	669.6897	3
465	$(Ala)_{29}$	693.3687	3
479	$(Ala)_{30}$	717.0478	3
490	$(Ala)_{31}$	740.7268	3
502	$(Ala)_{32}$	764.4058	3
516	$(Ala)_{33}$	788.0849	3
1670	Ubiquitin	951.1111	9
1730	Ubiquitin	856	10
1800	Ubiquitin	778.1818	11

1890	Ubiquitin	713.3333	12
1980	Ubiquitin	658.4615	13

Appendix 4: TWIMS CCS calibration curves

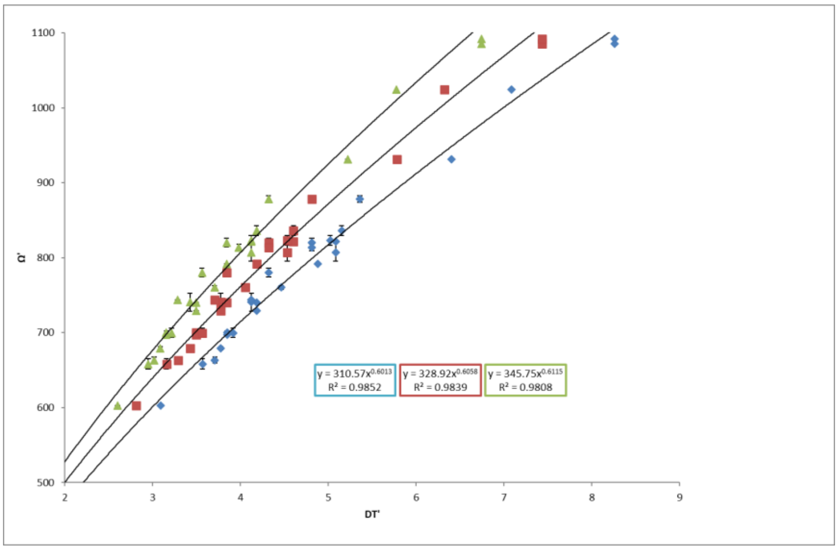


Figure A4-1 Graph used for $^{TW}CCS_{N2\rightarrow He}$ calibration of all 14-3-3-ζ-derived peptides

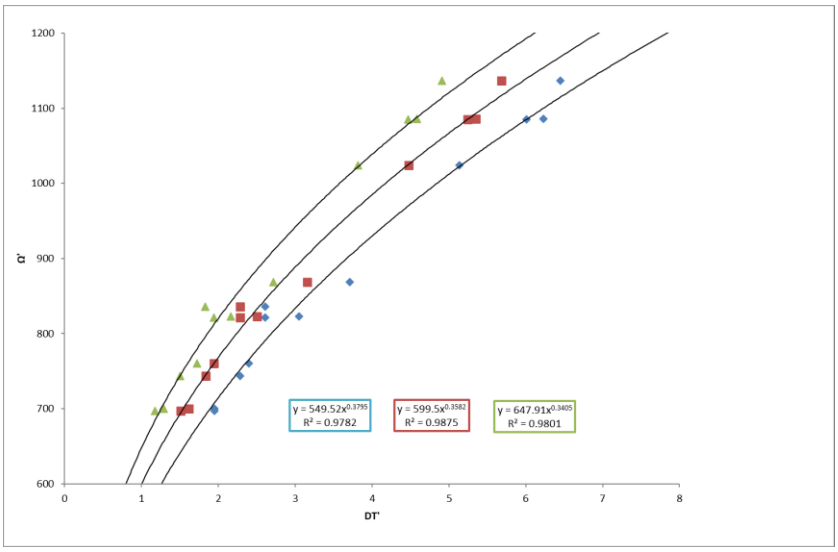


Figure A4-2 Graph used for $^{TW}CCS_{N2\rightarrow He}$ calibration of GSK3β-pY11 and GSK3β-Y11

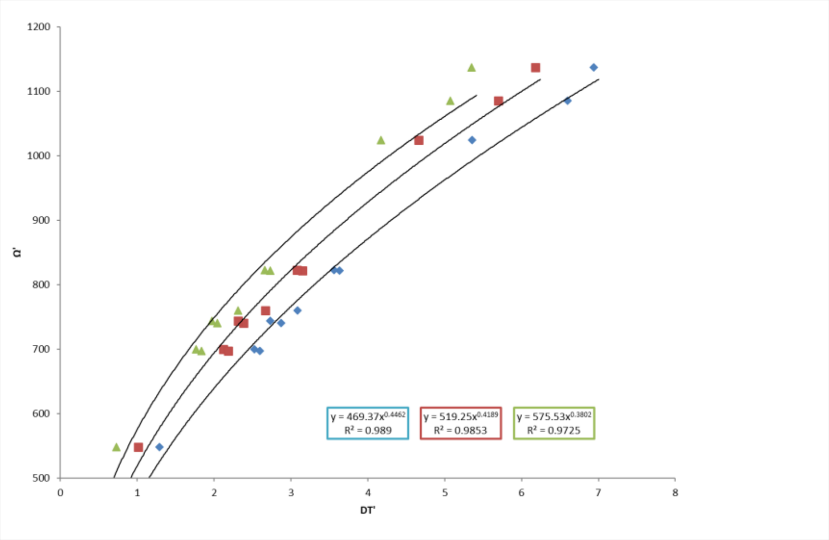


Figure A4-3 Graph used for $^{TW}CCS_{N2\rightarrow He}$ *calibration of GSK3β-pS11 and GSK3β-S11*

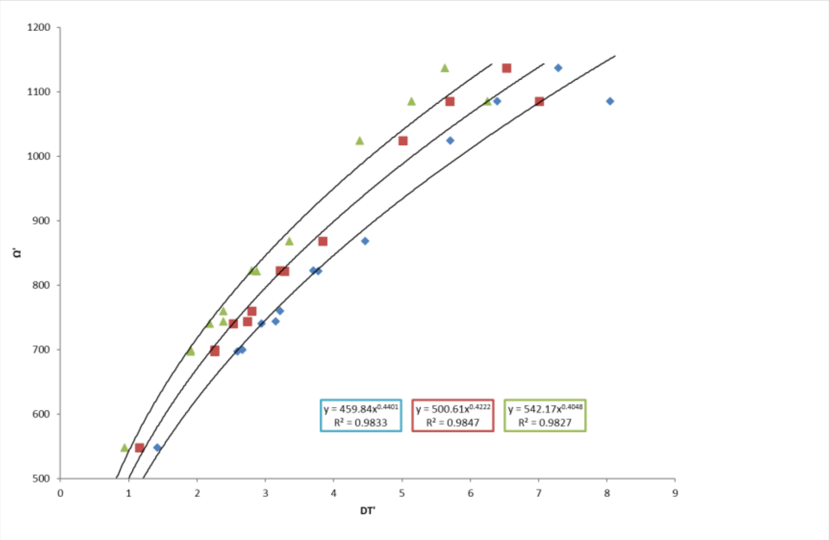


Figure A4-4 Graph used for $^{TW}CCS_{N2\rightarrow He}$ *calibration of GSK3β-pY11-noC and GSK3β-Y11-noC*

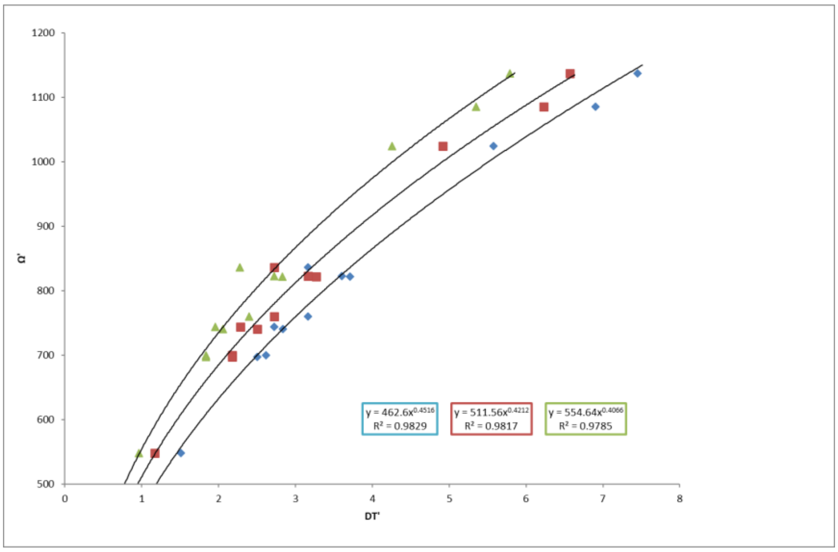


Figure A4-5 Graph used for $^{TW}CCS_{N2\rightarrow He}$ calibration of GSK3β-pY11-noEnoC- and GSK3β-Y11-noEnoC

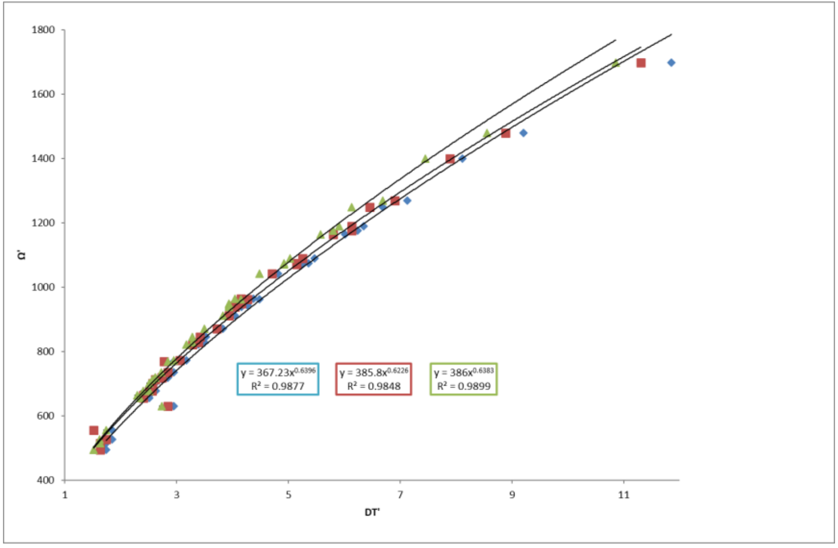


Figure A4-6 Graph used for $^{TW}CCS_{N2\rightarrow He}$ calibration of c^+ and z^+ fragments of GSK3β-pY11 $[M+2H]^{2+}$ and $[M+3H]^{3+}$

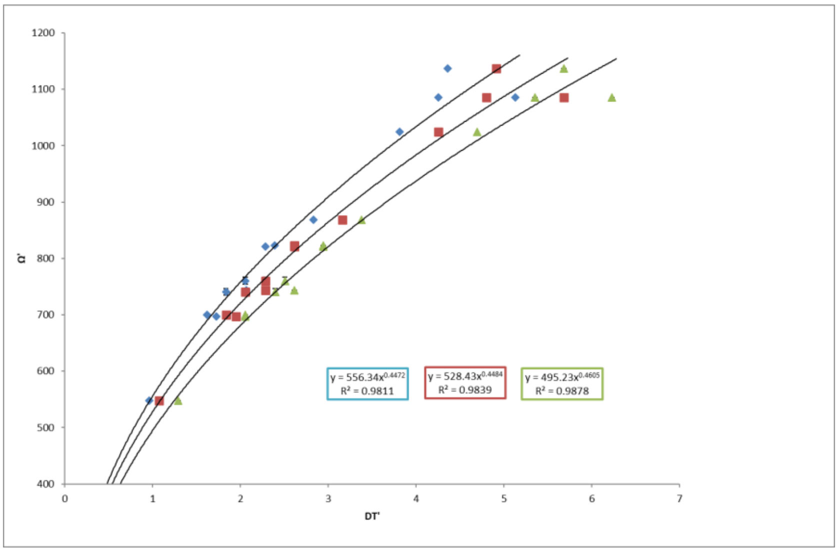


Figure A4-7 Graph used for $^{TW}CCS_{N2\rightarrow He}$ calibration of c^{2+} and z^{2+} fragments of GSK3β-pY11 $[M+3H]^{3+}$

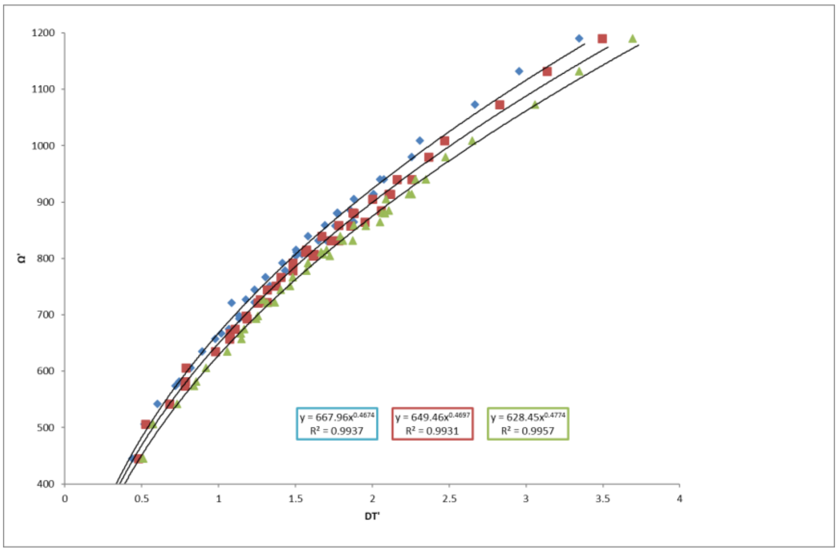
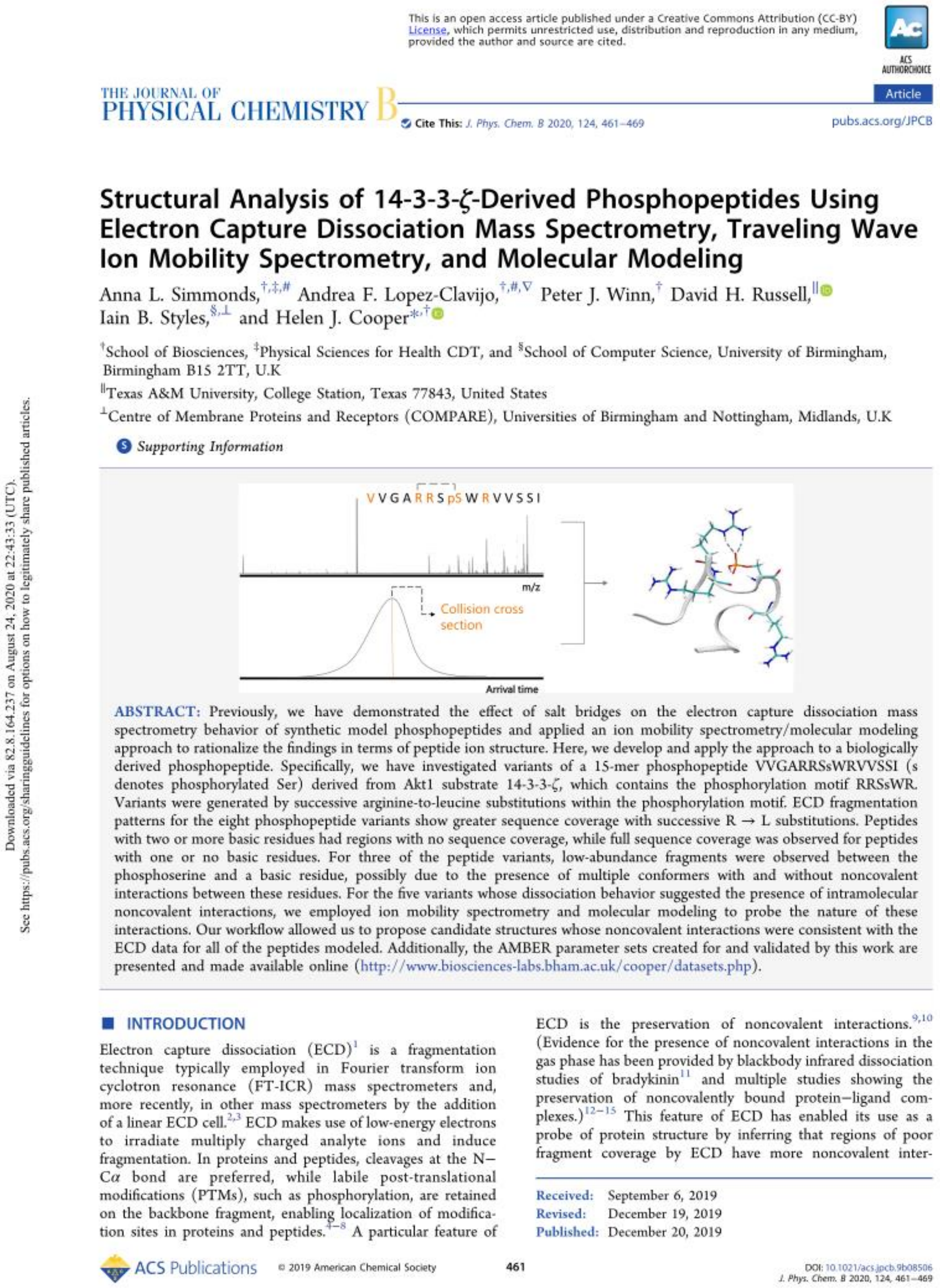


Figure A4-8 Graph used for $^{TW}CCS_{N2\rightarrow He}$ *calibration of c and z fragments of the β-casein* $[M+23H]^{23+}$ *ion*

Appendix 5: Copy of "Structural Analysis of 14-3-3-ζ-Derived Phosphopeptides Using Electron Capture Dissociation Mass Spectrometry, Traveling Wave Ion Mobility Spectrometry, and Molecular Modeling"



ACS AUTHORCHOICE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: *J. Phys. Chem. B* 2020, 124, 461−469

pubs.acs.org/JPCB

Structural Analysis of 14-3-3-ζ-Derived Phosphopeptides Using Electron Capture Dissociation Mass Spectrometry, Traveling Wave Ion Mobility Spectrometry, and Molecular Modeling

Anna L. Simmonds,[†,‡,#] Andrea F. Lopez-Clavijo,[†,#,∇] Peter J. Winn,[†] David H. Russell,[∥] Iain B. Styles,[§,⊥] and Helen J. Cooper[*,†]

[†]School of Biosciences, [‡]Physical Sciences for Health CDT, and [§]School of Computer Science, University of Birmingham, Birmingham B15 2TT, U.K

[∥]Texas A&M University, College Station, Texas 77843, United States

[⊥]Centre of Membrane Proteins and Receptors (COMPARE), Universities of Birmingham and Nottingham, Midlands, U.K

Supporting Information

ABSTRACT: Previously, we have demonstrated the effect of salt bridges on the electron capture dissociation mass spectrometry behavior of synthetic model phosphopeptides and applied an ion mobility spectrometry/molecular modeling approach to rationalize the findings in terms of peptide ion structure. Here, we develop and apply the approach to a biologically derived phosphopeptide. Specifically, we have investigated variants of a 15-mer phosphopeptide VVGARRSsWRVVSSI (s denotes phosphorylated Ser) derived from Akt1 substrate 14-3-3-ζ, which contains the phosphorylation motif RRSsWR. Variants were generated by successive arginine-to-leucine substitutions within the phosphorylation motif. ECD fragmentation patterns for the eight phosphopeptide variants show greater sequence coverage with successive R → L substitutions. Peptides with two or more basic residues had regions with no sequence coverage, while full sequence coverage was observed for peptides with one or no basic residues. For three of the peptide variants, low-abundance fragments were observed between the phosphoserine and a basic residue, possibly due to the presence of multiple conformers with and without noncovalent interactions between these residues. For the five variants whose dissociation behavior suggested the presence of intramolecular noncovalent interactions, we employed ion mobility spectrometry and molecular modeling to probe the nature of these interactions. Our workflow allowed us to propose candidate structures whose noncovalent interactions were consistent with the ECD data for all of the peptides modeled. Additionally, the AMBER parameter sets created for and validated by this work are presented and made available online (http://www.biosciences-labs.bham.ac.uk/cooper/datasets.php).

■ INTRODUCTION

Electron capture dissociation (ECD)[1] is a fragmentation technique typically employed in Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers and, more recently, in other mass spectrometers by the addition of a linear ECD cell.[2,3] ECD makes use of low-energy electrons to irradiate multiply charged analyte ions and induce fragmentation. In proteins and peptides, cleavages at the N−Cα bond are preferred, while labile post-translational modifications (PTMs), such as phosphorylation, are retained on the backbone fragment, enabling localization of modification sites in proteins and peptides.[4−8] A particular feature of ECD is the preservation of noncovalent interactions.[9,10] (Evidence for the presence of noncovalent interactions in the gas phase has been provided by blackbody infrared dissociation studies of bradykinin[11] and multiple studies showing the preservation of noncovalently bound protein−ligand complexes.)[12−15] This feature of ECD has enabled its use as a probe of protein structure by inferring that regions of poor fragment coverage by ECD have more noncovalent inter-

Received: September 6, 2019
Revised: December 19, 2019
Published: December 20, 2019

DOI: 10.1021/acs.jpcb.9b08506
J. Phys. Chem. B 2020, 124, 461−469

Table 1. Model Peptide Sequences[a]

peptide sequence	abbreviated name	monoisotopic mass (Da)	$[M + 2H]^{2+}$ (m/z)	$^{TW}CCS_{N2\rightarrow He}$ (Å^2)
VVGARRSsWRVVSSI	Pep-01	1737.27158	869.64307	354.82 ± 0.74
VVGARLSsWRVVSSI	Pep-02	1694.25453	848.13455	346.17 ± 1.05
VVGARRSsWLVVSSI	Pep-03	1694.25453	848.13455	329.55 ± 0.65
VVGALRSsWRVVSSI	Pep-04	1694.25453	848.13455	333.42 ± 1.10
VVGARLSsWLVVSSI	Pep-05	1609.19053	805.60255	332.47 ± 0.49
VVGALRSsWLVVSSI	Pep-06	1609.19053	805.60255	331.42 ± 1.91
VVGALLSsWRVVSSI	Pep-07	1609.19053	805.60255	336.19 ± 1.39
VVGALLSsWLVVSSI	Pep-08	1608.22043	805.11750	329.73 ± 0.65

[a] Arginine and phosphoserine residues are shown in bold.

actions and therefore more rigid structures. ECD studies of the IR-induced unfolding of ubiquitin ions showed the differences between solution- and gas-phase unfolding pathways,[16] with later work focusing on the unfolding of ubiquitin ions via desolvation.[17] In a study of the protein KIX, the presence of noncovalent interactions was shown to increase the longevity of the native solution-phase structure in the gas phase.[18] A study combining ECD and ion mobility spectrometry (IMS) data allowed the extent and location of collision-induced unfolding in human hemoglobin to be discerned,[19] while similar techniques were used to study the unfolding landscape of both the wild type and mutants of the protein Ltn.[20] Similarly, electron transfer dissociation (ETD), an electron-based dissociation method similar to ECD, has been used in conjunction with IMS to study the unfolding of several proteins.[21] ETD has also been used to predict intramolecular salt-bridge locations in several proteins by observing regions of the peptide sequence without fragment coverage. These proposed salt-bridge locations were then compared to all possible salt bridges and were found to closely resemble the salt bridges of the native protein structures.[22,23] It has been shown that both ECD and ETD can be used to determine the site of intermolecular interactions between acidic and basic peptides.[12]

We have shown previously that reduced ECD sequence coverage is observed in a suite of synthetic phosphopeptides based on the sequence APLSFRGSLPKSYVK, particularly in peptides containing two or more basic residues.[24] We concluded that this observation was the result of salt-bridge interactions between the phosphate and the basic residues. This conclusion was in agreement with prior ion mobility spectrometry experiments, which revealed that phosphopeptides have lower CCS than would be predicted for a random coil.[25−27] This compaction has been attributed to noncovalent interactions between the phosphate group and positively charged residues in the peptide sequence.[25] In more recent work, we showed that the presence of phosphorylated serine and basic amino acids alone (in peptides that otherwise only contained alanine and proline residues) was insufficient to inhibit ECD fragmentation and that the conformation of the peptide (and therefore the relative positions of the interacting residues) was important in determining ECD behavior.[28]

Further work on the APLSFRGSLPKSYVK phosphopeptides used an ion mobility spectrometry/molecular modeling (MM) approach to assess whether model structures could rationalize the observed ECD behavior.[29] The combined application of IMS and MM is a well-established approach for the investigation of gas-phase ion structure. That is, collision cross sections (CCS) are derived from IMS experiments and compared with those calculated for molecular models, enabling the viability of candidate structures to be assessed.[30] Traveling wave ion mobility spectrometry (TWIMS) is one such IMS platform[31] through which CCS can be calculated from an ion's arrival time by means of a calibration to reference standards.[32] TWIMS and MM have been successfully combined for structural studies of various molecular classes including drugs,[33] nucleic acids,[34] peptides,[35] and proteins and their complexes.[36]

Here, we have combined ECD, TWIMS, and MM for the study of the RRSsWR phosphorylation motif, from the 15-mer phosphopeptide VVGARRSsWRVVSSI (s denotes phosphorylated Ser) found in the Akt1 substrate 14-3-3-ζ. The S58 residue of 14-3-3-ζ has been shown to be phosphorylated in vivo by PKB/Akt[37] and SDK.[38] Successive arginine-to-leucine substitutions allowed us to examine the effect of multiple positive-charge carriers on ECD behavior. Further, we found that our approach allowed us to propose candidate structures consistent with the ECD and IMS data for all peptides investigated and interrogate their possible protonation patterns and conformers.

■ MATERIALS AND METHODS

Materials. Model peptides were synthesized by GenicBio (Shanghai, China) and used without further purification. Methanol (liquid chromatography−mass spectrometry (LC−MS) grade), water (LC−MS grade), and formic acid (LC−MS grade) were purchased from Fisher-Scientific (Leicestershire, U.K.). Stocks (1 mg/mL) of Pep-01 to Pep-08 (see Table 1) in methanol/water/formic acid (39.9:60:0.1) were diluted in methanol/water/formic acid (49.5:49.5:1) to a final concentration of ∼1.0 μM. Tryptic peptides were obtained from Thermo Scientific (Waltham, MA) and diluted to 100 μL in water.

Electron Capture Dissociation Mass Spectrometry. Samples were introduced to a 7 Tesla solariX-XR (Bruker Daltonics, Bremen, Germany) mass spectrometer equipped with a ParaCell via static positive ion mode nanospray ionization (nESI) with pulled glass capillaries (P-97 Flaming/Brown micropipette puller tip Sutter Instrument Company, Novato, CA). The capillary voltage was 800 V, and the temperature was 120 °C. Peptide ions were isolated in the quadrupole (isolation width 5 m/z) prior to fragmentation by ECD. ECD was performed by the use of thermal electrons at a current of 1.50 A, with a cathode bias of 0.6 V and a lens potential of 10 V. The pulse length was varied between 0.1 and 0.5 s according to each precursor ion. Data analysis was performed with Data Analysis 4.2 software (Bruker Daltonics) and manually searched for a, x, $c^{\bullet}/c$, b, y, $z/z^{\bullet}$ fragment ions.

Traveling Wave Ion Mobility Mass Spectrometry. TWIMS experiments were performed on a Synapt G2S mass

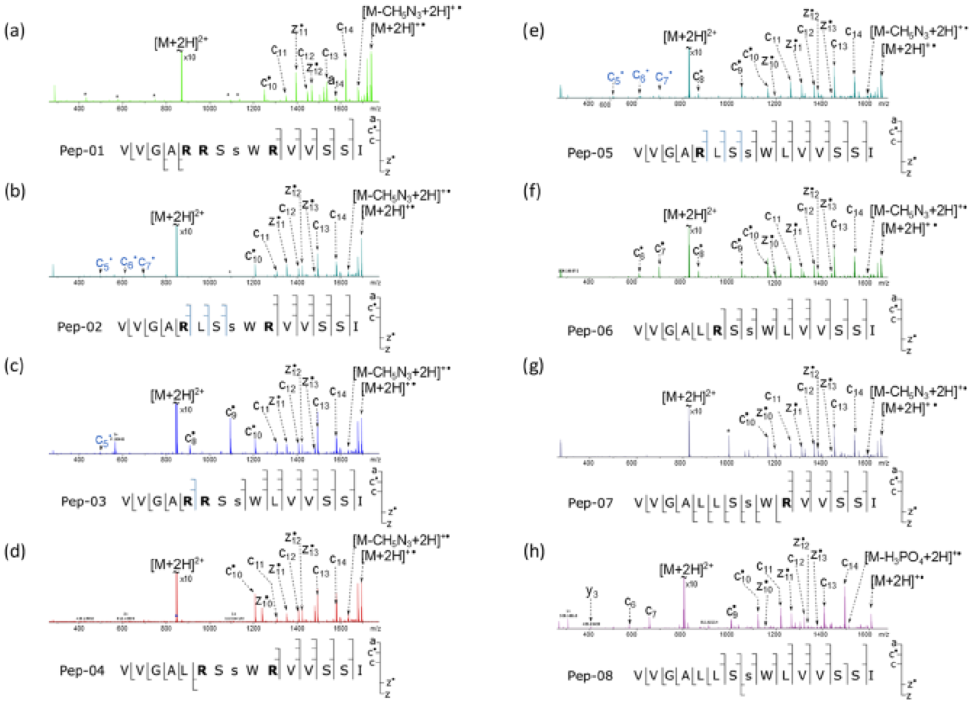


Figure 1. ECD mass spectra and corresponding fragmentation patterns observed for (a) Pep-01, (b) Pep-02, (c) Pep-03, (d) Pep-04, (e) Pep-05, (f) Pep-06, (g) Pep-07, and (h) Pep-08. Low abundance fragments are indicated with blue ticks in the fragmentation pattern and blue labels in the mass spectra.

spectrometer (Waters Corp., Milford, MA) equipped with a Triversa Nanomate nanoelectrospray (nESI) source (Advion, Ithaca, NY). Samples were introduced using nESI voltages in the range of 1.2−1.8 kV and gas pressures of 0.3−0.35 psi. TWIMS was operated at a wave velocity of 450 $\mathrm{ms^{-1}}$ and wave amplitudes of 19, 20, and 21 V, with the TWIMS cell maintained at ~2 mbar of nitrogen. Further information on instrument parameters can be found in Supporting Table S1, while details of the pressure readings observed during these experiments are in Supporting Table S2. Arrival time distributions were produced using a mass range that contained all of the isotopologues of each peptide, and collision cross sections (CCSs) were calculated from the apex value. CCS calibration was performed by an adaptation of the procedure described by Ruotolo et al.[32] Reduced drift time (DT′) of the calibrant ions was calculated by 1

$$\mathrm{DT}' = \mathrm{DT} - \frac{c\sqrt{m/z}}{1000} \tag{1}$$

where DT is the drift time and c is the "enhanced duty cycle" (EDC) delay coefficient (c = 1.43 for our instrument). Reduced cross section (Ω') of the calibrant ions was calculated by 2

$$\Omega' = \frac{{}^{\mathrm{DT}}\mathrm{CCS}_{\mathrm{He}}}{z\sqrt{\frac{1}{\mu}}} \tag{2}$$

where z is the charge of the ion and μ is the reduced mass. A power regression of $\Omega' = A{\cdot}DT'^{b}$ gives values for A and b, which can be used to calculate Ω for the analyte ions by 3

$$\mathrm{CCS}_{\mathrm{analyte}} = \frac{A{\cdot}z{\cdot}{}^{\mathrm{TW}}\mathrm{DT}'_{\mathrm{N_2}}{}^{b}}{\sqrt{\mu}} \tag{3}$$

This process was carried out over three wave amplitudes, from which a mean and standard deviation of $\mathrm{CCS}_{\mathrm{analyte}}$ was calculated. Standards used for calibration were $[\mathrm{M} + 2\mathrm{H}]^{2+}$ tryptic peptides of BSA, cytochrome c, and myoglobin, with reference drift-tube helium CCS obtained from the Clemmer database.[39] The list of calibrant ions used can be found in Supporting Table S3, and the calibration curves are in Supporting Figure S1. TWIMS data and data processing are reported according to the guidelines published by Gabelica et al.[40] Data acquisition and processing were carried out using MassLynx 4.1 and Driftscope 2.1 software (Waters Corporation, Manchester, U.K.).

Molecular Modeling. Parameters for neutral phosphoserine, neutral arginine, neutral C-terminal isoleucine, and neutral N-terminal valine residues were built in Avogadro 1.2.[41] with an acetyl-glycine cap on the N-terminus and an amidated glycine cap at the C-terminus and then optimized using Gaussian 09 (HF 6-31G (d,p)).[42] The electrostatic potential surface of the residue was calculated using the Merz−Singh−Kollman scheme,[43] and the residue was added to a .off library, using the prepgen, parmchk2, and tleap AMBER utility programs. The .off libraries and corresponding .frcmod files for these residues are available online (http://www.biosciences-labs.bham.ac.uk/cooper/datasets.php). Molecular modeling was performed in the AMBER16 suite[44] using the ff99sb and phos10aa forcefields. While these forcefields were developed for solvated proteins and are not as accurate for gas-phase simulations, they are considered suitable for semi-quantitative, gas-phase molecular modeling studies.[45] Simulations were performed using a three-tiered simulated annealing (SA) approach (t_{SA} = 25.2 fs, T = 1000 K; tier 1, n = 1000; tier 2, n = 12 000; tier 3, n = 12 000) from an extended starting structure to produce a 25 000-strong ensemble. MM ensembles were processed using cpptraj.[46] The simulated helium CCS for each of these structures was calculated in IMoS[47] using the trajectory method. A new CCS-matched ensemble was created from models whose simulated CCS was within the experimental CCS ± 5%. Onward analysis of CCS-matched ensembles was performed using MATLAB R2018b (Mathworks, Natick, MA).

Pairwise distances between all atoms were calculated for each model of the CCS-matched ensemble to create an $n_{\text{atoms}}{}^2 \times n_{\text{structures}}$ matrix, which was subjected to dimensionality reduction by random projection following the procedure described by Palmer et al.[48] The compression factor of the random projection was optimized on a subset of the ensemble to select the greatest compression factor, from which the ensemble could still be accurately reconstructed. For this optimization, the L1 norm of the difference between the original data (M) and the reconstructed data ($Q \times P$) was used as an error metric and calculated for a number of compression factors. Fitting a double exponential to these data and graphically solving for the compression factor at 2/3 the point of maximum curvature gives the optimum compression factor. The projections produced were clustered by k-means (n = 10, Euclidean), and the sum of least-squares method was used to define the structures closest to the centroids of each of the 10 clusters, which would serve as a representative of each cluster.

■ RESULTS AND DISCUSSION

The sequences, calculated monoisotopic masses, calculated mass-to-charge ratios, and experimentally derived CCS of the doubly protonated ions of the eight phosphopeptides derived from 14-3-3-ζ (Pep-01−Pep-08) are shown in Table 1. Pep-01 corresponds to the wild-type sequence, Pep-02−Pep-04 contains single Arg→Leu substitutions, and Pep-05 to Pep-07 contains two Arg→Leu substitutions. In Pep-08, all arginine residues have been substituted for Leu residues.

Electron Capture Dissociation. Figure 1a shows the ECD mass spectrum of the doubly charged [VVGARRSsWRVVSSI + 2H]$^{2+}$ (Pep-01) peptide ion. The observed fragment ions are summarized in Supporting Table S4. There is an absence of c/z-type ions within the **R**$_5$**R**Ss**WR**$_{10}$ phosphorylated motif, suggesting the presence of interactions between the phosphoserine and the side chain of Arg5 and Arg10. Figure 1b shows the ECD mass spectrum of the peptide with the motif **R**$_5$LSsW**R**$_{10}$ (Pep-02), in which Arg6 was replaced with Leu compared to the wild-type motif. A summary of the observed fragments can be found in Supporting Table S5. No fragment ions resulting from cleavage between the phosphoserine and the Arg at position 10 are observed, suggesting an interaction between these two residues. There are, however, very low abundance (i.e., an order of magnitude smaller than the next most abundant) fragment ions observed, resulting from cleavages between the Arg at position 5 and the phosphoserine ($c_5/c_5{}^\bullet$, $c_6/c_6{}^\bullet$, and $c_7/c_7{}^\bullet$), indicated by blue ticks in the fragmentation pattern in Figure 1b. We hypothesize that these fragments may derive from a minor peptide conformation, which lacks an interaction between Arg5 and pSer, and that a major conformation exists, which does contain this interaction. Fragmentation of the major conformer does not result in the formation of $c_5/c_5{}^\bullet$, $c_6/c_6{}^\bullet$, and $c_7/c_7{}^\bullet$, and these fragments are therefore detected in very low abundance as they originate only from the minor conformation. The ECD mass spectrum of Pep-03 (Figure 1c), in which Arg10 is replaced with Leu, lacks fragments between Arg6 and phosphoserine (see Supporting Table S6). There is a very low abundance fragment ($c_5/c_5{}^\bullet$) between Arg5 and Arg6 (see the blue tick in the fragmentation pattern in Figure 1c), which may again indicate the presence of major conformation(s) in which Arg5 is interacting with pSer and minor conformation(s) where it is not. Figure 1d shows the ECD mass spectrum of Pep-04, in which Arg5 is replaced with Leu. See Supporting Table S7. No fragments were observed within the R$_6$SsWR$_{10}$ motif in positions 6−9. It is notable that the substitution of Arg10 by Leu (Pep-03) results in the largest difference in the peptide's ECD behavior compared to that of the wild-type fragmentation spectrum, with two additional high abundance fragments being observed. When Arg5 or Arg6 is substituted, the region of the peptides' sequence in which fragments are not observed is more similar to the wild type, especially with low abundance fragments excluded. This observation implies that there is some redundancy between Arg5 and Arg6 in maintaining the fold of the peptide, i.e., if one is not present, the other is available to interact with pSer. The results obtained for these peptides suggest the presence of interactions between phosphoserine and Arg at position i ± 2, which are sufficiently strong to prevent separation of any ECD fragments, provided that another Arg residue is present.

To test this hypothesis, peptides in which two Arg residues were replaced by two Leu residues in the **R**$_5$**R**Ss**WR**$_{10}$ motif, see Table 1 (Pep-05, Pep-06, and Pep-07), were subjected to ECD. The resulting ECD mass spectra show complete fragment coverage, see Figure 1e−g and Supporting Tables S8−S10, although it is notable that for Pep-05, which has the motif **R**$_5$LSsWL, the fragment ions observed between the Arg and pSer residues ($c_5/c_5{}^\bullet$, $c_6/c_6{}^\bullet$, and $c_7/c_7{}^\bullet$) are of very low abundance. As with Pep-02 and Pep-03, these fragments (indicated by blue ticks in the fragmentation pattern in Figure 1e) may indicate that there are conformers of Pep-05 in which a noncovalent interaction is present between Arg5 and pSer. These results agree with the previous work in our laboratory, which showed that the presence of a single basic amino acid residue is insufficient to prevent the observation of full fragment coverage by ECD in phosphopeptides.[20] ECD of doubly charged ions of the peptide in which all Arg residues are replaced with Leu residues (Pep-08) also resulted in complete sequence coverage (Figure 1h and Supporting Table S11).

DOI: 10.1021/acs.jpcb.9b08506
J. Phys. Chem. B 2020, 124, 461−469

Ion Mobility Spectrometry and Molecular Modeling. To further investigate the gas-phase structures of these ions, we employed an ion mobility spectrometry−molecular modeling approach. TWIMS arrival time distributions of the doubly charged ions of Pep-01−Pep-08 are shown in Supporting Figure 2, and their corresponding CCS values are shown in Table 1. The focus of the IMS/MM study was placed on Pep-01−Pep-05 as the ECD behavior of these suggested the presence of intramolecular noncovalent interactions between the phosphoserine and arginine residue(s). It was necessary to model several protonation patterns for each peptide owing to the presence of multiple basic residues (i.e., the N-terminal valine and arginine residues), the acidic phosphoserine, and the necessity of maintaining a net charge of 2+ for all of the peptides. The locations of the charges for each protonation pattern are summarized in Table 2, in which the basic residues and phosphoserine are labeled ^{0},$^+$, or $^-$ to denote whether they are neutral, positive, or negative, respectively.

Table 2. Protonation Patterns Modeled for Each Peptide

peptide	model	sequence
Pep-01	A	V^0 V G A R^+ R^0 S s^0 W R^+ V V S S I
	B	V^+ V G A R^+ R^0 S s^- W R^+ V V S S I
	C	V^+ V G A R^0 R^+ S s^- W R^+ V V S S I
	D	V^0 V G A R^+ R^+ S s^- W R^+ V V S S I
Pep-02	E	V^0 V G A R^+ L^0 S s^0 W R^+ V V S S I
	F	V^+ V G A R^+ L^0 S s^- W R^+ V V S S I
Pep-03	G	V^0 V G A R^+ R^+ S s^0 W L^0 V V S S I
	H	V^+ V G A R^+ R^+ S s^- W L^0 V V S S I
Pep-04	I	V^0 V G A L^0 R^+ S s^0 W R^+ V V S S I
	J	V^+ V G A L^0 R^+ S s^- W R^+ V V S S I
Pep-05	K	V^0 V G A R^+ L^0 S s^0 W L^0 V V S S I

Each of these 11 protonation patterns was modeled by simulated annealing to produce eleven 25 000-strong molecular ensembles, and simulated CCS were calculated for each of those models. By filtering the SA ensembles by $CCS_{experimental} \pm 5\%$, we produced a CCS-matched ensemble for each protonation pattern: these ranged in size from 6477 structures to 21386 structures. Dimensionality reduction was performed on each of the CCS-matched ensembles to reduce the computational cost of the onward analysis and to improve the quality of the subsequent clustering. The projections produced by the dimensionality reduction were clustered by k-means (n = 10, Euclidean), a method that should find any clusters in the data or, in the absence of any clusters (i.e., the data existing as a continuum), should segment the data into 10 equally sized portions. By finding the structure closest to the centroid of each cluster, it was possible to analyze the noncovalent interaction patterns across each CCS-matched ensemble (assuming that the structure closest to the centroid of each cluster is representative of the whole cluster). These representative structures were grouped according to the noncovalent interactions between the phosphoserine and other pertinent residues (Val1, Arg5, Arg6, Arg10), with a noncovalent interaction being defined as occurring when a proton on one of these four residues was less than 4 Å distance from the phosphorus atom. Distances are measured between the phosphorus atom (as it is the center of the charge on the phosphate group) and the protons on the N-terminus (Val1) and the Arg side chains. We calculate that a P−H distance of 4 Å would bring the van der Waals radius of the proton within the van der Waals radius of at least one of the phosphate oxygen atoms, while also allowing for variations in bond angles and bond lengths from their equilibrium values. A noncovalent interaction defined on these terms was considered to be a salt bridge when it involved a negatively charged phosphorylated residue and a positively charged residue, while a noncovalent interaction between a neutral phosphorylated residue and a positively charged residue was considered an ionic hydrogen bond, as described by Kim et al.[29] Using this process, it was possible to reduce the complexity of the information in each 25 000-strong molecular modeling ensemble to 10 representative structures, i.e., one structure for each of the 10 clusters. Details of the 10 clusters for each CCS-matched ensemble, including the percentage of the CCS-matched ensemble that each representative structure represents, can be found in Supporting Table S12. The information on the noncovalent interactions with pSer across the representative structures for each of the 11 sequences modeled is summarized in Supporting Table S13, in which the proton−phosphorus distances for each of the 10 representative structures for model A−K are shown. For each representative structure, proton−phosphorus distances <4 Å are highlighted in red, while distances <4.5 Å are outlined with a red box.

As described above, for each peptide, we interpret a lack of fragments in the experimental ECD mass spectrum between any arginine and the phosphoserine as being indicative of the presence of a noncovalent interaction being present between these residues. For Pep-01, models A−D had no representative structures containing noncovalent interactions consistent with those predicted from the ECD fragmentation pattern (i.e., the absence of fragments between Arg5 and Arg10). Structures D2 and D6 from model D, in which all three arginine residues were protonated and pSer was deprotonated (i.e [Ser-$PO_4H]^-$), had noncovalent interactions between pSer and Arg5, while protons on Arg10 were just outside the range that we have defined as being a noncovalent interactions (at 4.2 and 4.1 Å, respectively). Although these two structures do not themselves have the noncovalent interactions predicted by the ECD data, they represent clusters that contain many structures in which both Arg5 and Arg10 form noncovalent interactions with pSer. Some 13% of the structures in cluster 2 contain noncovalent interactions consistent with the ECD data, while for cluster 6, 11% of the structures have those noncovalent interactions. These representative structures cumulatively represent 18% of the model D CCS-matched ensemble (i.e., cluster 2 and cluster 6) and are shown in Figure 2a. The interactions in these representative structures are characterized as being salt bridges, as the groups interacting are all charged. Model D is the only protonation pattern that produced any representative structures exhibiting noncovalent interactions that were close to being consistent with the ECD data, suggesting that Pep-01 exists predominantly in this protonation state, i.e., all arginine residues protonated and phosphoserine deprotonated, in the gas phase.

For Pep-02, model F produced a single representative structure with the Arg10−pSer noncovalent interaction consistent with the full ECD fragmentation pattern, including the low abundance $c_5/c_5^{\bullet}$, $c_6/c_6^{\bullet}$, and $c_7/c_7^{\bullet}$ fragments (Figure 2b). This representative structure (F4) represented 12% of the F CCS-matched ensemble. The noncovalent interactions in model F are between two ionic residues, the negative phosphoserine and positive arginine residues, and so are characterized as salt bridges. The representative structure F4

DOI: 10.1021/acs.jpcb.9b08506
J. Phys. Chem. B 2020, 124, 461−469

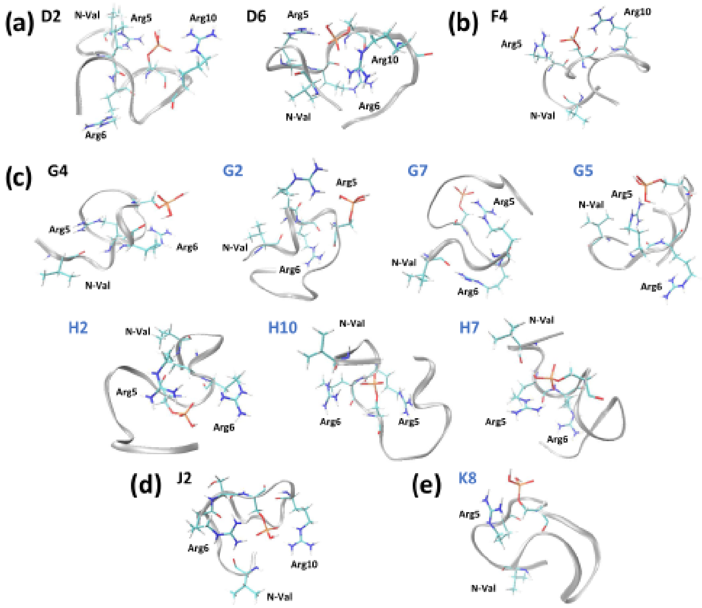


Figure 2. Representative structures with noncovalent interactions consistent with ECD fragmentation patterns for the clustered, CCS-matched modeling ensembles of (a) Pep-01, (b) Pep-02, (c) Pep-03, (d) Pep-04, and (e) Pep-05. Structures are labeled with their model letter and cluster number. Those structures labeled in blue have noncovalent interactions consistent with the ECD data once low abundance fragments have been excluded. Structure J8 only has one of the Arg−pSer interactions predicted by the ECD; the other (Arg6−pSer) is too far (5 Å) to be considered a noncovalent interaction.

does not have both the Arg5−pSer and Arg10−pSer noncovalent interactions that would be consistent with the ECD pattern once low abundance was excluded; however, Arg5 is in close proximity to pSer, with a proton−phosphorus distance of 4.3 Å. It is possible that this represents a weaker Arg5−pSer interaction that hinders the production of fragments in this region but is not sufficiently strong to prevent fragmentation altogether, resulting in the low abundance fragments observed. Alternatively, it is possible that the cluster that F4 represents contains structures with both Arg5 and Arg10 interacting with pSer and that these conformers are responsible for the low abundance of the $c_5/c_5^{\bullet}$, $c_6/c_6^{\bullet}$, and $c_7/c_7^{\bullet}$ fragments.

Model G produced one representative structure whose noncovalent interactions were consistent with the full fragmentation pattern of Pep-03 (Figure 2c). This representative structure (G4) contains an interaction between Arg6 and pSer and could be the conformer from which the c_5 fragment is generated. G4 represents 7% of the model G CCS-matched ensemble. Once the low abundance c_5 fragment is excluded, three further representative structures from model G and three representative structures from model H have the noncovalent interaction Arg5−pSer that is consistent with the ECD data. The three model G representative structures (G2, G7, and G5) all exhibit only the Arg5−pSer interaction and together represent 31% of the model G CCS-matched ensemble. The three representative structures from model H, representing 33% of the H CCS-matched ensemble, also had only the Arg5−pSer interaction, although the representative structure H10 also had protons from the N-terminal valine and Arg6 in close proximity with pSer. We propose G4 as the structure of the minor conformation from which the c_5 fragment arises, while all of the representative structures G2, G7, G5, H2, H10, and H7 are suitable candidates for the structure of the major conformation. As both models G and H produce several representative structures consistent with the proposed fragmentation pattern of the major conformer, we suggest that the two protonation patterns coexist in the gas phase. It should be noted that many of the representative structures produced from models G and H are very similar: G2, G7, and G5 are all structurally similar, as are H2 and H7. A lack of diversity in the representative structures may indicate that the G and H ensembles contain a narrower range of conformers to select from, which is consistent with the observed arrival time distribution for Pep-03, which is narrow relative to the other peptides (Supporting Figure 2).

For Pep-04, no representative structures from either model J or I were consistent with the ECD data, in which no fragments were observed between Arg6 and Arg10, indicating Arg6−pSer and Arg10−pSer noncovalent interactions. For model J, the representative structure of cluster 2 has a salt bridge between Arg6 and pSer and at least one proton on Arg10 within 4.5 Å. While a distance of 4.5 Å is too great to be considered a noncovalent interaction, it is possible that the cluster that this structure represents contains other structures that do have the Arg6−pSer and Arg10−pSer noncovalent interactions that are consistent with the ECD data. Examining the whole model J CCS-matched ensemble, 649 of the 14,442 (4%) structures have a proton on both Arg6 and Arg10 within 4 Å of the pSer phosphorus atom, while the protons on the N-terminal valine residue are further than 4 Å from the pSer phosphorus atom, i.e., the salt bridges are consistent with the ECD data. Cluster 2 contains 100 of these ECD-consistent structures; therefore, its representative structure is the most appropriate of those produced and has been selected as the candidate structure for model J (Figure 2d).

For Pep-05 (which contains a single arginine residue), only one sequence (model K) was modeled as there was only one likely combination of charges that produced a net charge of 2+. One representative structure (Figure 2e) was produced that was consistent with the ECD spectrum after the exclusion of low abundance fragments, representing 12% of the ensemble. Eight other representative structures (representing 84% of the K CCS-matched ensemble) had no noncovalent interactions between pertinent residues, which is consistent with the full ECD pattern with low abundance fragments included, in which full fragment coverage was observed and so would not predict the presence of any noncovalent interactions.

CONCLUSIONS

We have shown that the ECD behavior of the doubly charged ions of eight peptides derived from 14-3-3-ζ results in a reduced sequence fragment coverage when there are two or more arginine residues present. These missing regions allowed the locations of noncovalent interactions to be predicted. For the five peptides (Pep-01−Pep-05) for which reduced sequence coverage was observed, we employed an ion mobility spectrometry/molecular modeling workflow to probe the structures of their $[M+2H]^{2+}$ ions further. For Pep-02 where the ECD mass spectrum indicated the presence of conformers with two patterns of noncovalent interactions, the IMS/MM workflow produced a single candidate structure from the model F protonation pattern; this structure had noncovalent interactions consistent with the full ECD fragmentation pattern, though the Arg5 residue was sufficiently near the pSer residue to suggest that it may sometimes interact, leading to the reduced abundance of the $c_5/c_5^\bullet$, $c_6/c_6^\bullet$, and $c_7/c_7^\bullet$ fragments. One representative structure from model G was produced with noncovalent interactions consistent with the ECD data of Pep-03; however, with low abundance fragments excluded, six further representative structures had ECD-consistent noncovalent interactions. These seven representative structures could all serve as candidate structures for Pep-03. No representative structures generated for either Pep-01 or Pep-04 had noncovalent interactions that exactly matched those predicted from the ECD data; however, two structures from model D and one structure from model J had one of the predicted Arg−pSer and the other Arg just outside of interaction distance (<4.5 Å). Closer inspection of the associated clusters showed that they contained many structures that did have noncovalent interactions consistent with the ECD data, making them the most suitable representative structure to propose as candidates. For Pep-05, one structure was proposed as a candidate for the conformation containing the Arg5−pSer noncovalent interaction as indicated by the ECD fragmentation pattern with low abundance fragments removed. With low abundance fragments included, eight possible candidate structures were consistent with the predicted lack of noncovalent interactions between pS8 and positively charged residues.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.9b08506.

> Instrument parameters used and pressure readings observed during TWIMS experiments; list of calibrant ions; fragment ion assignments; information regarding the clusters of CCS-matched ensemble A−K; summary of proton−phosphorus distances; CCS calibration graphs; and arrival time distributions for peptides 1−8 (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: h.j.cooper@bham.ac.uk.

ORCID

David H. Russell: 0000-0003-0830-3914

Helen J. Cooper: 0000-0003-4590-9384

Present Address

$^\nabla$Babraham Institute, Cambridge CB22 3AT, U.K. (A.F.L.-C.).

Author Contributions

$^\#$A.L.S. and A.F.L.-C. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Lisa Perez of the TAMU Laboratory for Molecular Simulation for insightful discussions and guidance as related to these studies. H.J.C. is an EPSRC Established Career Fellow (EP/L023490/1 and EP/S002979/1). A.F.L.-C. was funded by EPSRC (EP/L023490/1). A.L.S. is funded by the EPSRC Physical Sciences for Health Doctoral Training Centre (EP/L013646/1). D.H.R. is funded by N.I.H. (R01GM121751 and P41GM128577). The Bruker SolariX-XR mass spectrometer was funded by the BBSRC (BB/M012492/1). The Advion Triversa Nanomate and the Waters Synapt G2S mass spectrometer were funded by EPSRC (EP/K039245/1). The authors thank the Centre of Membrane Proteins and Receptors (COMPARE), Universities of Birmingham and Nottingham, Midlands, U.K., for use of their computing facilities. The computations described in this paper were performed using the University of Birmingham's BlueBEAR HPC service, which provides a High Performance Computing service to the University's research community. See http://www.birmingham.ac.uk/bear for more details. Supplementary data supporting this research are openly available from the University of Birmingham data archive at DOI: https://doi.org/10.25500/edata.bham.00000416.

REFERENCES

(1) McLafferty, F. W.; Horn, D. M.; Breuker, K.; Ge, Y.; Lewis, M. A.; Cerda, B.; Zubarev, R. A.; Carpenter, B. K. Electron Capture Dissociation of Gaseous Multiply Charged Ions by Fourier-Transform Ion Cyclotron Resonance. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245−9.

(2) Fort, K. L.; Cramer, C. N.; Voinov, V. G.; Vasil'ev, Y. V.; Lopez, N. I.; Beckman, J. S.; Heck, A. J. R. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **2018**, *17*, 926−933.

(3) Voinov, V. G.; Deinzer, M. L.; Barofsky, D. F. Electron Capture Dissociation in a Linear Radiofrequency-Free Magnetic Cell. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3087−8.

(4) Shi, S. D.; Hemling, M. E.; Carr, S. A.; Horn, D. M.; Lindh, I.; McLafferty, F. W. Phosphopeptide/Phosphoprotein Mapping by Electron Capture Dissociation Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 19−22.

(5) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1793−800.

(6) Hakansson, K.; Cooper, H. J.; Emmett, M. R.; Costello, C. E.; Marshall, A. G.; Nilsson, C. L. Electron Capture Dissociation and Infrared Multiphoton Dissociation Ms/Ms of an N-Glycosylated Tryptic Peptic to Yield Complementary Sequence Information. *Anal. Chem.* **2001**, *73*, 4530−6.

(7) Kelleher, N. L.; Zubarev, R. A.; Bush, K.; Furie, B.; Furie, B. C.; McLafferty, F. W.; Walsh, C. T. Localization of Labile Posttranslational Modifications by Electron Capture Dissociation: The Case of Gamma-Carboxyglutamic Acid. *Anal. Chem.* **1999**, *71*, 4250−3.

(8) Chalmers, M. J.; Hakansson, K.; Johnson, R.; Smith, R.; Shen, J.; Emmett, M. R.; Marshall, A. G. Protein Kinase a Phosphorylation Characterized by Tandem Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Proteomics* **2004**, *4*, 970−81.

(9) Horn, D. M.; Ge, Y.; McLafferty, F. W. Activated Ion Electron Capture Dissociation for Mass Spectral Sequencing of Larger (42 Kda) Proteins. *Anal. Chem.* **2000**, *72*, 4778−84.

(10) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. Plasma Electron Capture Dissociation for the Characterization of Large Proteins by Top Down Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 1599−603.

(11) Schnier, P. D.; Price, W. D.; Jockusch, R. A.; Williams, E. R. Blackbody Infrared Radiative Dissociation of Bradykinin and Its Analogues: Energetics, Dynamics, and Evidence for Salt-Bridge Structures in the Gas Phase. *J. Am. Chem. Soc.* **1996**, *118*, 7178−89.

(12) Jackson, S. N.; Dutta, S.; Woods, A. S. The Use of ECD/ETD to Identify the Site of Electrostatic Interaction in Noncovalent Complexes. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 176−9.

(13) Robinson, C. V.; Chung, E. W.; Kragelund, B. B.; Knudsen, J.; Aplin, R. T.; Poulsen, F. M.; Dobson, C. M. Probing the Nature of Noncovalent Interactions by Mass Spectrometry. A Study of Protein−Coa Ligand Binding and Assembly. *J. Am. Chem. Soc.* **1996**, *118*, 8646−8653.

(14) Kitova, E. N.; Seo, M.; Roy, P. N.; Klassen, J. S. Elucidating the Intermolecular Interactions within a Desolvated Protein-Ligand Complex. An Experimental and Computational Study. *J. Am. Chem. Soc* **2008**, *130*, 1214−26.

(15) Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A. Top-Down Esi-Ecd-Ft-Icr Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites. *J. Am. Chem. Soc.* **2006**, *128*, 14432−3.

(16) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation. *J. Am. Chem. Soc.* **2002**, *124*, 6407−20.

(17) Skinner, O. S.; McLafferty, F. W.; Breuker, K. How Ubiquitin Unfolds after Transfer into the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1011−4.

(18) Breuker, K.; Bruschweiler, S.; Tollinger, M. Electrostatic Stabilization of a Native Protein Structure in the Gas Phase. *Angew. Chem., Int. Ed.* **2011**, *50*, 873−7.

(19) Cui, W.; Zhang, H.; Blankenship, R. E.; Gross, M. L. Electron-Capture Dissociation and Ion Mobility Mass Spectrometry for Characterization of the Hemoglobin Protein Assembly. *Protein Sci* **2015**, *24*, 1325−32.

(20) Harvey, S. R.; Porrini, M.; Tyler, R. C.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E. Electron Capture Dissociation and Drift Tube Ion Mobility-Mass Spectrometry Coupled with Site Directed Mutations Provide Insights into the Conformational Diversity of a Metamorphic Protein. *Phys. Chem. Chem. Phys.* **2015**, *17*, 10538−50.

(21) Lermyte, F.; Sobott, F. Electron Transfer Dissociation Provides Higher-Order Structural Information of Native and Partially Unfolded Protein Complexes. *Proteomics* **2015**, *15*, 2813−22.

(22) Zhang, Z.; Browne, S. J.; Vachet, R. W. Exploring Salt Bridge Structures of Gas-Phase Protein Ions Using Multiple Stages of Electron Transfer and Collision Induced Dissociation. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 604−13.

(23) Zhang, Z.; Vachet, R. W. Gas-Phase Protein Salt Bridge Stabilities from Collisional Activation and Electron Transfer Dissociation. *Int. J. Mass Spectrom.* **2017**, *420*, 51−56.

(24) Creese, A. J.; Cooper, H. J. The Effect of Phosphorylation on the Electron Capture Dissociation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1263−74.

(25) Ruotolo, B. T.; Gillig, K. J.; Woods, A. S.; Egan, T. F.; Ugarov, M. V.; Schultz, J. A.; Russell, D. H. Analysis of Phosphorylated Peptides by Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2004**, *76*, 6727−33.

(26) Ruotolo, B. T.; Verbeck; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H. Distinguishing between Phosphorylated and Nonphosphorylated Peptides with Ion Mobility−Mass Spectrometry. *J. Proteome Res.* **2002**, *1*, 303−306.

(27) Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. Characterization of Phosphorylated Peptides Using Traveling Wave-Based and Drift Cell Ion Mobility Mass Spectrometry. *Anal. Chem.* **2009**, *81*, 248−54.

(28) Lopez-Clavijo, A. F.; Duque-Daza, C. A.; Creese, A. J.; Cooper, H. J. Electron Capture Dissociation Mass Spectrometry of Phosphopeptides: Arginine and Phosphoserine. *Int. J. Mass Spectrom.* **2015**, *390*, 63−70.

(29) Kim, D.; Pai, P. J.; Creese, A. J.; Jones, A. W.; Russell, D. H.; Cooper, H. J. Probing the Electron Capture Dissociation Mass Spectrometry of Phosphopeptides with Traveling Wave Ion Mobility Spectrometry and Molecular Dynamics Simulations. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1004−13.

(30) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nat. Chem.* **2014**, *6*, 281−94.

(31) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a Travelling Wave-Based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401−14.

(32) Ruotolo, B. T.; Benesch, J. L.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3*, 1139−52.

(33) Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. Structural Characterization of Drug-Like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections. *Anal. Chem.* **2012**, *84*, 1026−33.

(34) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. Linking Molecular Models with Ion Mobility Experiments. Illustration with a Rigid Nucleic Acid Structure. *J. Mass Spectrom.* **2015**, *50*, 711−26.

(35) Moss, C. L.; Chamot-Rooke, J.; Nicol, E.; Brown, J.; Campuzano, I.; Richardson, K.; Williams, J. P.; Bush, M. F.; Bythell, B.; Paizs, B.; et al. Assigning Structures to Gas-Phase Peptide Cations and Cation-Radicals. An Infrared Multiphoton Dissociation, Ion Mobility, Electron Transfer, and Computational Study of a Histidine Peptide Ion. *J. Phys. Chem. B* **2012**, *116*, 3445−56.

(36) Bereszczak, J. Z.; Barbu, I. M.; Tan, M.; Xia, M.; Jiang, X.; van Duijn, E.; Heck, A. J. Structure, Stability and Dynamics of Norovirus P Domain Derived Protein Complexes Studied by Native Mass Spectrometry. *J. Struct. Biol.* **2012**, *177*, 273−82.

(37) Powell, D. W.; Rane, M. J.; Chen, Q.; Singh, S.; McLeish, K. R. Identification of 14-3-3zeta as a Protein Kinase B/Akt Substrate. *J. Biol. Chem.* **2002**, *277*, 21639−42.

(38) Megidish, T.; Cooper, J.; Zhang, L.; Fu, H.; Hakomori, S.; Novel Sphingosine-Dependent, A. Protein Kinase (Sdk1) Specifically

Phosphorylates Certain Isoforms of 14-3-3 Protein. *J. Biol. Chem.* **1998**, *273*, 21834−45.

(39) Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. Three-Dimensional Ion Mobility/Tofms Analysis of Electrosprayed Biomolecules. *Anal Chem* **1998**, *70*, 2236−42.

(40) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; et al. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrom Rev* **2019**, *38*, 291−320.

(41) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminfor.* **2012**, *4*, No. 17.

(42) Frisch, M. J.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.. et al.*Gaussian 09*, revision D.01; Gaussian, Inc.: Wallingford CT, 2016.

(43) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(44) Case, D. A.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H. et al. *Amber*; University of California: San Francisco, 2016.

(45) Lee, J. H.; Pollert, K.; Konermann, L. Testing the Robustness of Solution Force Fields for Md Simulations on Gaseous Protein Ions. *J. Phys. Chem. B* **2019**, *123*, 6705−6715.

(46) Roe, D. R.; Cheatham, T. E. Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.

(47) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *J. Comput. Phys.* **2013**, *251*, 344−363.

(48) Palmer, A. D.; Bunch, J.; Styles, I. B. The Use of Random Projections for the Analysis of Mass Spectrometry Imaging Data. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 315−22.

Appendix 6: EXD fragment assignment lists

Table A6-1 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-01

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
869.95924	869.95924	$[M+2H]^{2+}$	-0.770
1251.63442	1251.93826	$c_{10}^{\bullet}$	-1.278
1252.64224	1252.64209	c_{10}	-0.122
1351.71065	1351.71041	$c_{11}^{\bullet}$	-0.180
1351.71065	1351.71041	c_{11}	-0.180
1378.68653	1378.68792	$z_{11}^{\bullet}-H_20$	1.008
1450.77906	1450.77859	$c_{12}^{\bullet}$	2.182
1450.77906	1450.77859	c_{12}	-0.322
1467.73420	1467.73640	$z_{12}^{\bullet}$	1.499
1506.74510	1506.74599	$z_{13}^{\bullet}-H_20$	0.592
1524.75566	1524.75742	$z_{13}^{\bullet}$	1.152
1536.80327	1536.80398	$c_{13}^{\bullet}$	0.463
1537.81109	1537.81203	c_{13}	0.609
1580.82949	1580.83020	a_{14}	0.447
1623.82407	1623.82614	$z_{14}^{\bullet}$	1.273
1624.84312	1624.84391	c_{14}	0.484
1641.94214	1641.94322	$[M-H_3PO_4+2H]^{+\bullet}$	0.658
1680.87079	1680.87079	$[M-CH_5N_3+2H]^{+\bullet}$	1.061
1704.88192	1704.88272	$[M-H_2O-NH_3+2H]^{1+\bullet}$	0.469
1722.89249	1722.89137	$[M-NH_3+2H]^{1+\bullet}$	-0.648
1738.91121	1738.91036	$[M+H]^{+}$	-0.488
1739.91903	1739.91804	$[M+2H]^{+\bullet}$	-0.571

Table A6-2 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-02

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
499.32250	499.32197	$c_5^{\bullet}$	-1.051
500.33032	500.33001	c_5	-0.620
612.40656	612.40624	$c_6^{\bullet}$	-0.514
699.43858	699.43841	$c_7^{\bullet}$	-0.250
700.44641	700.44663	c_7	0.314
799.46227	799.46235	$[M- H_3PO_4+2H]^{2+}$	0.486
848.45072	848.45072	[M+2H]2+	0.000
1165.61101	1165.61041	a_{10}	-0.513
1208.61737	1208.61738	$c_{10}^{\bullet}$	0.010
1255.70315	1255.70447	$z_{11}^{\bullet}- H_3PO_4$	1.051
1264.67997	1264.67981	a_{11}	-0.129
1307.68578	1307.68579	$c_{11}^{\bullet}$	0.009
1308.69360	1308.69339	c_{11}	-0.163
1326.74026	1326.74052	$z_{12}^{\bullet}- H_3PO_4$	0.196
1335.66948	1335.66922	$z_{11}^{\bullet}-H_20$	-0.195
1353.68004	1353.68002	$z_{11}^{\bullet}$	-0.015
1363.74838	1363.74833	a_{12}	-0.039
1369.69876	1369.69871	y_{11}	-0.037
1380.72677	1380.72737	$z_{12}^{\bullet}-CO_2$	0.435
1383.76172	1383.76138	$z_{13}^{\bullet}- H_3PO_4$	-0.242
1406.70659	1406.70688	$z_{12}^{\bullet}$-H20	0.206
1406.75419	1406.75438	$c_{12}^{\bullet}$	0.136
1407.76201	1407.76228	c_{12}	0.190
1424.71715	1424.71697	$z_{12}^{\bullet}$	-0.126
1437.74823	1437.74974	$z_{13}^{\bullet}-CO_2$	1.048
1440.73587	1440.73646	y_{12}	0.410
1450.78041	1450.78050	a_{13}	0.060
1463.72805	1463.72820	$z_{13}^{\bullet}$-H20	0.104
1481.73861	1481.73892	$z_{13}^{\bullet}$	0.207
1482.83013	1482.83039	$z_{14}^{\bullet}- H_3PO_4$	0.179
1493.78622	1493.78514	$c_{13}^{\bullet}$	-0.722
1494.79404	1494.79389	c_{13}	-0.102
1497.75733	1497.75704	y_{13}	-0.196
1537.81244	1537.81217	a_{14}	-0.178
1562.79646	1562.79869	$z_{14}^{\bullet}-H_20$	1.428
1565.82852	1565.82767	$[M-C_9H_9N+2H]^{+\bullet}$	-0.543
1580.80702	1580.80709	$z_{14}^{\bullet}$	0.042
1581.82607	1581.82650	c_{14}	0.270
1581.89854	1581.89875	$[M-NH_3-H_3PO_4+2H]^{+\bullet}$	0.136
1595.80725	1595.80787	$[M-C_4H_{11}N_3+2H]^{+\bullet}$	0.389
1598.92454	1598.92523	$[M- H_3PO_4+2H]^{+\bullet}$	0.432
1661.86487	1661.86475	$[M- H_2O-NH_3+2H]^{1+\bullet}$	-0.074
1668.90707	1668.90753	$[M-CO+2H]^{1+\bullet}$	-0.277
1678.89142	1678.89195	$[M- H_2O +2H]^{1+\bullet}$	-0.317
1679.87544	1679.87554	$[M-NH_3+2H]^{1+\bullet}$	-0.061
1695.89416	1695.89408	$[M+H]^{+}$	-0.046
1696.90198	1696.90163	$[M+2H]^{+\bullet}$	-0.208

Table A6-3 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-03

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
499.32249	4999.3217	$c_5^{\bullet}$	-1.592
500.33032	500.32972	c_5	-1.199
848.45072	848.45074	$[M+2H]^{2+}$	0.024
909.45400	909.45426	$c_8^{\bullet}$	0.288
910.46182	910.46290	c_8	1.183
1052.52750	1052.52841	a_9	0.862
1095.53331	1095.53370	$c_9^{\bullet}$	0.358
1096.54113	1096.53701	c_9	-0.376
1165.61156	1165.61176	a_{10}	0.169
1208.61737	1208.61776	$c_{10}^{\bullet}$	0.324
1209.62519	1209.62475	c_{10}	-0.366
1255.70315	1255.70446	$z_{11}^{\bullet}-H_3PO_4$	1.043
1264.67997	1264.68019	a_{11}	0.172
1307.68578	1307.68610	$c_{11}^{\bullet}$	0.246
1308.69360	1308.69376	c_{11}	0.120
1326.74026	1326.74059	$z_{12}^{\bullet}- H_3PO_4$	0.249
1335.66948	1335.66980	$z_{11}^{\bullet}-H_20$	0.240
1353.68004	1353.68027	$z_{11}^{\bullet}$	0.170
1363.74838	1363.74856	a_{12}	0.130
1369.69876	1369.69869	y_{11}	-0.051
1383.76172	1383.76206	$z_{13}^{\bullet}- H_3PO_4$	0.249
1406.70659	1406.70808	$z_{12}^{\bullet}-H_20$	1.059
1406.75419	1406.75493	$c_{12}^{\bullet}$	0.527
1407.76201	1407.76243	c_{12}	0.296
1424.71715	1424.71709	$z_{12}^{\bullet}$	-0.042
1440.73587	1440.73710	y_{12}	0.854
1450.78041	1450.78058	a_{13}	0.115
1463.72805	1463.72805	$z_{13}^{\bullet}-H_20$	0.001
1481.73861	1481.73914	$z_{13}^{\bullet}$	0.356
1494.79404	1494.79413	$c_{13}^{\bullet}$	0.058
1497.75733	1497.75693	y_{13}	-0.269
1537.81244	1537.81322	$a_{14}^{\bullet}$	0.505
1562.79646	1562.79709	$z_{14}^{\bullet}-H_20$	0.404
1565.82852	1565.82813	$[M-C_9H_9N+2H]^{+\bullet}$	-0.249
1580.80702	1580.80728	$z_{14}^{\bullet}$	0.163
1581.82607	1581.82657	c_{14}	0.314
1581.89854	1581.89881	$[M-NH_3-H_3PO_4]^{+\bullet}$	0.174
1595.80725	1595.80825	$[M-C_4H_{11}N_3+2H]^{+\bullet}$	0.627
1598.92454	1598.92557	$[M- H_3PO_4+2H]^{+\bullet}$	0.644
1635.88505	1635.88371	$[M-NH_3-CO2+2H]^{+\bullet}$	-0.821
1661.86487	1661.86489	$[M- H_2O-NH_3+2H]^{1+\bullet}$	0.012
1668.90707	1668.90739	$[M-CO+2H]^{1+\bullet}$	0.192
1678.89142	1678.89190	$[M- H_2O +2H]^{1+\bullet}$	0.286
1679.87544	1679.87560	$[M-NH_3+2H]^{1+\bullet}$	0.095
1695.89416	1695.89419	$[M+H]^{+}$	0.017
1696.90198	1696.90118	$[M+2H]^{+\bullet}$	-0.471

Table A6-4 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-04

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
848.45072	848.45071	[M+2H]2+	-0.012
1208.61737	1208.61713	$c_{10}^{\bullet}$	-0.197
1209.62519	1209.62548	c_{10}	0.237
1222.58542	1222.58368	$z_{10}^{\bullet}-H_20$	-1.423
1240.59598	1240.59583	$z_{10}^{\bullet}$	-0.121
1241.60381	1241.60364	$z_{10}+H$	-0.137
1264.67942	1264.67925	a_{11}	-0.133
1307.68578	1307.68659	$c_{11}^{\bullet}$	0.621
1308.69360	1308.69333	c_{11}	-0.209
1326.74026	1326.74123	$z_{12}^{\bullet}-H_3PO_4$	0.731
1353.68004	1353.67978	$z_{11}^{\bullet}$	-0.192
1354.68787	1354.68865	$z_{11}+H$	0.576
1363.74783	1363.74766	a_{12}	-0.123
1383.76172	1383.76120	$z_{13}^{\bullet}-H_3PO_4$	-0.372
1406.75419	1406.75389	$c_{12}^{\bullet}$	-0.212
1407.76201	1407.76211	c_{12}	0.069
1424.71715	1424.71661	$z_{12}^{\bullet}$	-0.379
1440.73587	1440.73533	y_{12}	-0.375
1450.77986	1450.77995	a_{13}	0.063
1481.73861	1481.73849	$z_{13}^{\bullet}$	-0.083
1482.83013	1482.83090	$z_{14}^{\bullet}-H_3PO_4$	0.523
1494.79404	1494.79366	$c_{13}^{\bullet}$	-0.256
1497.75733	1497.75708	y_{13}	-0.169
1537.81244	1537.81284	a_{14}	0.258
1580.80702	1580.80632	$z_{14}^{\bullet}$	-0.445
1581.82607	1581.82601	c_{14}	-0.040
1581.89854	1581.89879	$[M-NH_3-H_3PO_4+2H]^{+\bullet}$	0.158
1598.92454	1598.92527	$[M- H_3PO_4+2H]^{+\bullet}$	0.116
1661.86487	1661.86496	$[M- H_2O-NH_3+2H]^{1+\bullet}$	-0.042
1668.90707	1668.90757	$[M-CO+2H]^{1+\bullet}$	0.299
1678.89142	1678.89139	$[M- H_2O +2H]^{1+\bullet}$	0.018
1679.87544	1679.87571	$[M-NH_3+2H]^{1+\bullet}$	0.161
1695.89416	1695.89402	$[M+H]^{+}$	0.083
1696.90198	1696.90234	$[M+2H]^{+\bullet}$	0.212

Table A6-5 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-05

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
499.32249	499.32261	$c_5^{\bullet}$	0.230
500.33032	500.33050	c_5	0.359
612.40656	612.40648	$c_6^{\bullet}$	-0.121
613.41438	613.41438	c_6	0.007
617.38684	617.38660	y_6	-0.397
656.43223	656.43244	a_7	0.329
699.43858	699.43845	$c_7^{\bullet}$	-0.196
700.44641	700.44652	c_7	0.161
743.38431	743.38383	$[M-2H_2O+H]^{2+}$	-0.644
808.93163	808.93079	$[M-2H_2O+H]^{2+}$	-1.030
817.93691	817.93647	$[M-H_2O+H]^{2+}$	-0.539
826.94219	826.94196	$[M+2H]^{2+}$	-0.281
848.42638	848.42560	$c_8^{\bullet}-H_2O$	-0.929
866.43695	866.43662	$c_8^{\bullet}$	-0.383
867.44477	867.44489	c_8	0.138
1009.50990	1009.50964	a_9	-0.257
1034.50569	1034.50560	$c_9^{\bullet}-H_2O$	-0.090
1052.51626	1052.51608	$c_9^{\bullet}$	-0.165
1053.52408	1053.52487	c_9	0.750
1122.59451	1122.59428	a_{10}	-0.203
1123.68603	1123.68562	$a_{11}-H_3PO_4$	-0.361
1149.58160	1149.58178	b_{10}	0.156
1165.60032	1165.60012	$c_{10}^{\bullet}$	-0.169
1166.60814	1166.60791	c_{10}	-0.201
1212.68610	1212.68565	$z_{11}^{\bullet}- H_3PO_4$	-0.374
1221.66292	1221.66300	a_{11}	0.064
1222.75444	1222.75455	$a_{12}- H_3PO_4$	0.096
1249.74152	1249.74208	$b_{12}- H_3PO_4$	0.450
1264.66873	1264.66843	$c_{11}^{\bullet}$	-0.237
1265.67655	1265.67623	c_{11}	-0.256
1283.72321	1283.72290	$z_{12}^{\bullet}- H_3PO_4$	-0.242
1292.65243	1292.65214	$z_{11}^{\bullet}-H_2O$	-0.227
1309.78646	1309.78622	$a_{13}- H_3PO_4$	-0.191
1310.66299	1310.66264	$z_{11}^{\bullet}$	-0.269
1320.73133	1320.73133	a_{12}	-0.001
1326.68171	1326.68160	y_{11}	-0.086
1340.74466	1340.74449	$z_{13}^{\bullet}- H_3PO_4$	-0.127
1363.68954	1363.68951	$z_{12}^{\bullet}- H_2O$	-0.020
1363.73714	1363.73733	$c_{12}^{\bullet}$	0.138
1364.74496	1364.74487	c_{12}	-0.066
1381.70010	1381.70011	$z_{12}^{\bullet}$	0.010
1389.75280	1389.75222	$a_{12}- H_2O$	-0.414
1397.71882	1397.71903	y_{12}	0.153
1407.76336	1407.76348	a_{13}	0.087
1420.71100	1420.71130	$z_{13}^{\bullet}- H_2O$	0.209
1433.76643	1433.76724	$c_{13}- H_2O$	0.563
1438.72156	1438.72200	$z_{13}^{\bullet}$	0.303
1439.81308	1439.81341	$z_{13}^{\bullet}- H_3PO_4$	0.233
1451.77699	1451.77700	c_{13}	0.004
1454.74028	1454.74014	y_{13}	-0.100
1476.78483	1476.78473	$a_{14}- H_2O$	-0.070

1481.72737	1481.72761	x_{13}	0.159
1493.79959	1493.80053	$z_{14}^{\bullet}- CO_2$	0.624
1494.79539	1494.79558	a_{14}	0.125
1510.88658	1510.88599	$[M-CO-NH_3- H_3PO_4+2H]^{+\bullet}$	0.382
1519.77941	1519.77969	$z_{14}^{\bullet}- H_2O$	0.189
1537.78997	1537.79018	$z_{14}^{\bullet}$	0.131
1538.80902	1538.80972	c_{14}	0.451
1538.88149	1538.88133	$[M-NH_3-H_3PO_4+2H]^{+\bullet}$	-0.104
1553.80869	1553.80844	y_{14}	-0.163
1555.90804	1555.90769	$[M- H_3PO_4+2H]^{+\bullet}$	-0.221
1580.79578	1580.79596	x_{14}	0.113
1592.86800	1592.86902	$[M-NH_3- CO_2+2H]^{+\bullet}$	0.637
1607.87945	1607.88001	$[M-COOH+2H]^{+\bullet}$	0.348
1625.89002	1625.88955	$[M-CO_2+2H]^{+\bullet}$	0.288
1635.87378	1635.87378	$[M- H_2O +2H]^{1+\bullet}$	0.358
1636.85742	1636.85742	$[M-NH_3+2H]^{1+\bullet}$	0.593
1652.87711	1652.87743	$[M+H]^{+}$	0.195
1653.88493	1653.88419	$[M+2H]^{+\bullet}$	-0.449

Table A6-6 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-06

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
612.40655	612.40824	$c_6^{\bullet}$	2.751
613.41438	613.41124	c_6	-0.228
699.43859	699.44000	$c_7^{\bullet}$	2.023
700.44641	700.44583	c_7	-0.831
817.93691	817.93755	$[M-2H_2O+H]^{2+}$	-0.003
826.94249	826.94304	$[M+2H]^{2+}$	1.017
866.43695	866.43754	$c_8^{\bullet}$	0.682
867.44477	867.44575	c_8	1.130
1009.50990	1009.50894	a_9	-0.946
1052.51626	1052.51469	$c_9^{\bullet}$	-1.491
1053.52408	1053.52348	c_9	-0.575
1122.59396	1122.59253	a_{10}	-1.274
1165.60318	1165.59872	$c_{10}^{\bullet}$	-1.371
1166.60814	1166.60624	c_{10}	-1.636
1221.66237	1221.66117	a_{11}	-0.984
1264.66873	1264.66681	$c_{11}^{\bullet}$	-1.518
1265.67655	1265.67453	c_{11}	-1.597
1267.60767	1267.60611	$z_{11}^{\bullet}-C_3H_7$	-1.231
1283.72321	1283.72124	$z_{12}^{\bullet}- H_3PO_4$	-1.538
1292.65243	1292.65051	$z_{11}^{\bullet}-H_2O$	-1.485
1309.78646	1309.78465	$a_{13}- H_3PO_4$	-1.383
1310.66299	1310.66129	$z_{11}^{\bullet}$	-1.297
1320.73078	1320.72984	a_{12}	-0.708
1326.68171	1326.68018	y_{11}	-1.154
1337.70972	1337.70938	$z_{12}^{\bullet}-CO_2$	-0.251
1340.74466	1340.74329	$z_{13}^{\bullet}- H_3PO_4$	-1.026
1363.68954	1363.68816	$z_{12}^{\bullet}- H_2O$	-1.014
1364.74496	1364.74371	c_{12}	-0.922
1381.70010	1381.69907	$z_{12}^{\bullet}$	-0.744
1389.75280	1389.75160	$a_{12}- H_2O$	-0.864
1394.73118	1394.72888	$z_{13}^{\bullet}- CO_2$	-1.650
1397.71882	1397.71805	y_{12}	-0.548
1407.76281	1407.76257	a_{13}	-0.173
1420.71100	1420.71036	$z_{13}^{\bullet}- H_2O$	-0.450
1438.72156	1438.72157	$z_{13}^{\bullet}$	0.006
1439.81308	1439.81284	$z_{14}^{\bullet}- H_3PO_4$	-0.164
1451.77699	1451.77689	c_{13}	-0.071
1454.74028	1454.74012	y_{13}	-0.109
1476.78483	1476.78474	$a_{14}- H_2O$	-0.056
1481.72737	1481.72886	x_{13}	1.004
1493.79959	1493.80035	$z_{14}^{\bullet}$-CO2	0.507
1494.79539	1494.79547	a_{14}	0.055
1509.90255	1509.90271	$[M-CO-NH_3- H_3PO_4+2H]^{+\bullet}$	0.106
1519.77941	1519.78032	$z_{14}^{\bullet}- H_2O$	0.597
1537.78997	1537.79141	$z_{14}^{\bullet}$	0.934
1538.80902	1538.81086	c_{14}	1.192
1538.88149	1538.88313	$[M-NH_3-H_3PO_4+2H]^{+\bullet}$	1.070
1555.90804	1555.90899	$[M- H_3PO_4+2H]^{+\bullet}$	0.613
1562.78522	1562.78614	$x_{14}- H_2O$	0.587
1580.79578	1580.79804	x_{14}	1.429
1592.86800	1592.86979	$[M-NH_3- CO_2+2H]^{+\bullet}$	v1.124

1607.87945	1607.88211	$[M\text{-}COOH+2H]^{+\bullet}$	1.655
1625.89002	1625.89230	$[M\text{-}CO_2+2H]^{+\bullet}$	1.403
1635.87378	1635.87681	$[M\text{-}\ H_2O\ +2H]^{1+\bullet}$	1.493
1636.85742	1636.86043	$[M\text{-}NH_3+2H]^{1+\bullet}$	1.248
1652.87711	1652.88016	$[M+H]^{+}$	-1.845
1653.88493	1653.88615	$[M+2H]^{+\bullet}$	-0.737

Table A6-7 Fragment ion assignments following ECD of $[M+2H]^{2+}$ *ions of Pep-07*

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
645.39300	645.39414	z_6+H	1.766
660.40389	660.40281	y_6	-1.635
808.93163	808.93268	$[M-2H_2O+H]^{2+}$	1.302
817.93691	817.93679	$[M-H_2O+H]^{2+}$	-0.145
826.94219	826.94229	$[M+2H]^{2+}$	0.119
830.46448	830.46470	$z_7^{\bullet}$	0.265
831.47231	831.47133	z_7+H	-1.179
846.48320	846.48269	y_7	-0.602
953.47246	953.47202	$z_8^{\bullet}$- CO_2	-0.461
979.45228	979.45099	$z_8^{\bullet}$- H_2O	-1.317
986.51798	986.51626	$z_9^{\bullet}$- H_3PO_4	-1.744
997.46284	997.46213	$z_8^{\bullet}$	-0.712
998.47067	998.46985	z_8+H	-0.821
1013.48156	1013.48127	y_8	-0.286
1040.50449	1040.50458	$z_9^{\bullet}$- CO_2	0.086
1066.48431	1066.48292	$z_9^{\bullet}$- H_2O	-1.303
1084.49487	1084.49352	$z_9^{\bullet}$	-1.245
1085.50270	1085.50176	z_9+H	-0.866
1099.60204	1099.60098	$z_{10}^{\bullet}$- H_3PO_4	-0.964
1100.51359	1100.51304	y_9	-0.500
1122.59396	1122.59349	a_{10}	-0.417
1123.68603	1123.68545	a_{11}- H_3PO_4	-0.512
1153.58855	1153.58845	$z_{10}^{\bullet}$- CO_2	-0.087
1165.60032	1165.59919	$c_{10}^{\bullet}$	-0.968
1166.60814	1166.60729	c_{10}	-0.731
1179.56837	1179.56744	$z_{10}^{\bullet}$- H_2O	-0.788
1197.57893	1197.57782	$z_{10}^{\bullet}$	-0.927
1198.58676	1198.58603	z_{10}+H	-0.609
1212.68610	1212.68425	$z_{11}^{\bullet}$- H_3PO_4	-1.526
1213.59765	1213.59699	y_{10}	-0.544
1221.66237	1221.66227	a_{11}	-0.080
1222.75444	1222.75365	a_{12}- H_3PO_4	-0.642
1264.66873	1264.66781	$c_{11}^{\bullet}$	-0.726
1265.67655	1265.67552	c_{11}	-0.816
1266.67261	1266.67081	$z_{11}^{\bullet}$- CO_2	-1.421
1283.72321	1283.72236	$z_{12}^{\bullet}$- H_3PO_4	-0.662
1292.65243	1292.65200	$z_{11}^{\bullet}$- H_2O	-0.333
1309.78646	1309.78658	a_{13}- H_3PO_4	0.088
1310.66299	1310.66209	$z_{11}^{\bullet}$	-0.687
1311.67082	1311.67128	z_{11}+H	0.351
1320.73078	1320.73077	a_{12}	-0.006
1326.68171	1326.68157	y_{11}	-0.106
1337.70972	1337.70957	$z_{12}^{\bullet}$- CO_2	-0.112
1340.74466	1340.74394	$z_{13}^{\bullet}$- H_3PO_4	-0.541
1363.68954	1363.68908	$z_{12}^{\bullet}$- H_2O	-0.337
1363.73714	1363.73779	$c_{12}^{\bullet}$	0.478
1364.74496	1364.74436	c_{12}	-0.442
1381.70010	1381.69973	$z_{12}^{\bullet}$	-0.268
1389.75280	1389.75234	a_{13}- H_2O	-0.330
1394.73118	1394.73002	$z_{13}^{\bullet}$- CO_2	-0.834
1397.71882	1397.71846	y_{12}	-0.258

1420.71100	1420.70989	$z_{13}^{\bullet}$- H_2O	-0.780
1433.76643	1433.76541	c_{13}- H_2O	-0.710
1438.72156	1438.72160	$z_{13}^{\bullet}$	0.026
1439.81308	1439.81312	$z_{14}^{\bullet}$- H_3PO_4	0.031
1451.77699	1451.77680	c_{13}	-0.133
1454.74028	1454.74000	y_{13}	-0.195
1476.78483	1476.78477	a_{14}- H_2O	-0.039
1481.72737	1481.72791	x_{13}	0.362
1493.79959	1493.80130	$z_{14}^{\bullet}$- CO_2	1.143
1512.89099	1512.89194	$[M- H_3PO_4+2H]^{+\bullet}$	-0.201
1519.77941	1519.77933	$z_{14}^{\bullet}$- H_2O	-0.051
1537.78997	1537.79068	$z_{14}^{\bullet}$	0.460
1538.80902	1538.80964	c_{14}	0.401
1580.79578	1580.79702	x_{14}	0.783
1592.86800	1592.86957	$[M-NH_3- CO_2+2H]^{+\bullet}$	0.984
1607.87945	1607.88060	$[M-COOH+2H]^{+\bullet}$	0.713
1625.89002	1625.89064	$[M-CO_2+2H]^{+\bullet}$	0.208
1635.87378	1635.87515	$[M- H_2O +2H]^{1+\bullet}$	0.398
1636.85742	1636.85897	$[M-NH_3+2H]^{1+\bullet}$	0.599
1652.87711	1652.87848	$[M+H]^{+}$	0.199
1653.88493	1653.88619	$[M+2H]^{+\bullet}$	-0.4090.212

Table A6-8 Fragment ion assignments following ECD of $[M+2H]^{2+}$ ions of Pep-08

Theoretical m/z	Experimental m/z	Assignment	Error (ppm)
405.23437	405.23495	y_4	1.436
504.30278	504.30289	y_5	0.218
553.37078	553.37075	b_6	-0.069
569.38950	569.38949	$c_6^\bullet$	-0.030
570.39733	570.39739	c_6	0.100
617.38684	617.38664	y_6	-0.322
638.41097	638.41093	$c_6^\bullet$- H_2O	-0.061
640.40282	640.40227	b_7	-0.859
656.42153	656.42146	$c_7^\bullet$	-0.122
657.42936	657.42933	c_7	-0.046
796.42839	796.42764	$[M- H_2O +2H]^{2+}$	-0.939
803.46615	803.46590	y_7	-0.311
805.43367	805.43350	$[M+2H]^{2+}$	-0.212
823.41990	823.42010	$c_8^\bullet$	0.243
824.42772	824.42746	c_8	-0.318
943.50093	943.49962	$z_9^\bullet$- H_3PO_4	-1.384
954.44579	954.44570	$z_8^\bullet$	-0.100
955.45362	955.45340	z_8+H	-0.229
966.49340	966.49347	a_9	0.065
970.46451	970.46560	y_8	1.128
981.60056	981.60086	a_{10}- H_3PO_4	0.303
991.48864	991.48807	$c_9^\bullet$- H_2O	-0.574
992.49647	992.49785	c_9- H_2O	1.392
1009.49921	1009.49895	$c_9^\bullet$	-0.255
1010.50703	1010.50738	c_9	0.340
1023.46726	1023.46616	$z_9^\bullet$- H_2O	-1.074
1057.49654	1057.49718	y_9	0.610
1079.57746	1079.57739	a_{10}	-0.068
1080.66898	1080.66874	a_{10}- H_3PO_4	-0.217
1105.58053	1105.58088	c_{10}- H_2O	0.320
1106.56455	1106.56402	b_{10}	-0.479
1122.58327	1122.58322	$c_{10}^\bullet$	-0.042
1123.59109	1123.59120	c_{10}	0.094
1134.55947	1134.55788	y_{10}-$2H_2O$	-1.406
1136.55132	1136.55178	$z_{10}^\bullet$- H_2O	0.406
1154.56188	1154.56134	$z_{10}^\bullet$	-0.466
1160.63531	1160.63407	a_{11}- H_2O	-1.071
1169.66905	1169.66899	$z_{11}^\bullet$- H_3PO_4	-0.055
1170.58060	1170.57948	y_{10}	-0.954
1178.64587	1178.64588	a_{11}	0.008
1179.73738	1179.73714	a_{12}- H_3PO_4	-0.205
1205.63296	1205.63242	b_{11}	-0.446
1221.65168	1221.65089	$c_{11}^\bullet$	-0.647
1222.65950	1222.65969	c_{11}	0.154
1240.70616	1240.70514	$z_{12}^\bullet$- H_3PO_4	-0.821
1249.63538	1249.63529	$z_{11}^\bullet$- H_2O	-0.074
1267.64594	1267.64586	$z_{11}^\bullet$	-0.065
1277.71428	1277.71468	a_{12}	0.311
1283.66466	1283.66461	y_{11}	-0.039
1297.72762	1297.72728	$z_{13}^\bullet$- H_3PO_4	-0.261
1303.71735	1303.71722	c_{12}- H_2O	-0.097

1304.70137	1304.70173	b_{12}	0.275
1320.67249	1320.67347	$z_{12}^{\bullet}- H_2O$	0.739
1321.72791	1321.72833	c_{12}	0.312
1338.68305	1338.68336	$z_{12}^{\bullet}$	0.233
1346.73575	1346.73588	$a_{13}- H_2O$	0.100
1354.70177	1354.70206	y_{12}	0.213
1359.68338	1359.68267	$z_{13}^{\bullet}-2H_2O$	-0.523
1363.67830	1363.67811	$x_{12}-H_2O$	-0.143
1377.69395	1377.69443	$z_{13}^{\bullet}-H_2O$	0.348
1390.74938	1390.74920	$c_{13}-H_2O$	-0.131
1391.73340	1391.73249	b_{13}	-0.654
1395.70451	1395.70484	$z_{13}^{\bullet}$	0.234
1396.79603	1396.79641	$z_{14}^{\bullet}- H_3PO_4$	0.273
1408.75994	1408.76060	c_{13}	0.469
1411.72323	1411.72379	y_{13}	0.396
1420.69976	1420.69992	$x_{13}-H_2O$	0.118
1433.76778	1433.76844	$a_{14}-H_2O$	0.461
1438.71032	1438.71101	x_{13}	0.475
1458.75179	1458.75184	$z_{14}^{\bullet}-2H_2O$	0.031
1476.76236	1476.76283	$z_{14}^{\bullet}-H_2O$	0.322
1477.78141	1477.78310	$c_{14}- H_2O$	1.146
1478.78659	1478.78660	$[M+1H-C_9H_9N]^{1+}$	0.005
1479.79442	1479.79574	$[M-C_9H_9N+2H]^{+\bullet}$	0.891
1494.77292	1494.77358	$z_{14}^{\bullet}$	0.438
1495.79197	1495.79274	c_{14}	0.511
1512.89044	1512.89194	$[M- H_3PO_4+2H]^{+\bullet}$	0.993
1564.86240	1564.86296	$[M-COOH+2H]^{+\bullet}$	0.354
1582.87297	1582.87405	$[M-CO+2H]^{+\bullet}$	0.681
1592.85732	1592.85848	$[M- H_2O +2H]^{1+\bullet}$	0.731
1593.84134	1593.84079	$[M-NH_3+2H]^{1+\bullet}$	-0.344
1609.86006	1609.86132	$[M+H]^{+}$	0.784
1610.86788	1610.86547	$[M+2H]^{+\bullet}$	-1.498

Table A6-9 Fragment ion assignments following ECD of GSK3B-pY11 $[M+2H]^{2+}$ with and without in-source CID (90 V)

Theoretical m/z	Experimental m/z	Error / ppm	In-source CID Experimental m/z	Error / ppm	Assignment
2381.122	2381.121	0.319177	2381.121	0.319177	$[M+2H]^{\bullet+}$
2381.122	2381.121	0.319177			$[2M+4H]^{\bullet+}$
	2373.608				
2364.096	2364.098	0.934818	2364.098	0.934818	$[M+2H-NH_3]^{\bullet+}$
2363.112	2363.115	1.629208			$[M+2H-H_2O]^{\bullet+}$
2360.611	2360.614	1.356349			$[2M+4H-CONH_2]^{2\bullet+}$
2354.127	2354.118	4.041999	2354.126	0.312379	$[M+2H-CO]^{\bullet+}$
2352.603	2352.6	1.56558			$[2M+4H-CH_3CONH_2]^{2\bullet+}$
	2337.113				
	2336.11		2336.108		
	2346.097				
2323.085	2323.086	0.281361	2323.085	0.117785	$[M+2H-CH_3CONH_2]^{\bullet+}$
	2310.703		2310.095		
	2308.608				
2303.064	2303.062	0.907487			$[2M+4H-R]^{2+}$
	2292.122		2292.117		
2288.576	2288.573	1.245316			$[2M+4H-Q]^{2\bullet+}$
	2281.037		2281.029		
	2276.102		2276.094		
	2259.563				
	2248.08		2248.078		
2232.033	2232.031	0.918445			$[2M+4H-QL]^{2\bullet+}$
2224.013	2224.018	2.275166	2224.01	1.249993	$[M+2H-R]^{+}$
2222.043	2222.045	1.044426			$[2M+4H-RY]^{2+}$
2195.034	2195.028	2.870115	2195.033	0.546689	$[M+2H-Q]^{\bullet+}$
2182.499	2182.498	0.799542			$[2M+4H-QLV]^{2\bullet+}$
	2153.981				
2103.945	2103.945	0.105754			$[2M+4H-QLVR]^{2\bullet+}$
2081.95	2081.945	2.363169	2081.952	0.936622	$[M+2H-QLV]^{\bullet+}$
2075.434	2075.435	0.579397			$[2M+4H-QLVRG]^{2\bullet+}$
1982.882	1982.881	0.489187	1982.881	0.489187	$[M+2H-QLV]^{\bullet+}$
	1588.084				$[2M+4H-NH3]^{3\bullet+}$
	1582.408				
	1573.075				
	1568.405				
	1558.084				
1525.717	1525.718	0.897938			$[2M+4H-Q]^{3\bullet+}$
1191.065	1191.064	0.268667	1191.065	0.008396	$[M+2H]^{2+}$
1140.62	1140.618	2.235626			c_{10}^{+}
1052.585	1052.586	0.864539			c_{9}^{+}

	788.429		788.428	Harmonic
	764.1015		764.1006	Harmonic
	731.7326		731.7318	Harmonic
			694.0319	Harmonic
	661.0051		661.0037	Harmonic
613.378	613.3791	1.85856		c_5^+
	595.5334		595.5321	Harmonic
556.3566	556.357	0.647067		c_4^+
321.1795	321.1797	0.684975		$z_2^{\bullet+}$

Table A6-10 Fragment ion assignments following ECD of GSK3B-Y11 $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2302.64	2302.667	11.82556	$[M+2H]^{\bullet+}$
	2284.925		
	2258.129		
	2244.13		
	2212.153		
2145.05	2145.053	1.249388	c_{17}^{+}
2115.07	2115.067	1.475128	$z_{17}^{\bullet+}$
2001.98	2001.979	0.409595	$z_{16}^{\bullet+}$
1902.92	1902.917	1.345301	$z_{15}^{\bullet+}$
1151.08	1151.08	0.399625	$[M+2H]^{2+}$

Table A6-11 Fragment ion assignments following ECD of GSK3B-pS11 $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2305.09	2305.084	2.611612	$[M+2H]^{\bullet+}$
	2289.068		
	2278.099		
	2261.072		
	2247.057		
	2234.065		
	2216.092		
2147.98	2147.979	1.415282	$c\text{-}1_{17}^{+}$
2119	2119.001	1.033505	$z_{17}^{\bullet+}$
2005.92	2005.914	3.095836	$z_{16}^{\bullet+}$
1985.93	1985.929	0.382692	c_{16}^{+}
1906.85	1906.848	0.907255	$z_{15}^{\bullet+}$
1822.86	1822.863	1.70611	c_{15}^{+}
1153.05	1153.051	0.45965	$[M+2H]^{2+}$
	461.2587		Harmonic
	329.4733		Harmonic
	230.6192		Harmonic
	164.7288		Harmonic

Table A6-12 Fragment ion assignments following ECD of GSK3B-S11 $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assigment
2225.124	2225.128	1.514522	$[M+2H]^{\bullet+}$
2208.098	2208.083	6.612026	$[M+2H-NH_3]^{\bullet+}$
	2182.094		
	2167.098		
	2136.137		
2069.023	2069.028	2.03961	c_{17}^{+}
2039.037	2039.038	0.88277	$z_{17}^{\bullet+}$
1925.953	1925.966	6.82779	$z_{16}^{\bullet+}$
1905.96	1905.968	4.380994	c_{16}^{+}
	1865.99		
1826.884	1826.881	1.472452	$z_{15}^{\bullet+}$
	1742.894	1.577833	c_{15}^{+}
1113.066	1113.067	1.39255	$[M+2H]^{2+}$
2225.124	2225.128	1.514522	$[M+2H]^{\bullet+}$

Table A6-13 Fragment ion assignments following ECD of GSK3B-pY11-noC $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2292.129	2292.136	3.224077	$[M+2H]^{\bullet +}$
2275.1	2275.133	14.35981	$[M+2H-NH_3]^{\bullet +}$
	2265.156		
	2248.136		
	2234.112		
	2221.128		
	2207.13		
	2193.069		
	2162.08		
2135.02	2135.024	1.957827	$c\text{-}1_{17}^{+}$
	2122.075		
2106.041	2106.041	0.052231	$z_{17}^{\bullet +}$
	2049.998		
	2008.992		
1992.957	1992.963	3.01562	$z_{16}^{\bullet +}$
1971.96	1971.954	0.993937	$c\text{-}1_{16}^{+}$
1893.888	1893.888	0.274567	$z_{15}^{\bullet +}$
1809.901	1809.902	0.67407	c_{15}^{+}
1528.086	1528.099	8.625169	$[2M+4H]^{3\bullet +}$
1146.568	1146.573	4.064303	$[M+2H]^{2+}$
613.378	613.3811	4.988767	c_5^{+}
	573.289		Harmonic
556.3566	556.3588	4.008221	c_4^{+}
484.2429	484.2458	5.968079	$z_3^{\bullet +}$
	458.6725		Harmonic
	327.6252		Harmonic
321.1795	321.1814	5.94683	$z_2^{\bullet +}$
	229.3259		Harmonic
	163.8048		Harmonic
158.1162	158.1173	7.083398	$z_1^{\bullet +}$
	127.404		Harmonic

Table A6-14 Fragment ion assignments following ECD of GSK3B-Y11-noC $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2212.162	**2212.17**	3.584728	$[M+2H]^{\bullet+}$
2195.14	**2195.152**	5.243401	$[M+2H-NH_3]^{\bullet+}$
	2185.181		
	2169.141		
	2154.14		
	2141.15		
	2127.153		
2056.061	**2056.058**	1.400737	c_{17}^{+}
2026.074	**2026.065**	4.422345	$z_{17}^{\bullet+}$
1912.99	**1912.998**	3.836925	$z_{16}^{\bullet+}$
1892.998	**1892.997**	0.40148	c_{16}^{+}
	1864.985		
1813.922	**1813.924**	1.075019	$z_{15}^{\bullet+}$
	1777.953		
	1748.951		
1729.934	1729.94	3.37007	c_{15}^{+}
1657.821	1657.816	2.889335	$z_{14}^{\bullet+}$
1600.799	1600.8	0.412294	$z_{13}^{\bullet+}$
1106.585	1106.587	2.195945	$[M+2H]^{2+}$
	1049.567		
613.378	613.3777	0.456488	c_{5}^{+}
556.3566	556.3575	1.545771	c_{4}^{+}
	553.3081		Harmonic
	442.6989		Harmonic
	221.3539		Harmonic

Table A6-15 Fragment ion assignments following ECD of GSK3B-pY11-noEnoC $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2234.123	2234.124	0.380462	$[M+2H]^{\bullet+}$
2217.097	2217.094	1.105049	$[M+2H-NH_3]^{\bullet+}$
	2207.12		
	2191.083		
	2176.075		
	2164.079		
	2149.09		
	2136.035		
	2127.061		
2077.01	2077.015	0.245545	$c\text{-}1_{17}^{+}$
2048.035	2048.029	3.007761	$z_{17}^{\bullet+}$
	2020.027		
1934.951	1934.953	1.121475	$z_{16}^{\bullet+}$
1914.959	1914.965	3.195892	c_{16}^{+}
1835.883	1835.876	3.627683	$z_{15}^{\bullet+}$
1751.895	1751.894	0.850507	c_{15}^{+}
1679.782	1679.778	2.143136	$z_{14}^{\bullet+}$
1622.76	1622.761	0.529961	$z_{13}^{\bullet+}$
1454.67	1454.672	1.450501	$z_{11}^{\bullet+}$
1117.565	1117.563	1.655385	$[M+2H]^{2+}$
	1088.037		
1081.611	1081.608	2.681185	c_{10}^{+}
	1039.003		
994.5792	994.5766	2.664443	c_{9}^{+}
	837.5268		
613.378	613.3758	3.521483	c_{5}^{+}
	558.795		Harmonic
	558.7785		Harmonic
556.3566	556.3551	2.785983	c_{4}^{+}
	531.4036		
484.2429	484.2419	2.127032	$z_{3}^{\bullet+}$
	279.3967		Harmonic
	223.5212		Harmonic
158.1162	158.1154	5.375793	$z_{1}^{\bullet+}$

Table A6-16 Fragment ion assignments following ECD of GSK3B-Y11-noEnoC $[M+2H]^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2154.157	**2154.151**	2.71568	$[M+2H]^{\bullet +}$
2137.13	**2137.138**	3.836921	$[M+2H-NH_3]^{\bullet +}$
	2111.121		
	2096.117		
	2069.128		
1998.056	**1998.056**	0.405394	c_{17}^{+}
1968.069	**1968.069**	0.076217	$z_{17}^{\bullet +}$
1854.985	**1854.991**	3.363909	$z_{16}^{\bullet +}$
1834.992	**1834.998**	3.100831	c_{16}^{+}
	1806.963		
1755.916	**1755.91**	3.46258	$z_{15}^{\bullet +}$
	1719.934		
1671.929	**1671.932**	1.866108	c_{15}^{+}
1599.815	**1599.811**	2.819076	$z_{14}^{\bullet +}$
1542.794	**1542.797**	1.944524	$z_{13}^{\bullet +}$
1515.828	**1515.833**	3.384289	c_{14}^{+}
1428.796	**1428.793**	1.72173	c_{13}^{+}
1374.704	1374.708	3.171592	$z_{11}^{\bullet +}$
1357.759	1357.758	0.235683	c_{12}^{+}
1161.593	1161.592	0.060262	$z_{9}^{\bullet +}$
1081.611	**1081.61**	0.822846	c_{10}^{+}
1077.582	**1077.583**	0.510402	$[M+2H]^{2+}$
1074.561	1074.562	1.470369	$z_{8}^{\bullet +}$
	1061.804		
994.5792	**994.5784**	0.814415	c_{9}^{+}
	796.6803		
640.344	640.3428	1.858376	$z_{4}^{\bullet +}$
613.378	613.3786	0.994493	c_{5}^{+}
556.3566	556.356	1.006549	c_{4}^{+}
	538.7886		harmonic
	531.4053		
158.1162	158.1157	2.972497	$z_{1}^{\bullet +}$

Table A6-17 Fragment ion assignments following ETD-TWIMS of GSK3B-pY11 [M+2H]$^{2+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2381.122	2381.06	26.88	$[M+2H]^{\bullet+}$
2381.122	2381.06	26.88	$[2M+4H]^{2+}$
2364.0955	2364.05	20.67	$[MH-NH_3]^{\bullet+}$
2322.0849	2322.08	0.70	$[y_{18}-NH_3]^{+}$
2225.0209	2225.05	15.10	c_{17}^{+}
2195.0342	2195.10	29.57	$z_{17}^{\bullet+}$
2081.9501	2082.00	25.92	$z_{16}^{\bullet+}$
1982.8817	1982.97	45.61	$z_{15}^{\bullet+}$
1587.4147	1587.35	39.82	$[2M+4H]^{3+}$
1525.717	1525.70	10.12	$[2M+4H-Q]^{3+}$
1191.0647	1191.16	80.01	MH^{2+}
1182.5514	1182.54	12.09	$MH-NH_3^{2+}$
1170.0594	1170.10	32.22	y_{18}^{2+}
158.1162	158.11	38.27	$z_1^{\bullet+}$

Table A6-18 Fragment ion assignments following ETD-TWIMS of GSK3B-pY11 $[M+3H]^{3+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2381.122	2381.06	24.96	$[M+3H]^{\bullet+}$
2364.096	2364.07	11.95	$[M+3H-NH_3]^{\bullet+}$
2339.112	2339.07	17.85	y_{18}^{+}
2322.085	2322.06	8.90	$[y_{18}-NH_3]^{+}$
2225.021	2224.98	17.16	c_{17}^{+}
2211.053	2211.02	16.81	y_{17}^{+}
2195.034	2194.99	21.18	$z_{17}^{\bullet+}$
2180	2179.97	13.44	a_{17}^{+}
2097.969	2097.92	23.65	y_{16}^{+}
2081.95	2081.91	18.70	$z_{16}^{\bullet+}$
2061.958	2061.93	12.81	c_{16}^{+}
1998.9	1998.86	20.39	y_{15}^{+}
1982.882	1982.85	15.16	$z_{15}^{\bullet+}$
1898.894	1898.82	39.31	c_{15}^{+}
1842.799	1842.78	9.61	y_{14}^{+}
1826.781	1826.76	9.31	$z_{14}^{\bullet+}$
1785.778	1785.76	9.36	y_{13}^{+}
1769.759	1769.75	4.53	$z_{13}^{\bullet+}$
1742.793	1742.79	0.86	c_{14}^{+}
1655.761	1655.74	12.69	c_{13}^{+}
1559.683	1559.64	25.30	y_{11}^{+}
1543.664	1543.66	5.62	$z_{11}^{\bullet+}$
1429.621	1429.62	3.35	$z_{10}^{\bullet+}$
1382.646	1382.66	12.38	c_{11}^{+}
1330.552	1330.55	1.45	$z_{9}^{\bullet+}$
1243.52	1243.52	1.25	$z_{8}^{\bullet+}$
1191.065	1191.07	0.87	$[M+3H]^{2+}$
1182.551	1182.56	8.52	$[M+3H-NH_3]^{2+}$
1170.059	1170.06	3.56	y_{18}^{2+}
1161.546	1161.55	4.20	$[y_{18}-NH_3]^{2+}$
1139.617	1139.62	4.66	c_{10}^{+}
1113.014	1113.01	6.64	c_{17}^{2+}
1106.03	1106.03	2.82	y_{17}^{2+}
1098.021	1098.03	7.75	$z_{17}^{2\bullet+}$
1094.595	1094.61	11.14	a_{10}^{+}
1052.585	1052.59	8.35	c_{9}^{+}
1049.488	1049.49	6.62	y_{16}^{2+}
1041.479	1041.46	17.05	$z_{16}^{2\bullet+}$
999.9538	999.94	13.62	y_{15}^{2+}
991.9445	991.93	18.48	$z_{15}^{2\bullet+}$
953.5163	953.47	51.75	c_{8}^{+}

839.4734	839.47	0.55	c_7^+
794.3789	794.38	1.90	$[M+3H]^{3+}$
788.7033	788.73	39.62	$[M+3H-NH_3]^{3+}$
780.3753	780.34	44.74	y_{18}^{3+}
613.378	613.37	8.33	c_5^+
568.3566	568.36	10.17	a_5^+
556.3566	556.35	18.37	c_4^+
511.3351	511.33	0.45	a_4^+
484.2429	484.22	48.69	$z_3^{\bullet+}$
366.2023	366.17	79.19	$[b_3-NH_3]^+$
321.1795	321.16	70.71	$z_2^{\bullet+}$
239.139	239.13	38.60	$[a_2-NH_3]^+$
188.103	188.09	78.41	c_1^+
187.0951	187.08	66.02	$c-1_1^+$
171.0764	171.09	54.58	b_1^+
159.124	159.12	26.63	$z+1_1^{\bullet+}$

Table A6-19 Fragment ion assignments following ETD-ECD of the $z_{17}^{2\bullet+}$ *fragment of GSK3B-pY11* $[M+3H]^{3+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2196.038	2196.037	0.250451	MS^3 $z_{17}^{\bullet+}$
2178.008	2178.002	2.773177	MS^3 [z_{17}-NH3] $^{\bullet+}$
	2168.031		
	2150.004		
	2136.995		
	2105.006		
2080.947	2080.949	1.191765	$z_{16}^{\bullet+}$
1981.878	1981.878	0.03532	z-$1_{15}^{\bullet+}$
1875.87	1875.864	2.777378	$c_{15}^{\bullet+}$
1825.777	1825.757	11.18975	z-$1_{14}^{\bullet+}$
1768.756	1768.741	8.604919	z-$1_{13}^{\bullet+}$
1713.81	1713.808	0.787719	c+$1_{14}^{\bullet+}$
	1114.006		
	1105.001		
1098.021	1098.015	5.628309	MS^2 $z_{17}^{2\bullet+}$
	1095.592		
	966.394		
	953.8603		
	732.054		Harmonic
	725.385		
	712.3757		Harmonic
	613.3689		
	549.0023		Harmonic
	366.0135		Harmonic
	219.6091		Harmonic
	190.4122		
159.1195	159.1212	10.49526	z+$1_1^{\bullet+}$

Table A6-20 Fragment ion assignments following ECD-ECD of the $z_{17}^{2\bullet+}$ *fragment of GSK3B-pY11 [M+3H]*$^{3+}$

Theoretical m/z	Experimental m/z	Error / ppm	Assignment
2196.038	2196.037	0.250451	MS^3 $z_{17}^{\bullet+}$
2179.011	2179.012	0.569066	MS^3 [z_{17}-NH3] $^{\bullet+}$
	2136.995		
	2105.006		
2080.947	2080.93	8.03961	$z_{16}^{\bullet+}$
1981.878	1981.861	8.759367	z-$1_{15}^{\bullet+}$
1876.873	1876.864	4.747258	c+$1_{15}^{\bullet+}$
1825.777	1825.772	3.089095	z-$1_{14}^{\bullet+}$
1768.756	1768.754	0.757595	z-$1_{13}^{\bullet+}$
1713.81	1713.808	0.787719	c+$1_{14}^{\bullet+}$
1543.664	1543.659	3.083573	$z_{11}^{\bullet+}$
1330.552	1330.542	8.124445	$z_{9}^{\bullet+}$
1243.52	1243.513	6.015181	$z_{8}^{\bullet+}$
	1106.015		
1098.021	1098.015	5.628309	MS^2 $z_{17}^{2\bullet+}$
	1095.592		
	1088.002		
	732.054		Harmonic
	549.0009		Harmonic
	366.0129		Harmonic
	219.6089		Harmonic
	60.05448		

Table A6-21 Fragment ion assignments of the $z_{17}^{2\bullet+}$ fragment of GSK3B-pY11-noC $[M+3H]^{3+}$ following ETD-ECD, ECD-ECD and the respective isolation spectra ("ETD iso" and "ECD iso")

Theoretical m/z	Experimental m/z	Error / ppm	Assignment	ETD-ECD	ECD-ECD	ETD iso	ECD iso
2107.049	2107.05	0.792578	MS^3 $z_{17}^{\bullet+}$	YES	YES	NO	NO
2089.038	2089.027	5.323025	MS^3 [z_{17}-NH3] $^{\bullet+}$	YES	YES	NO	NO
	2078.28			YES	NO	NO	NO
	2049.017			YES	YES	NO	NO
	2047.011			YES	YES	NO	NO
2008.975	2009.005	14.85832	y_{16}^{+}	NO	YES	NO	NO
	2008.916			NO	YES	NO	NO
1990.95	1991.944	4.357532	$z_{16}^{\bullet+}$	NO	YES	NO	NO
1892.885	1892.887	1.352433	$z_{15}^{\bullet+}$	YES	YES	NO	NO
1891.896	1891.839	30.46679	$[y_{15}\text{-}H_2O]^{+}$	NO	YES	NO	NO
	1788.314			NO	YES	NO	NO
1786.876	1786.863	7.476735	c_{15}^{+}	NO	YES	NO	NO
1753.806	1753.803	1.556615	y_{14}^{+}	NO	YES	NO	NO
1737.787	1737.781	3.274279	$z_{14}^{\bullet+}$	NO	YES	NO	NO
1680.766	1680.769	1.784901	$z_{13}^{\bullet+}$	YES	YES	NO	NO
	1679.767			NO	YES	NO	NO
1624.816	1624.809	4.388189	c_{14}^{+}	NO	YES	NO	NO
1453.667	1453.66	4.40266	$z\text{-}1_{11}^{\bullet+}$	NO	YES	NO	NO
	1301.712			NO	YES	NO	NO
1240.555	1240.547	6.58576	$z\text{-}1_{9}^{\bullet+}$	NO	YES	NO	NO
1154.527	1154.523	2.910283	$z_{8}^{\bullet+}$	NO	YES	NO	NO
1153.519	1153.513	5.0541	$[y_{8}\text{-}NH_3]^{+}$	NO	YES	NO	NO
	1147.068			YES	YES	NO	NO
	1139.61			NO	YES	NO	NO
	1117.548			NO	YES	NO	NO
	1061.527			YES	YES	NO	NO
	1056.59			NO	YES	NO	NO
1053.524	1053.518	5.400921	MS^2 $z_{17}^{\bullet+}$	YES	YES	NO	NO
	1052.579		MS^2 c_{10}^{+}	YES	YES	NO	NO
	1017.51			NO	YES	NO	NO
947.9516	947.9506	1.107651	$z_{15}^{2\bullet+}$	NO	YES	NO	NO
867.5002	867.5017	1.706051	c_{8}^{+}	NO	YES	NO	NO
	764.7103			NO	YES	NO	NO
725.3842	725.3862	2.688231	$[y_{5}\text{-}H_2O]^{+}$	YES	YES	NO	NO
	613.3754			YES	YES	NO	NO
539.3062	539.3214	28.18436	b_{5}^{+}	YES	NO	NO	YES
512.3148	512.339	47.23658	a_{5}^{+}	NO	YES	NO	NO
500.2616	500.2571	9.055262	y_{3}^{+}	YES	NO	NO	NO
383.2722	383.2729	1.800287	a_{4}^{+}	YES	NO	NO	YES
	372.1999			YES	NO	NO	NO

	337.2293			YES	NO	NO	NO
	326.177			YES	NO	YES	YES
	313.0894			YES	NO	NO	NO
	303.1759			YES	NO	NO	YES
	299.1698			YES	NO	YES	YES
	284.191			YES	YES	YES	YES
	280.1707			YES	NO	NO	NO
	278.1803			YES	NO	NO	NO
	266.1798			YES	NO	NO	NO
	260.169			YES	NO	NO	NO
	256.1952			YES	NO	NO	YES
	242.1577			YES	NO	NO	NO
	216.0695			YES	NO	NO	YES
	212.1301			YES	YES	YES	YES
197.141	197.1295	58.13098	b_2^+	YES	NO	NO	YES
	185.1899			YES	NO	NO	YES
	171.0756			YES	YES	YES	YES
157.1127	157.1076	32.77902	$z\text{-}1_1^{\bullet +}$	YES	YES	YES	YES
136.0757	136.0751	4.629776	Y^+	YES	YES	YES	YES
115.1035	115.1046	9.469738	$c\text{-}1_1^+$	YES	YES	YES	YES
112.0869	112.0864	4.639258	R^+	YES	YES	YES	YES
100.0869	100.0864	4.995659	R^+	NO	YES	NO	YES
86.0964	86.096	4.645955	$I^+/[R\text{-}H]^+$	YES	YES	YES	YES
72.0808	72.08043	5.133128	V^+	YES	YES	YES	YES
70.0651	70.06479	4.424457	R^+/P^+	YES	YES	YES	YES
60.0444	60.05534	182.1985	S^+	NO	YES	NO	YES

Table A6-22 Fragment ion assignments of the $z_{17}^{2\bullet+}$ fragment of GSK3B-Y11-noC $[M+3H]^{3+}$ following ETD-ECD, ECD-ECD and the respective isolation spectra (“ETD iso” and “ECD iso”)

Theoretical m/z	Experimental m/z	Error / ppm	Assignment	ETD-ECD	ECD-ECD	ETD iso	ECD iso
2027.082	2027.09	3.6703	MS^3 $z_{17}^{\bullet+}$	YES	YES	NO	NO
2009.072	2009.077	2.573328	MS^3 $[z_{17}\text{-}H_2O]^{\bullet+}$	YES	YES	NO	NO
	1997.049			NO	YES	NO	NO
	1991.047			NO	YES	NO	NO
	1968.015			YES	YES	NO	NO
	1950.002			NO	YES	NO	NO
1911.987	1911.977	5.005265	$z\text{-}1_{16}^{\bullet+}$	NO	YES	NO	NO
1869.978	1869.977	0.379684	$c\text{-}1_{16}^{+}$	YES	YES	NO	NO
1812.918	1812.917	0.524017	$z\text{-}1_{15}^{\bullet+}$	YES	YES	NO	NO
	1794.887			NO	YES	NO	NO
	1729.913			NO	YES	NO	NO
1707.918	1707.919	0.837277	c_{15}^{+}	YES	YES	NO	NO
1673.839	1673.842	1.445778	y_{14}^{+}	NO	YES	NO	NO
1656.817	1656.815	1.49081	$z\text{-}1_{14}^{\bullet+}$	NO	YES	NO	NO
1599.796	1599.797	1.050134	$z\text{-}1_{13}^{\bullet+}$	YES	YES	NO	NO
	1573.828			NO	YES	NO	NO
1544.854	1544.857	1.527652	c_{14}^{+}	YES	YES	NO	NO
1487.775	1487.788	8.294263	y_{12}^{+}	NO	YES	NO	NO
1388.753	1388.749	3.0747	c_{13}^{+}	NO	YES	NO	NO
1373.7	1373.693	5.328673	$z\text{-}1_{11}^{\bullet+}$	NO	YES	NO	NO
	1314.656			NO	YES	NO	NO
1301.721	1301.719	1.413513	c_{12}^{+}	NO	YES	NO	YES
1276.68	1276.675	3.650093	y_{10}^{+}	NO	YES	NO	NO
1259.658	1259.646	9.304116	$z\text{-}1_{10}^{\bullet+}$	NO	YES	NO	NO
1230.684	1230.679	4.038404	c_{11}^{+}	NO	YES	NO	NO
1160.589	1160.581	7.254937	$z\text{-}1_{9}^{\bullet+}$	NO	YES	NO	YES
	1142.574			NO	YES	NO	YES
	1139.616			NO	YES	NO	NO
	1124.561			NO	YES	NO	NO
1117.6	1117.598	2.210093	c_{10}^{+}	NO	YES	NO	NO
1089.576	1089.567	8.425298	y_{8}^{+}	NO	YES	NO	NO
1073.557	1073.548	8.569643	$z\text{-}1_{8}^{\bullet+}$	NO	YES	NO	NO
1055.552	1055.541	10.47793	$[a_{10}\text{-}NH_3]^{+}$	NO	YES	NO	NO
	1052.579			NO	YES	NO	NO
	1018.516			NO	YES	NO	NO
1013.041	1013.042	0.799573	MS^2 $z_{17}^{\bullet+}$	YES	YES	NO	NO
	1009.571			NO	YES	NO	NO
	1004.532			NO	YES	NO	NO
	985.5071			NO	YES	NO	NO
	977.5238			NO	YES	NO	NO

	960.5127			NO	YES	NO	NO
954.5367	954.5337	3.100981	c_9^+	NO	YES	NO	NO
910.4937	910.4884	5.799052	$z\text{-}1_7^{\bullet+}$	NO	YES	NO	NO
867.5047	867.4983	7.331372	c_8^+	NO	YES	NO	YES
	738.0499		MS^1 $[M+3H]^{3+}$	NO	YES	NO	NO
726.3725	726.3655	9.595628	$z\text{-}1_5^{\bullet+}$	NO	YES	NO	YES
707.3918	707.3947	4.156113	$[a_7\text{-}NH_3]^+$	NO	YES	NO	YES
	697.3618			NO	YES	NO	YES
	688.0386			YES	YES	YES	YES
640.344	640.3406	5.356496	$z_4^{\bullet+}$	NO	YES	NO	NO
	613.3749			NO	YES	NO	NO
	556.353			NO	YES	NO	NO
484.2429	484.2404	5.224651	$z_3^{\bullet+}$	NO	YES	NO	NO
321.1795	321.1769	8.064027	$z_2^{\bullet+}$	NO	YES	NO	NO
	311.1685			NO	YES	NO	NO
	298.1485		$z_6^{\bullet+}$	NO	YES	NO	NO
	284.1598			YES	YES	YES	YES
	256.1636		$c\text{-}3_2^+$	NO	YES	NO	NO
	239.1484			NO	YES	NO	NO
	212.1009			YES	YES	NO	YES
	185.163		$a\text{-}1_{2+}$	NO	YES	NO	YES
171.1574	171.0748	482.4799	a_2^+	YES	YES	YES	YES
158.1162	158.1156	4.047656	$z\text{-}1_1^{\bullet+}$	YES	YES	NO	YES
136.0757	136.0752	4.041868	Y^+	YES	YES	YES	YES
	115.086			YES	YES	NO	YES
112.0869	112.0865	3.657876	R^+	YES	YES	YES	YES
100.0869	100.086	8.592533	R^+	NO	YES	YES	YES
86.0964	86.09613	3.13602	$I^+/[R\text{-}H]^+$	YES	YES	NO	YES
72.0808	72.08052	3.88453	V^+	YES	YES	NO	YES
70.0651	70.06486	3.425386	R^+/P^+	YES	YES	YES	YES

Table A6-23 Fragment ion assignments of the $z_{17}^{2\bullet+}$ fragment of GSK3B-pY11-noEnoC $[M+3H]^{3+}$ following ETD-ECD, ECD-ECD and the respective isolation spectra (“ETD iso” and “ECD iso”)

Theoretical m/z	Experimental m/z	Error / ppm	Assignment	ETD-ECD	ECD-ECD	ETD iso	ECD iso
2050.047	2050.042	2.409701	MS^3 $z_{17}^{\bullet+}$	YES	YES	NO	NO
2032.016	2032.019	1.348414	MS^3 [z_{17}-NH3] $^{\bullet+}$	YES	YES	NO	NO
	2022.023			YES	YES	NO	NO
	2020.023			NO	YES	NO	NO
	2014.013			NO	YES	NO	NO
	2005.024			YES	YES	NO	NO
	1990.994			YES	YES	NO	NO
	1972.993			NO	YES	NO	NO
	1952.063			NO	YES	NO	NO
1934.951	1934.954	1.674461	$z_{16}^{\bullet+}$	YES	YES	NO	NO
1892.942	1892.935	3.724362	c_{16}^{+}	YES	YES	NO	NO
	1891.934			YES	NO	NO	NO
1835.883	1835.877	3.028516	$z_{15}^{\bullet+}$	YES	YES	NO	NO
	1751.895			NO	YES	NO	NO
1729.879	1729.874	2.653366	c_{15}^{+}	YES	YES	NO	NO
1695.8	1695.803	1.491921	y_{14}^{+}	NO	YES	NO	NO
	1693.79			NO	YES	NO	NO
1679.782	1679.779	1.470429	$z_{14}^{\bullet+}$	YES	YES	NO	NO
1638.779	1638.783	2.562884	y_{13}^{+}	NO	YES	NO	NO
1622.76	1622.762	1.226306	$z_{13}^{\bullet+}$	YES	YES	NO	NO
1566.815	1566.816	0.453149	c_{14}^{+}	NO	YES	NO	NO
1453.667	1453.67	2.043109	$z\text{-}1_{11}^{\bullet+}$	NO	YES	NO	NO
	1394.626			NO	YES	NO	NO
	1372.773			NO	YES	NO	NO
	1243.714			NO	YES	NO	NO
1240.555	1240.554	1.088222	$z\text{-}1_{9}^{\bullet+}$	NO	YES	NO	NO
1170.546	1170.543	2.503961	y_{8}^{+}	NO	YES	NO	NO
1153.523	1153.525	1.413062	$z\text{-}1_{8}^{\bullet+}$	NO	YES	NO	NO
	1142.585			NO	YES	NO	NO
	1118.068			NO	YES	NO	NO
	1109.557			NO	YES	NO	NO
	1088.549			NO	YES	NO	NO
	1081.612			NO	YES	NO	NO
	1037.597			NO	YES	NO	YES
	1032.529			YES	YES	NO	NO
	1031.563			NO	YES	NO	YES
1024.523	1024.525	1.805718	MS^2 $z_{17}^{\bullet+}$	YES	YES	NO	NO
	1020.596			NO	YES	NO	NO
	1018.006			NO	YES	NO	NO
	994.5797			NO	YES	NO	YES

	937.5585			NO	YES	NO	YES
918.9489	918.9475	1.479952	$z+1_{15}^{2\bullet+}$	NO	YES	NO	NO
910.4937	910.4921	1.768271	$z-1_{7}^{\bullet+}$	NO	YES	NO	NO
896.5312	896.5304	0.847712	c_9^+	NO	YES	NO	NO
809.4992	809.4987	0.679432	c_8^+	NO	YES	NO	YES
	781.4678			NO	YES	NO	NO
	745.3789		MS^1 $[M+3H]^{3+}$	NO	YES	NO	NO
728.3838	728.384	0.329497	$z_5^{\bullet+}$	NO	YES	NO	NO
641.3518	641.3517	0.155921	$z_4^{\bullet+}$	NO	YES	NO	YES
	613.3782			YES	YES	NO	NO
	597.3465			NO	YES	NO	YES
	569.3643			YES	YES	NO	NO
	556.3566			NO	YES	NO	NO
	539.3304			YES	NO	NO	NO
484.2429	484.2428	0.165206	$z_3^{\bullet+}$	NO	YES	NO	NO
	383.2292			YES	YES	NO	NO
	331.0683			YES	NO	NO	NO
321.1795	321.1797	0.498164	$z_2^{\bullet+}$	YES	YES	NO	NO
	320.171			YES	NO	NO	NO
	313.057			YES	YES	NO	NO
	311.1711			YES	YES	NO	NO
	303.1453			YES	YES	NO	NO
	299.1388			YES	NO	NO	NO
	284.1606			YES	YES	YES	YES
	275.1483			YES	NO	NO	NO
	239.1486			YES	NO	NO	NO
	227.1122			YES	YES	NO	NO
	216.0405			YES	NO	NO	NO
	212.1015			YES	YES	YES	YES
	171.0753			YES	YES	YES	YES
158.1162	158.116	1.075159	$z_1^{\bullet+}$	YES	YES	YES	NO
136.0757	136.0756	1.102328	Y^+	YES	YES	YES	NO
	123.0094			NO	YES	NO	YES
	115.0865			YES	YES	YES	NO
112.0869	112.0869	0.356866	R^+	YES	YES	YES	YES
100.0869	100.0869	0.499566	R^+	NO	YES	NO	YES
86.0964	86.09638	0.232298	$I^+/[R-H]^+$	YES	YES	YES	NO
72.0808	72.08073	0.971132	V^+	YES	YES	YES	NO
70.0651	70.06508	0.285449	R^+/P^+	YES	YES	YES	NO
	60.0556			NO	YES	YES	YES

Table A6- 24 Fragment ion assignments of the $z_{17}^{2\bullet+}$ fragment of GSK3B-Y11-noEnoC $[M+3H]^{3+}$ following ETD-ECD, ECD-ECD and the respective isolation spectra ("ETD iso" and "ECD iso")

Theoretical m/z	Experimental m/z	Error / ppm	Assignment	ETD-ECD	ECD-ECD	ETD iso	ECD iso
1970.08	1970.079	0.630939	MS^3 $z_{17}^{\bullet+}$	YES	YES	NO	NO
1952.05	1952.046	2.248918	MS^3 $[z_{17}\text{-}NH3]^{\bullet+}$	YES	YES	NO	NO
	1940.067			NO	YES	NO	NO
	1911.021			YES	YES	NO	NO
1853.981	1853.982	0.129451	$z_{16}^{\bullet+}$	YES	YES	NO	NO
	1835.982			YES	YES	NO	NO
1811.972	1811.956	9.078507	$c\text{-}1_{16}^{+}$	YES	YES	NO	NO
1755.916	1755.907	5.33055	$z_{15}^{\bullet+}$	YES	YES	NO	NO
	1671.918			NO	YES	NO	NO
1649.912	1649.915	1.394014	c_{15}^{+}	YES	YES	NO	NO
1616.838	1616.826	7.143575	$y+1_{14}^{+}$	NO	YES	NO	NO
1599.815	1599.809	4.044217	$z_{14}^{\bullet+}$	YES	YES	NO	NO
	1596.783			NO	YES	NO	NO
1588.868	1588.893	15.90441	$[a+1_{15}\text{-}NH_3]^{+}$	NO	YES	NO	NO
1541.79	1541.792	1.413941	$z\text{-}1_{13}^{\bullet+}$	YES	YES	NO	NO
	1525.773			NO	YES	NO	NO
	1515.831			NO	YES	NO	NO
1486.849	1486.846	2.185831	c_{14}^{+}	YES	YES	NO	YES
	1428.801			NO	YES	NO	YES
1373.7	1373.693	5.328673	$z\text{-}1_{11}^{\bullet+}$	YES	YES	NO	NO
	1371.779			NO	YES	NO	YES
	1357.765			NO	YES	NO	NO
1330.748	1330.745	2.44975	c_{13}^{+}	YES	YES	NO	NO
	1314.633		$b+1_{13}^{+}$	NO	YES	NO	NO
1276.68	1276.675	3.650093	y_{10}^{+}	NO	YES	NO	NO
1259.658	1259.653	3.754989	$z\text{-}1_{10}^{\bullet+}$	NO	YES	NO	NO
1243.716	1243.716	0.482425	c_{12}^{+}	YES	YES	NO	YES
1172.679	1172.68	1.28765	c_{11}^{+}	NO	YES	NO	NO
1160.589	1160.581	7.254937	$z\text{-}1_{9}^{\bullet+}$	YES	YES	NO	YES
	1142.568			NO	YES	NO	YES
	1131.578			NO	YES	NO	NO
	1124.561			NO	YES	NO	YES
	1081.609			NO	YES	NO	NO
	1078.082			NO	YES	NO	NO
1074.561	1074.556	4.215677	$z_{8}^{\bullet+}$	YES	YES	NO	YES
	1070.077			NO	YES	NO	NO
1059.595	1059.59	4.652723	c_{10}^{+}	NO	YES	NO	NO
	1056.065			NO	YES	NO	NO
	1048.559			NO	YES	NO	NO
	1040.054			NO	YES	NO	NO

	1031.558			NO	YES	NO	YES
	1024.581		$[b_{10}-H_2O]^+$	NO	YES	NO	NO
	999.5266			NO	YES	NO	NO
	994.5775			YES	YES	NO	YES
	993.044			YES	NO	NO	NO
984.5403	984.5343	6.04343	MS^2 $z_{17}^{\bullet+}$	YES	YES	NO	NO
	979.5671			NO	YES	NO	YES
	978.0243			YES	YES	NO	NO
	976.0255			NO	YES	NO	NO
	937.5575			NO	YES	NO	YES
	935.4914			NO	YES	NO	YES
	932.0103			NO	YES	NO	YES
911.4972	911.4935	4.048285	$z_7^{\bullet+}$	NO	YES	NO	NO
896.5312	896.5286	2.866604	c_9^+	YES	YES	NO	YES
878.9636	878.9621	1.672424	$z_{15}^{2\bullet+}$	NO	YES	NO	NO
809.4992	809.4947	5.509579	c_8^+	YES	YES	NO	YES
798.4131	798.4102	3.64473	$z_6^{\bullet+}$	NO	YES	NO	NO
	781.4624			NO	YES	NO	NO
727.3803	727.3734	9.499845	$z_5^{\bullet+}$	NO	YES	NO	YES
	718.7183		MS^1 $[M+3H]^{3+}$	YES	YES	NO	NO
	707.3447			NO	YES	NO	YES
	697.3618			NO	YES	NO	YES
	691.3241			NO	YES	NO	YES
	688.0344			YES	YES	YES	YES
640.3483	640.3424	9.229352	$z_4^{\bullet+}$	NO	YES	NO	NO
	613.3749			NO	YES	NO	NO
596.3877	596.3856	3.487664	c_6^+	NO	YES	NO	YES
	556.353			YES	YES	NO	NO
484.2429	484.2404	5.224651	$z_3^{\bullet+}$	NO	YES	NO	YES
	383.2246			YES	YES	NO	YES
	380.1887			YES	NO	NO	NO
	364.1824			YES	YES	NO	NO
	355.23			YES	YES	NO	NO
	352.2303			YES	YES	NO	NO
	327.129		$a+1_3^+$	YES	YES	NO	NO
321.1795	321.1774	6.631806	$z_2^{\bullet+}$	YES	YES	NO	YES
	311.169			YES	YES	YES	YES
	303.1428			YES	YES	NO	YES
	299.1368			YES	YES	NO	NO
	298.1475			NO	YES	NO	YES
	284.1571			YES	YES	YES	YES
	280.1368			YES	NO	NO	NO
	277.1636			YES	YES	NO	YES
	275.1479			YES	YES	NO	NO

	266.1481			YES	YES	NO	YES
	260.1361			YES	NO	NO	NO
	256.1625			YES	YES	NO	YES
	251.0997			YES	YES	NO	NO
	249.1569			YES	YES	NO	NO
	239.1473			YES	NO	NO	YES
	229.1268			YES	YES	NO	NO
	227.1111			YES	YES	NO	YES
	223.105			YES	NO	NO	NO
	212.1004			YES	YES	YES	YES
	184.1059			YES	YES	NO	NO
	171.0744			YES	YES	YES	YES
	169.1063			YES	YES	NO	NO
	167.0796			YES	YES	NO	NO
	159.1222			YES	NO	YES	NO
158.1162	158.1151	6.83042	$z\text{-}1_1^{\bullet+}$	YES	YES	YES	YES
	140.0802			YES	NO	NO	NO
	138.0643			YES	NO	NO	NO
136.0757	136.0747	7.539921	Y^+	YES	YES	YES	YES
	115.0856			YES	YES	YES	YES
112.0869	112.0861	7.137319	R^+	YES	YES	YES	YES
100.0869	100.0862	7.193749	R^+	NO	YES	NO	YES
	94.72027			YES	YES	NO	YES
86.0964	86.0958	6.968932	$I^+/[R\text{-}H]^+$	YES	YES	YES	YES
72.0808	72.08024	7.769059	V^+	YES	YES	YES	YES
70.0651	70.06464	6.565323	R^+/P^+	YES	YES	YES	YES
	60.05521			YES	YES	NO	YES
	57.02532			YES	YES	NO	YES

Appendix 7: GSK3B dimer possibilities

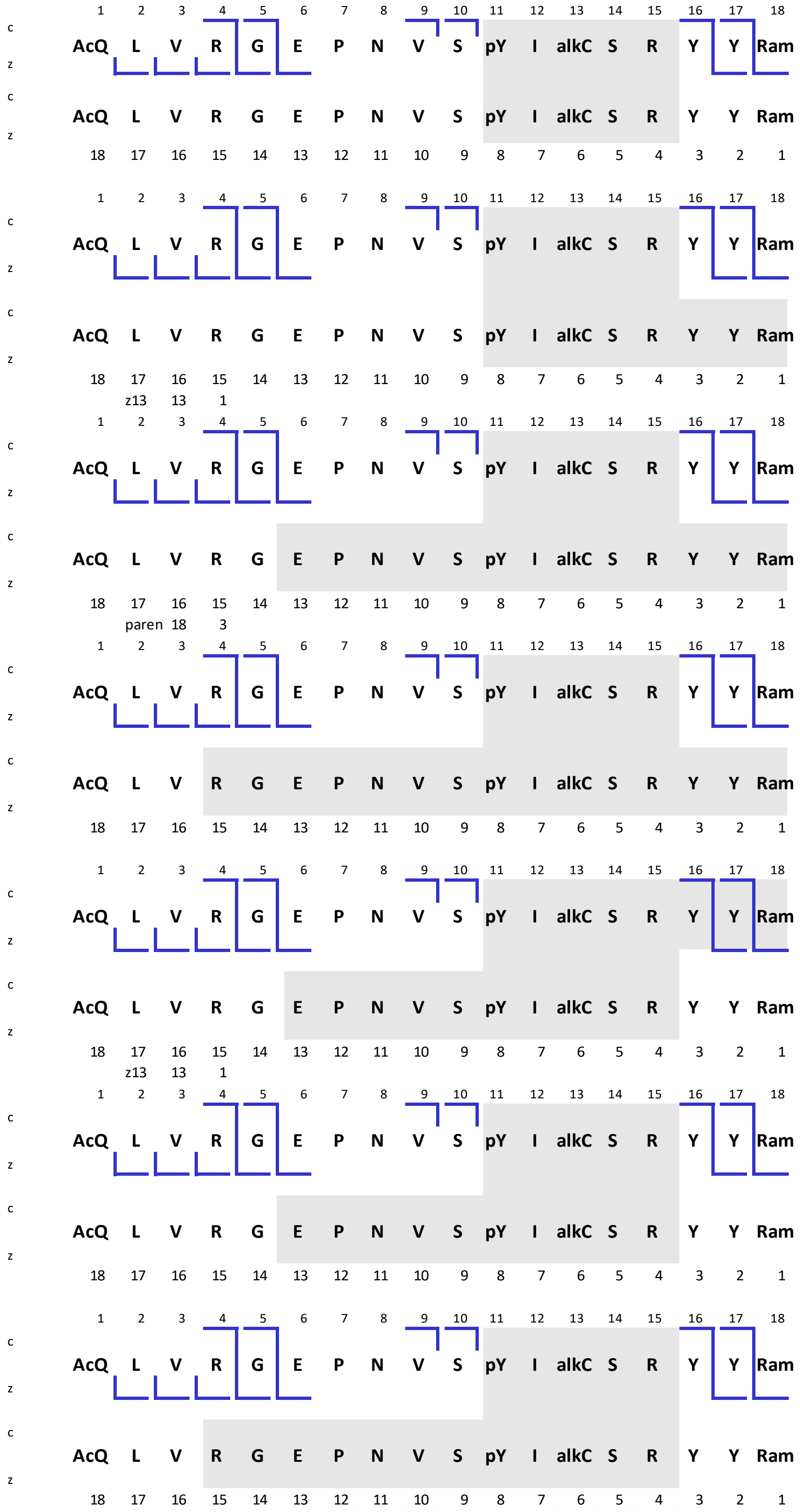


Figure A7-1 Fragmentation patterns for the dimeric species $[2M+4H]^{4+}$ GSK3B-pY11 showing possible interfaces region between the two peptide units in grey. These patterns assume that all fragments in the ECD MS/MS spectrum without in-source CID are formed from the dimer except $c_{17}^{\bullet+}$, $z_{17}^{\bullet+}$, $z_{16}^{\bullet+}$ and $z_{15}^{\bullet+}$. Fragments are only marked on one of the peptide units, but could form from any part of the dimer outside of the interface region

Appendix 8: Activated-ion ECD MS/MS spectra

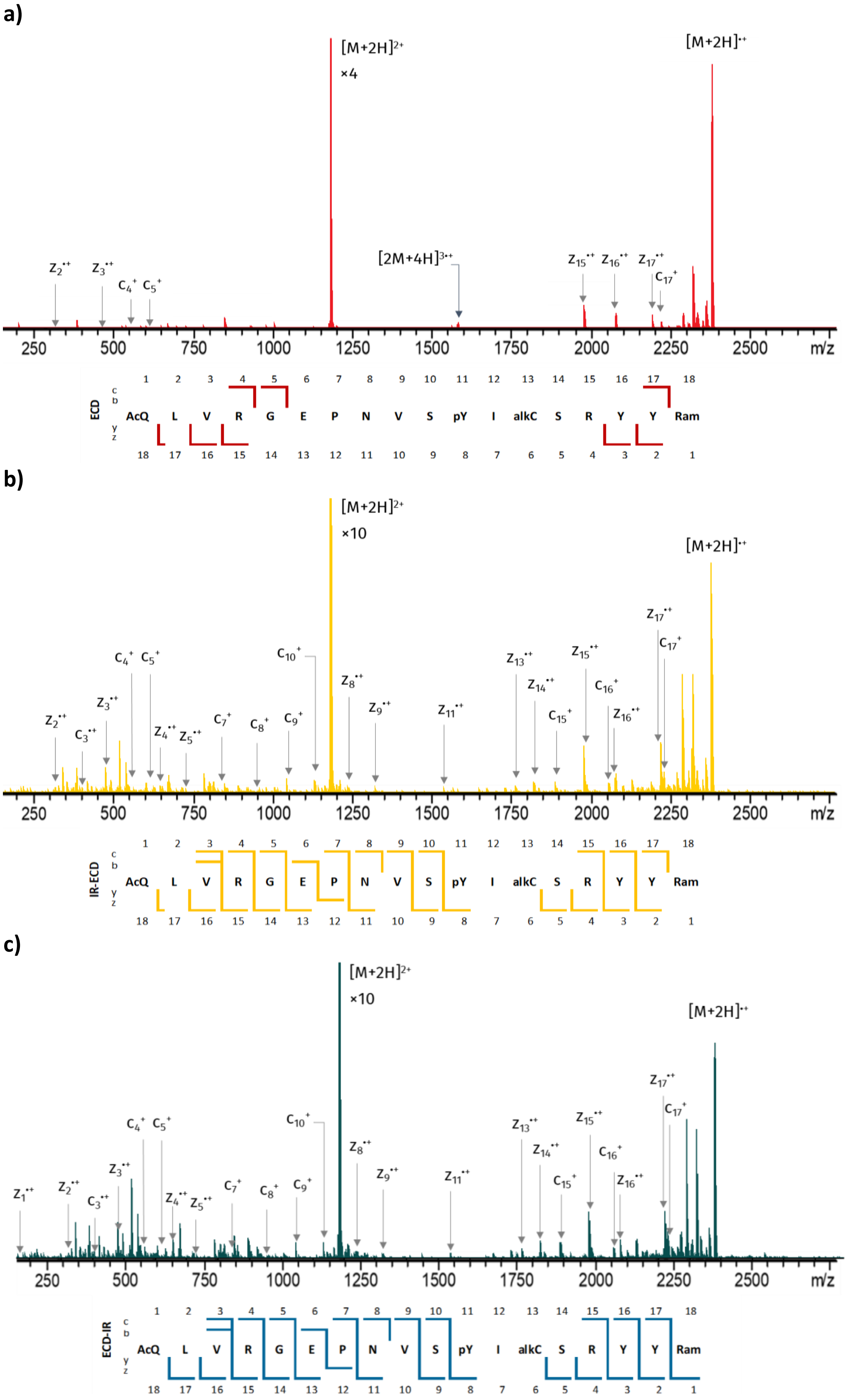


Figure A8-1 Activated-ion electron capture dissociation spectra of GSK3B-pY11 [M+2H]$^{2+}$: a) ECD, b) IR-ECD and c) ECD-IR

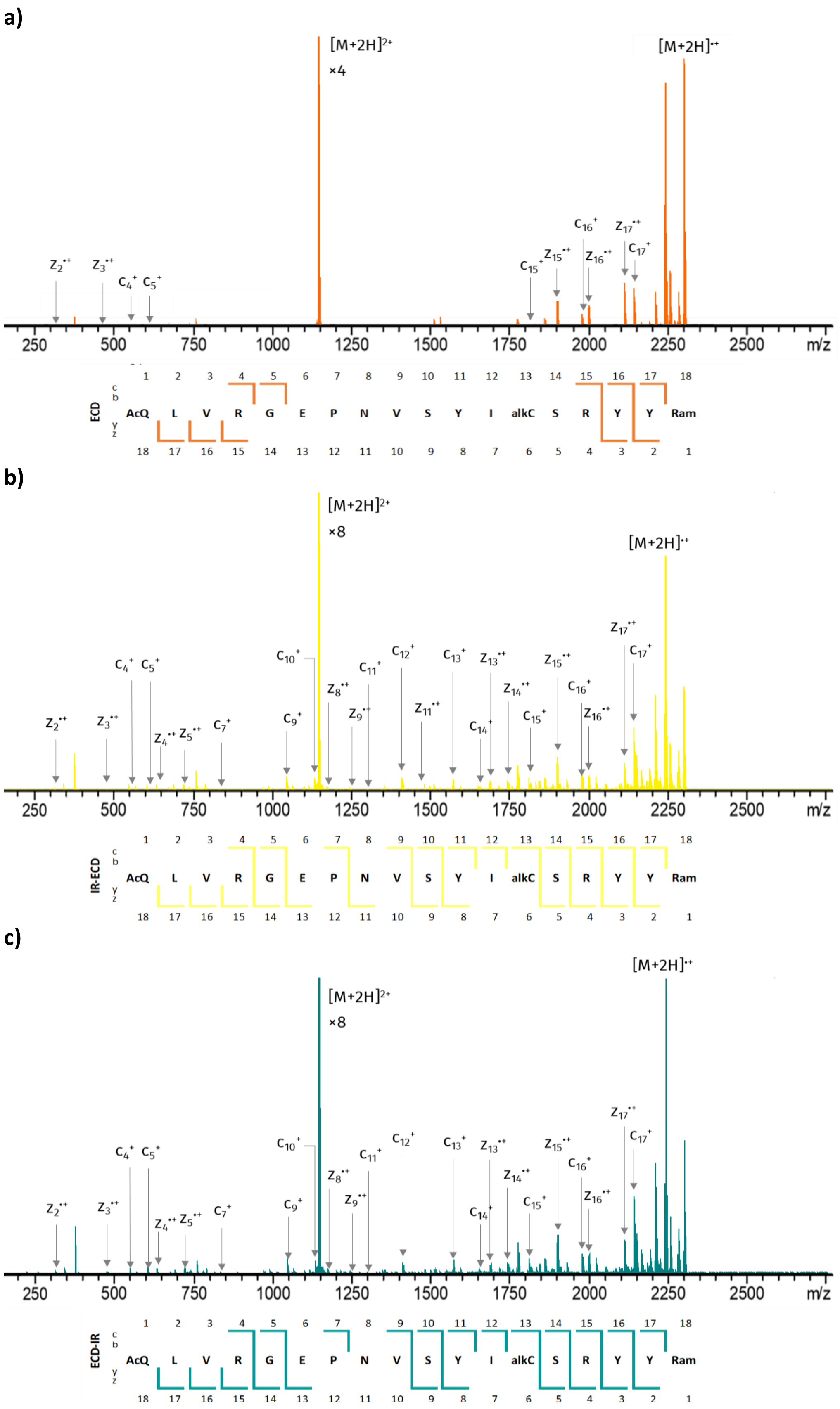


Figure A8-2 Activated-ion electron capture dissociation spectra of GSK3B-Y11 $[M+2H]^{2+}$: a) ECD, b) IR-ECD and c) ECD-IR

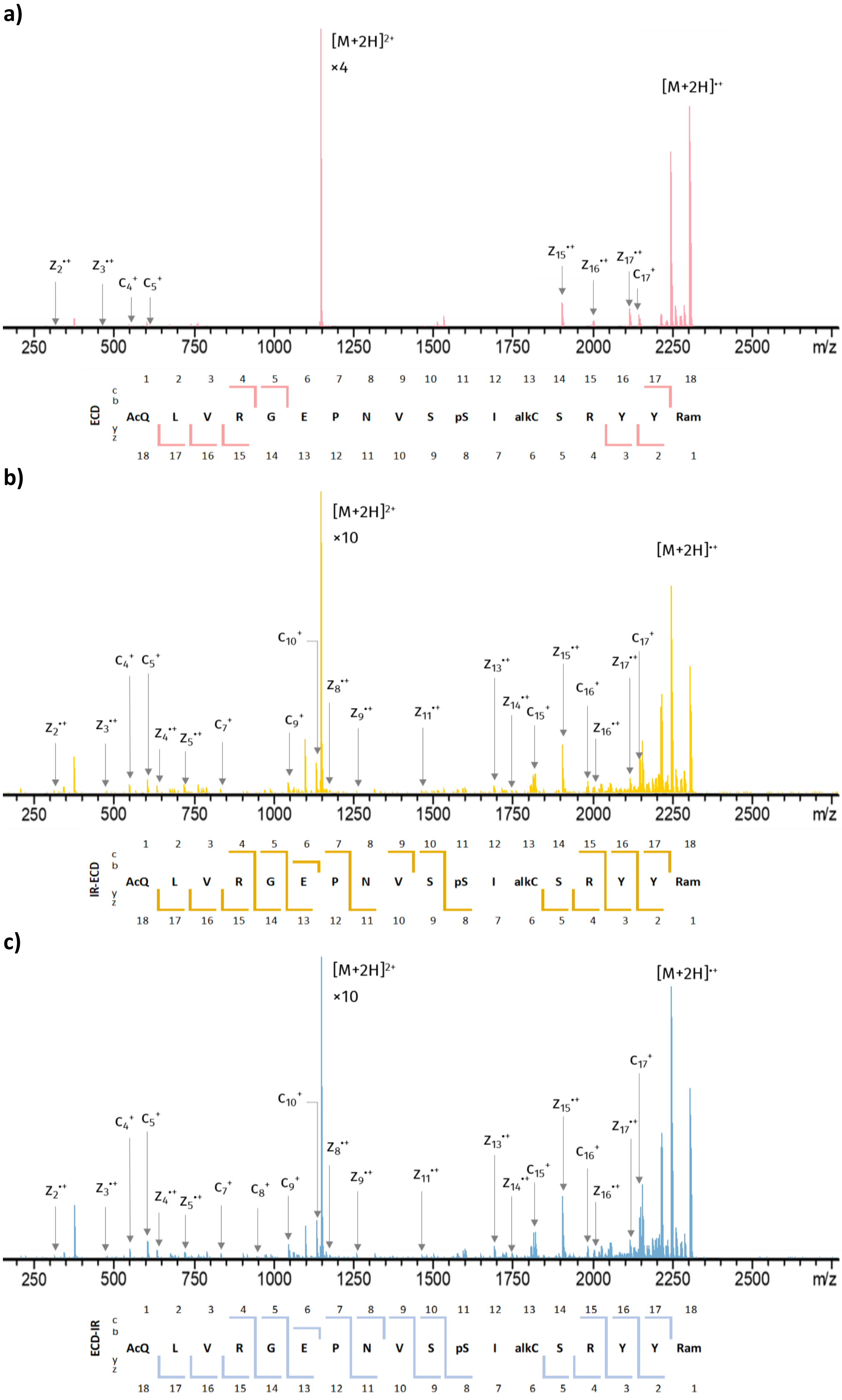


Figure A8-3 Activated-ion electron capture dissociation spectra of GSK3B-pS11 $[M+2H]^{2+}$: a) ECD, b) IR-ECD and c) ECD-IR